Carly Diss

August 2, 2023

[1]:
```
#Reading metadata
install.packages('readxl')
library('readxl')
dsoutput<- read_excel('ds_output.xls')
excel_sheets('ds_output.xls')
ds_output<- read_excel('ds_output.xls', sheet = 1)
lama_ds_output<- read_excel('ds_output.xls', sheet = 'lama_ds_output')
#renaming metadata
x<- data.frame(BRCID= lama_ds_output$brcid, Gender = as.
 ↪factor(lama_ds_output$Gender_ID), Ethnicity = as.
 ↪factor(lama_ds_output$ethnicitycleaned), Age = lama_ds_output$age, Doc_Date␣
 ↪= lama_ds_output$document_date, CN_DOC_ID = lama_ds_output$CN_Doc_ID)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

1. 'lama_ds_output' 2. 'diagnosis_output'

[2]:
```
#Reading text data
path_file<- '18_093_attachment_text_Aug21.csv'
file_text<- readLines(path_file, encoding = 'UTF-8')
file_utf8<- iconv(file_text, to = 'UTF-8', sub = '')
temp_file<- tempfile()
writeLines(file_utf8, temp_file)
textdata<- read.csv(temp_file)
```

[3]:
```
#merge the two datasets
y<- data.frame(Summary = textdata$Summary, Text = textdata$Attachment_text,␣
 ↪CN_DOC_ID = textdata$cn_doc_id)
merged.data<- merge(x, y, by= 'CN_DOC_ID')
data<- merged.data[,c( 'Gender', 'Ethnicity', 'Text')]
```

[4]:
```
table(data$Gender)
#makes sense to focus on just male and female as it represents the large␣
 ↪majority of the gender distribution, female being more than twice of male
```

```
      Female         Male Not Specified        Other
       21902        11166            29            5
```

[5]:
```
#separate the dataset by gender
#check the levels of variables within gender vector
factor_levels<- levels(data$Gender)
print(factor_levels) #needs to observe the lingusitc changes between 1. Female␣
 ↪2. Male 3. Not Specified 4. Other

#female data
data.female<- data[data$Gender=='Female',]

#male data
data.male<- data[data$Gender=='Male',]
```

```
[1] "Female"        "Male"          "Not Specified" "Other"
```

[6]:
```
table(data$Ethnicity)
```

```
                  African (N) Any other Asian background (L)
                         2330                             610
Any other black background (P)     Any other ethnic group (S)
                         2345                            3671
Any other mixed background (G) Any other white background (C)
                          313                            2347
               Bangladeshi (K)                    British (A)
                          266                           10039
                Caribbean (M)                    Chinese (R)
                         1156                             169
                   Indian (H)                      Irish (B)
                          515                             502
                Not Stated (Z)                             NULL
                         1818                            5989
                Pakistani (J)           White and Asian (F)
                          505                              35
  White and Black African (E)  White and black Caribbean (D)
                          141                             351
```

[7]:
```
#for easy management, the project reclasses these 18 catgories into 7 main␣
 ↪categories

#africian data
data.african <- data[data$Ethnicity=='African (N)',]

#any black background data
```

```
data.aob <- data[data$Ethnicity=='Any other black background (P)',]

#any other white background data
data.aow <- data[data$Ethnicity=='Any other white background (C)',]

#3mixed ethnic data
data.othermixed<- data[data$Ethnicity== 'Any other mixed background (G)',]

#White and black caribbean data
data.wabc<- data[data$Ethnicity== 'White and black Caribbean (D)',]

#white and black african data
data.waba<- data[data$Ethnicity== 'White and Black African (E)',]

#white and asian data
data.was<- data[data$Ethnicity== 'White and Asian (F)',]

#British data
data.british <- data[data$Ethnicity=='British (A)',]

#Irish data
data.irish <- data[data$Ethnicity=='Irish (B)',]

#Caribbean data
data.caribbean <- data[data$Ethnicity=='Caribbean (M)',]

#other asian data
data.otherasian <- data[data$Ethnicity== 'Any other Asian background (L)',]

#chinese data
data.chinese <- data[data$Ethnicity== 'Chinese (R)',]

#indian data
data.indian<- data[data$Ethnicity=='Indian (H)',]

#pakistani data
data.pakistani<- data[data$Ethnicity=='Pakistani (J)',]

#bangladeshi data
data.bangladeshi<- data[data$Ethnicity=='Bangladeshi (K)',]
```

[8]:
```
#for easy management, the project reclasses them into 7 main categories

#1 africian
data.african <- data[data$Ethnicity=='African (N)',]

#2 any black background data
```

```
data.black <- data[data$Ethnicity=='Any other black background (P)',]

#3 any white blackground data
data.white <- data[data$Ethnicity=='Any other white background (C)',]

# 4 mixed ethnic data
data.mixed<- c(data.othermixed, data.wabc, data.waba, data.was)

# 5 European data
data.eu <- c(data.british, data.irish)

#6 Caribbean data
data.caribbean <- data[data$Ethnicity=='Caribbean (M)',]

#7 Asian data
data.asian <- c(data.otherasian, data.chinese, data.indian, data.bangladeshi,␣
 ↪data.bangladeshi)
```

[8]:
```
#download tm packages
install.packages('tm')
library(tm)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Loading required package: NLP
```

[90]:
```
#create function
create_corpus = function(input) {
    text_vector <- as.character(input$Text)
my_corpus<- Corpus(VectorSource(text_vector))}
```

[91]:
```
female_corpus<-  create_corpus(data.female)
```

[92]:
```
male_corpus<-  create_corpus(data.male)
```

[95]:
```
african_corpus<- create_corpus(data.african)
```

[96]:
```
black_corpus<- create_corpus(data.black)
```

[97]:
```
white_corpus<- create_corpus(data.white)
```

[98]:
```
mixed_corpus<- create_corpus(data.mixed)
```

[99]:
```
eu_corpus<- create_corpus(data.eu)
```

[100]:
```
caribbean_corpus<- create_corpus(data.caribbean)
```

[101]:
```
asian_corpus<- create_corpus(data.asian)
```

[18]:
```
#downloading the library for cleaning the text
install.packages('SnowballC')
library(SnowballC)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[19]:
```
#test old working code
#create new function
create_archive_dtm = function(corpus){
    archive_dtm = DocumentTermMatrix(corpus, control = list(stopwords = TRUE,␣
 ↪removePunctuation = TRUE, stemming = TRUE, removeNumbers = TRUE, bounds =␣
 ↪list(global=c(20,Inf))))
wordstoremove<-c('zzzzz', 'qqqqq', 'also', 'can', 'will', 'yes','may', 'one',␣
 ↪'now', 'monday', 'tuesday', 'wednesday', 'thursday', 'friday', 'saturday',␣
 ↪'sunday', 'whilst', 'known', 'tel', 'floor', 'due', 'day', 'october',␣
 ↪'november', 'january', 'febuary', 'march', 'april', 'may', 'june', 'july',␣
 ↪'august', 'september', 'december', 'hill', 'south', 'southwark', 'lambeth',␣
 ↪'lewis', 'alex', 'look', 'road', 'formcheckbox', 'use', 'ongo', 'taken',␣
 ↪'someone', 'way', 'ref', 'croydon', 'year', 'week', 'referral', 'two',  ␣
 ↪'regard', 'headquarter', 'fax', 'tel', 'seen', 'chris', 'date', 'helen',␣
 ↪'sydenham')
dtm_filtered<- removeSparseTerms(archive_dtm, 0.99)
dtm_filtered<-dtm_filtered[,!(colnames(dtm_filtered) %in% wordstoremove)]
                            }
```

[20]:
```
#female corpus
archive_dtm_female<- create_archive_dtm(female_corpus)
```

[293]:
```
#male corpus
archive_dtm_male<- create_archive_dtm(male_corpus)
```

[102]:
```
#african
archive_dtm_african<- create_archive_dtm(african_corpus)
```

[103]:
```
#black background
archive_dtm_black<- create_archive_dtm(black_corpus)
```

[104]:
```
#white background
archive_dtm_white<- create_archive_dtm(white_corpus)
```

[117]:
```
#mixed corpus
archive_dtm_mixed<- create_archive_dtm(mixed_corpus)
```

[105]:
```
#European
archive_dtm_eu<- create_archive_dtm(eu_corpus)
```

[106]:
```
#caribbean corpus
archive_dtm_caribbean<- create_archive_dtm(caribbean_corpus)
```

[107]:
```
#asian corpus
archive_dtm_asian<- create_archive_dtm(asian_corpus)
```

[111]:
```
install.packages('slam')
library(slam)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[112]:
```
#create new function
filter_dtm<- function(dtm_filtered){
    sel_idx<-row_sums(dtm_filtered)>0
    filtered_dtm<-dtm_filtered[sel_idx, ]
   }
```

[301]:
```
#female corpus filtered
female_done<- filter_dtm(archive_dtm_female)
```

[302]:
```
#male corpus filtered
male_done<- filter_dtm(archive_dtm_male)
```

[113]:
```
#african corpus filtered
african_done<- filter_dtm(archive_dtm_african)
```

[114]:
```
#black corpus filtered
black_done<- filter_dtm(archive_dtm_black)
```

[115]:
```
#white corpus filtered
white_done<- filter_dtm(archive_dtm_white)
```

[118]:
```
#mixed ethcicity corpus filtered
mixed_done<- filter_dtm(archive_dtm_mixed)
```

[119]:
```
#european corpus filtered
eu_done<- filter_dtm(archive_dtm_eu)
```

[120]:
```
#caribbean corpus filtered
caribbean_done<- filter_dtm(archive_dtm_caribbean)
```

[121]:
```
#asian corpus filtered
asian_done<- filter_dtm(archive_dtm_asian)
```

[122]:
```
#libs for LDA
install.packages('topicmodels')
library(topicmodels)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[123]:
```
install.packages('topicdoc')
library(topicdoc)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[124]:
```
#define function find best k to see which has the highest averge topic coherence
find_best_k = function(dtm){
    k_coherence = data.frame(matrix(nrow = 5, ncol = 2))
colnames(k_coherence) = c('k', 'average_topic_coherence')
i=0
for (k in c(2, 5, 10, 15, 20)){
    i=i+1
    print(paste(k, 'topic', sep = ''))
    filtered_lda = LDA(dtm, k, method='Gibbs', control=list(iter=500,␣
 ↪verbose=25))
    print(terms(filtered_lda))
    print(topic_diagnostics(filtered_lda, dtm))
    k_coherence[i,] = c(k, mean(topic_diagnostics(filtered_lda,␣
 ↪dtm)$topic_coherence))
    }
k_coherence}
```

[309]:
```
#apply function to female
k_coherence_female = find_best_k(female_done)
```

```
[1] "2topic"
K = 2; V = 1496; M = 15322
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
Topic 1 Topic 2
 "team"  "feel"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   465.5193               6.3        0.5341916   126.3184
2         2  1030.4807               5.8        0.3395445   141.7346
  doc_prominence topic_coherence topic_exclusivity
1          13541       -40.51593          9.415176
2          14653       -34.79700          7.719668
[1] "5topic"
K = 5; V = 1496; M = 15322
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
```

```
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
 "assess"    "feel" "appoint"    "team"  "health"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   389.4080               5.8        0.5203405   119.5490
2         2   505.9247               4.6        0.4903459   122.0784
3         3   184.4110               6.6        0.6702890   109.3674
4         4   204.8324               6.1        0.6296746   106.1137
5         5   211.4239               6.0        0.5503110   131.3365
  doc_prominence topic_coherence topic_exclusivity
1           3550       -46.66519          8.280405
2           5909       -42.73639          8.674093
3           5580       -50.85811          9.527483
4           7013       -46.22514          9.472547
5           3328       -46.22096          9.626928
[1] "10topic"
K = 10; V = 1496; M = 15322
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "detail"  "report" "appoint"   "medic"  "hospit"   "pleas"    "feel"  "mental"
```

```
  Topic 9  Topic 10
   "need"  "servic"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  123.52238               6.1        0.6093460  128.98665
2          2  220.04837               5.3        0.5689651  124.48806
3          3  101.61621               6.5        0.7034804   98.18865
4          4  213.45898               6.2        0.6222657   84.34761
5          5   88.34904               7.1        0.7461093   85.75385
6          6  168.87658               5.5        0.6353027  105.81170
7          7  232.23575               4.7        0.6056969   89.48419
8          8   85.64831               5.9        0.7046340  123.99324
9          9  162.95346               6.0        0.5745509  122.07968
10        10   99.29091               6.5        0.6765999   97.04997
   doc_prominence topic_coherence topic_exclusivity
1             751       -41.13840          9.563060
2            2480       -37.69461          9.473865
3            2069       -34.29609          9.604725
4            1839       -52.06709          9.389873
5            1777       -48.94296          9.796689
6            1115       -80.91915          9.312038
7            1570       -63.66306          9.441258
8            1360       -43.40095          9.736004
9             785       -44.59254          9.376761
10           1955       -59.94120          9.696535
[1] "15topic"
K = 15; V = 1496; M = 15322
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
```

```
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
   "feel"  "mental"   "medic"   "pleas"  "health"  "number"    "said"    "need"
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
"appoint"  "report" "therapi"  "hospit"     "mrs"  "servic"  "detail"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  157.00694               5.4        0.6283211   89.58163
2          2   76.83938               5.6        0.6494108  122.13304
3          3  125.95955               6.4        0.6716625   79.65146
4          4  118.93052               6.0        0.6511601   88.11526
5          5   52.56853               5.5        0.7427521   96.33049
6          6   74.77661               5.9        0.6938187  117.95211
7          7  149.99400               5.8        0.6168375  110.66324
8          8  100.61419               6.0        0.6104413  126.16761
9          9   81.99280               5.8        0.7243569   97.39660
10        10  133.17796               5.3        0.6348033  110.49570
11        11   73.90194               6.1        0.6851124  108.15840
12        12   61.56363               6.9        0.7683423   89.31367
   doc_prominence topic_coherence topic_exclusivity
1             785       -62.97028          9.500743
2             605       -51.03379          9.705542
3             948       -57.37667          9.674797
4             480       -60.23673          9.493821
5             696       -46.12616          9.828767
6            1170       -45.66344          9.642610
7            1191       -40.51119          9.472959
8             420       -43.60915          9.493719
9            1657       -48.86669          9.681938
10           1007       -37.64571          9.725026
11           1079       -47.83495          9.697882
12           1104       -56.69030          9.823372
 [ reached 'max' / getOption("max.print") -- omitted 3 rows ]
[1] "20topic"
K = 20; V = 1496; M = 15322
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
```

```
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
  "appoint"   "appoint"    "famili"     "medic"      "feel"    "health"
    Topic 7     Topic 8     Topic 9    Topic 10    Topic 11    Topic 12
   "report"       "mrs"    "servic"   "patient"   "support"   "therapi"
   Topic 13    Topic 14    Topic 15    Topic 16    Topic 17    Topic 18
   "clinic"    "detail"      "feel"    "mental"    "referr"  "discharg"
   Topic 19    Topic 20
"interpret"      "team"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   49.80756               5.4        0.7298473   90.16545
2          2   51.68285               7.1        0.7775692   86.26668
3          3   70.85027               5.7        0.6419648  114.41014
4          4  109.71415               5.5        0.6784437   86.59985
5          5  119.81348               3.9        0.6919274   86.91753
6          6   40.41032               5.5        0.7461856   96.73309
7          7  107.87398               5.3        0.6243767  121.97903
8          8   93.74653               6.2        0.6205996   92.33886
9          9   45.53303               6.1        0.7180782  108.47263
10        10   90.59673               5.2        0.6753976   60.44759
11        11   79.31195               5.0        0.6674351   98.39430
12        12   58.16472               6.1        0.6874419  108.15840
   doc_prominence topic_coherence topic_exclusivity
1             864       -67.93660          9.830184
2             929       -51.72570          9.834761
3             210       -54.47011          9.662571
4             858       -44.88576          9.646116
5             424       -65.58229          9.603127
6             661       -57.02643          9.882041
7            1120       -33.23824          9.552709
8             255       -76.63520          9.633583
9             682       -52.56165          9.800372
10            193       -86.15782          9.825526
11            279       -52.54711          9.608702
12            806       -47.61974          9.692493
 [ reached 'max' / getOption("max.print") -- omitted 8 rows ]
```

[310]:
```
#apply function to male
k_coherence_male = find_best_k(male_done)
```

```
[1] "2topic"
K = 2; V = 1465; M = 7023
```

```
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2
  "feel" "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1  1031.4983               5.2        0.3444605   88.73376
2         2   433.5017               6.1        0.5294559   84.41135
  doc_prominence topic_coherence topic_exclusivity
1           6615       -34.51198          9.065057
2           6274       -30.75953          9.574931
[1] "5topic"
K = 5; V = 1465; M = 7023
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
```

```
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
  "medic" "appoint"  "referr"    "feel"  "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   352.1087               5.5        0.5445360   79.57086
2         2   183.2058               6.1        0.6660236   74.62978
3         3   324.9742               5.6        0.5455035   85.52171
4         4   423.8867               5.0        0.5193744   79.54986
5         5   180.8247               5.9        0.6179924   82.55976
  doc_prominence topic_coherence topic_exclusivity
1           2511       -37.85740          9.280045
2           3199       -26.91546          9.492451
3           1703       -43.30380          8.984328
4           2197       -48.51953          9.121152
5           3026       -40.86020          9.482908
[1] "10topic"
K = 10; V = 1465; M = 7023
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
   "team"  "referr"  "servic"  "assess" "appoint"    "feel" "support" "appoint"
  Topic 9  Topic 10
  "medic"  "report"
```

```
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  104.02049               5.7        0.6794317   69.19236
2          2  114.09542               5.9        0.6387381   71.63919
3          3   86.90382               5.8        0.6859623   75.45261
4          4  198.61816               5.2        0.5990879   73.29645
5          5  105.07422               6.2        0.6859488   68.65887
6          6  228.32401               4.3        0.6087287   65.90970
7          7  113.20791               5.2        0.6392681   73.21430
8          8  110.10905               5.8        0.7021581   64.68657
9          9  202.86059               6.1        0.6339193   56.17042
10        10  201.78634               5.3        0.5858134   81.39974
   doc_prominence topic_coherence topic_exclusivity
1             408       -67.23110          9.766676
2             423       -61.43737          9.785519
3             828       -46.24297          9.704731
4             303       -57.87951          9.087518
5             985       -38.92450          9.612441
6             773       -57.49778          9.389577
7             417       -42.57553          9.610969
8            1396       -51.49160          9.674889
9             730       -72.24771          9.496144
10           1098       -31.93827          9.550686
[1] "15topic"
K = 15; V = 1465; M = 7023
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
   "team"  "report" "support"  "referr"   "medic"  "servic" "patient"  "report"
```

```
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
"appoint"    "feel"  "servic"    "team" "therapi" "appoint"    "risk"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   55.24008               5.6        0.6935471   71.00763
2          2  146.73141               5.3        0.6174597   79.43013
3          3  106.02015               4.9        0.6520456   67.45420
4          4   81.28284               5.7        0.6577814   70.69472
5          5  110.83796               6.2        0.6688544   58.94300
6          6   89.13668               4.9        0.6956948   57.21952
7          7  110.98764               6.0        0.6670768   55.44847
8          8  149.11261               6.4        0.6159202   69.42623
9          9   60.86486               6.2        0.7493678   53.64889
10        10  168.73326               4.3        0.6426683   63.60659
11        11   49.35031               6.1        0.7198838   77.90338
12        12   71.28387               6.8        0.7137719   53.23015
   doc_prominence topic_coherence topic_exclusivity
1             360       -35.93982          9.761783
2             772       -34.34255          9.625340
3             181       -44.72612          9.598945
4             313       -58.29617          9.787570
5             448       -49.68981          9.756317
6             207       -71.84531          9.662736
7             218       -81.20163          9.794340
8             171       -69.99504          9.413247
9             439       -41.06115          9.785887
10            517       -58.23272          9.542551
11            375       -48.52862          9.823097
12            629       -49.76016          9.733828
 [ reached 'max' / getOption("max.print") -- omitted 3 rows ]
[1] "20topic"
K = 20; V = 1465; M = 7023
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
```

```
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
       Topic 1        Topic 2        Topic 3        Topic 4        Topic 5
        "feel"      "appoint"     "discharg"       "servic"      "therapi"
       Topic 6        Topic 7        Topic 8        Topic 9       Topic 10
      "servic"          "nhs"       "suicid"       "report"       "assess"
      Topic 11       Topic 12       Topic 13       Topic 14       Topic 15
      "memori"         "work"       "number"       "mother"        "centr"
      Topic 16       Topic 17       Topic 18       Topic 19       Topic 20
"psychiatrist"      "support"       "referr"        "medic"      "appoint"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  103.14469               4.0        0.6957573   59.49056
2          2   52.04358               6.3        0.7491720   51.96100
3          3   65.28979               5.9        0.6771604   66.71853
4          4   41.01145               6.4        0.7219710   74.69115
5          5   53.30455               6.3        0.6936621   70.56846
6          6   39.38875               5.8        0.7141979   77.26558
7          7   71.94371               5.4        0.7088817   56.67466
8          8   70.34458               5.2        0.6674310   83.35596
9          9  113.69700               5.4        0.6518250   69.94000
10        10  103.49663               5.8        0.6435837   67.08911
11        11  107.27320               5.6        0.6621799   49.38569
12        12  112.92464               5.3        0.6659113   57.90535
   doc_prominence topic_coherence topic_exclusivity
1             160       -62.31368          9.707933
2             296       -48.10710          9.806965
3             184       -61.33293          9.791730
4             318       -45.69724          9.793536
5             252       -49.82285          9.745619
6              60       -39.86532          9.804146
7             125       -71.21607          9.703083
8             151       -35.73972          9.773704
9             299       -40.31870          9.631341
10            130       -94.30787          9.498114
11            167       -82.87766          9.738285
12            137       -58.96753          9.369991
 [ reached 'max' / getOption("max.print") -- omitted 8 rows ]
```

[125]:
```
#apply function to african
k_coherence_african = find_best_k(african_done)
```

```
[1] "2topic"
K = 2; V = 1504; M = 1598
Sampling 500 iterations!
Iteration 25 …
```

```
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
Topic 1 Topic 2
 "feel"  "team"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1  1037.1607               5.9        0.3345894   47.63870
2         2   466.8393               5.7        0.5340726   43.50042
  doc_prominence topic_coherence topic_exclusivity
1           1517       -35.99427          6.794040
2           1409       -32.30778          9.239469
[1] "5topic"
K = 5; V = 1504; M = 1598
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
```

```
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
 "servic"    "feel" "appoint"   "medic"  "referr"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   213.8595               5.6        0.6145862   45.11167
2         2   446.0033               4.6        0.5250338   34.94219
3         3   214.9025               6.6        0.6630228   30.51157
4         4   341.4788               5.7        0.5610924   34.64314
5         5   287.7559               5.7        0.5479308   45.93274
  doc_prominence topic_coherence topic_exclusivity
1            772       -35.70743          9.407108
2            567       -46.32482          9.086836
3            736       -35.15360          9.455217
4            490       -42.76391          9.143239
5            326       -43.40844          9.219326
[1] "10topic"
K = 10; V = 1504; M = 1598
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
 "appoint"   "report" "children"  "patient"   "servic"    "medic"     "team"
   Topic 8    Topic 9   Topic 10
 "support"     "feel"  "appoint"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  115.64303               6.6        0.6878285   33.98807
```

```
2          2  198.05902              5.3        0.6063339   36.11217
3          3  145.35064              5.8        0.5979651   42.30017
4          4  148.21172              5.4        0.6482676   26.71290
5          5  142.91813              5.9        0.6351440   34.22444
6          6  156.63912              5.5        0.6286059   33.44541
7          7   97.98043              5.8        0.6872787   40.90360
8          8  164.68996              5.5        0.6141655   38.41504
9          9  214.75036              3.9        0.6296147   25.87232
10        10  119.75758              7.1        0.7190617   27.10140
   doc_prominence topic_coherence topic_exclusivity
1             338       -35.52007          9.624151
2             217       -35.71392          9.571615
3             116       -49.30886          9.584559
4              42       -70.32416          9.621954
5              91       -81.87431          9.622306
6             194       -43.43769          9.532391
7             169       -50.41809          9.743850
8              83       -41.31375          9.218628
9             111       -59.86227          9.370365
10            228       -58.00943          9.745874
[1] "15topic"
K = 15; V = 1504; M = 1598
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
 "patient"     "babi"     "feel"   "referr"  "support"   "servic"    "medic"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
  "report"   "servic"   "health"   "hospit"  "appoint" "children"     "team"
```

```
  Topic 15
    "mood"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  127.83607               5.7        0.6738457   26.06706
2          2   90.29836               6.2        0.6093479   38.92908
3          3  163.69344               4.2        0.6603137   27.06705
4          4   98.84802               5.6        0.6283167   36.47165
5          5  120.71406               5.2        0.6443148   29.02675
6          6   79.34077               5.7        0.6732477   33.65493
7          7  115.55779               6.2        0.6443047   33.94040
8          8  122.53115               5.8        0.6446477   36.50929
9          9   91.92767               6.1        0.6775529   26.49402
10        10   62.86828               6.6        0.6947289   40.71606
11        11   74.67100               6.5        0.7455411   27.21718
12        12   60.81086               5.3        0.7039587   31.08775
13        13   98.02138               5.7        0.6149763   41.22627
14        14   71.65730               6.2        0.7352812   23.50276
15        15  125.22386               5.2        0.6528341   27.13183
   doc_prominence topic_coherence topic_exclusivity
1              45       -92.67940          9.727652
2              45       -55.54277          9.648960
3              74       -65.65347          9.543728
4              84       -50.09298          9.690158
5              45       -56.47885          9.583019
6             115       -56.13412          9.718841
7              89       -60.60428          9.541813
8             125       -36.07592          9.705613
9              77       -76.60512          9.717890
10             86       -50.77191          9.809727
11            111       -60.17670          9.813001
12            128       -60.97787          9.823552
13             39       -52.78566          9.612947
14            119       -49.71900          9.806941
15             86       -48.64935          9.670011
[1] "20topic"
K = 20; V = 1504; M = 1598
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
```

```
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
    "babi"   "report"     "feel" "children"   "servic"     "hous"   "inform"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
  "mental"   "famili"  "therapi" "discharg"   "servic"  "consult"  "appoint"
  Topic 15   Topic 16   Topic 17   Topic 18   Topic 19   Topic 20
  "hospit"   "memori"  "appoint"   "referr"  "support"    "medic"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   69.61034                6.1        0.6331524   39.37526
2          2   97.54540                5.3        0.6513711   35.55663
3          3  120.45582                4.0        0.6866911   26.34843
4          4   73.78692                6.0        0.6342429   32.93734
5          5   54.59619                5.8        0.7195750   33.07278
6          6  102.91648                5.7        0.6803163   32.91587
7          7   42.10127                5.5        0.7329274   36.04331
8          8   61.93186                6.2        0.7180654   31.88400
9          9   80.00722                5.8        0.6199090   40.99273
10        10   89.72957                6.3        0.6695551   24.34708
11        11   87.85424                6.1        0.6569618   30.35214
12        12   59.34535                5.5        0.7061192   29.63970
13        13   57.86145                5.9        0.7416783   24.19374
14        14   49.28170                5.7        0.7015813   31.22266
15        15   56.97440                6.9        0.7531686   27.20549
16        16  109.32299                5.5        0.6815878   24.24665
17        17   58.21772                5.2        0.7058678   27.57590
18        18   63.01525                5.5        0.6874227   35.41871
19        19   70.75294                5.5        0.6568992   38.71232
20        20   98.69290                5.7        0.6834398   28.90760
   doc_prominence topic_coherence topic_exclusivity
1              36       -55.43189          9.728474
2             102       -34.23891          9.737290
3              44       -72.13948          9.702985
4              23       -51.69376          9.776310
5              39       -56.16344          9.799955
6              19       -77.13942          9.493374
7               9       -55.99155          9.878337
8             116       -59.73798          9.835856
9              31       -59.06368          9.623130
10             33       -65.95200          9.746142
```

```
11              33       -66.19578          9.601172
12              60       -73.65955          9.829593
13             103       -54.75595          9.816907
14             108       -66.71270          9.844279
15              83       -56.55386          9.859524
16              35       -79.68471          9.816592
17              80       -62.03663          9.780896
18              66       -55.28728          9.843904
19              20       -66.40990          9.699424
20              77       -44.93447          9.627049
```

[126]:
```
#apply function to black
k_coherence_black = find_best_k(black_done)
```

```
[1] "2topic"
K = 2; V = 1524; M = 1637
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
Topic 1 Topic 2
 "team"  "feel"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   481.0245                5.9        0.5296294   43.08552
2         2  1042.9755                5.5        0.3342560   46.11004
  doc_prominence topic_coherence topic_exclusivity
1           1456       -34.54381          9.073174
2           1584       -35.84801          7.070530
[1] "5topic"
K = 5; V = 1524; M = 1637
```

```
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
 "referr"  "servic"  "report"    "feel" "appoint"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   278.9746               5.6        0.5440207   40.73869
2         2   238.6622               5.5        0.6263727   36.88637
3         3   416.3365               5.4        0.5415594   38.10320
4         4   399.6107               4.7        0.5547540   34.89475
5         5   190.4161               6.2        0.6379529   37.23849
  doc_prominence topic_coherence topic_exclusivity
1            388       -43.80328          9.374571
2            799       -41.12950          9.260233
3            559       -37.55775          8.861415
4            503       -51.42873          9.064473
5            698       -56.68412          9.568687
[1] "10topic"
K = 10; V = 1524; M = 1637
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
```

```
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "servic"  "person" "appoint"  "report"   "medic"    "feel"    "team"  "health"
  Topic 9  Topic 10
 "mental"  "assess"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  118.82243               6.2        0.6614664   36.85716
2          2  177.61291               5.3        0.5922423   41.53904
3          3   91.13491               6.6        0.7262444   28.91002
4          4  215.61451               5.2        0.5873955   36.86538
5          5  196.46877               5.6        0.6526716   27.78772
6          6  206.38063               4.6        0.6269399   27.76115
7          7  117.84909               5.8        0.6753335   32.01422
8          8  119.73619               6.2        0.5950913   39.79059
9          9  180.44644               5.2        0.5810526   40.00514
10        10   99.93412               5.6        0.6592230   35.40303
   doc_prominence topic_coherence topic_exclusivity
1             210       -49.41644          9.567574
2             103       -41.66814          9.259052
3             196       -64.32961          9.801074
4             221       -43.44306          9.504625
5             149       -54.85499          9.388233
6             142       -65.69256          9.556080
7             261       -44.15423          9.617349
8             129       -54.83302          9.737401
9              86       -56.69612          9.226761
10            226       -51.18917          9.694647
[1] "15topic"
K = 15; V = 1524; M = 1637
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
```

```
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
   "work"  "person" "appoint"  "number"  "servic"    "risk" "therapi"  "health"
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
 "referr" "thought"   "medic"    "feel"  "mother" "patient" "appoint"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  110.96925               4.7        0.6694482   26.16753
2          2  111.01900               5.4        0.6235257   40.39469
3          3   80.92458               6.4        0.6901815   33.01629
4          4   72.29784               6.3        0.6703944   33.07054
5          5   81.16832               5.4        0.7006506   29.07316
6          6  142.13389               5.4        0.5973527   39.42182
7          7   78.32974               6.1        0.6815086   35.55169
8          8   53.01846               6.1        0.7050584   35.69637
9          9   82.19730               6.0        0.6172168   38.72959
10        10  127.40599               5.4        0.6379921   32.83272
11        11  110.70958               5.8        0.6783574   30.21487
12        12  149.52031               4.7        0.6622383   28.09637
13        13  110.77867               6.4        0.6309046   30.29181
14        14  135.17691               5.3        0.6374638   25.35072
15        15   78.35016               6.8        0.7263560   27.12412
   doc_prominence topic_coherence topic_exclusivity
1              31       -62.76434          9.653890
2              58       -48.91236          9.422625
3             138       -38.82469          9.673576
4             150       -59.81118          9.758731
5              65       -68.68690          9.730049
6              48       -59.14721          9.242525
7             132       -47.51513          9.721162
8              76       -50.25949          9.841857
9              69       -54.65375          9.718294
10            115       -36.25403          9.704742
11             86       -51.45347          9.625047
12             81       -56.65683          9.557691
13             84       -46.13222          9.800583
```

```
14             36       -86.57978          9.565867
15            143       -62.49333          9.821419
[1] "20topic"
K = 20; V = 1524; M = 1637
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
 "support"   "servic"  "appoint"  "therapi"  "appoint"   "person" "discharg"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
  "number"     "need"   "report"  "patient"   "report"   "mother"     "feel"
  Topic 15   Topic 16   Topic 17   Topic 18   Topic 19   Topic 20
  "health"   "referr"     "miss"   "health"    "medic"   "servic"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   90.51013               4.3        0.6761836   30.79529
2          2   68.53521               5.9        0.7151944   20.45864
3          3   54.48329               7.1        0.7438454   27.20130
4          4   63.78743               6.8        0.6773890   35.85809
5          5   70.54655               6.6        0.6957713   32.74494
6          6   94.72134               5.6        0.6278448   36.83750
7          7   67.30699               5.7        0.6917442   35.64884
8          8   55.16253               5.6        0.6849881   29.93416
9          9   67.51274               5.8        0.6204809   38.27162
10        10  106.99561               5.4        0.6681191   32.05829
11        11  107.80693               5.5        0.6643213   24.21174
12        12   98.56108               5.0        0.6038851   32.61477
13        13  111.39224               5.9        0.6485479   31.38943
14        14  109.81265               4.6        0.6595002   29.10745
15        15   47.85141               5.7        0.7076350   31.65394
```

```
16        16   66.85906              5.5        0.6806299   34.37062
17        17   44.94847              5.5        0.6842282   32.43688
18        18   65.21469              6.4        0.6334364   35.52268
19        19   82.78076              5.8        0.6898784   22.91862
20        20   49.21089              6.0        0.6864004   34.01006
   doc_prominence topic_coherence topic_exclusivity
1              19       -56.12511          9.502894
2              45       -62.33364          9.755818
3              81       -60.38287          9.860601
4             111       -49.88420          9.696144
5             110       -53.28581          9.694259
6              22       -53.03729          9.476628
7              24       -41.51329          9.721344
8              87       -54.09695          9.791032
9              23       -48.54771          9.647203
10             74       -38.55409          9.711940
11             31       -89.35880          9.694820
12             21       -76.81655          9.613654
13             61       -44.13205          9.666964
14             61       -57.70460          9.594844
15             72       -58.39192          9.860455
16             88       -52.10302          9.813268
17             51       -55.54648          9.824911
18             48       -61.29052          9.791357
19             56       -63.72269          9.821495
20             69       -59.11812          9.826385
```

[127]:
```
#apply function to white
k_coherence_white = find_best_k(white_done)
```

```
[1] "2topic"
K = 2; V = 1556; M = 1600
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
```

```
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2
  "feel" "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   1028.827               5.7        0.3502139   44.84504
2         2    527.173               6.2        0.5179961   39.76327
  doc_prominence topic_coherence topic_exclusivity
1           1529       -35.46924          8.164227
2           1427       -33.16834          9.329544
[1] "5topic"
K = 5; V = 1556; M = 1600
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
"appoint"   "medic"    "feel"  "servic"  "health"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   226.4918               5.7        0.6413350   33.89053
2         2   358.8857               6.0        0.5559827   34.61748
3         3   440.9165               4.9        0.5146040   38.75621
4         4   259.5449               6.1        0.5951782   39.32532
5         5   270.1611               6.0        0.5450644   44.34774
  doc_prominence topic_coherence topic_exclusivity
1            912       -34.15349          9.462527
```

```
2            395       -74.14822          9.193619
3            539       -43.48766          9.122732
4            417       -57.21080          9.327072
5            383       -42.61476          9.344162
[1] "10topic"
K = 10; V = 1556; M = 1600
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
"interpret"    "referr"    "memori"    "servic"    "health"   "appoint"
    Topic 7     Topic 8     Topic 9    Topic 10
   "assess"     "medic"      "feel"    "hospit"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  150.38059               5.4        0.6432921   34.38803
2          2  132.00536               5.4        0.6073939   38.05468
3          3  208.72349               5.5        0.6061903   33.33135
4          4  116.63342               6.4        0.6552555   33.59692
5          5   95.97017               5.8        0.6763612   39.50758
6          6  120.07333               6.6        0.6797765   32.30366
7          7  167.06384               5.9        0.5732583   36.78646
8          8  207.40908               5.4        0.6101757   34.59846
9          9  258.04349               5.6        0.5790425   33.28843
10        10   99.69723               6.1        0.7157265   28.10761
   doc_prominence topic_coherence topic_exclusivity
1              58       -59.92489          9.520079
2              97       -44.75211          9.596948
3              97       -81.23585          9.556116
4             156       -47.03163          9.593339
```

```
5             160       -41.60684          9.691202
6             239       -40.40248          9.571874
7              75       -49.64691          9.436179
8             255       -38.49749          9.510334
9             281       -46.54004          9.381315
10            240       -47.10002          9.765452
[1] "15topic"
K = 15; V = 1556; M = 1600
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
  "therapi"    "detail" "interpret"   "appoint"    "famili"    "hospit"
    Topic 7     Topic 8     Topic 9    Topic 10    Topic 11    Topic 12
  "support"    "memori"    "referr"      "feel"    "mental"     "medic"
   Topic 13    Topic 14    Topic 15
   "clinic"    "report"    "assess"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   78.98010              6.4        0.6703221   33.59692
2          2   93.66163              6.4        0.6198809   38.29878
3          3  113.88672              5.8        0.6558893   32.85549
4          4   60.93900              5.2        0.7107896   33.35715
5          5   98.14671              5.8        0.6184739   36.01411
6          6   56.37211              6.9        0.7323345   28.88473
7          7  118.61386              5.3        0.6319181   34.83285
8          8  150.20737              5.4        0.6337000   33.44335
9          9  110.74337              5.5        0.6190782   38.58251
10        10  169.00645              4.5        0.6330692   31.04376
11        11   53.08452              5.8        0.7192797   36.35957
```

```
12        12  121.40805              6.3        0.6714237   26.93285
13        13   83.94201              5.4        0.7018320   27.80211
14        14  153.23302              5.6        0.6320667   34.56187
15        15   93.77506              5.5        0.6921296   30.10285
   doc_prominence topic_coherence topic_exclusivity
1             117       -41.07980          9.670139
2              56       -49.81213          9.645210
3              55       -60.09124          9.556371
4              86       -45.85085          9.775165
5              87       -54.18024          9.707825
6             125       -61.52947          9.900468
7              24       -54.89473          9.479664
8              66       -83.36679          9.570159
9              62       -42.96403          9.465623
10            132       -54.86899          9.495376
11             26       -39.15842          9.838017
12            104       -45.64341          9.658701
13            110       -46.74752          9.653640
14            151       -39.97068          9.641111
15            180       -51.19599          9.669944
[1] "20topic"
K = 20; V = 1556; M = 1600
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
  "therapi"    "famili"    "detail"   "appoint"    "referr"    "report"
    Topic 7     Topic 8     Topic 9    Topic 10    Topic 11    Topic 12
   "memori"     "medic"      "feel" "interpret"   "appoint"    "clinic"
```

```
   Topic 13    Topic 14    Topic 15    Topic 16    Topic 17    Topic 18
   "servic"    "mother"    "health"   "patient"      "team"    "mental"
   Topic 19    Topic 20
   "servic"    "hospit"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   87.18936                6.6        0.6777679   22.12799
2          2   75.88428                5.7        0.6011888   38.88701
3          3   73.09455                5.9        0.6313808   36.52957
4          4   72.68174                6.1        0.7032675   33.74079
5          5   79.19575                5.2        0.6243923   38.82264
6          6  109.95316                5.3        0.6605537   35.50843
7          7  102.85918                5.9        0.6283267   32.08651
8          8   89.70603                5.8        0.6866797   25.46149
9          9  112.54700                4.0        0.6762539   27.55657
10        10   92.89468                5.9        0.6567115   32.77782
11        11   46.75619                7.3        0.7367525   25.81435
12        12   58.77443                5.8        0.7118857   25.51969
13        13   54.11093                6.9        0.6660790   34.93018
14        14   99.76161                6.4        0.6601980   29.98079
15        15   52.82611                6.0        0.7049875   33.36186
16        16   97.71725                5.1        0.6856766   20.48013
17        17   53.38521                5.6        0.6831324   25.92827
18        18   60.47984                5.7        0.6621844   37.13991
19        19   49.47790                5.9        0.6947801   35.17384
20        20   86.70481                6.4        0.6485444   26.67099
   doc_prominence topic_coherence topic_exclusivity
1              40       -74.31055          9.724414
2              21       -47.17857          9.586104
3              44       -49.94471          9.674218
4             138       -59.67078          9.691375
5              41       -39.26672          9.469278
6              89       -37.60977          9.683791
7              45       -77.24371          9.686185
8              75       -54.21267          9.712268
9              53       -60.30949          9.664252
10             54       -59.06400          9.559769
11            102       -59.11806          9.893677
12             66       -57.45436          9.765652
13             85       -47.99103          9.704088
14             46       -48.03785          9.690104
15             56       -48.56378          9.801429
16             27       -87.84026          9.782514
17             62       -50.31112          9.802776
18             29       -54.14662          9.797244
19             59       -62.13888          9.818434
20             35       -71.86377          9.670741
```

[128]:
```
#apply function to mixed
k_coherence_mixed = find_best_k(mixed_done)
```

```
[1] "2topic"
K = 2; V = 371; M = 242
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2
"mental"   "team"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1    218.978               5.4        0.3798665   17.29782
2         2    152.022               5.8        0.5249377   14.02729
  doc_prominence topic_coherence topic_exclusivity
1            242       -40.10843          9.178319
2            224       -34.88036          8.936045
[1] "5topic"
K = 5; V = 371; M = 242
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
```

```
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
 "servic"  "health"    "team"    "feel" "appoint"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   64.17132               5.5        0.5669393  15.943329
2         2   67.62044               5.9        0.5787836  15.312699
3         3   66.70361               6.3        0.6110866  13.125259
4         4  107.37818               4.7        0.5697567  12.629451
5         5   65.12645               6.2        0.6604362   7.813763
  doc_prominence topic_coherence topic_exclusivity
1             60       -48.16239          9.523202
2             66       -41.31770          9.430906
3            113       -54.85641          9.500136
4             74       -42.15721          9.557979
5            122       -31.30616          9.406131
[1] "10topic"
K = 10; V = 371; M = 242
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
```

```
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "referr" "appoint"    "feel"  "health" "therapi" "support"  "servic" "current"
  Topic 9  Topic 10
 "hospit"  "assess"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   32.46359               5.6        0.6264771  14.352447
2          2   23.34482               5.7        0.7046544   9.173878
3          3   50.37972               5.4        0.6519946  13.020316
4          4   31.57684               6.2        0.6752967  11.624394
5          5   34.22307               6.6        0.6393548  11.860046
6          6   34.03155               4.9        0.6180887  13.690816
7          7   31.53318               5.9        0.6964118  11.729214
8          8   43.72084               6.3        0.6478104  12.565742
9          9   40.01759               6.5        0.7204954   6.149055
10        10   49.70881               5.9        0.6450468  10.068293
   doc_prominence topic_coherence topic_exclusivity
1              21       -47.50808          9.706422
2              20       -37.17241          9.760964
3              29       -44.48701          9.650221
4              17       -54.78401          9.677656
5              16       -66.53010          9.655809
6              11       -46.81973          9.627559
7              30       -60.63155          9.708146
8              23       -38.37865          9.636522
9              52       -44.40335          9.616318
10             19       -51.15792          9.535450
[1] "15topic"
K = 15; V = 371; M = 242
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
```

```
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "mental"  "report"  "servic"  "recent" "appoint"    "wing" "histori"  "servic"
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
 "assess"    "feel"  "person" "support" "therapi"  "detail"  "clinic"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   18.01265               4.8        0.7118207  12.731359
2          2   32.60982               5.7        0.6788418  11.376659
3          3   17.49056               5.3        0.6523986  14.311673
4          4   37.57218               5.2        0.6777129   8.951876
5          5   24.00752               7.2        0.7126977   8.564471
6          6   28.51603               6.2        0.7224161   4.817191
7          7   27.95190               6.4        0.6632289   9.690766
8          8   21.78465               5.2        0.6736899  12.052277
9          9   23.32054               6.3        0.6708388   9.149273
10        10   30.79457               4.3        0.6806465  11.574865
11        11   22.01799               5.3        0.6240617  11.356786
12        12   23.96334               4.9        0.6412252  13.690816
13        13   21.70096               7.1        0.6532638  12.248442
14        14   21.20767               6.0        0.6608709  13.012477
15        15   20.04963               5.4        0.7176457   7.263319
   doc_prominence topic_coherence topic_exclusivity
1              12       -57.20265          9.770488
2              16       -35.25673          9.659715
3               2       -49.51196          9.747902
4               9       -47.44638          9.659662
5              11       -60.33040          9.715522
6              33       -30.34664          9.696822
7               4       -44.81364          9.739438
8              12       -55.52016          9.744080
9              12       -63.03667          9.776272
10              9       -49.61298          9.749511
11              1       -58.68200          9.719244
12              4       -46.21921          9.594944
13              9       -67.67319          9.755711
14             17       -43.49892          9.752265
15              9       -45.05792          9.720155
[1] "20topic"
K = 20; V = 371; M = 242
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
```

```
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
 "appoint"  "therapi"  "support"    "medic"   "clinic"   "report"   "assess"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
"lewisham"  "concern"   "servic"     "hous"     "mood"  "histori"   "hospit"
  Topic 15   Topic 16   Topic 17   Topic 18   Topic 19   Topic 20
   "centr"   "health"     "name"     "care"     "feel"   "referr"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   14.44152               5.6        0.7158915   7.912818
2          2   19.05358               6.7        0.7185125   4.603622
3          3   17.67378               5.7        0.6545733  12.399905
4          4   21.86839               6.1        0.7079877   7.703670
5          5   13.51948               5.7        0.7000217   7.716641
6          6   23.75383               5.4        0.6618107  10.982521
7          7   17.88427               6.4        0.7101145   9.231183
8          8   15.57138               7.0        0.6432412  12.387723
9          9   16.55354               5.5        0.6343078  12.413350
10        10   16.12685               5.2        0.6844961  10.047140
11        11   15.37804               4.7        0.6649670  13.380375
12        12   29.35527               6.0        0.6509443  11.208923
13        13   23.86151               6.6        0.6844032  10.288208
14        14   17.20264               6.9        0.7339879   6.226028
15        15   20.44898               6.9        0.7348322   4.859171
16        16   13.24914               6.0        0.7324924  10.794031
17        17   18.50871               5.9        0.6594181  13.465871
18        18   16.65846               5.6        0.6989782   7.392565
19        19   24.73959               3.9        0.6940639  10.386728
20        20   15.15103               5.7        0.6562479  15.309261
   doc_prominence topic_coherence topic_exclusivity
1               2       -53.44973          9.803433
2              14       -56.07208          9.784127
3               3       -53.16582          9.787383
4               3       -54.25337          9.780728
```

```
5               9       -56.03671          9.804965
6              12       -39.43652          9.772263
7               6       -57.33341          9.793247
8               6       -57.54004          9.745683
9               1       -57.47233          9.677819
10             11       -70.55090          9.840175
11              2       -64.81430          9.720969
12              7       -54.10457          9.530116
13              5       -41.75937          9.749013
14              7       -57.41718          9.844178
15             15       -32.08849          9.798393
16              3       -72.70258          9.856772
17             10       -42.53650          9.688919
18              0       -75.02035          9.813420
19              4       -52.12748          9.805023
20              2       -44.42644          9.675569
```

[129]:
```
#apply function to eu
k_coherence_eu = find_best_k(eu_done)
```

```
[1] "2topic"
K = 2; V = 1455; M = 6522
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2
  "feel" "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   983.2492               5.1        0.3443919   85.24183
```

```
2         2   471.7508               6.0        0.5202961   79.05595
  doc_prominence topic_coherence topic_exclusivity
1           6292       -32.71533          8.313171
2           5905       -29.83545          9.259871
[1] "5topic"
K = 5; V = 1455; M = 6522
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
   "feel" "appoint"   "medic"    "team"  "referr"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   408.3997               5.7        0.5142016   76.76709
2         2   208.1330               6.7        0.6273681   73.43024
3         3   369.1499               5.5        0.5319961   78.24341
4         4   195.9254               5.7        0.6356060   69.21502
5         5   273.3920               5.6        0.5423586   81.80721
  doc_prominence topic_coherence topic_exclusivity
1           2045       -48.13719          9.039048
2           2668       -46.82270          9.465234
3           2090       -39.43800          8.940810
4           2989       -40.77491          9.491097
5           1562       -43.12502          9.311791
[1] "10topic"
K = 10; V = 1455; M = 6522
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
```

```
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2  Topic 3  Topic 4  Topic 5  Topic 6  Topic 7  Topic 8
"referr" "report" "health"   "feel"   "team"  "medic" "hospit" "servic"
 Topic 9 Topic 10
"assess" "health"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  183.54517               5.5        0.5973659   76.08093
2          2  214.62953               5.3        0.5830000   73.12476
3          3  105.66650               6.0        0.6227464   67.74435
4          4  237.18597               4.8        0.5904611   62.78376
5          5   98.29579               6.1        0.7107656   49.46100
6          6  214.76533               6.3        0.6257529   55.88650
7          7   98.67192               6.8        0.7227994   55.39142
8          8  105.07586               5.9        0.6579295   73.11345
9          9  101.21467               6.3        0.6913064   73.72924
10        10   95.94926               5.7        0.6727760   81.98730
   doc_prominence topic_coherence topic_exclusivity
1             500       -45.26760          9.157342
2             936       -38.43798          9.557729
3             260       -63.97502          9.769136
4             852       -56.95833          9.340132
5             736       -36.81931          9.676800
6             672       -58.12560          9.317491
7             757       -62.38855          9.765144
8             682       -52.54071          9.630848
9             867       -45.50323          9.656173
10            468       -34.23104          9.683166
[1] "15topic"
K = 15; V = 1455; M = 6522
Sampling 500 iterations!
Iteration 25 …
```

```
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "health" "thought"  "report"     "nhs"  "referr"  "memori"    "feel"  "hospit"
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
"appoint"    "team" "therapi"   "medic"  "detail"  "servic"  "servic"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   61.88574               5.2        0.6976180   78.28598
2          2  128.91251               4.9        0.6282679   68.60274
3          3  133.50502               6.6        0.6311056   69.87004
4          4   81.33238               5.7        0.7009852   57.07047
5          5  117.01438               5.2        0.6198188   67.28880
6          6  134.03904               5.2        0.6354009   57.38668
7          7  139.21840               4.0        0.6752396   58.13368
8          8   67.47726               6.7        0.7497750   55.01979
9          9   86.91637               5.6        0.7227788   63.11271
10        10   52.00869               5.3        0.7249005   45.59737
11        11  124.84106               6.1        0.6321154   58.23604
12        12  118.38288               6.4        0.6730509   53.37408
13        13   77.06731               6.0        0.6420691   72.17204
14        14   58.75895               6.1        0.7154186   69.27565
15        15   73.64001               5.9        0.6775727   72.59240
   doc_prominence topic_coherence topic_exclusivity
1             235       -39.43267          9.779445
2             428       -32.47088          9.563276
3             403       -47.09658          9.673916
4             192       -64.87817          9.727579
5             320       -56.76137          9.471305
6             174       -94.78898          9.606194
7             228       -60.41450          9.666320
```

```
8             486       -59.30667          9.817436
9             754       -59.30462          9.677843
10            219       -51.63361          9.832622
11            261       -55.05220          9.525603
12            401       -54.49649          9.746851
13            157       -54.06898          9.729911
14            412       -50.49435          9.782769
15            509       -53.87110          9.716902
[1] "20topic"
K = 20; V = 1455; M = 6522
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
  "hospit" "discharg"  "depress"   "inform"   "referr"     "team"  "support"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
     "nhs"  "current"  "thought"    "medic"     "feel"   "memori"  "therapi"
  Topic 15   Topic 16   Topic 17   Topic 18   Topic 19   Topic 20
  "health"     "team"  "appoint" "lewisham"   "number"     "said"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   52.14607               7.2        0.7640947   47.17154
2          2   64.50628               5.8        0.6713979   69.69404
3          3  105.36523               6.5        0.6649523   51.85220
4          4   93.34477               5.8        0.6359789   59.14039
5          5   63.10978               5.8        0.6809818   69.44009
6          6   46.64127               6.3        0.7223678   58.40572
7          7   71.88156               4.8        0.6773272   61.33673
8          8   68.40709               5.6        0.7170308   61.19099
9          9   83.04023               7.2        0.6384513   68.17759
```

```
10        10   98.59310              5.3        0.6586284   73.12476
11        11   65.01670              5.5        0.7064673   49.25265
12        12  107.54812              4.4        0.6489642   56.61727
13        13  104.71413              5.4        0.6629510   57.43590
14        14   58.59035              7.4        0.6904262   62.58246
15        15   59.99855              6.2        0.6836077   66.71388
16        16   57.83834              5.9        0.7554624   34.74349
17        17   42.00472              5.8        0.7541695   53.46514
18        18   48.28628              6.1        0.7322773   71.18692
19        19   42.29342              5.3        0.7147363   67.22276
20        20  121.67404              5.3        0.6528719   70.67413
   doc_prominence topic_coherence topic_exclusivity
1             380       -65.33554          9.852726
2             134       -46.58980          9.655645
3              99       -55.35855          9.614832
4             122       -70.79731          9.456378
5             241       -58.57547          9.759680
6             292       -71.11377          9.828052
7             127       -54.06027          9.649029
8             139       -61.38045          9.707161
9             111       -59.75890          9.638286
10            318       -35.24986          9.703104
11            109       -60.07117          9.869851
12            222       -59.53715          9.604057
13            129       -88.09963          9.673983
14            364       -58.51115          9.766524
15            236       -72.92932          9.815940
16            455       -50.77848          9.780705
17            299       -56.42504          9.852150
18            357       -50.94335          9.801494
19            202       -33.92282          9.772715
20            316       -45.67954          9.495934
```

[130]:
```
#apply function to caribbean
k_coherence_caribbean = find_best_k(caribbean_done)
```

```
[1] "2topic"
K = 2; V = 1085; M = 809
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
```

```
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2
"current"  "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   722.0174               5.3        0.3445512   30.11588
2         2   362.9826               5.9        0.5175891   31.33070
  doc_prominence topic_coherence topic_exclusivity
1            791       -41.63389          8.589210
2            740       -28.46800          8.939557
[1] "5topic"
K = 5; V = 1085; M = 809
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
  "medic" "appoint"    "team"  "referr"    "feel"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   233.2887               5.7        0.5626207   23.81325
```

```
2         2   159.9518              5.9        0.6372101   26.53953
3         3   181.4143              5.3        0.5963115   29.74011
4         4   191.0051              5.9        0.5519666   31.72524
5         5   319.3400              5.1        0.5268055   27.75438
  doc_prominence topic_coherence topic_exclusivity
1            184       -68.53143          9.172566
2            420       -37.40418          9.456231
3            366       -38.36410          9.283745
4            159       -39.73758          9.301010
5            270       -43.53533          8.800475
[1] "10topic"
K = 10; V = 1085; M = 809
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
 "report"  "detail"  "servic"  "hospit" "patient"    "help"   "medic"  "assess"
  Topic 9  Topic 10
"support"  "referr"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1  163.83595              5.6        0.6033199   27.36848
2          2   92.77558              5.9        0.5920450   32.04717
3          3  102.88778              5.2        0.6670574   23.60583
4          4   74.60487              6.1        0.6890771   22.11493
5          5  126.02312              6.0        0.6270779   22.33580
6          6  137.91260              5.2        0.6435567   18.33695
7          7  105.60625              6.1        0.6376875   17.26340
8          8   83.89560              6.3        0.6565277   25.72098
9          9   92.30786              5.4        0.6238084   27.42159
```

```
10        10  105.15038               5.4        0.6207968   25.88111
   doc_prominence topic_coherence topic_exclusivity
1             105       -44.35983          9.316042
2              35       -38.70833          9.481180
3             127       -42.53065          9.468014
4              96       -56.26783          9.742033
5              42       -75.82709          9.621453
6              48       -58.23380          9.583972
7              95       -55.42401          9.686213
8             149       -42.75783          9.609799
9              37       -40.95768          9.650894
10             61       -49.70595          9.612113
[1] "15topic"
K = 15; V = 1085; M = 809
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5   Topic 6   Topic 7   Topic 8
   "team"  "detail"  "mental"  "report"  "memori"    "team"  "referr"  "hospit"
  Topic 9  Topic 10  Topic 11  Topic 12  Topic 13  Topic 14  Topic 15
   "said" "support"   "medic"  "servic" "appoint"  "servic" "therapi"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   56.77605               5.4        0.7089746   18.92100
2          2   70.99860               6.2        0.6111098   28.19338
3          3   63.13483               5.9        0.6280614   27.02039
4          4   91.85569               5.1        0.6433625   22.56590
5          5   91.41597               5.5        0.6363103   18.42225
6          6   52.44713               5.5        0.6893224   23.31293
7          7   76.20772               5.4        0.6387597   22.47497
```

```
8          8   52.10575              6.6        0.7184136   15.22460
9          9  113.58368              5.8        0.6459614   23.14593
10        10   71.51408              5.1        0.6467448   24.40339
11        11   76.59917              6.6        0.6723924   17.00053
12        12   51.21718              5.7        0.6826139   24.96087
13        13   49.58243              5.3        0.7195243   15.50633
14        14   70.94938              5.6        0.6728816   21.59417
15        15   96.61234              5.6        0.6702611   18.74353
   doc_prominence topic_coherence topic_exclusivity
1              79       -52.61815          9.777821
2              24       -41.44564          9.571422
3              34       -55.30112          9.698313
4              46       -39.76809          9.735140
5              32       -81.35436          9.813742
6              36       -53.93927          9.755731
7              41       -48.23601          9.670003
8              51       -59.81183          9.880824
9              48       -50.28731          9.474063
10             13       -48.60866          9.612161
11             48       -54.37168          9.704471
12             40       -52.27414          9.770228
13             73       -52.43751          9.845143
14             31       -73.41250          9.665054
15             30       -59.48143          9.642785
[1] "20topic"
K = 20; V = 1085; M = 809
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
```

```
   Topic 1    Topic 2    Topic 3    Topic 4    Topic 5    Topic 6    Topic 7
 "support"   "report"   "health"   "detail"  "patient"   "mental"   "servic"
   Topic 8    Topic 9   Topic 10   Topic 11   Topic 12   Topic 13   Topic 14
  "inform"   "hospit"  "increas"   "referr"   "memori"     "said"  "appoint"
  Topic 15   Topic 16   Topic 17   Topic 18   Topic 19   Topic 20
"lewisham"  "therapi"  "contact"    "medic"   "assess"   "famili"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   52.19642               5.3        0.6616787   27.18806
2          2   67.64802               5.0        0.6619637   23.51880
3          3   32.28438               5.7        0.7140306   21.20164
4          4   56.83463               6.0        0.6128334   26.63719
5          5   61.13409               5.1        0.6660353   18.11033
6          6   47.87269               5.8        0.6495395   28.34324
7          7   49.23351               5.4        0.6772737   22.42451
8          8   63.92867               5.9        0.6305617   21.98958
9          9   41.59455               6.6        0.7322521   15.22460
10        10   60.41771               5.5        0.6652636   15.14094
11        11   36.47645               5.6        0.6912404   26.91161
12        12   67.52712               6.2        0.6578781   19.51566
13        13   82.21492               6.1        0.6572195   24.24409
14        14   42.36751               5.7        0.7013404   20.67960
15        15   40.43748               5.8        0.6948787   25.88875
16        16   71.61440               6.0        0.6847101   16.58830
17        17   44.67407               5.2        0.7009859   18.07477
18        18   53.68395               6.5        0.6951390   17.58810
19        19   47.25624               5.9        0.7162096   19.54045
20        20   65.60321               5.4        0.6420970   21.93111
   doc_prominence topic_coherence topic_exclusivity
1              11       -37.25721          9.632223
2              31       -38.28214          9.716484
3              15       -57.92292          9.877761
4              24       -43.64907          9.569334
5              16       -69.42410          9.778915
6              22       -51.13889          9.712801
7              33       -71.04631          9.699721
8              25       -52.56608          9.548840
9              41       -59.81183          9.893532
10             20       -48.37898          9.754622
11             14       -45.57255          9.800655
12             18       -81.62428          9.741162
13             40       -45.17321          9.584577
14             53       -57.22236          9.780864
15             38       -45.95200          9.753164
16             22       -68.81010          9.787582
17              8       -68.02050          9.841163
18             21       -60.20311          9.767333
19             64       -40.08874          9.742267
20              7       -52.25343          9.502338
```

[131]:
```
#apply function to asian
k_coherence_asian = find_best_k(asian_done)
```

```
[1] "2topic"
K = 2; V = 642; M = 361
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
 Topic 1  Topic 2
 "medic" "servic"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   407.4512               5.3        0.3772115   21.80595
2         2   234.5488               6.3        0.5166657   20.40808
  doc_prominence topic_coherence topic_exclusivity
1            354       -35.05833          8.398926
2            343       -34.36578          9.450540
[1] "5topic"
K = 5; V = 642; M = 361
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
```

```
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
  Topic 1   Topic 2   Topic 3   Topic 4   Topic 5
  "medic" "appoint"  "referr"  "assess"  "report"
  topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1         1   152.8713               5.5        0.5855057   15.70928
2         2   101.0989               5.7        0.6571658   13.12429
3         3   107.5189               5.5        0.5671978   20.10636
4         4   100.9259               6.3        0.6269192   18.39245
5         5   179.5849               5.1        0.5449535   17.83331
  doc_prominence topic_coherence topic_exclusivity
1            127       -47.18282          9.485796
2            171       -35.09117          9.560445
3             94       -44.38709          9.524955
4            159       -40.68171          9.479512
5            131       -41.49139          9.246418
[1] "10topic"
K = 10; V = 642; M = 361
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
```

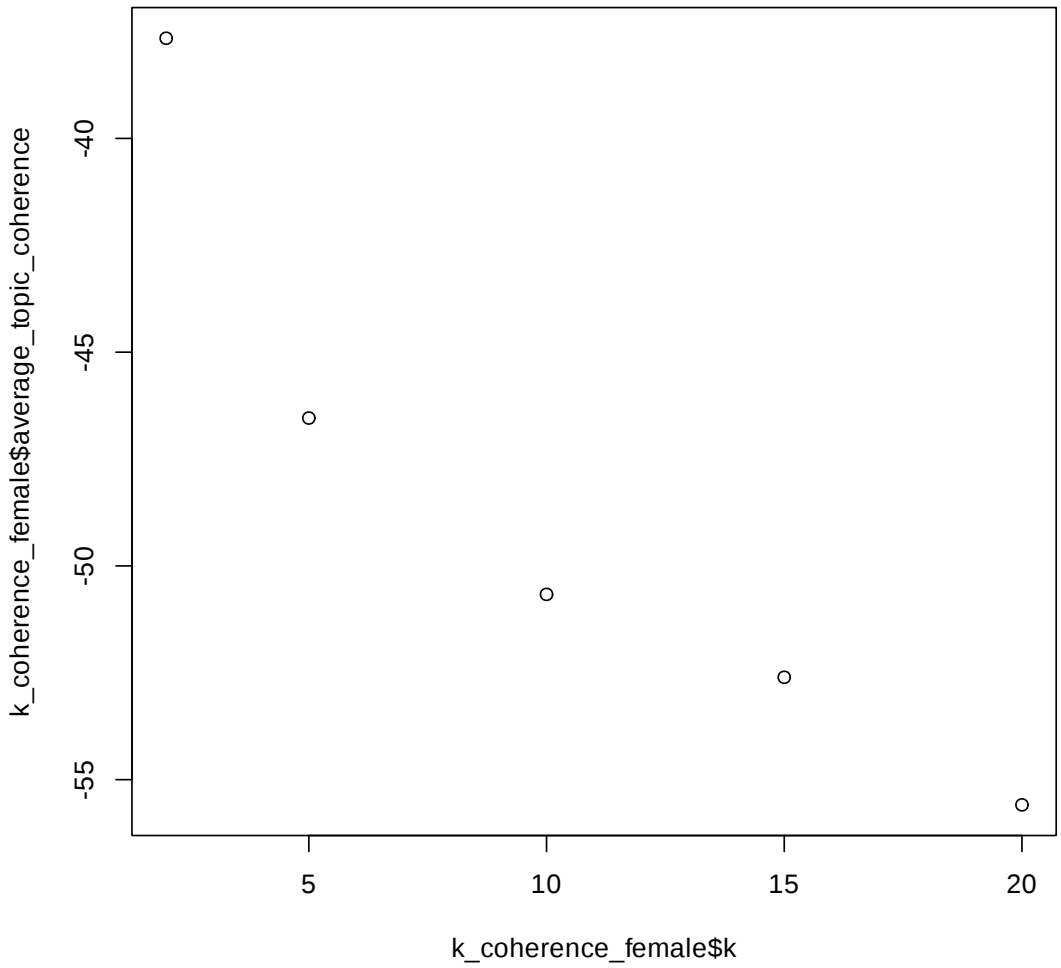
```
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
"interpret"     "medic"      "help"    "servic"    "referr"    "hospit"
    Topic 7     Topic 8     Topic 9    Topic 10
     "feel"   "appoint"    "assess"    "report"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   54.42422               5.4        0.6647900   13.23742
2          2   77.06432               5.3        0.6431673   14.51757
3          3   59.27734               5.6        0.6222283   19.05354
4          4   59.76426               5.2        0.6479466   13.41964
5          5   55.24567               5.6        0.5952429   20.55992
6          6   57.01793               6.1        0.6409835   16.63663
7          7   89.68640               4.2        0.6421721   15.07200
8          8   51.25468               6.2        0.6860128   13.23551
9          9   53.85091               6.2        0.6516236   17.87864
10        10   84.41428               5.8        0.6269445   17.58681
   doc_prominence topic_coherence topic_exclusivity
1              28       -60.31930          9.681201
2              41       -40.17860          9.489017
3              10       -39.74048          9.469544
4              31       -46.11443          9.518884
5              32       -45.91315          9.597938
6              28       -63.24521          9.614445
7              41       -52.54357          9.613321
8              58       -35.27038          9.632214
9              42       -43.05298          9.544199
10             42       -38.87122          9.545147
[1] "15topic"
K = 15; V = 642; M = 361
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
```

```
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
  "current"    "referr" "interpret"  "lewisham"   "current"   "appoint"
    Topic 7     Topic 8     Topic 9    Topic 10    Topic 11    Topic 12
     "feel"     "medic"      "help"    "servic"   "consult"    "health"
   Topic 13    Topic 14    Topic 15
  "thought"   "therapi"   "patient"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   43.04037                6.1        0.6433083  17.190910
2          2   34.79127                5.5        0.6658816  16.794868
3          3   40.78340                6.1        0.7111482  12.921710
4          4   31.20482                5.9        0.7166790  11.567750
5          5   47.19654                5.3        0.6207837  18.294775
6          6   36.10406                6.0        0.6928370  16.472109
7          7   63.89089                4.8        0.6503591  14.671485
8          8   51.37709                5.2        0.6734594  15.466260
9          9   39.02521                5.8        0.6035126  16.991465
10        10   38.56209                6.1        0.6600516  19.268657
11        11   30.43386                6.0        0.7212792   9.129967
12        12   33.29110                5.4        0.6446112  18.623923
13        13   58.10307                5.8        0.6412163  14.947230
14        14   46.55066                5.8        0.6714224  12.896233
15        15   47.64556                5.7        0.6225459  15.363682
   doc_prominence topic_coherence topic_exclusivity
1              19       -48.01424          9.737281
2              24       -52.35933          9.840978
3              30       -80.20764          9.792540
4              10       -52.73645          9.794996
5               9       -45.45590          9.584097
6              34       -43.31374          9.691916
7              26       -47.61677          9.585391
8              16       -40.61391          9.673923
9               3       -57.03068          9.646269
10             25       -32.77811          9.555882
11             20       -42.28709          9.821879
12             12       -39.01326          9.706512
13             13       -47.28721          9.478973
14             10       -59.40342          9.686957
15             12       -75.40991          9.616996
[1] "20topic"
K = 20; V = 642; M = 361
Sampling 500 iterations!
Iteration 25 …
Iteration 50 …
Iteration 75 …
```

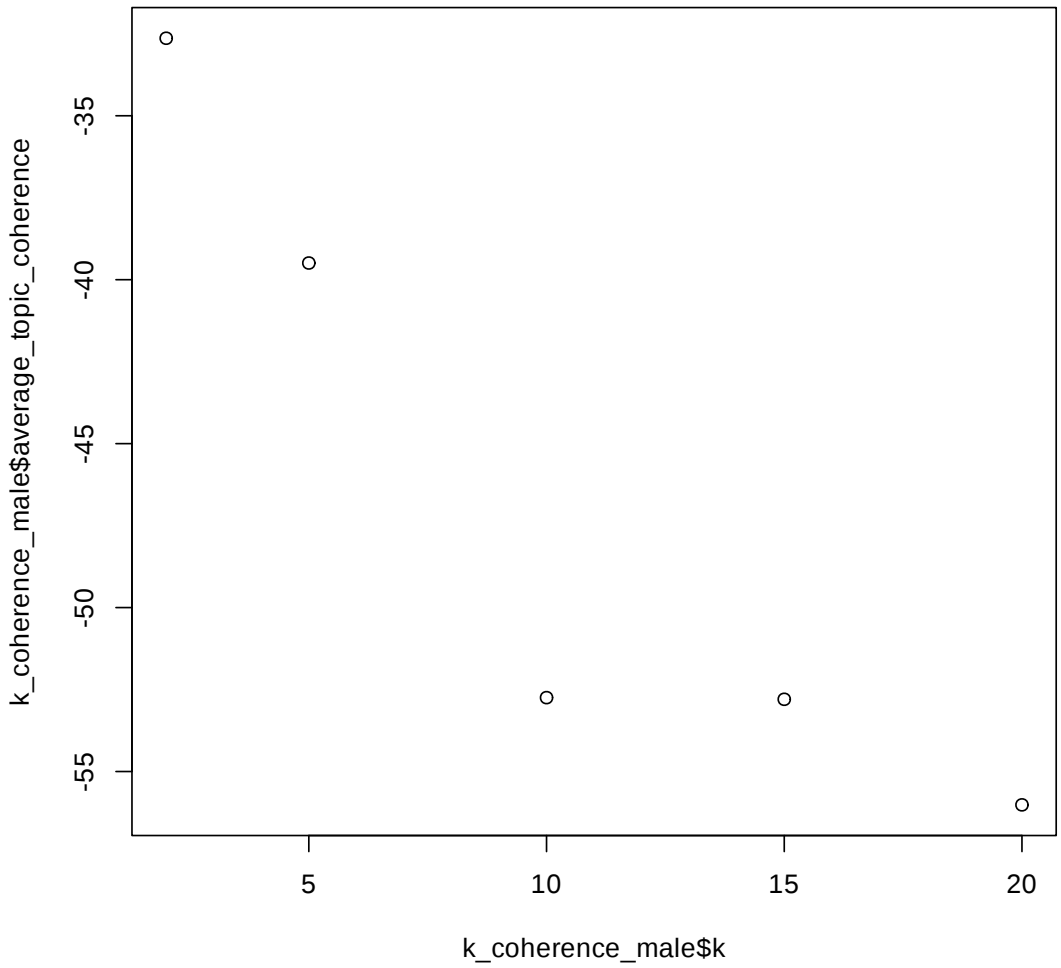
```
Iteration 100 …
Iteration 125 …
Iteration 150 …
Iteration 175 …
Iteration 200 …
Iteration 225 …
Iteration 250 …
Iteration 275 …
Iteration 300 …
Iteration 325 …
Iteration 350 …
Iteration 375 …
Iteration 400 …
Iteration 425 …
Iteration 450 …
Iteration 475 …
Iteration 500 …
Gibbs sampling completed!
    Topic 1     Topic 2     Topic 3     Topic 4     Topic 5     Topic 6
   "referr"    "servic"    "servic"   "thought"   "appoint"      "feel"
    Topic 7     Topic 8     Topic 9    Topic 10    Topic 11    Topic 12
  "current"  "lewisham"      "babi" "interpret"   "consult"     "medic"
   Topic 13    Topic 14    Topic 15    Topic 16    Topic 17    Topic 18
   "mental"    "health"    "report"   "therapi"    "report"   "appoint"
   Topic 19    Topic 20
   "assess"      "help"
   topic_num topic_size mean_token_length dist_from_corpus tf_df_dist
1          1   29.74405               5.8        0.6536809  18.675167
2          2   25.68321               5.6        0.6404860  17.695691
3          3   32.60207               5.3        0.6908358  13.102097
4          4   43.08931               5.7        0.6351010  13.107280
5          5   24.77087               6.6        0.7257575  10.704300
6          6   45.19279               4.2        0.6764208  14.898888
7          7   32.78853               5.5        0.6313745  16.683568
8          8   24.42506               5.9        0.6647471  12.441259
9          9   35.56010               5.6        0.6253514  14.558091
10        10   28.68092               5.1        0.6947377  10.684077
11        11   24.77839               5.7        0.7025972   8.984604
12        12   34.88231               5.1        0.6851501  12.910627
13        13   32.89186               6.3        0.6638876  17.124089
14        14   28.71696               5.9        0.6666598  17.661488
15        15   38.95265               5.5        0.6758890  16.028343
16        16   36.13306               5.8        0.6852794  10.480439
17        17   43.34526               4.8        0.6693281  14.249114
18        18   24.70040               6.4        0.6963530  14.876701
19        19   27.10434               6.1        0.6801673  16.031050
20        20   27.95788               5.6        0.6310012  17.581481
   doc_prominence topic_coherence topic_exclusivity
```

1	17	-51.85880	9.704306
2	7	-53.10805	9.735673
3	11	-74.80940	9.736343
4	10	-51.95804	9.703010
5	15	-64.53025	9.852705
6	13	-49.79651	9.686361
7	4	-45.78688	9.627785
8	14	-58.04855	9.783248
9	8	-68.92299	9.700644
10	17	-56.05104	9.804795
11	19	-47.78431	9.829272
12	6	-47.20006	9.732262
13	17	-47.99113	9.782586
14	15	-47.33515	9.718076
15	9	-51.29826	9.726788
16	9	-70.88832	9.742574
17	8	-49.44556	9.674285
18	14	-52.41220	9.771661
19	19	-52.52470	9.718197
20	3	-47.86888	9.702849

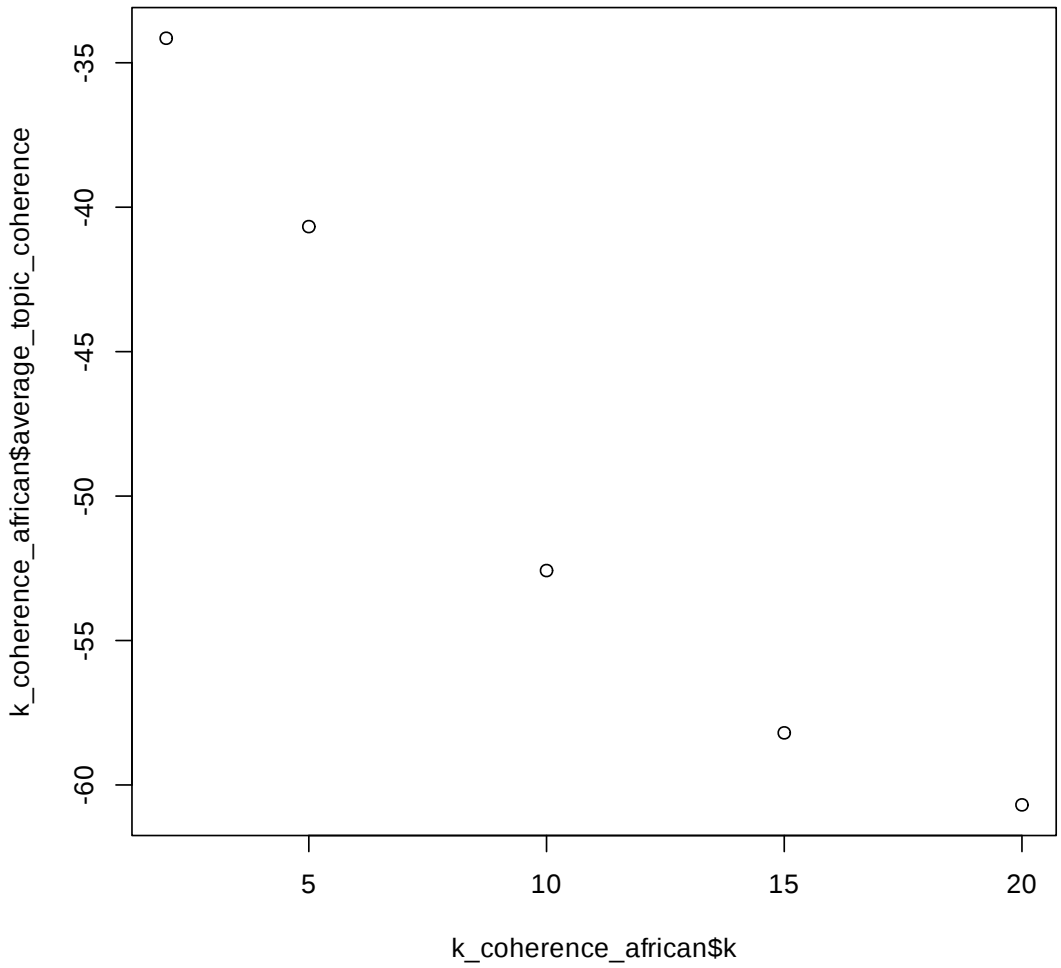
[316]:
```
#to plot and decide what k works best for female
plot(k_coherence_female$k, k_coherence_female$average_topic_coherence)
```



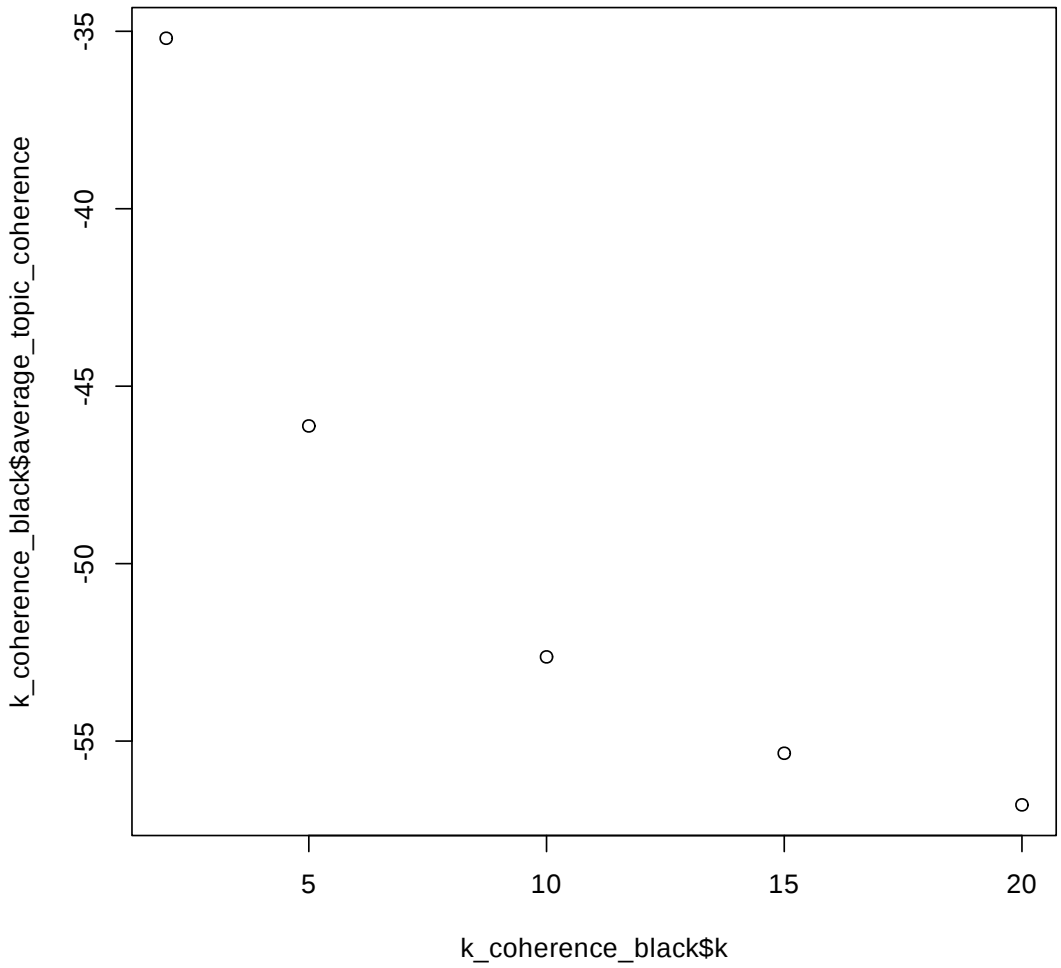
[317]:
```
#to plot and decide what k works best for male
plot(k_coherence_male$k, k_coherence_male$average_topic_coherence)
```



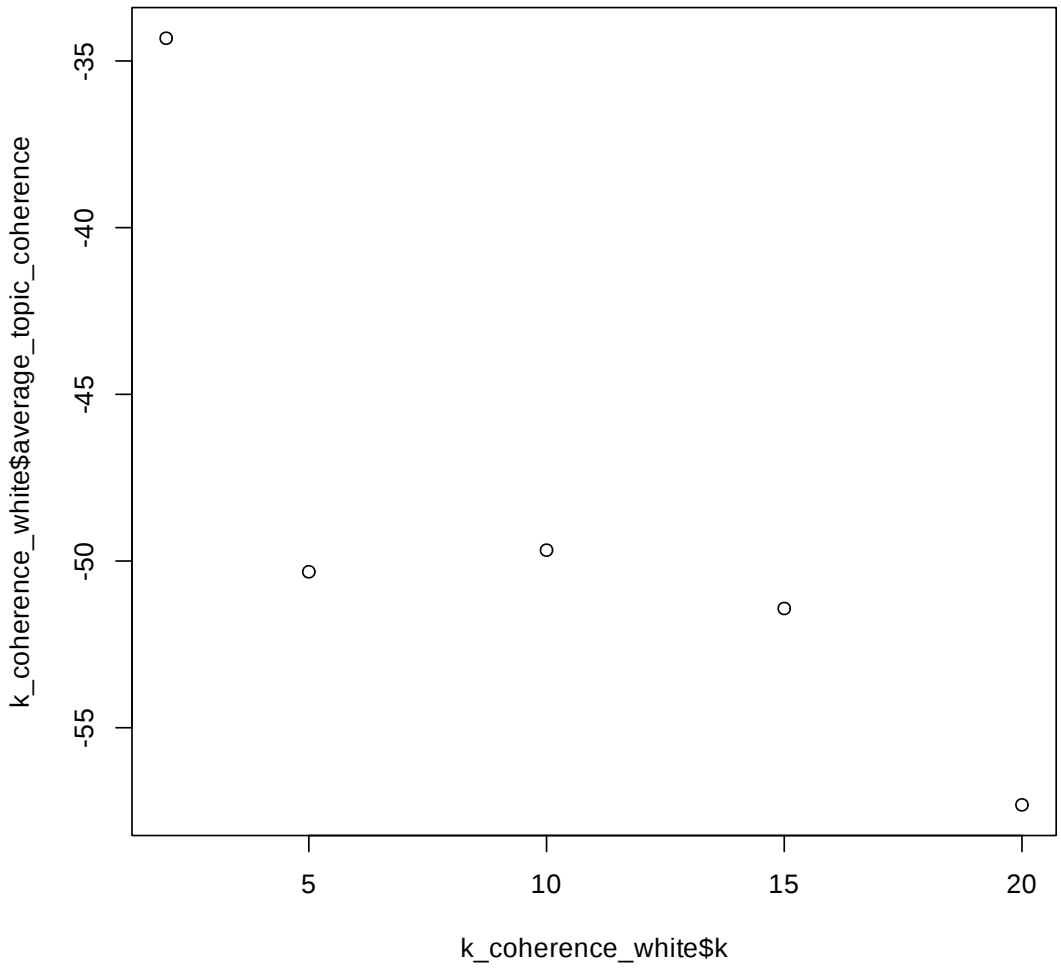
[132]:
```
#to plot and decide what k works best for african
plot(k_coherence_african$k, k_coherence_african$average_topic_coherence)
```



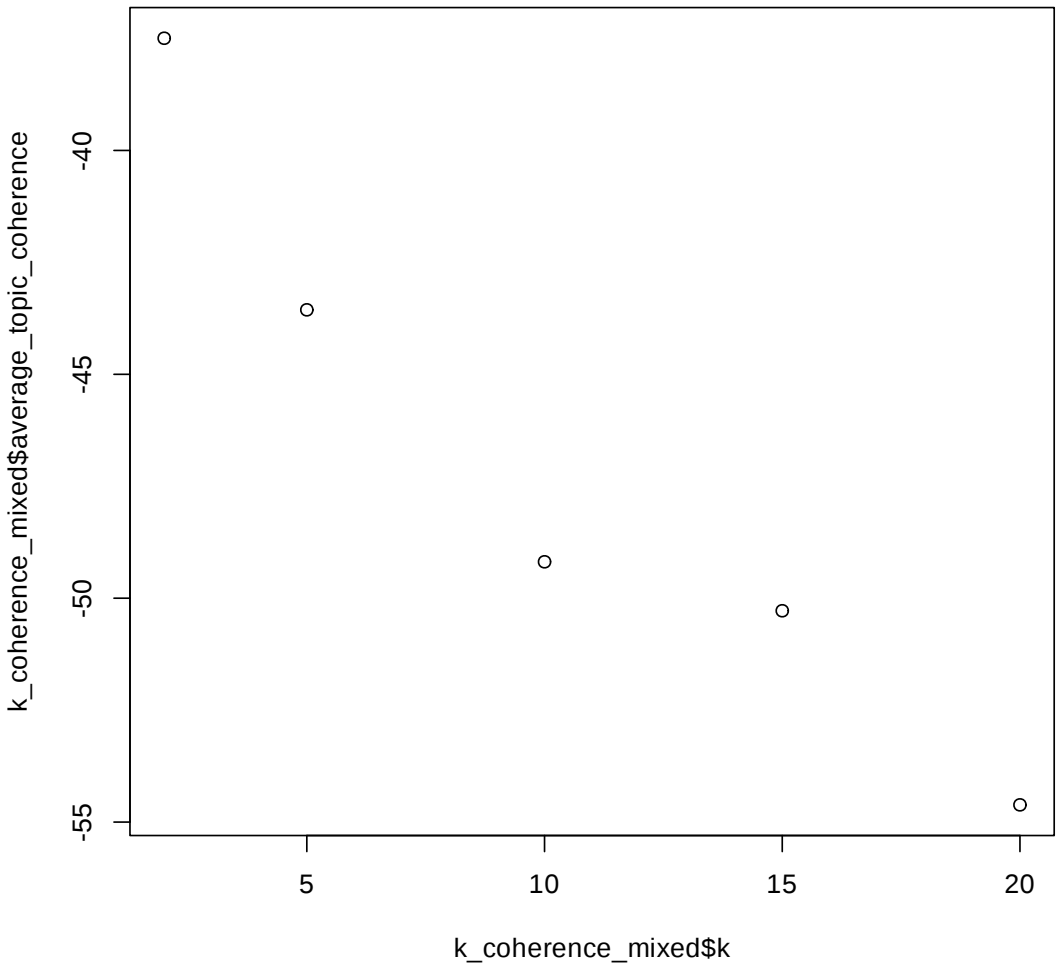
[133]:
```
#to plot and decide what k works best for black
plot(k_coherence_black$k, k_coherence_black$average_topic_coherence)
```



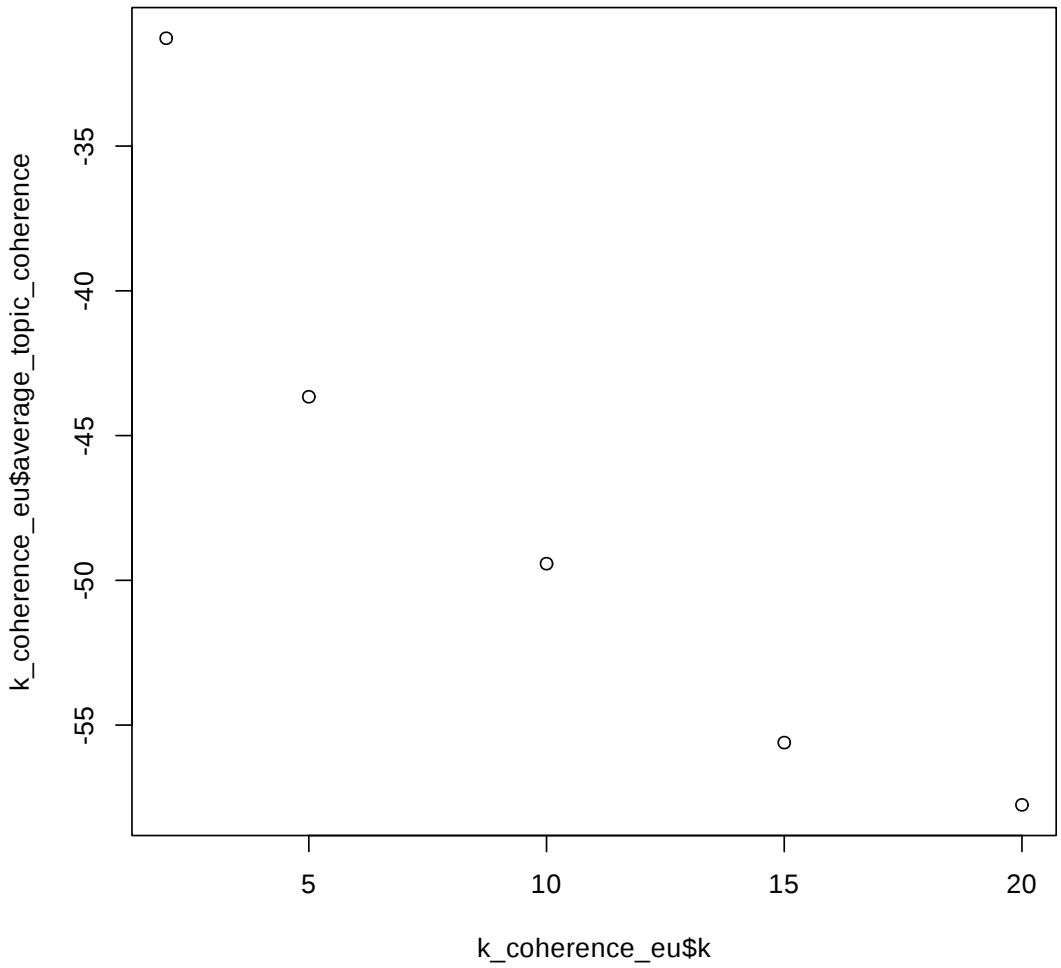
[134]:
```
#to plot and decide what k works best for white
plot(k_coherence_white$k, k_coherence_white$average_topic_coherence)
```



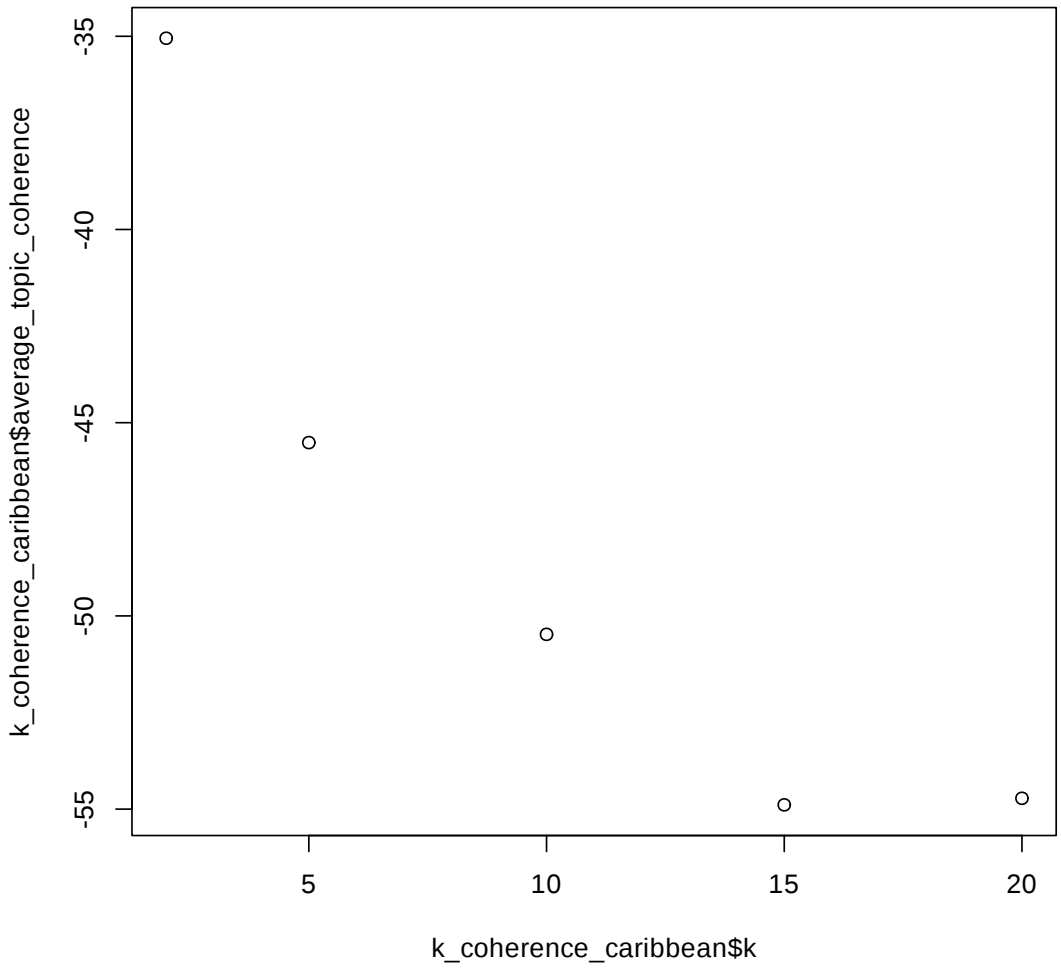
[135]:
```
#to plot and decide what k works best for mixed
plot(k_coherence_mixed$k, k_coherence_mixed$average_topic_coherence)
```



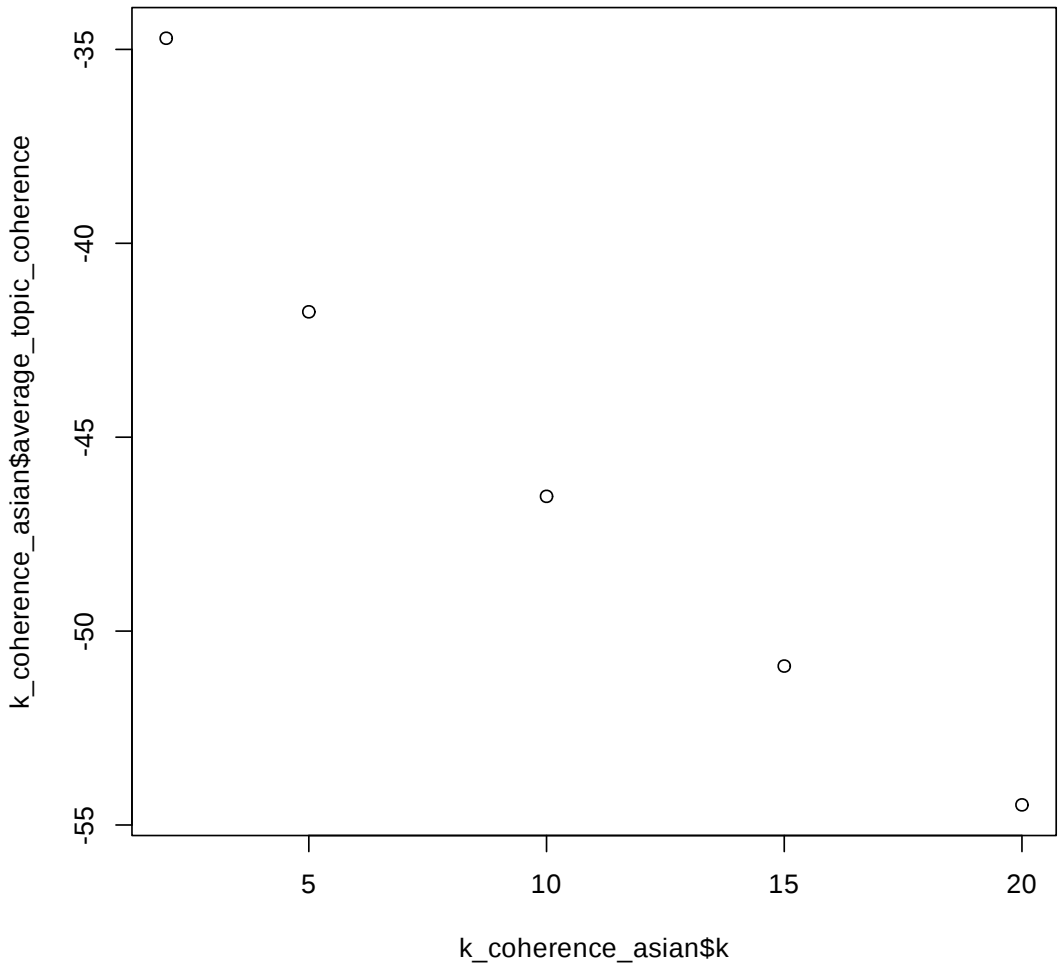
[136]:
```
#to plot and decide what k works best for eu
plot(k_coherence_eu$k, k_coherence_eu$average_topic_coherence)
```



[137]:
```
#to plot and decide what k works best for caribbean
plot(k_coherence_caribbean$k, k_coherence_caribbean$average_topic_coherence)
```



[138]:
```
#to plot and decide what k works best for asian
plot(k_coherence_asian$k, k_coherence_asian$average_topic_coherence)
```



[323]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_female = 2
lda_best_number_female = LDA(female_done, k_best_female, control =␣
 ↪list(verbose=25))
```

```
final e step document 15322
```

[324]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_male = 2
lda_best_number_male = LDA(male_done, k_best_male, control = list(verbose=25))
```

```
final e step document 7023
```

[139]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_african = 2
lda_best_number_african = LDA(african_done, k_best_african, control =␣
 ↪list(verbose=25))
```

```
**** em iteration 25 ****
document 1598
new alpha = 0.28737
final e step document 1598
```

[140]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_black = 2
lda_best_number_black = LDA(black_done, k_best_black, control =␣
 ↪list(verbose=25))
```

```
final e step document 1637
```

[141]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_white = 2
lda_best_number_white = LDA(white_done, k_best_white, control =␣
 ↪list(verbose=25))
```

```
final e step document 1600
```

[142]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_mixed = 2
lda_best_number_mixed = LDA(mixed_done, k_best_mixed, control =␣
 ↪list(verbose=25))
```

```
final e step document 242
```

[143]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_eu = 2
lda_best_number_eu = LDA(eu_done, k_best_eu, control = list(verbose=25))
```

```
final e step document 6522
```

[144]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_caribbean = 2
```

```
lda_best_number_caribbean = LDA(caribbean_done, k_best_caribbean, control =␣
 ↪list(verbose=25))
```

```
final e step document 809
```

[145]:
```
#look at the 100 most frequent terms per topic
#the best number for the topic is 2
k_best_asian = 2
lda_best_number_asian = LDA(asian_done, k_best_asian, control =␣
 ↪list(verbose=25))
```

```
final e step document 361
```

[146]:
```
#set number to 100, in printing the 100 most used terms
f<- data.frame(terms(lda_best_number_female, 100))
```

```
Error in h(simpleError(msg, call)): error in evaluating the argument 'x' in␣
 ↪selecting a method for function 'terms': object 'lda_best_number_female' not␣
 ↪found
Traceback:

1. data.frame(terms(lda_best_number_female, 100))
2. terms(lda_best_number_female, 100)
3. .handleSimpleError(function (cond)
 . .Internal(C_tryCatchHelper(addr, 1L, cond)), "object 'lda_best_number_female␣
 ↪not found",
 .     base::quote(terms(lda_best_number_female, 100)))
4. h(simpleError(msg, call))
```

[337]:
```
options(max.print=200)
print(f [1:50,])
print(f[51:100,])
```

```
      Topic.1      Topic.2
1        team       referr
2      servic      appoint
3      health        medic
4      assess       hospit
5     contact         care
6        time        pleas
7      mental         feel
8       pleas         risk
9      report      support
10       feel       servic
11    appoint       mental
12    current      thought
13    depress       follow
```

14	plan	inform
15	refer	current
16	support	therapi
17	help	detail
18	address	relationship
19	mother	discuss
20	famili	state
21	clinic	sincer
22	concern	anxieti
23	mood	mood
24	psycholog	like
25	number	histori
26	children	need
27	confidenti	nhs
28	need	work
29	perinat	report
30	problem	health
31	inform	dear
32	suicid	miss
33	see	group
34	present	assess
35	patient	consult
36	social	attend
37	email	name
38	medic	perinat
39	telephon	famili
40	lewisham	low
41	review	last
42	work	hous
43	care	clinic
44	continu	babi
45	attend	child
46	person	centr
47	abl	said
48	live	psycholog
49	hour	physic
50	histori	crisi
	Topic.1	Topic.2
51	describ	liaison
52	dear	children
53	first	psychiatrist
54	outpati	engag
55	centr	communiti
56	appear	home
57	make	manag
58	treatment	help
59	well	depress
60	know	number

```
61      howev         note
62      sleep          mrs
63       take         felt
64       name       father
65        dob         deni
66        low         well
67       good       includ
68    possibl   difficulti
69       back       parent
70       home       privat
71       want       person
72   discharg       social
73   diagnosi     telephon
74       made         take
75       past    treatment
76     memori          age
77       emot         past
78   pregnanc        place
79      sever         plan
80       issu      practic
81    liaison       integr
82     privat       receiv
83     detail      increas
84     provid         meet
85     parent       depart
86       sinc         slam
87      offer       friend
88       form        offer
89        get         harm
90       high       requir
91       hous       affect
92       said        other
93      month         wish
94    symptom         life
95     includ      alcohol
96   maudsley        month
97        old     previous
98        new      complet
99   previous       recent
100   medicin         find
```

[339]:
```
#male set number to 100, in printing the 100 most used terms
m<- data.frame(terms(lda_best_number_male, 600))
print(m[1:50,])
print(m[51:100,])
print(m[101:150,])
print(m[151:200,])
```

	Topic.1	Topic.2
1	appoint	report
2	servic	feel
3	pleas	medic
4	team	current
5	assess	risk
6	liaison	thought
7	contact	mood
8	dear	time
9	number	depress
10	clinic	assess
11	telephon	plan
12	care	work
13	consult	health
14	sincer	mental
15	psycholog	suicid
16	health	histori
17	referr	low
18	inform	support
19	time	state
20	attend	said
21	group	referr
22	psychiatrist	person
23	mental	care
24	centr	help
25	hour	need
26	confidenti	present
27	offer	servic
28	support	contact
29	see	refer
30	integr	patient
31	privat	famili
32	medicin	inform
33	hospit	home
34	academ	live
35	therapi	social
36	follow	sleep
37	communiti	discuss
38	nhs	continu
39	map	review
40	letter	treatment
41	crisi	month
42	cag	problem
43	lewisham	describ
44	medic	team
45	duti	past
46	trust	last
47	email	take

48	line	symptom
49	need	well
50	advic	life
	Topic.1	Topic.2
51	meet	previous
52	direct	follow
53	like	physic
54	maudsley	alcohol
55	depart	discharg
56	help	abl
57	arrang	howev
58	call	other
59	treatment	difficulti
60	refer	hospit
61	hous	sinc
62	bring	like
63	write	anxieti
64	north	therapi
65	slam	deni
66	resourc	get
67	detail	manag
68	know	appear
69	discharg	concern
70	complet	name
71	ask	start
72	east	detail
73	switchboard	mother
74	practition	includ
75	lead	recent
76	wish	relationship
77	address	increas
78	administr	engag
79	enclos	good
80	discuss	want
81	thank	hous
82	sent	address
83	hyperlink	self
84	person	issu
85	forward	activ
86	let	harm
87	place	attend
88	practic	felt
89	manag	drug
90	name	friend
91	anxieti	peopl
92	headquart	think
93	request	wife
94	receiv	improv

95	tamworth	talk
96	leaflet	back
97	concern	clinic
98	reablement	disord
99	camberwel	agre
100	confirm	ago

	Topic.1	Topic.2
101	mobil	diagnosi
102	grate	see
103	wwwslamnhsuk	old
104	take	sever
105	lane	normal
106	phone	children
107	unabl	behaviour
108	mood	chang
109	streatham	make
110	client	pleas
111	patient	say
112	form	tri
113	purley	thing
114	outpati	ward
115	coordin	age
116	street	benefit
117	denmark	dob
118	social	stop
119	emerg	end
120	surgeri	admiss
121	note	consid
122	high	find
123	requir	level
124	hope	requir
125	abl	note
126	user	complet
127	suitabl	daili
128	review	consult
129	arriv	memori
130	altern	attempt
131	text	given
132	senior	psycholog
133	wait	signific
134	back	around
135	conveni	made
136	current	prescrib
137	someon	father
138	possibl	poor
139	dob	episod
140	assist	number
141	offic	form

```
142   profession    client
143         find     relat
144       wallac    tablet
145        initi  although
146         juli difficult
147       recept    provid
148      jeanett    school
149      septemb     today
150 somasunderam    experi
       Topic.1    Topic.2
151       book        nhs
152     doctor       term
153    patrick      drink
154       list       told
155       work  sertralin
156 southbrook  experienc
157     novemb      first
158       kind     effect
159   februari       leav
160       copi      adult
161     edridg        act
162  therapist     reason
163      octob      futur
164       unit      ideat
165       make     intent
166     mercer      place
167     messag     factor
168     london      polic
169     access      score
170   lordship      night
171      speak       pain
172   therefor     affect
173     provid       move
174    januari  psychiatr
175     reason   identifi
176       awar    session
177     remind     stress
178      anoth mirtazapin
179       hear        job
180       give     better
181      thoma   appropri
182       west       much
183  psychiatr   selfharm
184     decemb     cognit
185        get      initi
186       nurs      child
187       case      found
188       best       long
```

```
189       well      still
190     worker       evid
191       soon       left
192       plan     return
193    foundat     result
194      locum       voic
195    borough  communiti
196      avail       test
197  sanjeevan      admit
198      hesit     parent
199    dulwich     remain
200  interpret       awar
```

[147]:
```
#african set number to 100, in printing the 100 most used terms
a<- data.frame(terms(lda_best_number_african, 600))
print(a[1:50,])
print(a[51:100,])
#101-200
print(a[101:150,])
print(a[151:200,])
```

```
    Topic.1      Topic.2
1    mental      appoint
2    health         team
3   support        pleas
4      feel       servic
5    report       clinic
6  children       assess
7   current         dear
8    servic      contact
9    referr       hospit
10    medic       sincer
11   assess      liaison
12     need    psycholog
13     care      consult
14   famili       number
15  thought       inform
16     time         time
17     plan       health
18     risk     telephon
19  depress       mental
20     mood         care
21    state      perinat
22     team   confidenti
23  contact psychiatrist
24  histori        medic
25     name         hour
26   inform          see
```

```
27   social      medicin
28   detail        group
29     work       attend
30     help       privat
31     live     lewisham
32      low        offer
33   person        centr
34    child    communiti
35     home      patient
36  present       letter
37    pleas       follow
38  concern       depart
39  address        trust
40    sleep       academ
41     well          map
42   mother    treatment
43     said          cag
44    refer         help
45     hous      denmark
46   parent         slam
47   suicid        crisi
48   includ      outpati
49      abl        north
50     babi       memori
         Topic.1      Topic.2
51          take       referr
52       continu          nhs
53       symptom     discharg
54  relationship        write
55      previous       arrang
56       describ       integr
57          past         line
58        attend         miss
59        follow         east
60        review         duti
61         howev         like
62           dob      support
63       discuss      therapi
64        appear    administr
65        provid     maudsley
66        physic         hous
67         engag         call
68          sinc    headquart
69          deni       direct
70    difficulti         meet
71          last         need
72         manag   psychiatri
73         other  switchboard
```

74	problem	refer
75	month	bring
76	email	wwwslamnhsuk
77	harm	wing
78	want	coordin
79	discharg	advic
80	get	note
81	old	review
82	first	king
83	therapi	high
84	start	serum
85	hospit	discuss
86	recent	anxieti
87	like	place
88	complet	camberwel
89	friend	colleg
90	number	forward
91	diagnosi	thank
92	requir	mood
93	life	resourc
94	good	ask
95	treatment	know
96	perinat	surgeri
97	sever	practition
98	voic	wish
99	psychot	practic
100	anxieti	hope

	Topic.1	Topic.2
101	make	street
102	patient	email
103	self	park
104	difficult	take
105	centr	possibl
106	daughter	normal
107	agre	sent
108	son	confirm
109	talk	juli
110	made	person
111	issu	grate
112	receiv	enclos
113	admiss	lane
114	increas	lead
115	signific	west
116	experi	let
117	father	unabl
118	school	back
119	pregnanc	best
120	husband	rearrang

```
121        see   schedul
122      peopl      find
123       form     blood
124       back   concern
125      thing     cheyn
126    alcohol     phone
127     around      nurs
128  sertralin      test
129       find    requir
130     client      ward
131     improv      unit
132  experienc   request
133     reason      soon
134       poor streatham
135      activ     hesit
136       felt    novemb
137       emot   januari
138     impact     thoma
139        say      valu
140      place    social
141       told     emerg
142       meet    doctor
143       leav     first
144    section   suitabl
145      daili  februari
146  psychiatr     offic
147       abus      make
148      adult    decemb
149 understand     level
150    benefit      text
      Topic.1   Topic.2
151     chang      venu
152     think    receiv
153     relat     manag
154     earli hyperlink
155       age      give
156     group     arriv
157    remain       dob
158    involv   complet
159      note  tamworth
160    stress     count
161     visit   conveni
162     birth    recept
163     given      kind
164       nhs     anoth
165    situat    detail
166      move      user
167     worri   leaflet
```

```
168 communiti    senior
169    worker  recoveri
170    episod      feel
171   partner  therefor
172  telephon   septemb
173       end     local
174 behaviour     picup
175  appropri       mrs
176  although      work
177      drug    action
178      nurs    messag
179      awar    unless
180    consid       via
181      stop      page
182       tri       nov
183     carer      form
184     addit  lordship
185   consent      last
186      pain       abl
187   possibl    someon
188      stay    assist
189    disord    altern
190   develop     initi
191   protect      well
192     today       get
193    period     octob
194      post   pearson
195  accommod      name
196   chaucer       ann
197      term   dulwich
198      much     mobil
199    access      hear
200   request   current
```

[148]:
```
#black set number to 100, in printing the 100 most used terms
b<- data.frame(terms(lda_best_number_black, 600))
print(b[1:50,])
print(b[51:100,])
#101-200
print(b[100:150,])
print(b[151:200,])
```

```
       Topic.1      Topic.2
1       appoint         feel
2        servic       report
3          team       health
4         pleas       mental
5        assess      current
```

6	dear	medic
7	contact	support
8	health	referr
9	sincer	risk
10	clinic	time
11	hospit	mood
12	psycholog	famili
13	number	assess
14	mental	thought
15	liaison	servic
16	telephon	depress
17	centr	work
18	time	plan
19	confidenti	mother
20	attend	name
21	perinat	care
22	lewisham	histori
23	privat	person
24	consult	team
25	care	state
26	therapi	detail
27	inform	need
28	offer	inform
29	group	children
30	depart	contact
31	hour	low
32	outpati	pleas
33	support	help
34	psychiatrist	social
35	follow	patient
36	medicin	relationship
37	referr	child
38	see	concern
39	integr	present
40	meet	well
41	need	live
42	administr	address
43	miss	refer
44	like	said
45	letter	home
46	nhs	past
47	academ	suicid
48	resourc	previous
49	arrang	continu
50	crisi	take

	Topic.1	Topic.2
51	call	includ
52	write	follow

53	trust	sinc
54	discharg	hous
55	wwwslamnhsuk	discuss
56	know	problem
57	switchboard	other
58	bring	howev
59	duti	like
60	refer	month
61	discuss	sleep
62	medic	describ
63	slam	babi
64	communiti	hospit
65	line	want
66	help	treatment
67	direct	abl
68	hous	last
69	let	engag
70	high	get
71	wish	attend
72	denmark	symptom
73	map	issu
74	complet	discharg
75	psychiatri	good
76	possibl	harm
77	wing	father
78	email	appear
79	maudsley	manag
80	streatham	anxieti
81	street	parent
82	ask	deni
83	grate	physic
84	best	felt
85	treatment	complet
86	hyperlink	difficulti
87	headquart	number
88	forward	review
89	note	old
90	practic	dob
91	ann	therapi
92	septemb	life
93	soon	make
94	juli	requir
95	place	peopl
96	tamworth	recent
97	practition	self
98	king	start
99	enclos	agre
100	advic	client

	Topic.1	Topic.2
100	advic	client
101	thank	ward
102	hesit	see
103	therapist	age
104	concern	provid
105	octob	thing
106	cag	form
107	abl	experi
108	colleg	abus
109	unit	behaviour
110	schedul	email
111	chaucer	think
112	unabl	admiss
113	rearrang	diagnosi
114	hope	session
115	phone	pregnanc
116	review	first
117	thoma	sever
118	park	talk
119	novemb	partner
120	receiv	chang
121	confirm	activ
122	senior	episod
123	conveni	school
124	recept	adult
125	wait	consid
126	leaflet	disord
127	back	difficult
128	surgeri	around
129	univers	alcohol
130	sent	reason
131	social	communiti
132	address	increas
133	someon	experienc
134	anxieti	group
135	initi	note
136	coordin	leav
137	arriv	clinic
138	east	find
139	find	friend
140	januari	birth
141	take	receiv
142	emerg	nhs
143	therefor	drug
144	akanuma	involv
145	detail	made
146	lane	improv

```
147      hear       tri
148      feel       son
149     anoth  although
150     north     relat
        Topic.1   Topic.2
151   specialist      back
152        first      emot
153         venu psycholog
154     februari   benefit
155         mood     given
156         form  signific
157         make     level
158        cheyn  daughter
159         kind     black
160         nurs psychiatr
161 psychologist     place
162          dob   section
163         case     earli
164       provid   protect
165         well    normal
166         cipt      meet
167       nozomi  selfharm
168         work      give
169       decemb      stop
170       altern     centr
171         home  telephon
172       person  prescrib
173   profession      awar
174       ground     futur
175      request   perinat
176   rotherhith      much
177   southbrook       end
178         last   respons
179      ladywel     initi
180       purley     visit
181         list       ago
182       parent      told
183         copi  appropri
184       longer    sister
185        manag safeguard
186        visit       act
187         moss       ask
188      patient   regular
189      brixton      nurs
190      suitabl    period
191          via   struggl
192    psychiatr       say
193       andrew   psychot
```

```
194         lead     daili
195         awar     found
196       requir    worker
197       report      miss
198         iptt    record
199    necessari    memori
200        cours      poor
```

[149]:
```
#white set number to 100, in printing the 100 most used terms
w<- data.frame(terms(lda_best_number_white, 600))
print(w[1:50,])
print(w[51:100,])
#101-200
print(w[100:150,])
print(w[151:200,])
```

```
        Topic.1      Topic.2
1          feel      appoint
2        report        pleas
3        health       servic
4        mental         team
5       current      contact
6         medic       assess
7       depress         dear
8        assess       clinic
9        referr    interpret
10       famili       number
11         mood         time
12         time       hospit
13      histori       sincer
14      support       health
15       servic    psycholog
16         risk       inform
17      thought   confidenti
18         care       attend
19         plan       mental
20          low      perinat
21         work      liaison
22        state     lewisham
23       detail       privat
24         said     telephon
25         need       depart
26     children      consult
27       mother        centr
28         name         hour
29         help      outpati
30         team      medicin
31         live        offer
```

32	present	care
33	social	need
34	suicid	therapi
35	inform	group
36	relationship	integr
37	person	book
38	contact	see
39	refer	administr
40	problem	psychiatrist
41	well	client
42	patient	follow
43	sleep	direct
44	month	email
45	past	help
46	dob	arrang
47	describ	referr
48	home	academ
49	address	crisi
50	discuss	refer

	Topic.1	Topic.2
51	review	line
52	continu	write
53	concern	denmark
54	pleas	letter
55	last	slam
56	hospit	nhs
57	parent	cag
58	previous	know
59	symptom	possibl
60	anxieti	like
61	child	support
62	take	meet
63	treatment	wing
64	memori	discuss
65	includ	maudsley
66	follow	map
67	babi	ask
68	difficulti	trust
69	like	complet
70	therapi	note
71	abl	call
72	sinc	bring
73	good	duti
74	attend	medic
75	start	address
76	get	confirm
77	manag	hous
78	howev	resourc

79	physic	treatment
80	deni	psychiatri
81	appear	discharg
82	other	wish
83	life	street
84	daughter	communiti
85	want	patient
86	alcohol	let
87	husband	requir
88	engag	detail
89	old	high
90	first	receiv
91	chang	soon
92	number	mrs
93	discharg	king
94	recent	enclos
95	father	request
96	hous	memori
97	agre	name
98	nhs	unit
99	normal	forward
100	mrs	colleg

	Topic.1	Topic.2
100	mrs	colleg
101	felt	user
102	clinic	miss
103	issu	unabl
104	self	north
105	age	wwwslamnhsuk
106	partner	provid
107	perinat	venu
108	see	concern
109	diagnosi	grate
110	talk	park
111	increas	thank
112	harm	advic
113	back	someon
114	psychiatr	practic
115	complet	languag
116	improv	find
117	ago	switchboard
118	centr	arriv
119	email	leaflet
120	think	novemb
121	consult	person
122	friend	schedul
123	requir	anoth
124	sever	form

125	pregnanc	practition
126	level	take
127	form	hyperlink
128	find	profession
129	reason	iptt
130	difficult	copi
131	psycholog	rearrang
132	emot	speak
133	provid	avail
134	admiss	streatham
135	experienc	prefer
136	move	sent
137	disord	place
138	daili	recept
139	behaviour	altern
140	tablet	hope
141	birth	cours
142	thing	januari
143	episod	wait
144	signific	thoma
145	activ	emerg
146	made	abl
147	son	conveni
148	abus	lead
149	make	octob
150	remain	best
	Topic.1	Topic.2
151	benefit	feedback
152	relat	psychotherapi
153	consid	surgeri
154	sertralin	make
155	school	cheyn
156	cognit	juli
157	group	anxieti
158	drug	coordin
159	score	headquart
160	adult	septemb
161	prescrib	therapist
162	poor	visit
163	communiti	decemb
164	awar	februari
165	period	feel
166	affect	phone
167	given	mood
168	tri	suitabl
169	stop	univers
170	peopl	ann
171	around	question

```
172        new          code
173       note    southbrook
174      initi      tamworth
175      visit          hear
176       leav          case
177       pain       ladywel
178      today        wallac
179       told          text
180   identifi          work
181   telephon       jeanett
182    section       central
183   although     camberwel
184    practic       session
185     sister        gender
186   selfharm        doctor
187       test       adamson
188       high          home
189       ward  questionnair
190    protect          give
191     record          east
192       much        ground
193   appropri        messag
194       post    understand
195       evid         hesit
196        say          nurs
197 understand          list
198      carer           new
199       term          dept
200    session         offic
```

[150]:
```
#mixed set number to 100, in printing the 100 most used terms
mix<- data.frame(terms(lda_best_number_mixed, 600))
print(mix[1:50,])
print(mix[51:100,])
#101-200
print(mix[100:150,])
print(mix[151:200,])
```

```
       Topic.1    Topic.2
1       servic      pleas
2      contact     assess
3      appoint       care
4       health    support
5         plan     mental
6         time       team
7         team       feel
8         need     hospit
9       person     attend
```

10	centr	referr
11	referr	psycholog
12	concern	number
13	mental	current
14	risk	therapi
15	like	social
16	support	health
17	liaison	follow
18	see	medic
19	clinic	help
20	privat	servic
21	hospit	mood
22	famili	clinic
23	inform	dear
24	live	hous
25	meet	email
26	detail	appoint
27	work	lewisham
28	sincer	work
29	problem	low
30	thought	refer
31	name	address
32	current	inform
33	report	trust
34	present	depart
35	depress	previous
36	histori	client
37	treatment	provid
38	dear	perinat
39	continu	telephon
40	dob	detail
41	hour	depress
42	month	need
43	medicin	difficulti
44	pleas	name
45	discuss	mother
46	mother	nhs
47	relationship	offer
48	call	sincer
49	assess	child
50	consult	possibl

	Topic.1	Topic.2
51	place	state
52	manag	home
53	state	report
54	confidenti	unabl
55	integr	communiti
56	help	manag

57	patient	confidenti
58	requir	adamson
59	recent	direct
60	adult	back
61	life	friend
62	said	consult
63	feel	risk
64	discharg	contact
65	line	start
66	telephon	complet
67	good	histori
68	number	birth
69	get	parent
70	suicid	switchboard
71	make	review
72	emerg	wing
73	abl	form
74	letter	felt
75	made	thoma
76	ask	surgeri
77	well	suicid
78	hous	children
79	nhs	time
80	address	describ
81	last	famili
82	crisi	write
83	mood	group
84	medic	practition
85	parent	agre
86	review	self
87	offer	advic
88	group	well
89	wish	tri
90	futur	said
91	forward	local
92	give	past
93	sleep	centr
94	thoma	relationship
95	know	symptom
96	reason	kind
97	duti	last
98	psychiatrist	includ
99	westminst	peopl
100	confirm	harm

	Topic.1	Topic.2
100	confirm	harm
101	perinat	abl
102	thing	iptt

103	lewisham	return
104	includ	denmark
105	slam	take
106	miss	first
107	describ	understand
108	take	met
109	southbrook	liaison
110	administr	person
111	first	anxieti
112	children	worker
113	benefit	call
114	academ	see
115	next	relat
116	appropri	hour
117	hope	engag
118	express	unit
119	note	difficult
120	session	develop
121	cognit	find
122	situat	disord
123	harm	arrang
124	low	post
125	profession	crisi
126	physic	around
127	diagnosi	although
128	email	request
129	talk	talk
130	outpati	abus
131	write	psychiatrist
132	son	receiv
133	might	want
134	other	thank
135	cope	meet
136	told	administr
137	wing	other
138	maudsley	mobil
139	conveni	king
140	end	sinc
141	activ	phone
142	bridg	headquart
143	colleg	ann
144	sinc	best
145	given	practic
146	remain	discuss
147	want	make
148	receiv	colleg
149	novemb	made
150	anxieti	check

	Topic.1	Topic.2
151	home	resourc
152	close	recent
153	difficulti	think
154	age	get
155	arrang	ground
156	sertralin	partner
157	appear	initi
158	someon	suffer
159	past	sleep
160	old	slam
161	resourc	soon
162	attend	drug
163	map	psychiatr
164	cag	issu
165	increas	thought
166	north	flat
167	practic	memori
168	psychiatri	awar
169	ill	monitor
170	daili	found
171	around	coordin
172	scutari	record
173	juli	improv
174	leav	longer
175	copi	lower
176	today	activ
177	think	howev
178	respons	good
179	client	session
180	case	note
181	issu	treatment
182	februari	selfharm
183	abus	job
184	neuropsychiatri	increas
185	new	addit
186	summari	today
187	access	honor
188	therefor	discharg
189	impact	attempt
190	effect	sever
191	stress	consid
192	experienc	new
193	back	neuropsychiatri
194	partner	signific
195	period	therapist
196	view	case
197	involv	given

```
198         alcohol             dob
199       psycholog             end
200           howev            term
```

[151]:
```
#eu set number to 100, in printing the 100 most used terms
e<- data.frame(terms(lda_best_number_eu, 600))
print(e[1:50,])
print(e[51:100,])
#101-200
print(e[100:150,])
print(e[151:200,])
```

```
      Topic.1   Topic.2
1      servic    assess
2     appoint   current
3      health      care
4        feel    hospit
5      mental      team
6      referr     pleas
7        team    report
8       pleas     medic
9     thought treatment
10       time    sincer
11    support      mood
12    therapi      time
13    contact   depress
14      medic   contact
15       risk      need
16      group      dear
17      refer      work
18     clinic   histori
19       plan    mother
20     report      hour
21  psycholog    inform
22     number      home
23     detail      said
24      offer      past
25     inform       low
26     attend   concern
27       help psycholog
28    patient    social
29     assess   discuss
30     famili   problem
31     recent    clinic
32       mood  telephon
33       good      live
34 confidenti    servic
35    address      help
```

```
36       take   medicin
37    consult   continu
38      sleep    attend
39     follow    mental
40     includ      feel
41       name      plan
42        abl    review
43    describ   support
44      state    follow
45       know     state
46      email      like
47    liaison     centr
48       work   consult
49  communiti       see
50     person   perinat
        Topic.1      Topic.2
51      depress     lewisham
52          get       health
53      anxieti       letter
54  psychiatrist         well
55       suicid      liaison
56  relationship       suicid
57      complet     discharg
58     maudsley       academ
59         like         find
60        other         hous
61      present         self
62         miss       person
63          see        engag
64        centr        write
65      histori      appoint
66       integr       number
67     discharg      anxieti
68          nhs       privat
69       disord          nhs
70         deni       famili
71        month      present
72          dob         back
73         well         harm
74       privat         risk
75          ask         talk
76         life        manag
77      session       memori
78       receiv         felt
79        trust        activ
80         note       parent
81       physic         make
82      symptom       provid
```

83	need	last
84	thing	direct
85	meet	children
86	emot	requir
87	dear	sinc
88	lewisham	previous
89	discuss	howev
90	manag	first
91	difficulti	relationship
92	crisi	referr
93	arrang	difficulti
94	children	appropri
95	last	mrs
96	relat	name
97	howev	take
98	form	confidenti
99	line	alcohol
100	made	place

	Topic.1	Topic.2
100	made	place
101	telephon	crisi
102	previous	case
103	improv	sever
104	hous	prescrib
105	appear	psychiatrist
106	chang	start
107	unabl	psychiatr
108	want	detail
109	although	abl
110	selfharm	symptom
111	end	given
112	today	old
113	depart	administr
114	low	benefit
115	care	period
116	initi	new
117	drug	month
118	father	therapi
119	practic	depart
120	stress	increas
121	friend	ago
122	age	normal
123	issu	call
124	consid	enclos
125	start	high
126	agre	east
127	come	peopl
128	sister	street

129	stop	slam
130	possibl	wish
131	better	episod
132	next	best
133	sinc	map
134	difficult	email
135	think	drink
136	duti	meet
137	increas	unit
138	behaviour	outpati
139	cognit	therefor
140	call	refer
141	say	reason
142	review	thank
143	tri	futur
144	admiss	patient
145	emerg	abus
146	awar	ward
147	signific	client
148	long	school
149	resourc	told
150	access	advic

	Topic.1	Topic.2
151	diagnosi	within
152	perinat	respons
153	hope	diagnosi
154	around	address
155	tamworth	forward
156	babi	iapt
157	sertralin	attempt
158	client	partner
159	post	dob
160	child	agre
161	adult	denmark
162	old	understand
163	explain	think
164	regular	psychiatri
165	remain	prior
166	protect	london
167	north	factor
168	record	want
169	outpati	level
170	keep	cag
171	alcohol	experienc
172	novemb	possibl
173	often	point
174	first	form
175	slam	juli

```
176      coordin     friend
177     identifi       give
178          act       note
179          map      advis
180        local    counsel
181       affect    practic
182      brother     experi
183          cag        ill
184      partner        get
185        offic       lead
186       experi    describ
187      surgeri      child
188      suggest    complet
189      summari      happi
190        bring    confirm
191        birth       term
192     research    septemb
193  switchboard    anxious
194          eat      trust
195 psychologist     return
196       requir     affect
197        speak  intervent
198         quit    surgeri
199      psychot    request
200        peopl       poor
```

[152]:
```
#caribbean set number to 100, in printing the 100 most used terms
c<- data.frame(terms(lda_best_number_caribbean, 600))
print(c[1:50,])
print(c[51:100,])
#101-200
print(c[101:150,])
print(c[151:200,])
```

```
      Topic.1    Topic.2
1        team      pleas
2      health     mental
3     current     assess
4        care     servic
5      servic    appoint
6     support      medic
7      inform       need
8     contact     famili
9      clinic     attend
10     report     referr
11       feel       time
12     referr       name
13       risk    depress
```

```
14     number   children
15        low     hospit
16    patient       plan
17       live    thought
18     sincer   telephon
19     detail       said
20    present       work
21     review  psycholog
22       time       hous
23        nhs    contact
24 confidenti       feel
25     person       mood
26    consult     social
27    anxieti    perinat
28       help     mother
29      medic       take
30      state     detail
31       mood      centr
32     assess       dear
33        abl     health
34     follow    histori
35      refer     includ
36    appoint     memori
37    histori      refer
38   lewisham       last
39     requir    support
40    liaison       back
41      group  communiti
42        get     report
43       dear    problem
44  treatment       like
45       hour     number
46     tablet   daughter
47    address      month
48    concern difficulti
49    perinat        dob
50    discuss    address
        Topic.1      Topic.2
51          see psychiatrist
52         good     previous
53        trust         help
54         miss        manag
55         home      describ
56       suicid        howev
57        sleep      complet
58       letter        offer
59         line          see
60       continu         agre
```

```
61       increas      concern
62       therapi      liaison
63         child     lewisham
64         centr         past
65  relationship       physic
66          well         home
67      discharg      therapi
68         activ        visit
69       possibl      discuss
70          like       privat
71          call         well
72          plan      partner
73      diagnosi        other
74          harm         meet
75        depart      medicin
76         email         babi
77          form         hour
78     administr         note
79          meet        state
80          make         want
81          life       direct
82          deni        group
83          slam relationship
84         crisi         risk
85         initi        email
86        improv       person
87        arrang       suicid
88         today         sinc
89          talk       follow
90        social        think
91         birth         pain
92       practic         awar
93         unabl          old
94          issu          son
95         mobil       client
96          nurs       inform
97         carer     discharg
98       problem        score
99        hospit   profession
100 wwwslamnhsuk        first
       Topic.1    Topic.2
101     receiv     father
102     mental     parent
103       made     colleg
104        ask     provid
105     normal      adult
106    depress     school
107    outpati    denmark
```

108	find	post
109	next	practic
110	map	sertralin
111	experienc	mrs
112	affect	poor
113	write	level
114	test	benefit
115	high	issu
116	self	child
117	give	integr
118	jamaica	find
119	manag	ask
120	recent	continu
121	experi	say
122	appropri	bring
123	practition	none
124	record	know
125	academ	chang
126	enter	tri
127	engag	psychiatr
128	period	coordin
129	place	maudsley
130	dob	respons
131	work	reason
132	said	form
133	appear	access
134	phone	enclos
135	identifi	crisi
136	start	close
137	felt	come
138	pregnant	sincer
139	emot	treatment
140	struggl	symptom
141	psychiatr	move
142	result	reablement
143	pleas	age
144	first	ill
145	past	receiv
146	symptom	appear
147	chang	recent
148	worker	miss
149	reason	given
150	provid	abus

	Topic.1	Topic.2
151	advic	engag
152	medicin	consid
153	factor	surgeri
154	client	within

155	new	copi
156	sever	iapt
157	although	duti
158	relat	thank
159	session	street
160	privat	still
161	relev	seem
162	psychiatri	request
163	father	wish
164	children	signific
165	keep	soon
166	diabet	thing
167	complet	friend
168	explain	resourc
169	someon	self
170	pregnanc	act
171	friend	start
172	around	depart
173	offer	staff
174	psycholog	deterior
175	rate	leav
176	ago	doctor
177	case	difficult
178	episod	evid
179	earli	lane
180	hope	cognit
181	know	lot
182	king	stress
183	futur	clinic
184	want	develop
185	behaviour	juli
186	forward	impact
187	worri	offic
188	involv	behaviour
189	sinc	suffer
190	duti	worker
191	advis	alcohol
192	area	low
193	howev	happi
194	born	page
195	peopl	place
196	local	made
197	lead	write
198	adult	unit
199	univers	grate
200	block	nhs

[153]:
```
#asain set number to 100, in printing the 100 most used terms
a<- data.frame(terms(lda_best_number_asian, 600))
print(a[1:50,])
print(a[51:100,])
#101-200
print(a[101:150,])
print(a[151:200,])
```

```
        Topic.1  Topic.2
1       appoint    medic
2        servic  current
3         pleas     feel
4          team   report
5        assess   referr
6      lewisham   mental
7        number  thought
8       liaison     risk
9       contact   health
10        dear  support
11      health     mood
12      sincer     plan
13      clinic  depress
14      mental     time
15   psycholog   servic
16        time    state
17       centr   assess
18    telephon    pleas
19       group     care
20        care      low
21       offer     said
22         see  histori
23     consult     need
24      attend   famili
25   interpret  contact
26        hour  patient
27      hospit    sleep
28        book     help
29     medicin   suicid
30      integr     work
31  confidenti   inform
32      inform     live
33      privat   social
34        duti  discuss
35 psychiatrist   mother
36      academ     name
37   treatment   detail
38     support  present
```

39	slam	home
40	medic	babi
41	resourc	children
42	email	includ
43	map	team
44	hous	improv
45	high	well
46	nhs	symptom
47	depart	refer
48	perinat	abl
49	write	howev
50	help	problem
	Topic.1	Topic.2
51	follow	continu
52	miss	past
53	need	address
54	communiti	month
55	referr	previous
56	outpati	take
57	meet	last
58	street	attend
59	bring	anxieti
60	letter	harm
61	east	deni
62	detail	physic
63	novemb	hous
64	crisi	review
65	hyperlink	person
66	refer	like
67	therapi	child
68	administr	engag
69	call	describ
70	line	discharg
71	cag	client
72	direct	treatment
73	name	husband
74	like	life
75	ask	concern
76	client	follow
77	advic	activ
78	west	self
79	arrang	increas
80	enclos	get
81	park	parent
82	switchboard	dob
83	wwwslamnhsuk	want
84	mrs	hospit
85	complet	agre

86	streatham	good
87	know	appear
88	helplin	therapi
89	trust	back
90	purley	manag
91	practic	sinc
92	memori	tri
93	possibl	other
94	address	number
95	lane	sever
96	confirm	partner
97	place	peopl
98	altern	complet
99	forward	perinat
100	person	clinic
	Topic.1	Topic.2
101	ladywel	disord
102	octob	difficulti
103	practition	alcohol
104	jeanett	see
105	wallac	prescrib
106	lead	reduc
107	schedul	old
108	univers	relationship
109	prefer	experienc
110	unit	telephon
111	maudsley	think
112	nurs	provid
113	southbrook	recent
114	anxieti	felt
115	grate	ago
116	form	admiss
117	coordin	start
118	arriv	email
119	discuss	poor
120	concern	given
121	thank	form
122	emerg	today
123	denmark	diagnosi
124	let	daili
125	manag	although
126	discharg	made
127	hope	consult
128	offic	say
129	decemb	nhs
130	conveni	make
131	leaflet	difficult
132	mood	around

133	venu	psycholog
134	recept	iapt
135	current	leav
136	septemb	friend
137	soon	tablet
138	psychiatri	talk
139	take	issu
140	order	first
141	phone	father
142	senior	night
143	session	age
144	januari	normal
145	king	requir
146	avail	cognit
147	dulwich	daughter
148	wish	awar
149	rearrang	still
150	well	better

	Topic.1	Topic.2
151	initi	chang
152	unabl	move
153	adult	benefit
154	lordship	signific
155	psychiatr	regular
156	edridg	hour
157	addit	end
158	surgeri	stop
159	receiv	consid
160	note	abus
161	requir	distress
162	profession	drug
163	reason	sertralin
164	colleg	find
165	suitabl	reason
166	februari	period
167	code	centr
168	speak	crisi
169	chang	anxious
170	director	pregnanc
171	urgent	birth
172	therapist	high
173	patient	mind
174	best	remain
175	locum	appropri
176	gender	ward
177	make	behaviour
178	access	note
179	someon	understand

```
180                              back       come
181                              awar        new
182                            cancel    attempt
183                               abl     affect
184                             anoth     involv
185                             provid      initi
186                            timehin       cope
187                             london       user
188                             social    session
189                               camh     school
190                            languag     intent
191                               type     episod
192                               kind       pain
193                                ann        sri
194                                end     access
195                              local    consent
196 httpsslamsmartsurveycouksnavusu        ill
197                              last    protect
198                            holiday      level
199                              cheyn      relat
200                              clive    sometim
```

[9]:
```
#downloading libs for sentiment analysis
install.packages('stringr')
library(stringr)
install.packages('stopwords')
library(stopwords)
install.packages('dplyr')
library(dplyr)
install.packages('tidytext')
library(tidytext)
install.packages('syuzhet')
library(syuzhet)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

```
Attaching package: ‘dplyr’


The following objects are masked from ‘package:stats’:

    filter, lag


The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union


Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done
```

[14]:
```
#function in cleaning the text
cleaning_function = function(data) {
    cleanedText<- tolower(data$Text)
    cleanedText<- gsub('[[:punct:]]','', cleanedText)
    cleanedText<- gsub('\\d+','',  cleanedText)
    stopwords_en<- stopwords::stopwords('en')
    cleanedText<- cleanedText[!(cleanedText%in%stopwords_en)]
    custom<-c('zzzzz', 'qqqqq', 'ie', 'also', 'can', 'will', 'yes','may',␣
 ↪'one', 'now', 'monday', 'tuesday', 'wednesday', 'thursday', 'friday',␣
 ↪'saturday', 'sunday', 'whilst', 'known', 'tel', 'floor', 'due', 'day',␣
 ↪'october', 'november', 'january', 'febuary', 'march', 'april', 'may',␣
 ↪'june', 'july', 'august', 'september', 'december', 'hill', 'south',␣
 ↪'southwark', 'lambeth', 'lewis', 'alex', 'look', 'road', 'formcheckbox',␣
 ↪'use', 'ongo', 'taken', 'someone', 'way', 'ref', 'croydon', 'year', 'week',␣
 ↪'referral', 'two',   'regard', 'headquarter', 'fax', 'tel', 'seen', 'chris',␣
 ↪'date', 'helen', 'sydenham')
    cleanedText<- cleanedText[!(cleanedText%in%custom)]
    data$Text<- cleanedText
    return(data)
    }
```

[15]:
```
#clean female data for sentiment analysis
clean.data.female<- cleaning_function(data.female)
```

[72]:
```
#clean male data for sentiment analysis
clean.data.male<- cleaning_function(data.male)
```

[17]:
```
#clean african data for sentiment analysis
clean.data.african<- cleaning_function(data.african)
```

[18]:
```
#clean black data for sentiment analysis
clean.data.black<- cleaning_function(data.black)
```

[19]:
```
#clean white data for sentiment analysis
clean.data.white<- cleaning_function(data.white)
```

[20]:
```
#clean mixed data for sentiment analysis
clean.data.mixed<- cleaning_function(data.mixed)
```

[21]:
```
#clean eu data for sentiment analysis
clean.data.eu<- cleaning_function(data.eu)
```

[22]:
```
#clean caribbean data for sentiment analysis
clean.data.caribbean<- cleaning_function(data.caribbean)
```

[23]:
```
#clean asian data for sentiment analysis
clean.data.asian<- cleaning_function(data.asian)
```

[24]:
```
#creates function for sentiment
get_sentiment_scores <- function(text_vector) {
  scores <- get_nrc_sentiment(text_vector)
}
```

[26]:
```
#female sentiment
female_sentiment<- get_sentiment_scores(clean.data.female$Text)
```

```
Warning message:
"`spread_()` was deprecated in tidyr 1.2.0.
 Please use `spread()` instead.
 The deprecated feature was likely used in the syuzhet
package.
  Please report the issue to the authors."
```

[73]:
```
#male sentiment
male_sentiment<- get_sentiment_scores(clean.data.male$Text)
```

[47]:
```
#african sentiment
african_sentiment<- get_sentiment_scores(clean.data.african$Text)
```

[48]:
```
#black sentiment
black_sentiment<- get_sentiment_scores(clean.data.black$Text)
```

[49]:
```
#white sentiment
white_sentiment<- get_sentiment_scores(clean.data.white$Text)
```

[50]:
```
#mixed sentiment
mixed_sentiment<- get_sentiment_scores(clean.data.mixed$Text)
```

[51]:
```
#eu sentiment
eu_sentiment<- get_sentiment_scores(clean.data.eu$Text)
```

[52]:
```
#caribbean sentiment
caribbean_sentiment<- get_sentiment_scores(clean.data.caribbean$Text)
```

[53]:
```
#asian sentiment
asian_sentiment<- get_sentiment_scores(clean.data.asian$Text)
```

[36]:
```
#create a function to plot sentiment into a histogram
sentiment_hist = function(sentiment_text){
for (c in colnames(sentiment_text)){
    print(c)
    hist(sentiment_text[,c], main =c)
    }
    }
```

[69]:
```
#create function for barplots
barplot_function = function (text){
    x2<- get_nrc_sentiment(text)
    print(summary(x2))
    barplot(colSums(x2),
            las=2,
            col=rainbow(10),
            ylab='Count',
            main='Sentiment Score')
    }
```

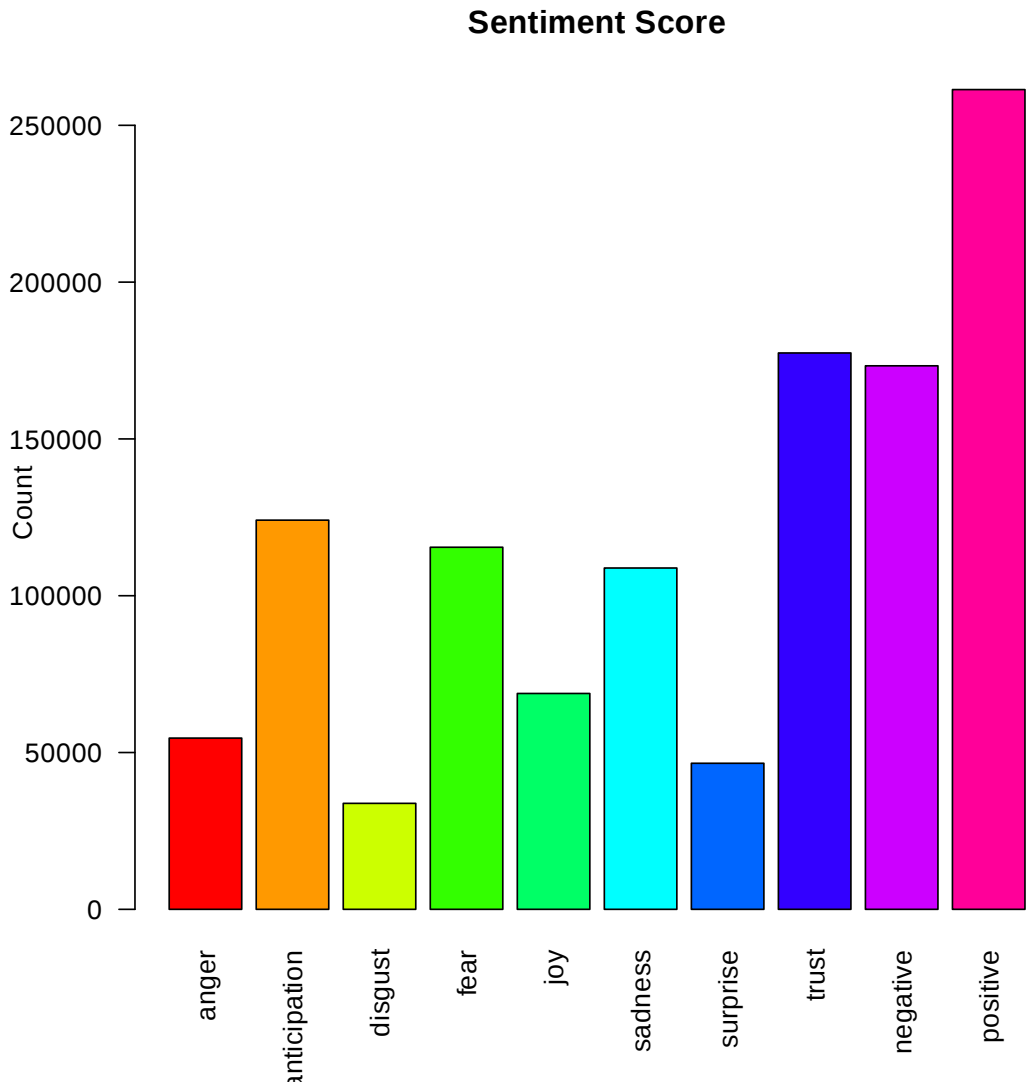
[ ]:
```
#create function for barplots
barplot_function = function (text){
    x1<- as.character(text)
    x2<- get_nrc_sentiment(x1)
    print(summary(x2))
    barplot(colSums(x2),
            las=2,
            col=rainbow(10),
            ylab='Count',
            main='Sentiment Score')
```

```
    }
```

[44]:
```
#average sentiment function
average_sentiment = function(sentiment_text){
     sentiment_text$sentiment = sentiment_text$positive -␣
 ↪sentiment_text$negative
    }
```

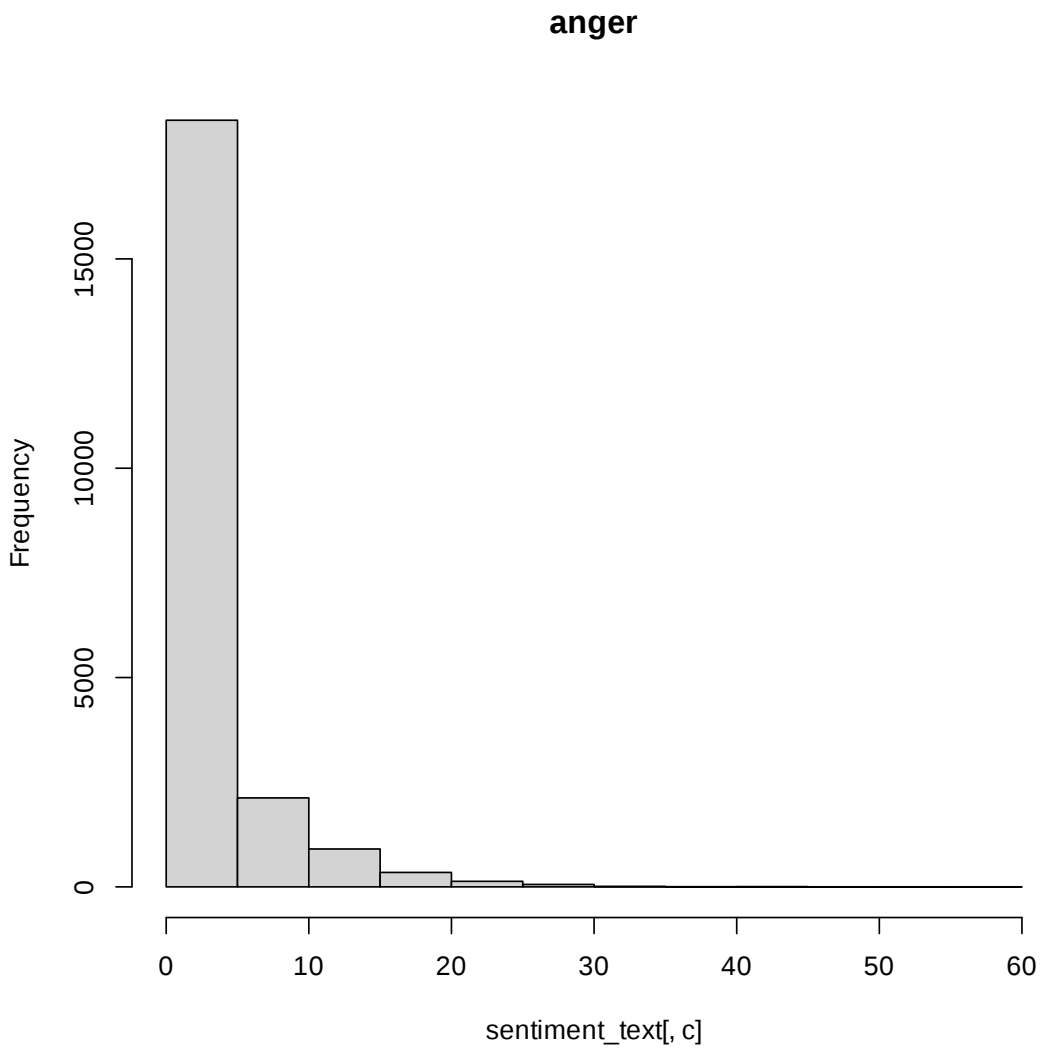
[42]:
```
#female barplot
barplot_function(clean.data.female$Text)
```

```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
 Median : 0.000   Median : 3.000   Median : 0.000   Median : 2.000
 Mean   : 2.493   Mean   : 5.666   Mean   : 1.543   Mean   : 5.272
 3rd Qu.: 3.000   3rd Qu.: 8.000   3rd Qu.: 2.000   3rd Qu.: 7.000
 Max.   :56.000   Max.   :53.000   Max.   :33.000   Max.   :69.000
      joy            sadness          surprise          trust
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.0
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.0
 Median : 1.000   Median : 2.000   Median : 1.000   Median : 6.0
 Mean   : 3.143   Mean   : 4.969   Mean   : 2.126   Mean   : 8.1
 3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 3.000   3rd Qu.:11.0
 Max.   :41.000   Max.   :66.000   Max.   :25.000   Max.   :78.0
    negative          positive
 Min.   :  0.000   Min.   :  0.00
 1st Qu.:  0.000   1st Qu.:  0.00
 Median :  2.000   Median :  7.00
 Mean   :  7.913   Mean   : 11.93
 3rd Qu.: 10.000   3rd Qu.: 16.00
 Max.   :108.000   Max.   :135.00
```

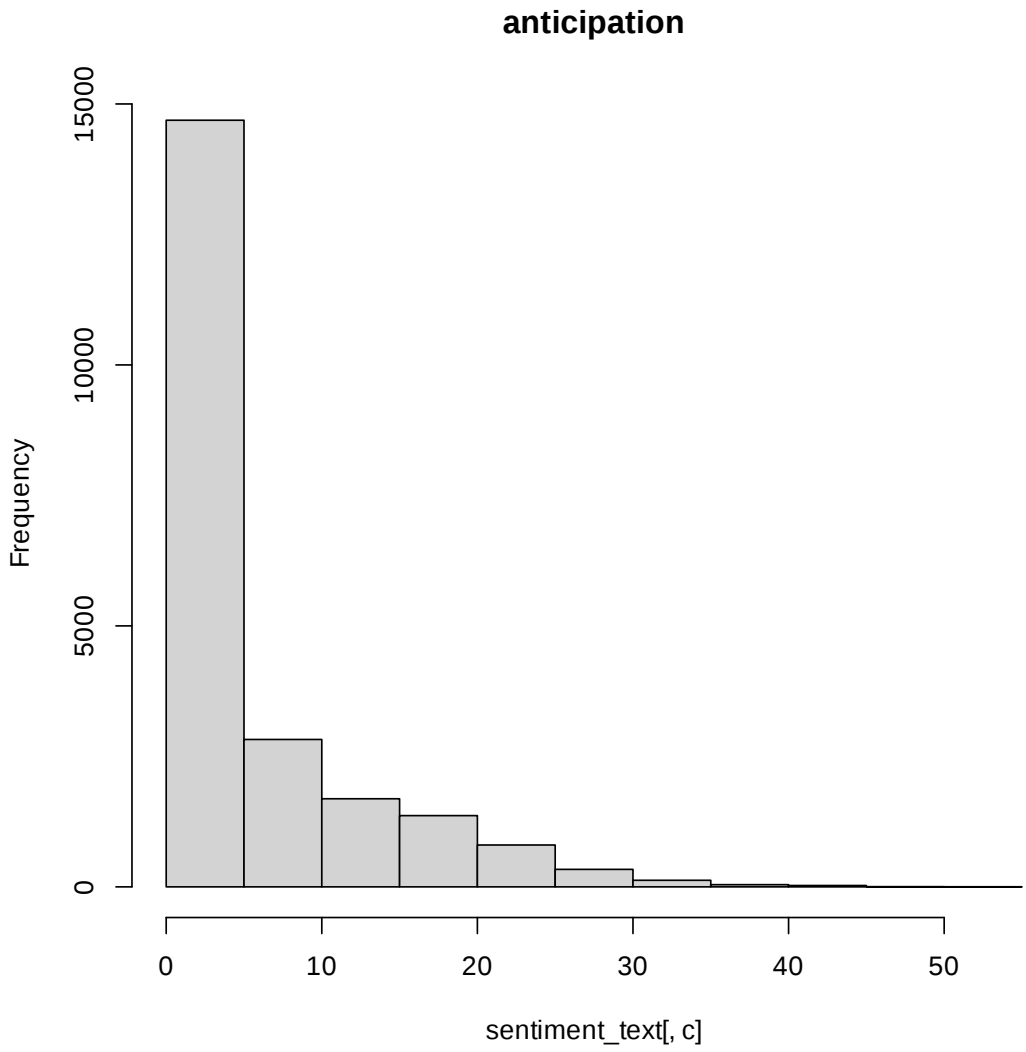


[43]:
```
#female sentiment histogram
sentiment_hist(female_sentiment)
```

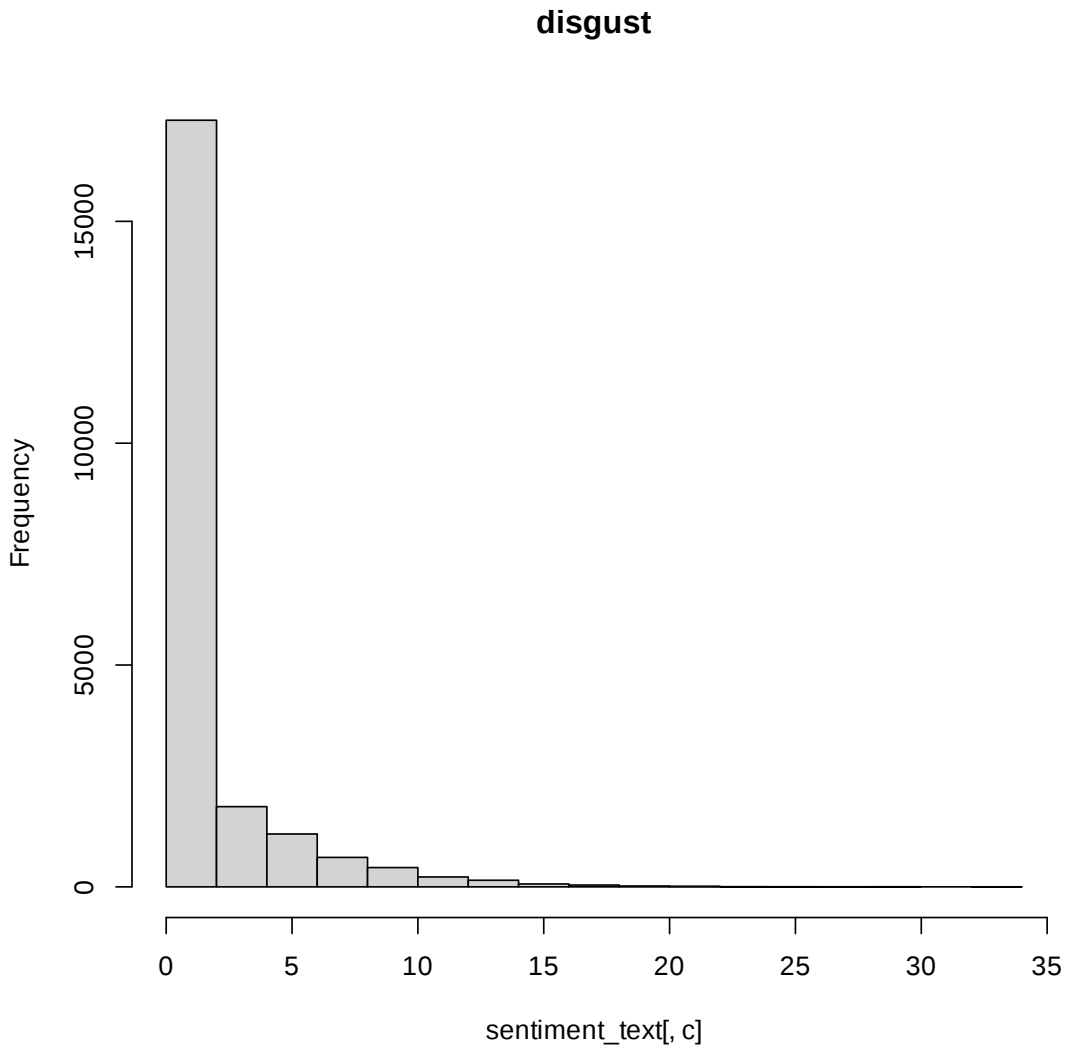
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
5000
10000
15000
0
10
20
30
40
50
60
sentiment_text[, c]

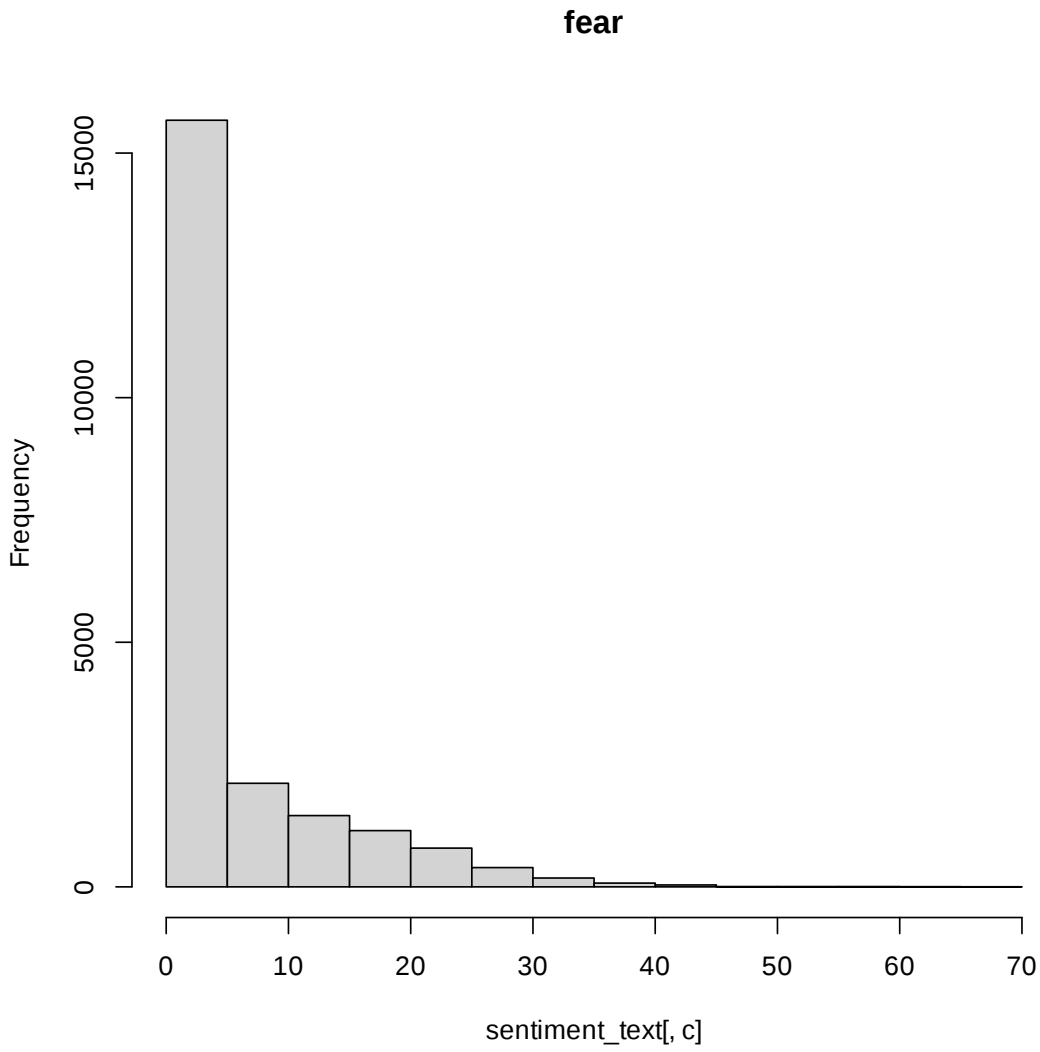
```
[1] "disgust"
```


anticipation
Frequency
0
5000
10000
15000
0
10
20
30
40
50
sentiment_text[, c]

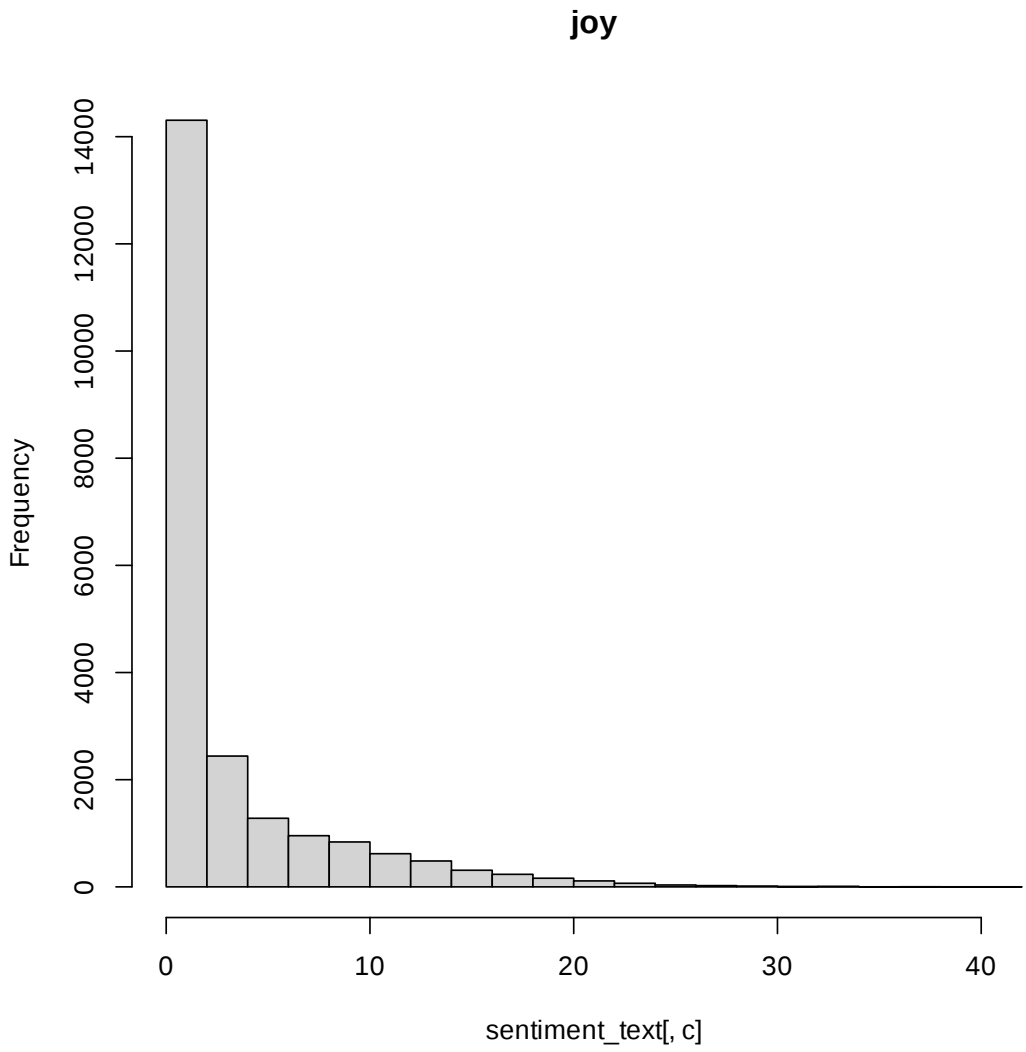
```
[1] "fear"
```


disgust
Frequency
0
5000
10000
15000
0
5
10
15
20
25
30
35
sentiment_text[, c]

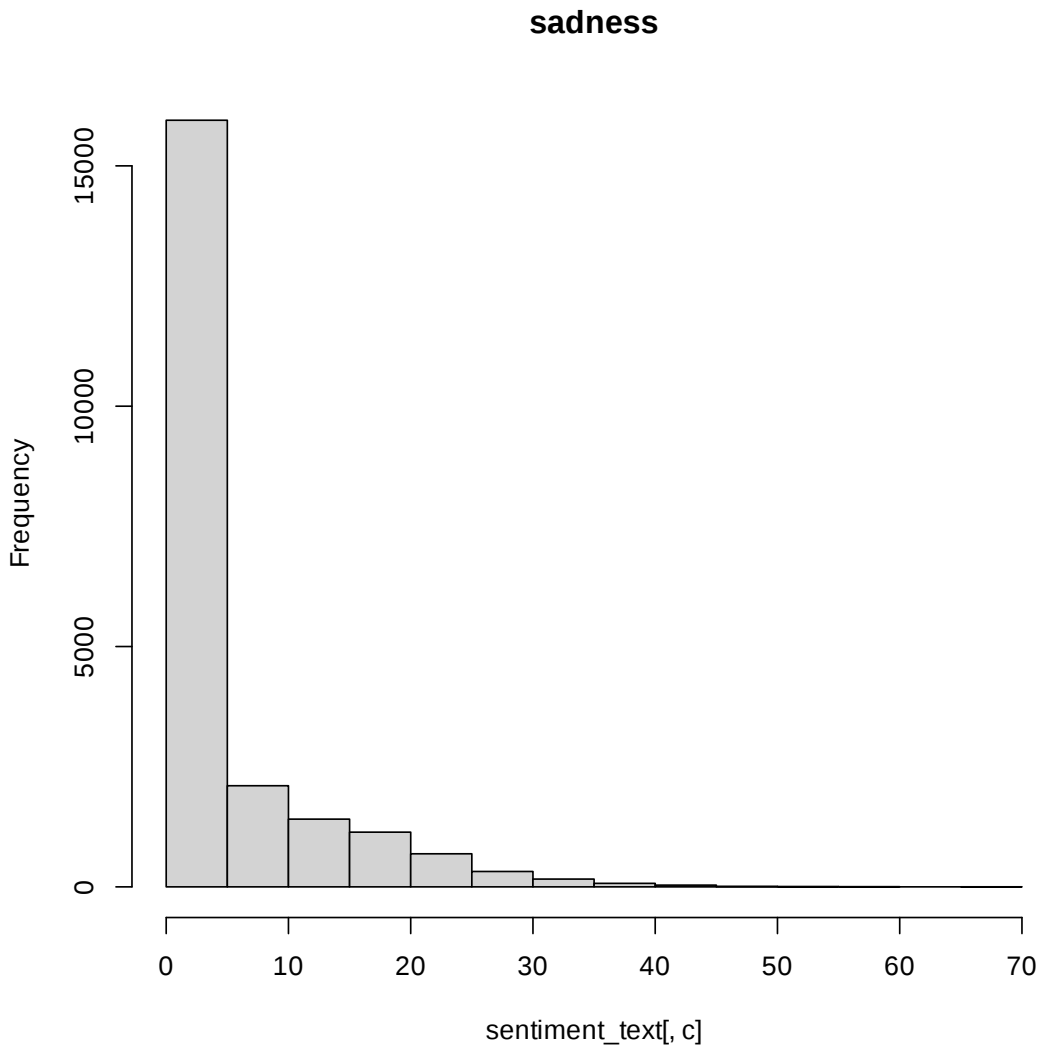
```
[1] "joy"
```


fear
Frequency
0
5000
10000
15000
0
10
20
30
40
50
60
70
sentiment_text[, c]

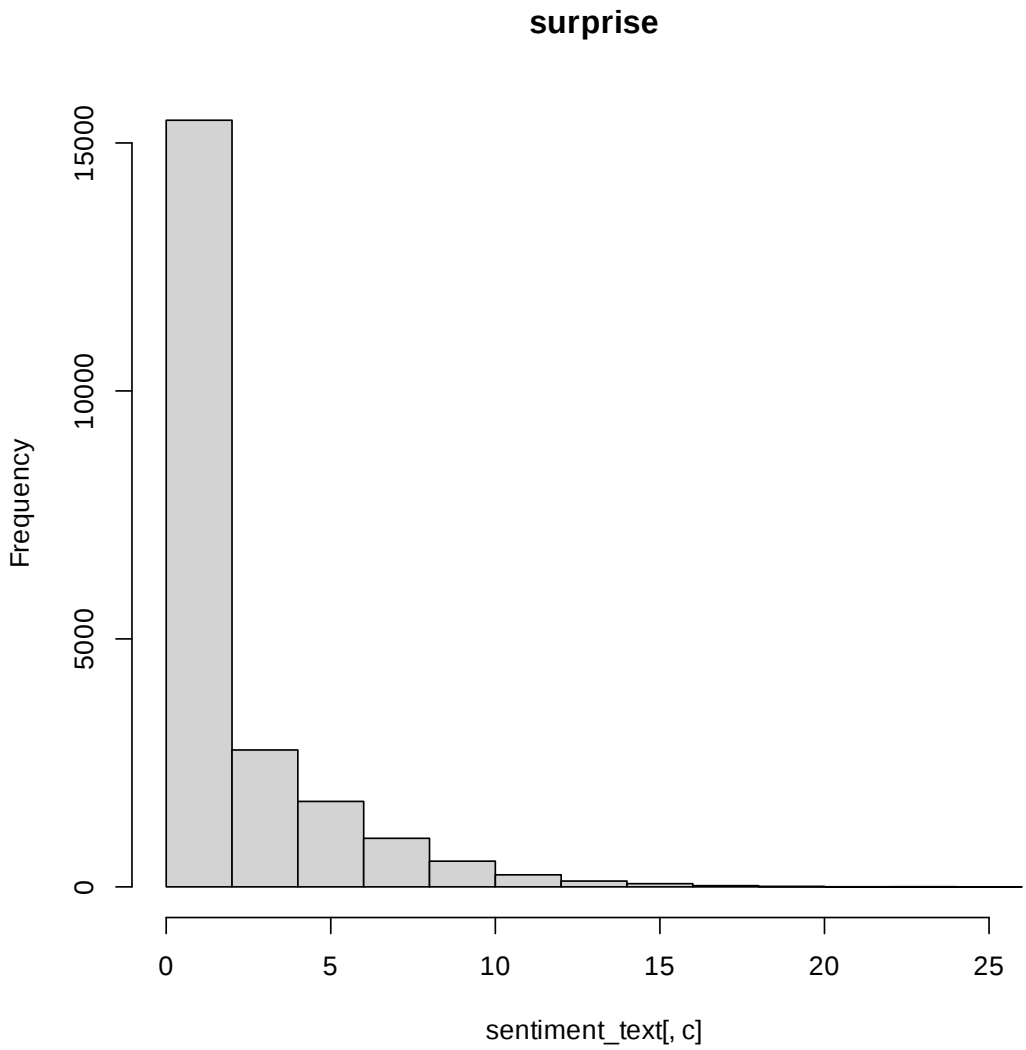
```
[1] "sadness"
```


joy
Frequency
0
2000
4000
6000
8000
10000
12000
14000
0
10
20
30
40
sentiment_text[, c]

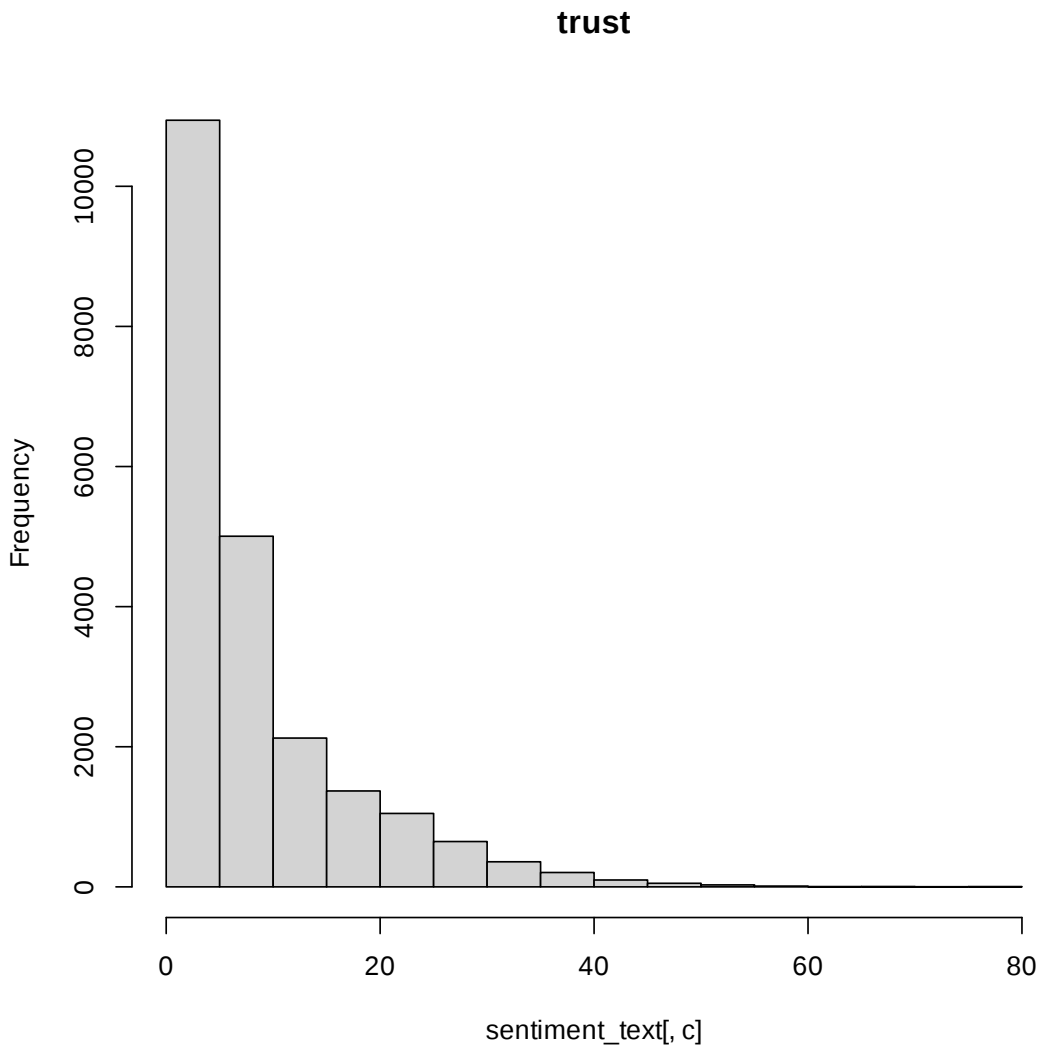
```
[1] "surprise"
```


sadness
Frequency
0
5000
10000
15000
0
10
20
30
40
50
60
70
sentiment_text[, c]

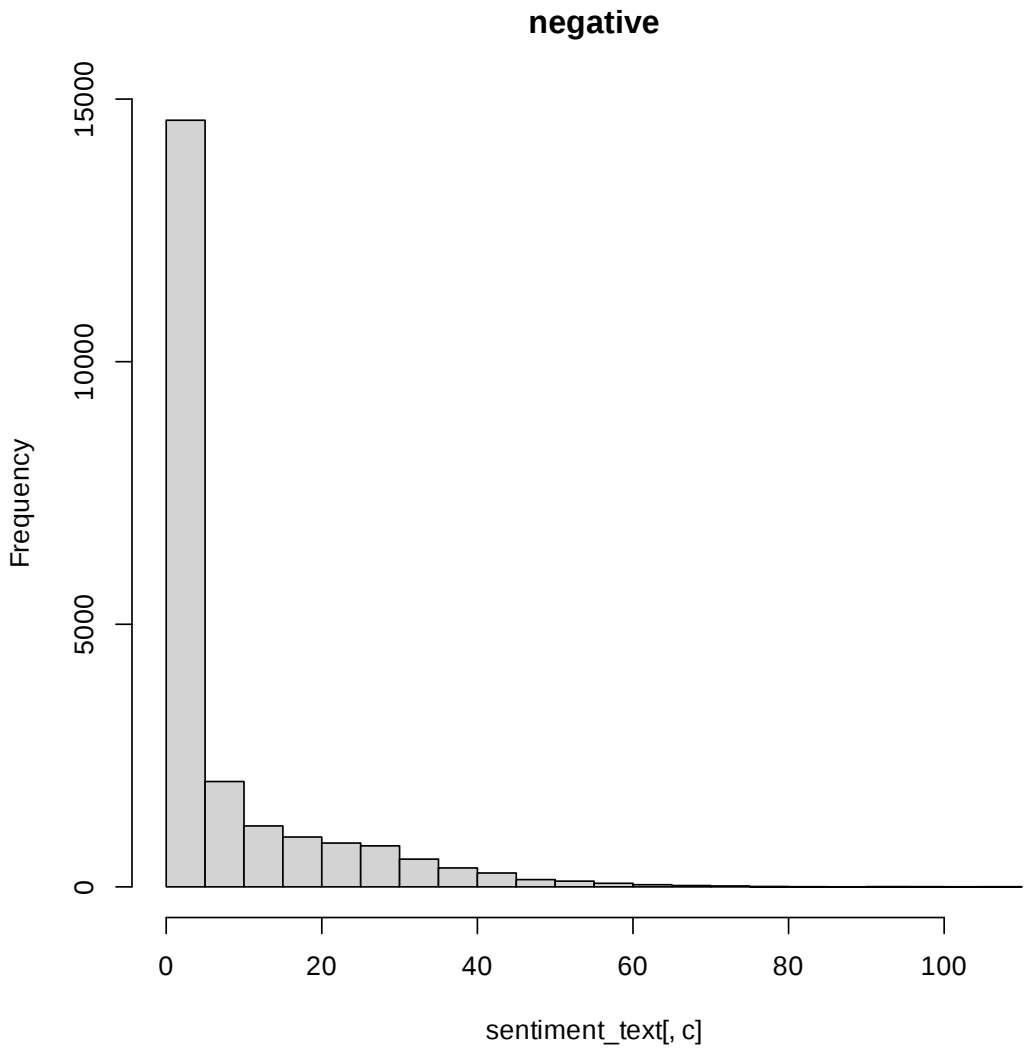
```
[1] "trust"
```

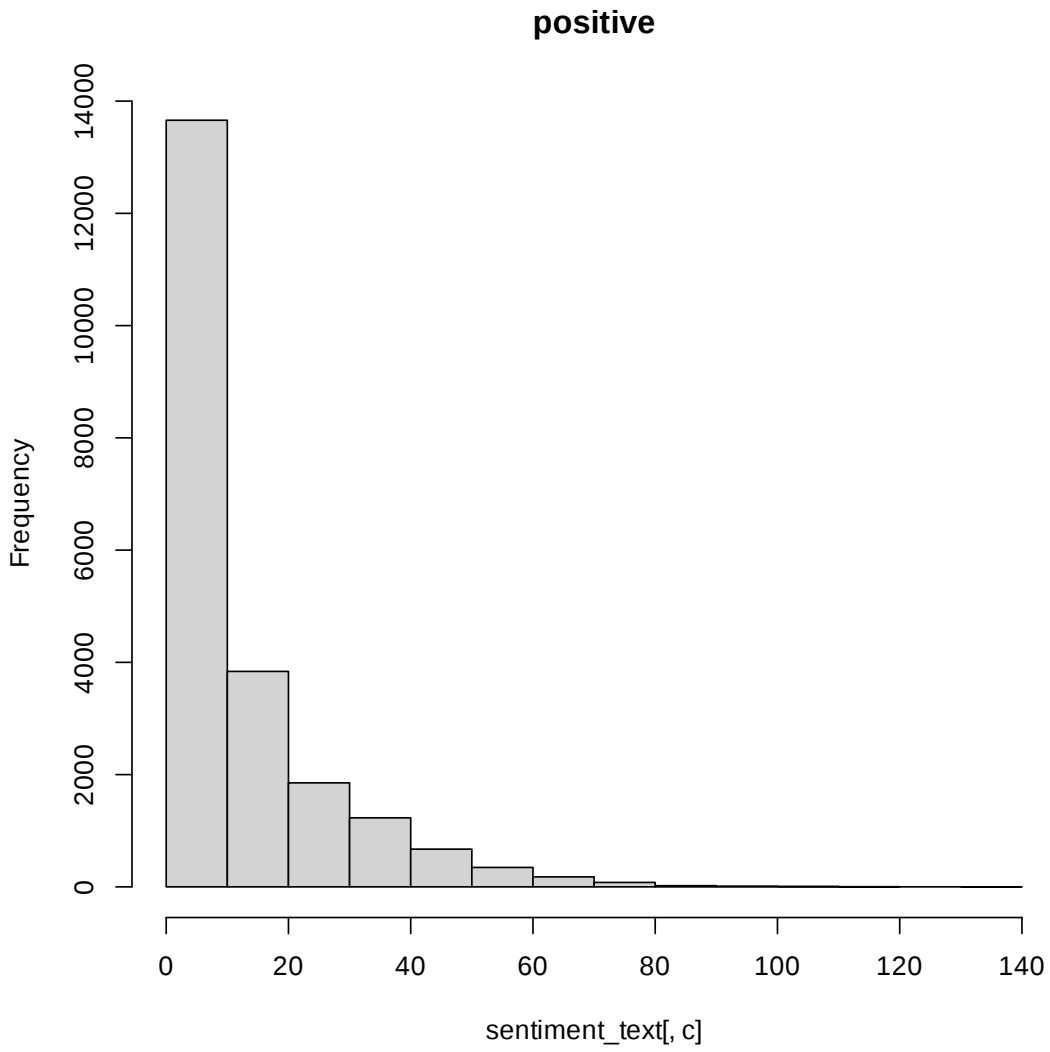

surprise
Frequency
0
5000
10000
15000
0
5
10
15
20
25
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
2000
4000
6000
8000
10000
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
5000
10000
15000
0
20
40
60
80
100
sentiment_text[, c]



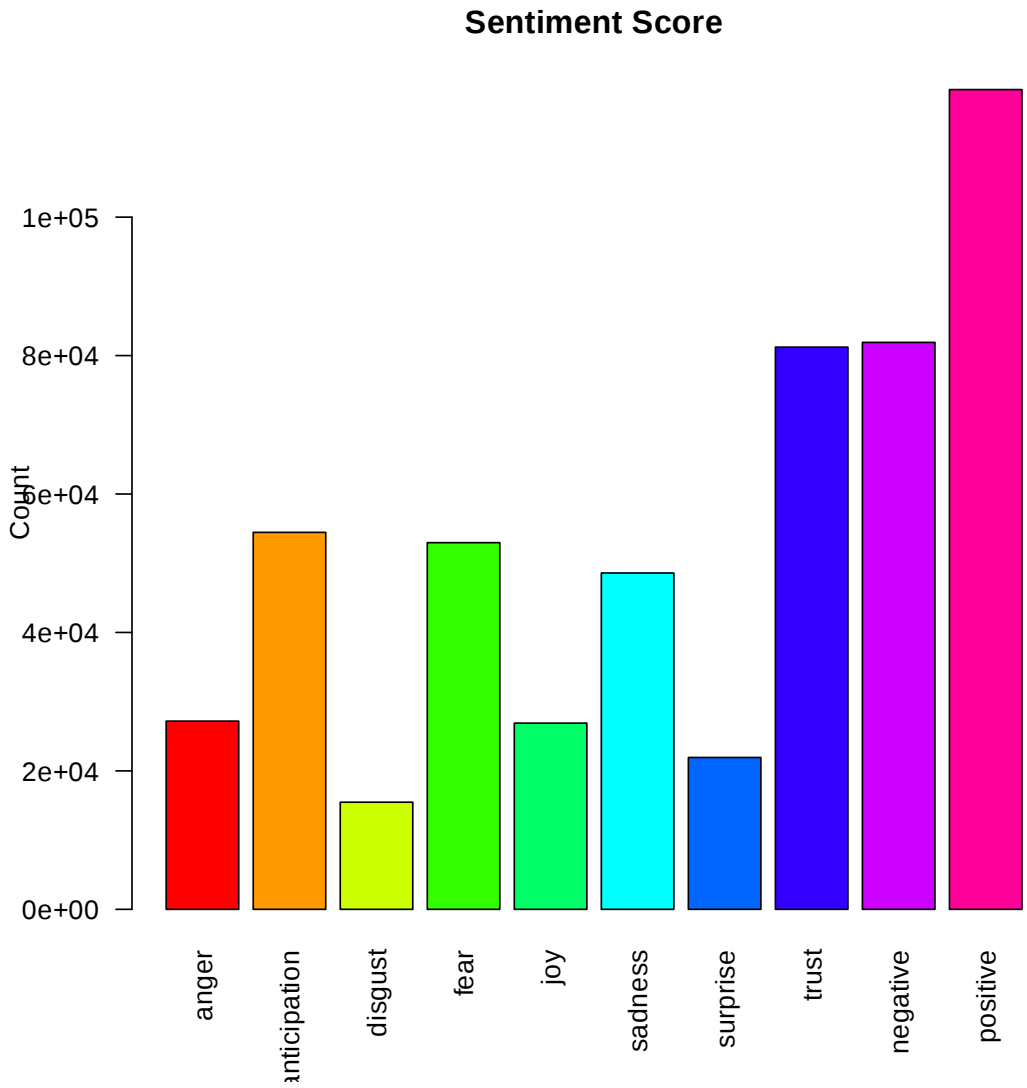
[45]:
```
#female average sentiment
average_female<- average_sentiment(female_sentiment)
summary(average_female)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-26.000   0.000   3.000   4.022   7.000  80.000
```

[81]:
```
#male barplot
barplot_function(clean.data.male$Text)
```

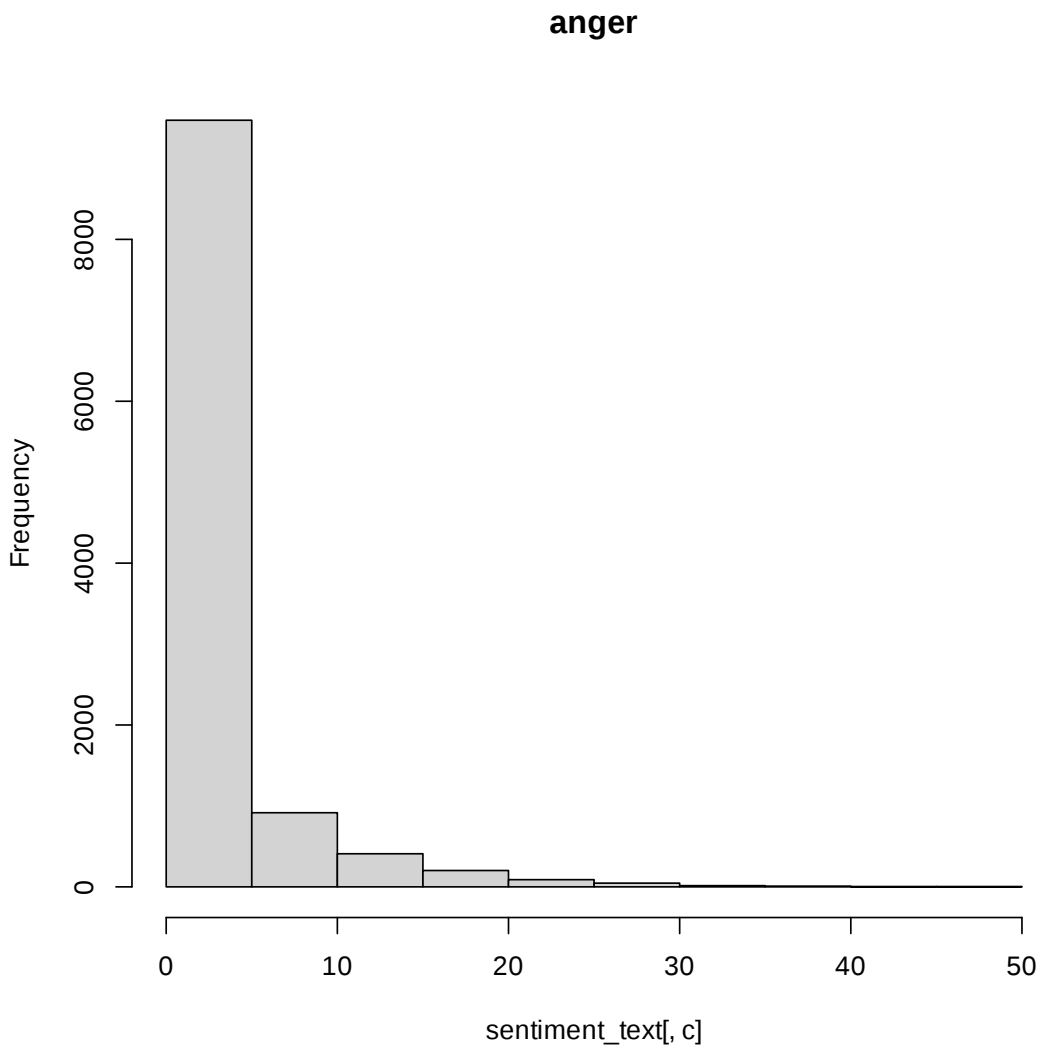
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 2.000   Median : 0.000   Median : 1.000
Mean   : 2.435   Mean   : 4.877   Mean   : 1.386   Mean   : 4.744
3rd Qu.: 2.000   3rd Qu.: 7.000   3rd Qu.: 1.000   3rd Qu.: 6.000
Max.   :50.000   Max.   :50.000   Max.   :38.000   Max.   :73.000
     joy              sadness          surprise         trust
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
Median : 0.000   Median : 1.000   Median : 1.000   Median : 5.000
Mean   : 2.409   Mean   : 4.352   Mean   : 1.965   Mean   : 7.274
3rd Qu.: 3.000   3rd Qu.: 5.000   3rd Qu.: 3.000   3rd Qu.:10.000
Max.   :38.000   Max.   :64.000   Max.   :26.000   Max.   :77.000
   negative          positive
Min.   :  0.000   Min.   :  0.00
1st Qu.:  0.000   1st Qu.:  0.00
Median :  2.000   Median :  7.00
Mean   :  7.335   Mean   : 10.61
3rd Qu.:  8.000   3rd Qu.: 14.00
Max.   :112.000   Max.   :120.00
```

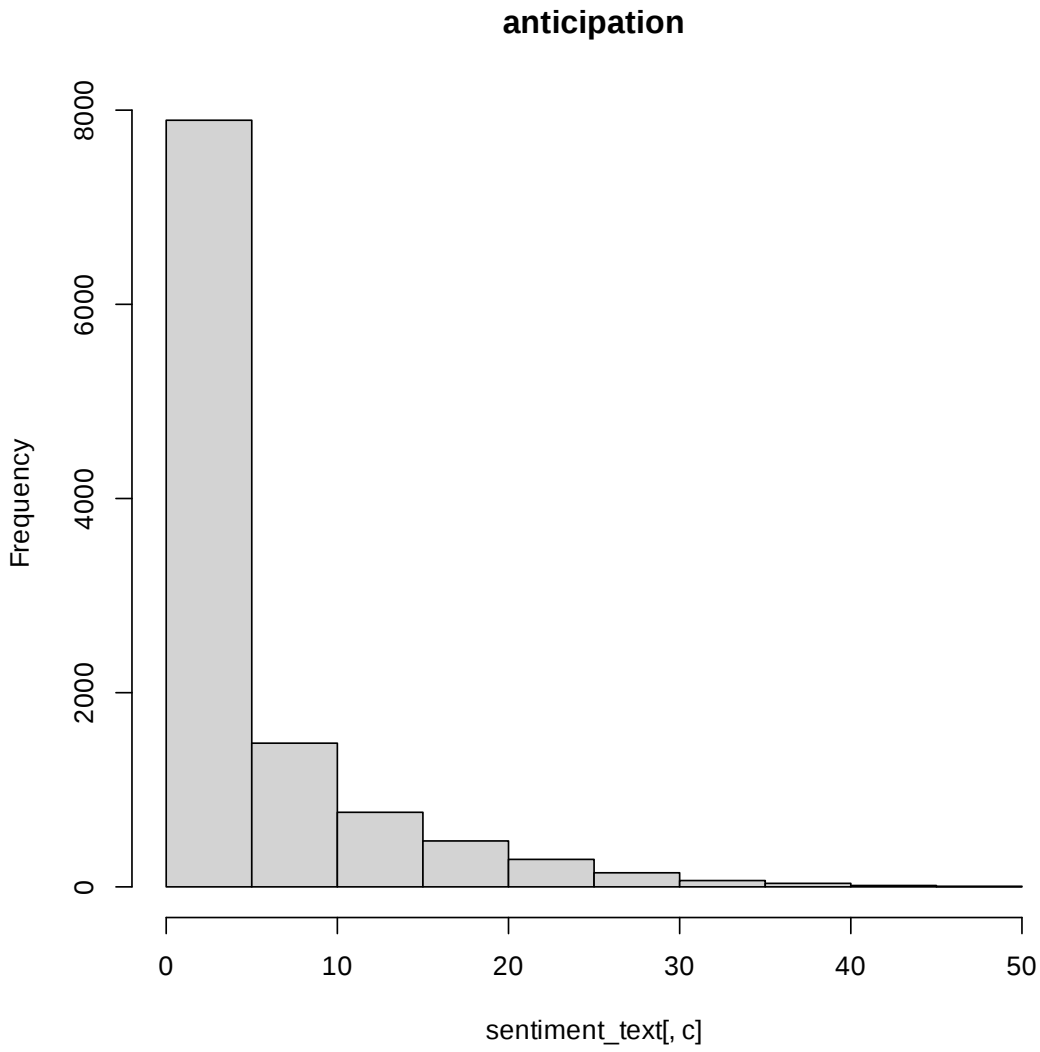


[75]:
```
#male sentiment histogram
sentiment_hist(male_sentiment)
```

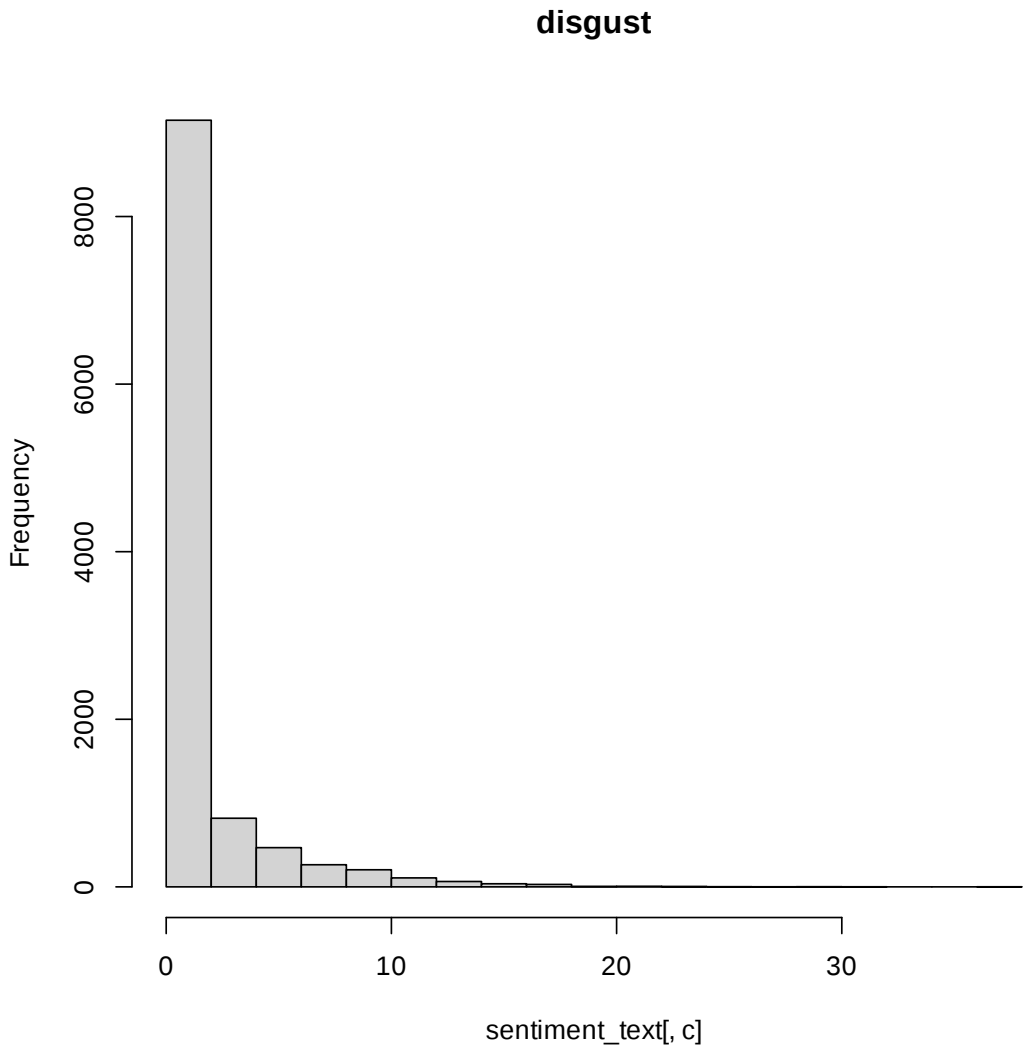
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
2000
4000
6000
8000
0
10
20
30
40
50
sentiment_text[, c]

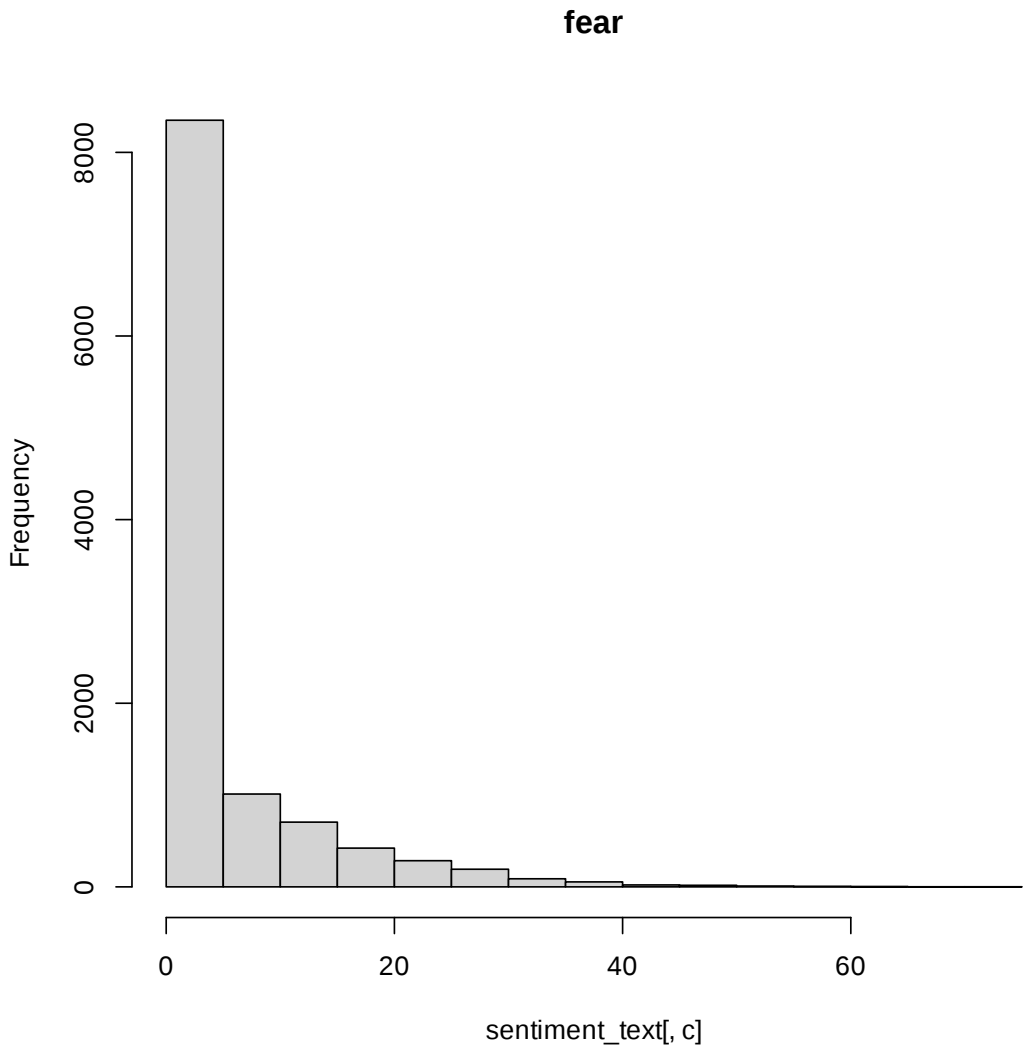
```
[1] "disgust"
```


anticipation
Frequency
0
2000
4000
6000
8000
0
10
20
30
40
50
sentiment_text[, c]

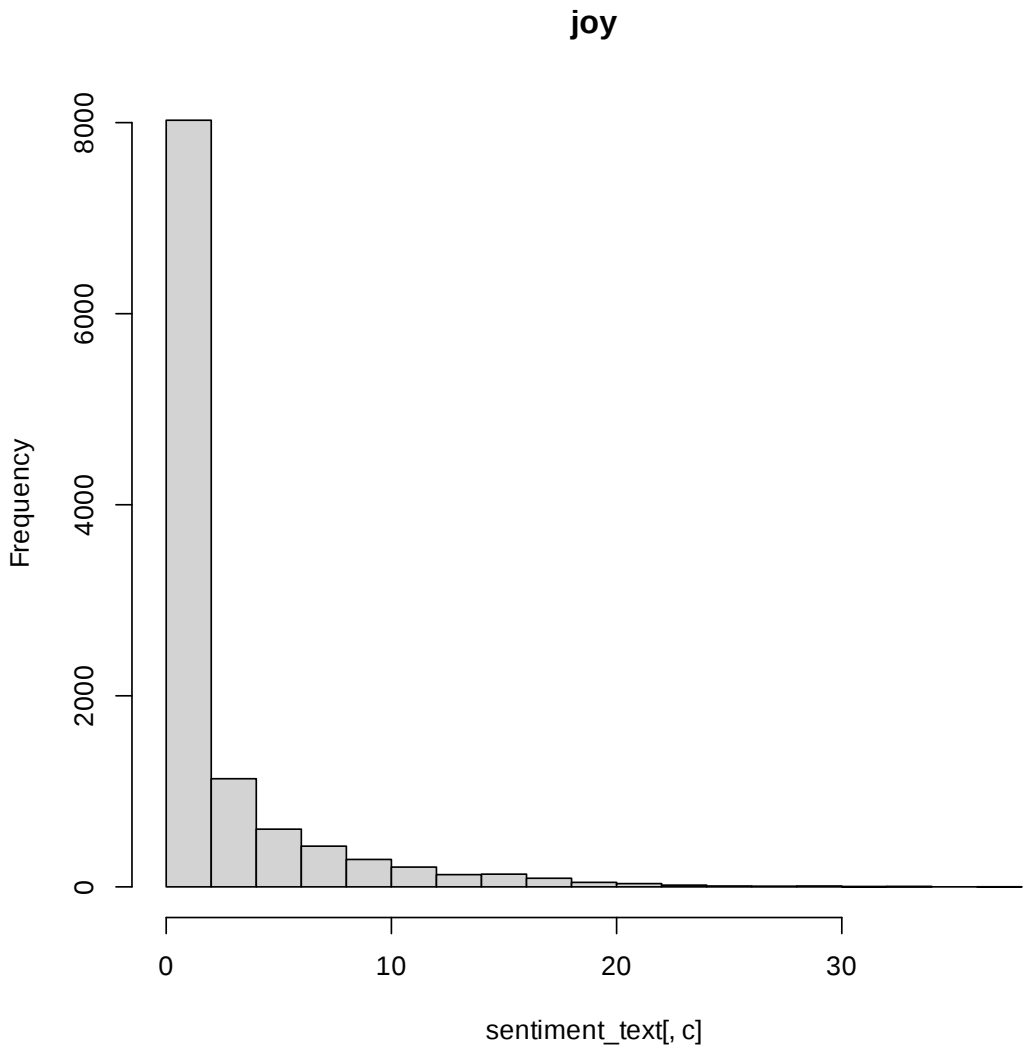
```
[1] "fear"
```


disgust
Frequency
0
2000
4000
6000
8000
0
10
20
30
sentiment_text[, c]

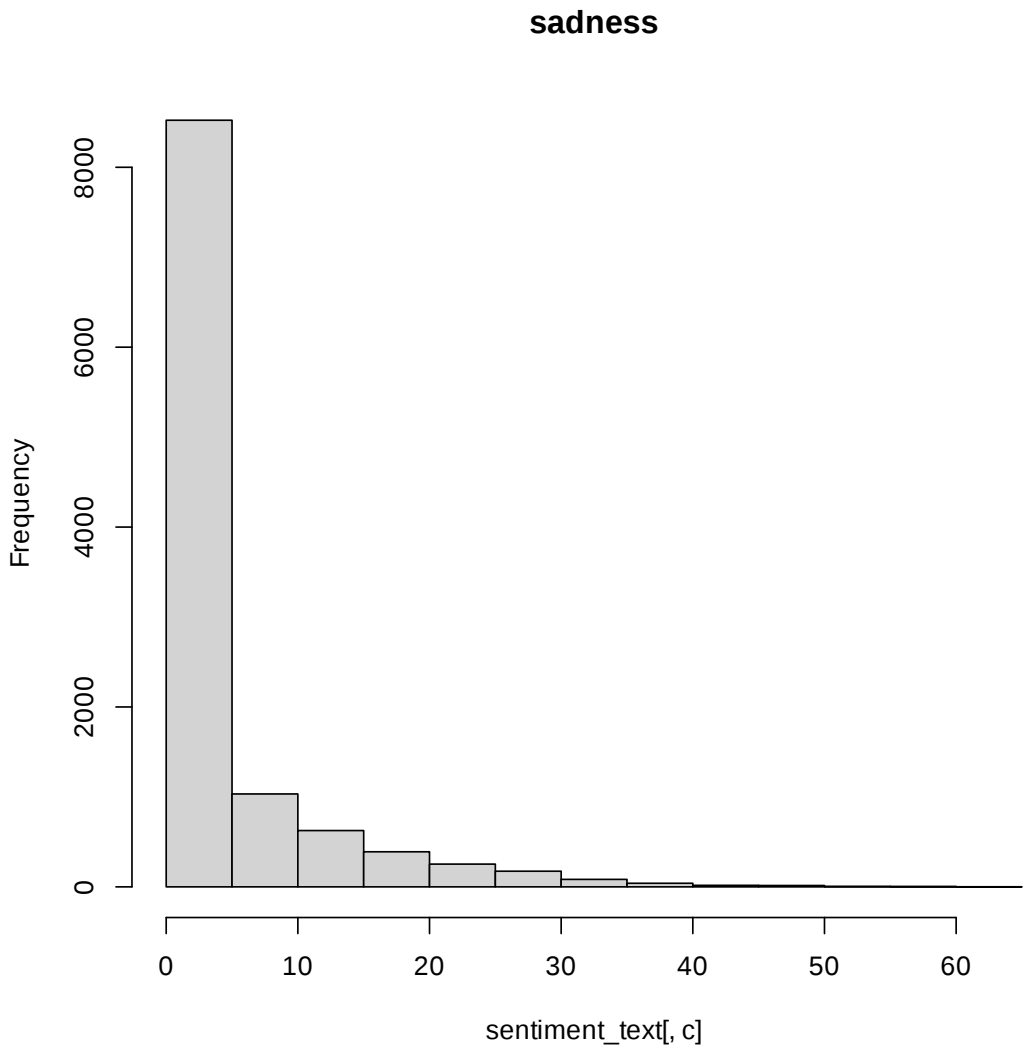
```
[1] "joy"
```


fear
Frequency
0
2000
4000
6000
8000
0
20
40
60
sentiment_text[, c]

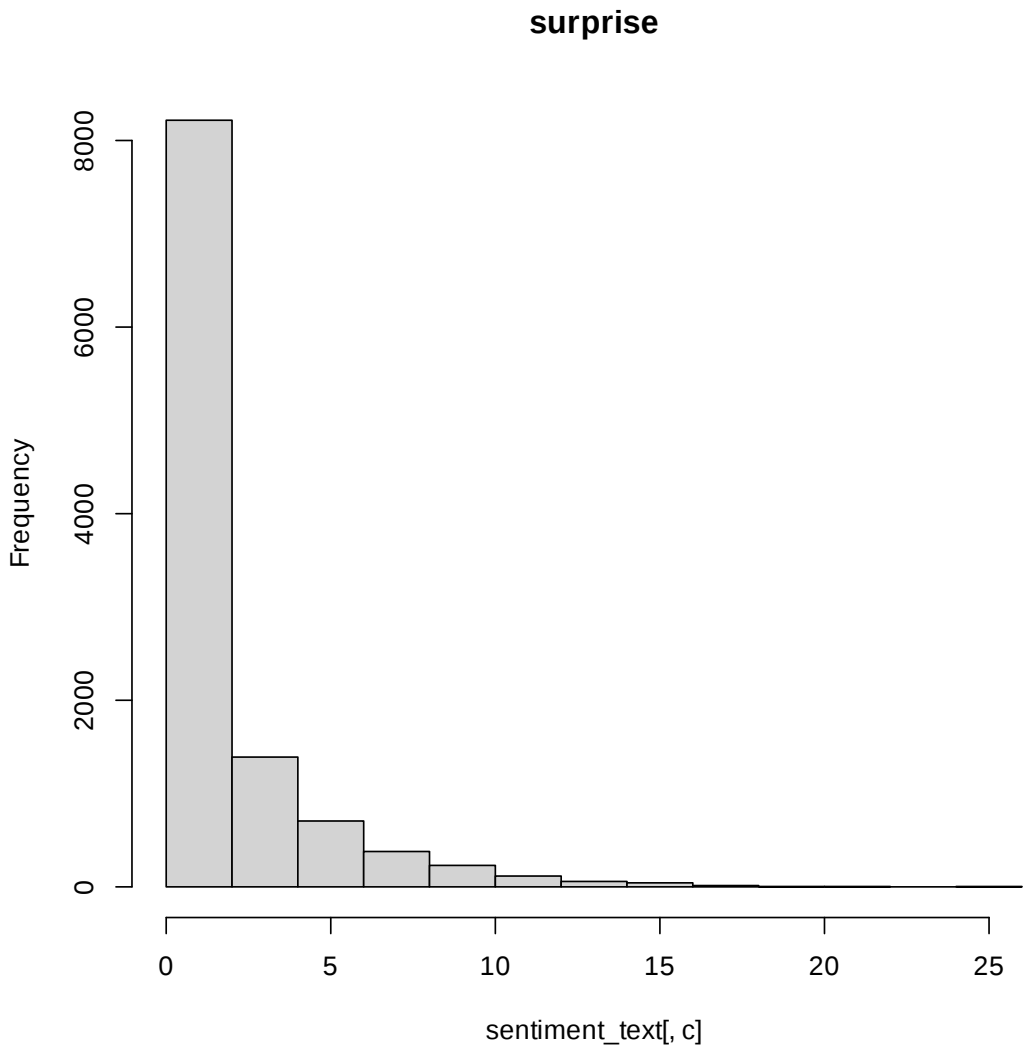
```
[1] "sadness"
```


joy
Frequency
0
2000
4000
6000
8000
0
10
20
30
sentiment_text[, c]

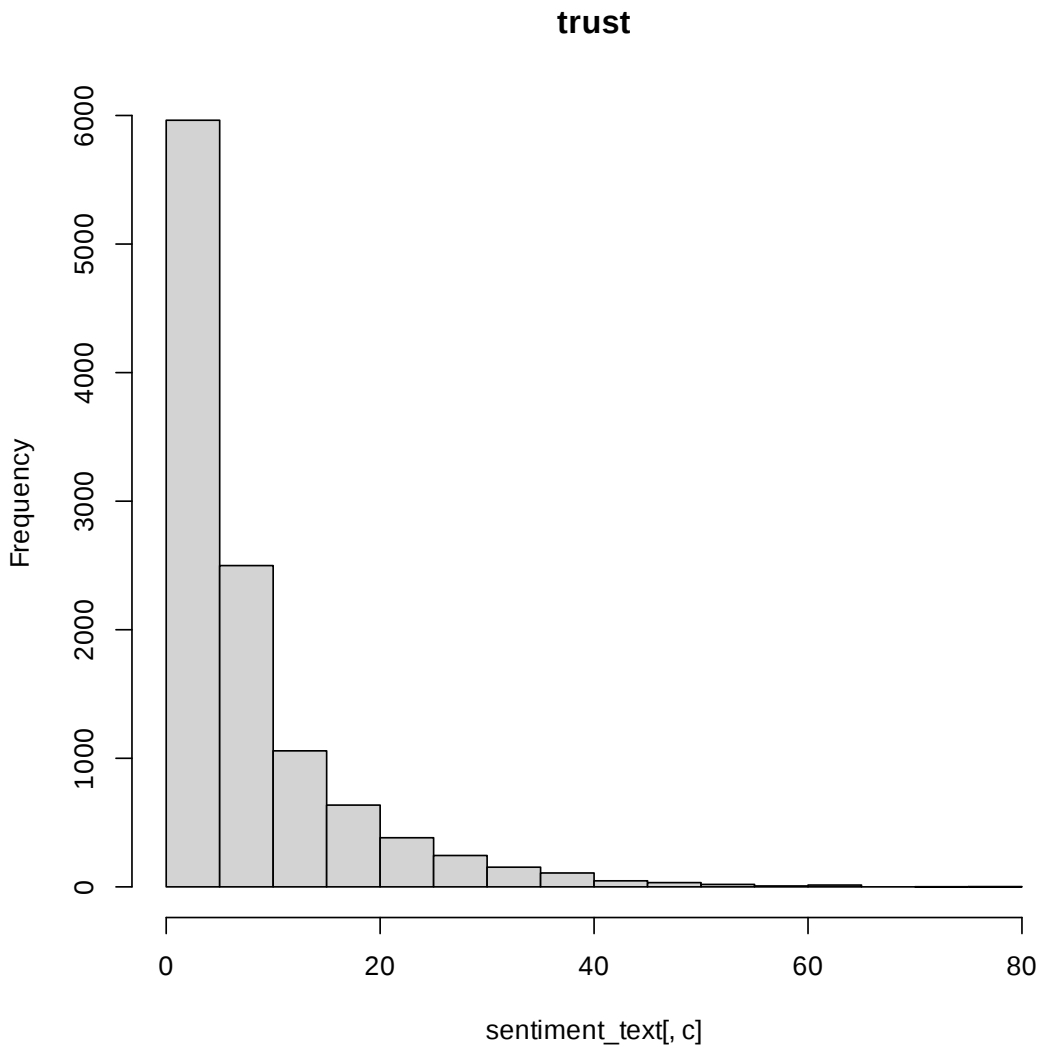
```
[1] "surprise"
```


sadness
Frequency
0
2000
4000
6000
8000
0
10
20
30
40
50
60
sentiment_text[, c]

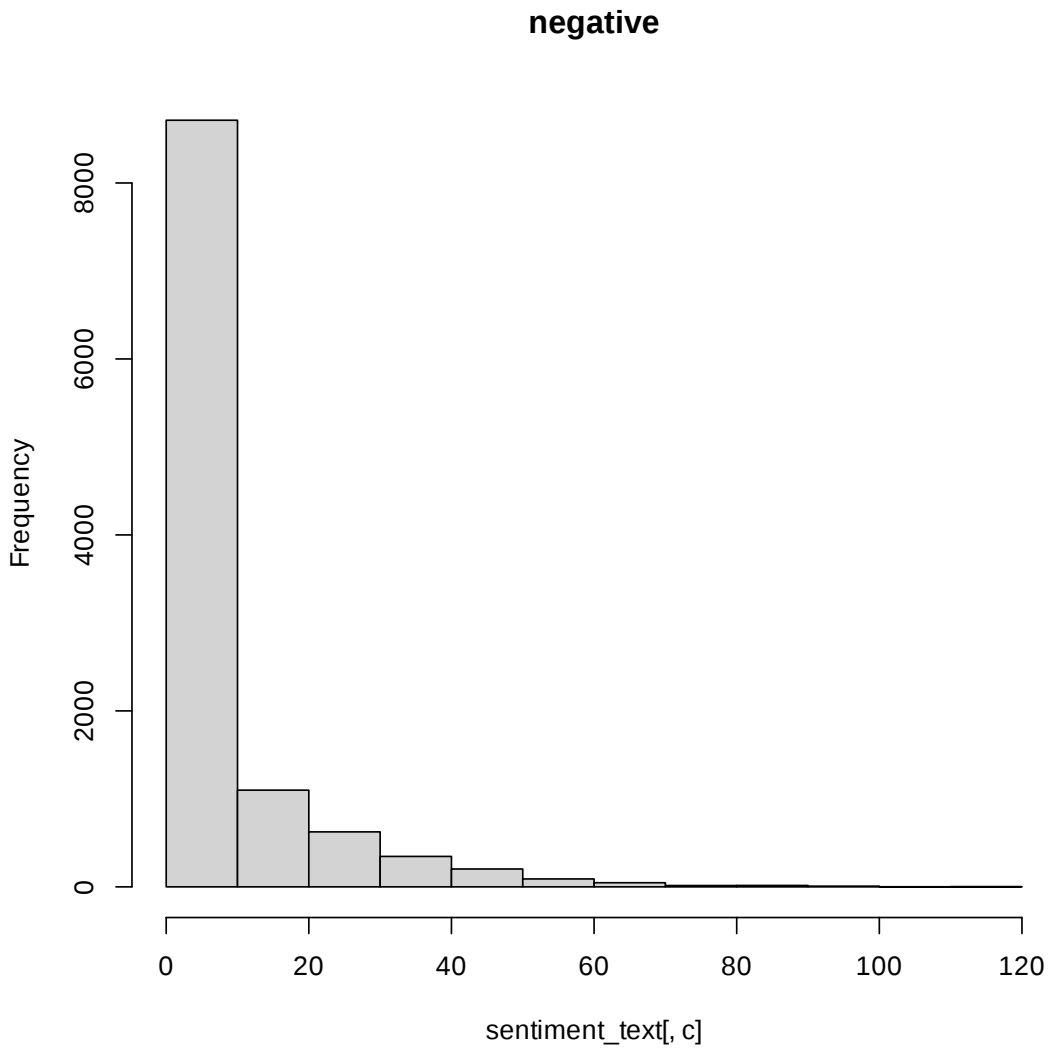
```
[1] "trust"
```

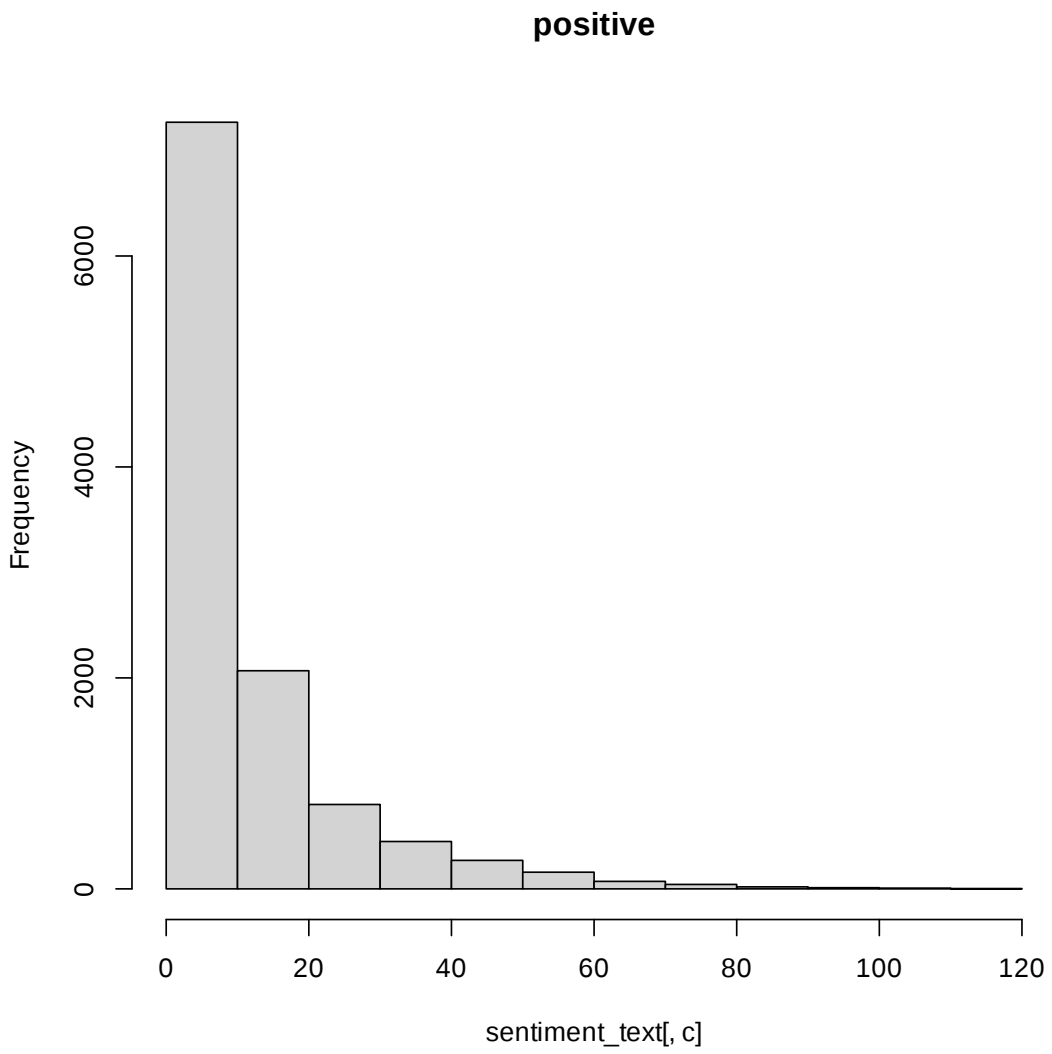

surprise
Frequency
0
2000
4000
6000
8000
0
5
10
15
20
25
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
1000
2000
3000
4000
5000
6000
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
2000
4000
6000
8000
0
20
40
60
80
100
120
sentiment_text[, c]



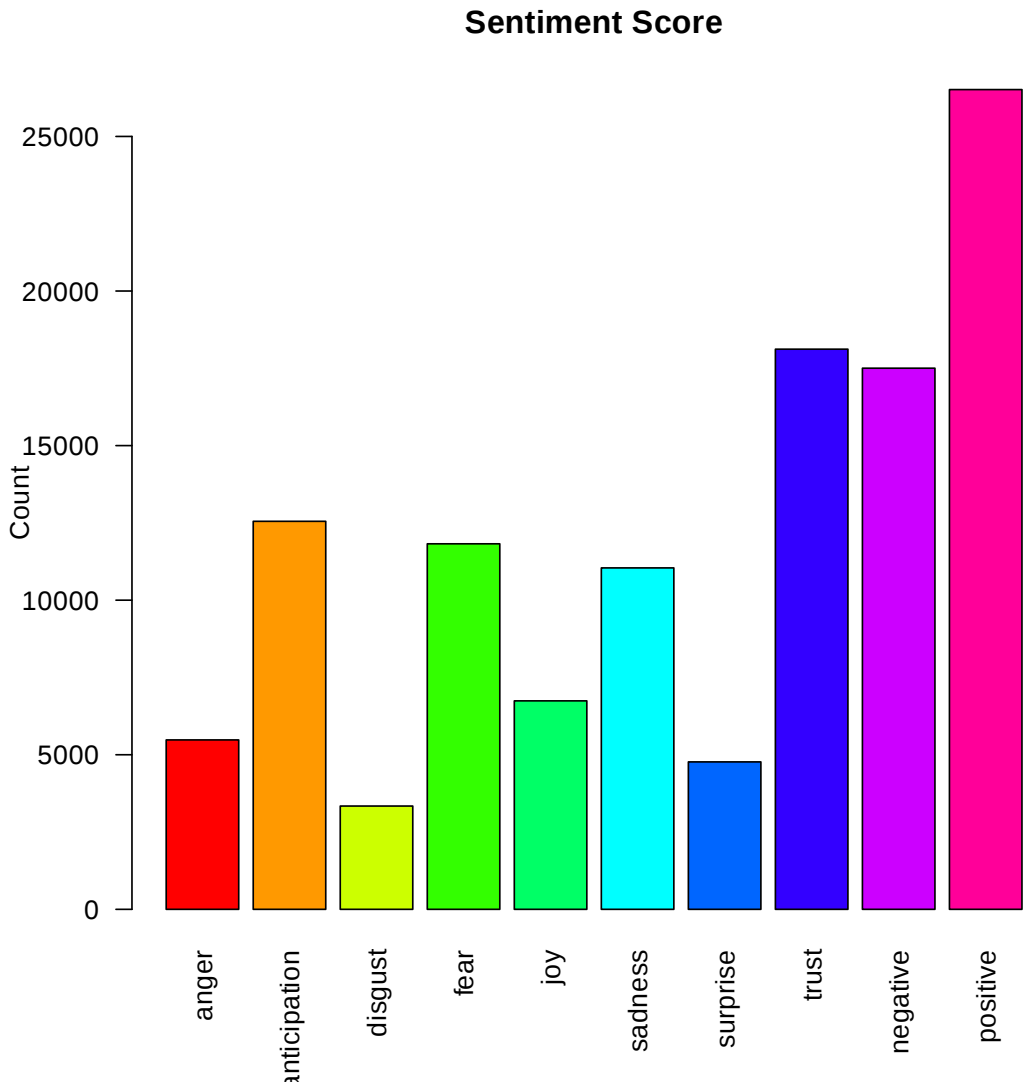
[64]:
```
#male average sentiment
average_male<- average_sentiment(male_sentiment)
summary(average_male)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-46.000   0.000   1.000   3.271   6.000  64.000
```

[95]:
```
#african barplot
barplot_function(clean.data.african$Text)
```

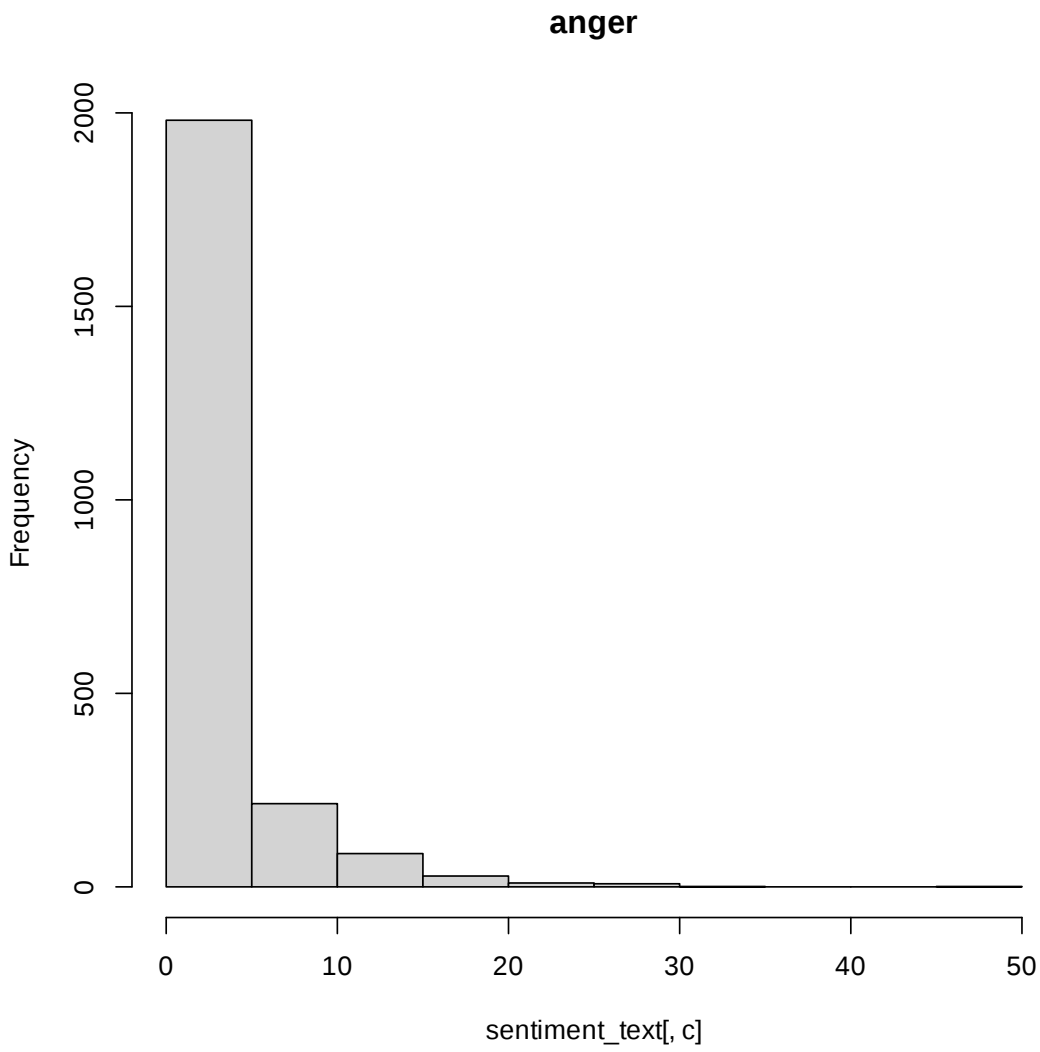
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 3.000   Median : 0.000   Median : 2.000
Mean   : 2.352   Mean   : 5.387   Mean   : 1.434   Mean   : 5.074
3rd Qu.: 3.000   3rd Qu.: 7.000   3rd Qu.: 2.000   3rd Qu.: 6.000
Max.   :46.000   Max.   :48.000   Max.   :26.000   Max.   :67.000
     joy            sadness         surprise          trust
Min.   : 0.000   Min.   : 0.00   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000
Median : 1.000   Median : 2.00   Median : 1.000   Median : 5.000
Mean   : 2.894   Mean   : 4.74   Mean   : 2.047   Mean   : 7.776
3rd Qu.: 4.000   3rd Qu.: 6.00   3rd Qu.: 3.000   3rd Qu.:11.000
Max.   :40.000   Max.   :58.00   Max.   :19.000   Max.   :77.000
   negative          positive
Min.   :  0.000   Min.   :  0.00
1st Qu.:  0.000   1st Qu.:  0.00
Median :  2.000   Median :  7.00
Mean   :  7.512   Mean   : 11.38
3rd Qu.:  9.000   3rd Qu.: 15.00
Max.   :112.000   Max.   :135.00
```

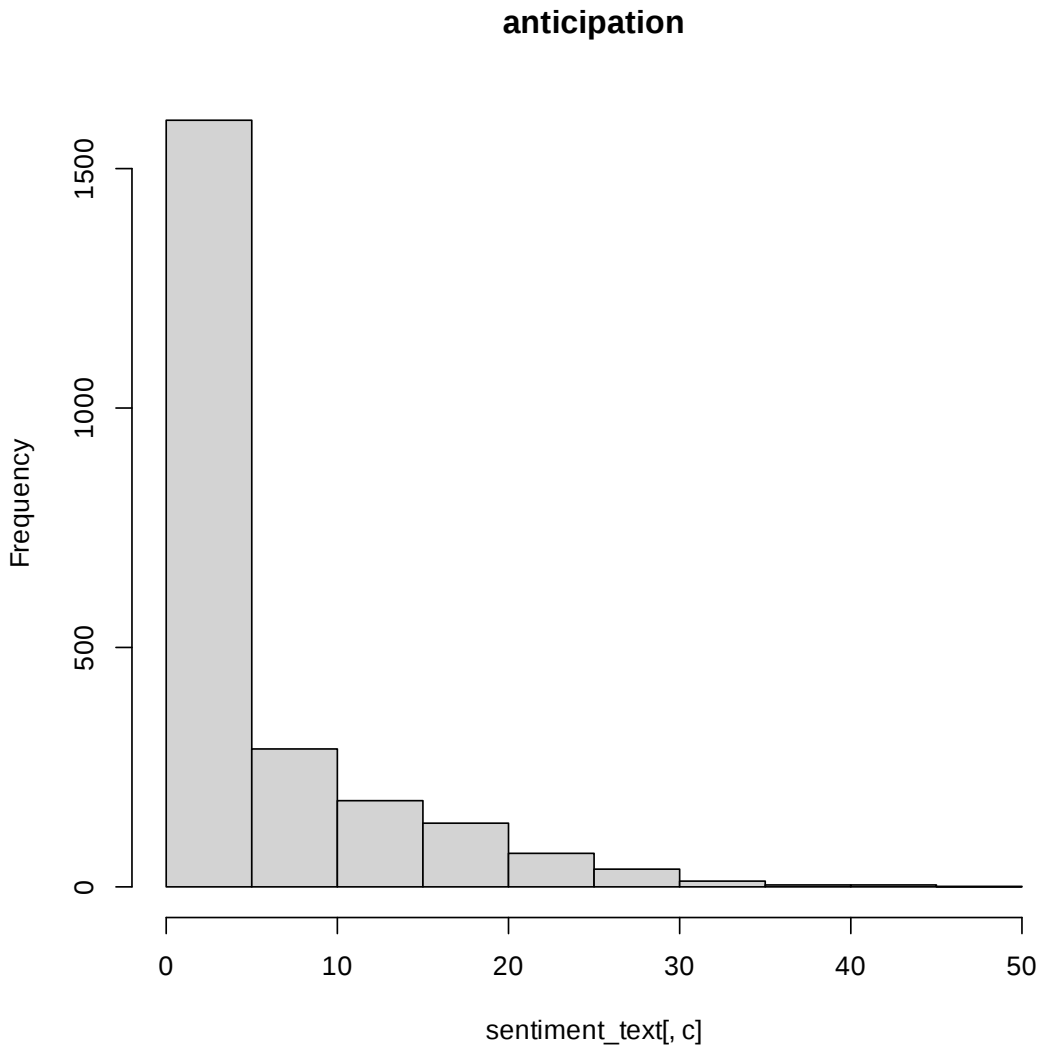


[96]:
```
#african sentiment histogram
sentiment_hist(african_sentiment)
```

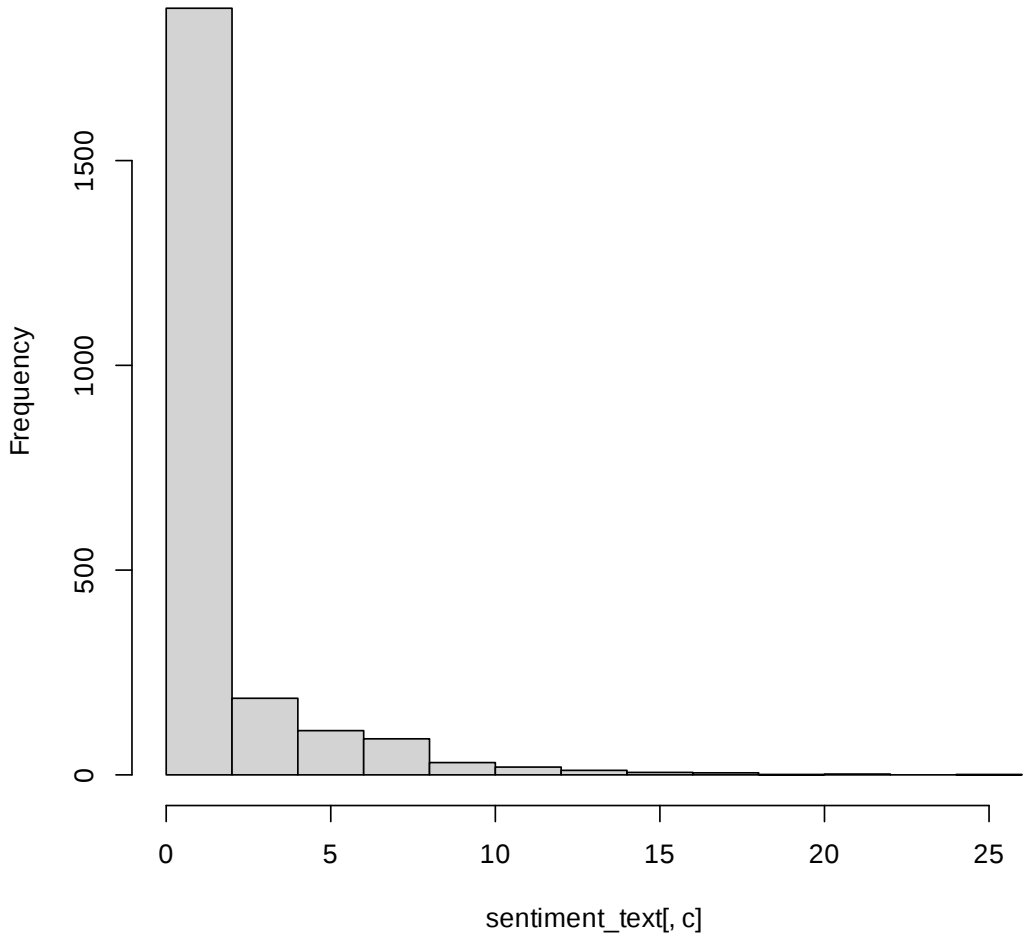
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
500
1000
1500
2000
0
10
20
30
40
50
sentiment_text[, c]

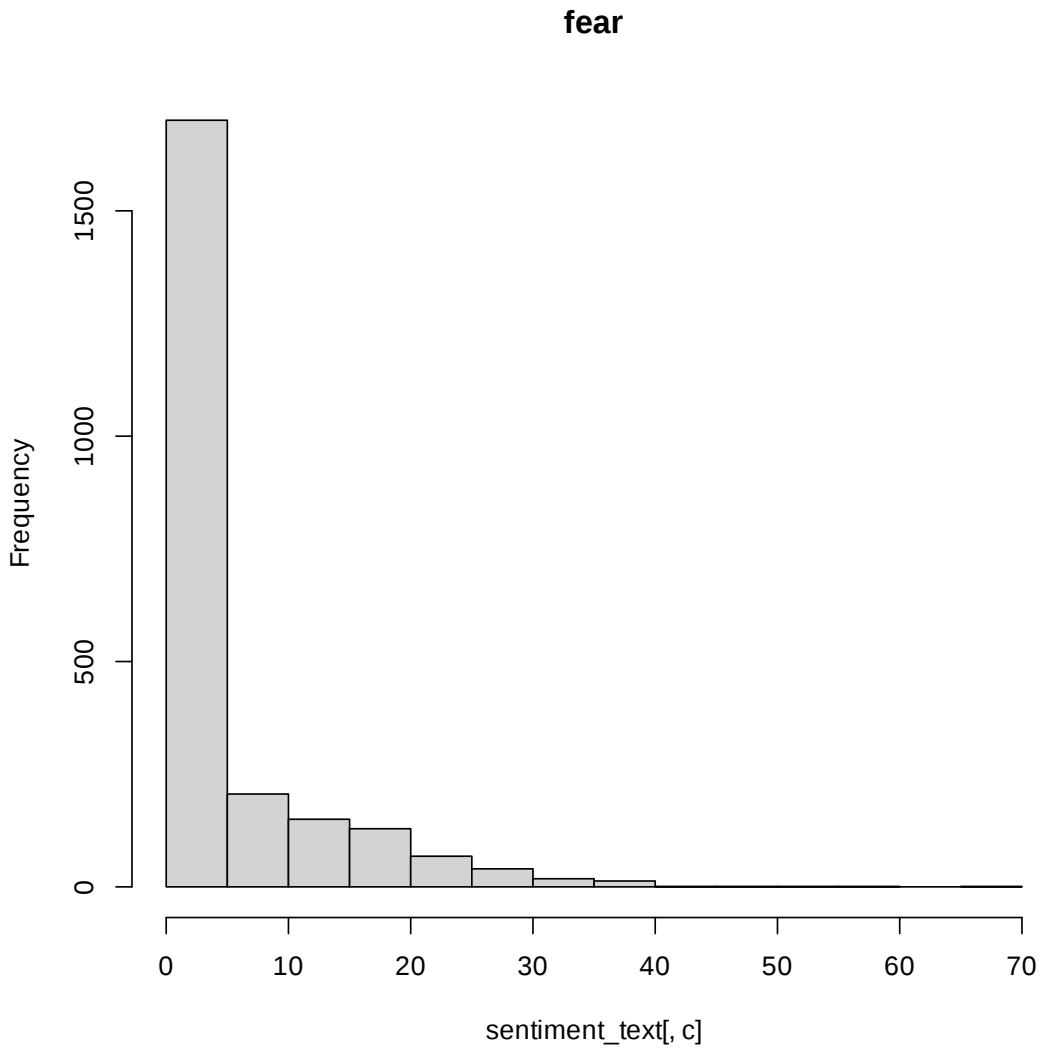
```
[1] "disgust"
```


anticipation
Frequency
0
500
1000
1500
0
10
20
30
40
50
sentiment_text[, c]

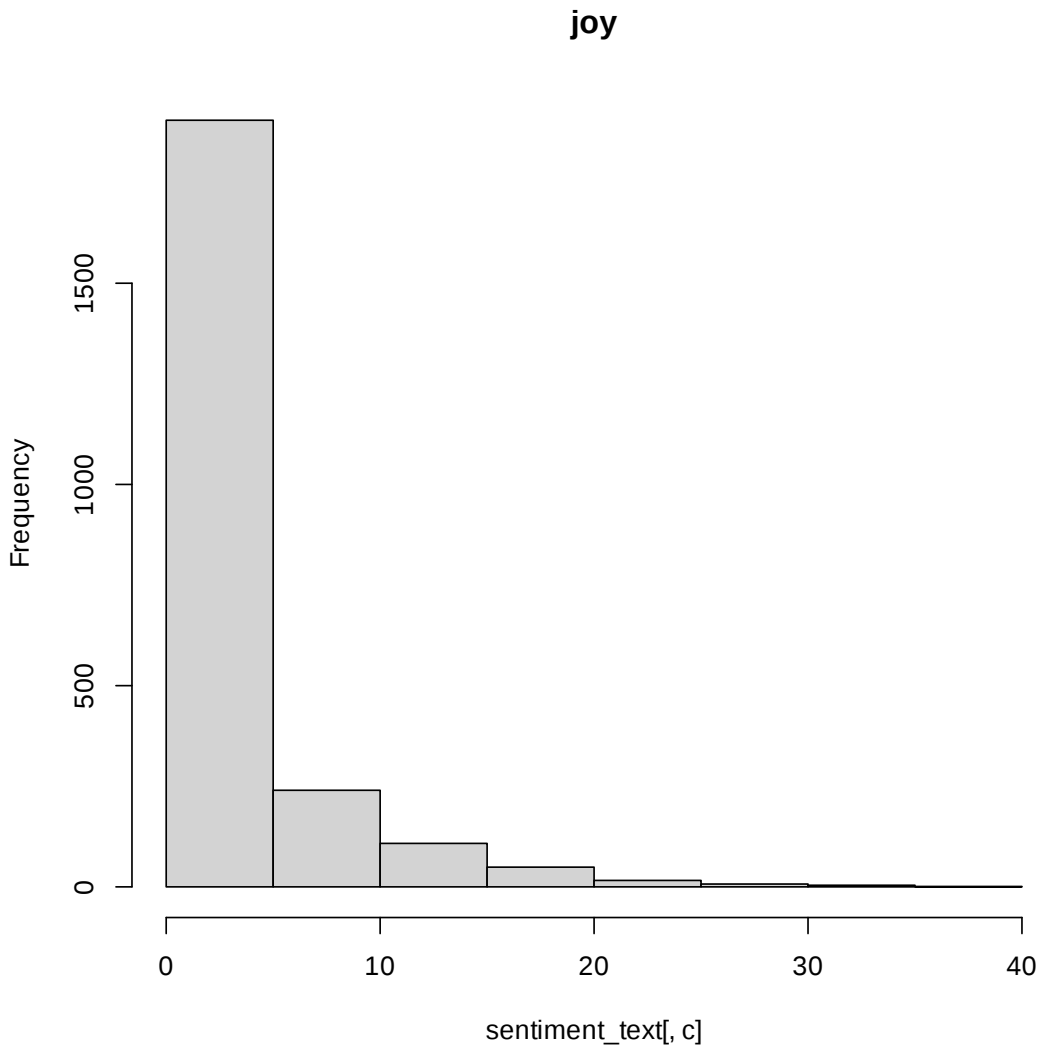
```
[1] "fear"
```


disgust
Frequency
0
500
1000
1500
0
5
10
15
20
25
sentiment_text[, c]

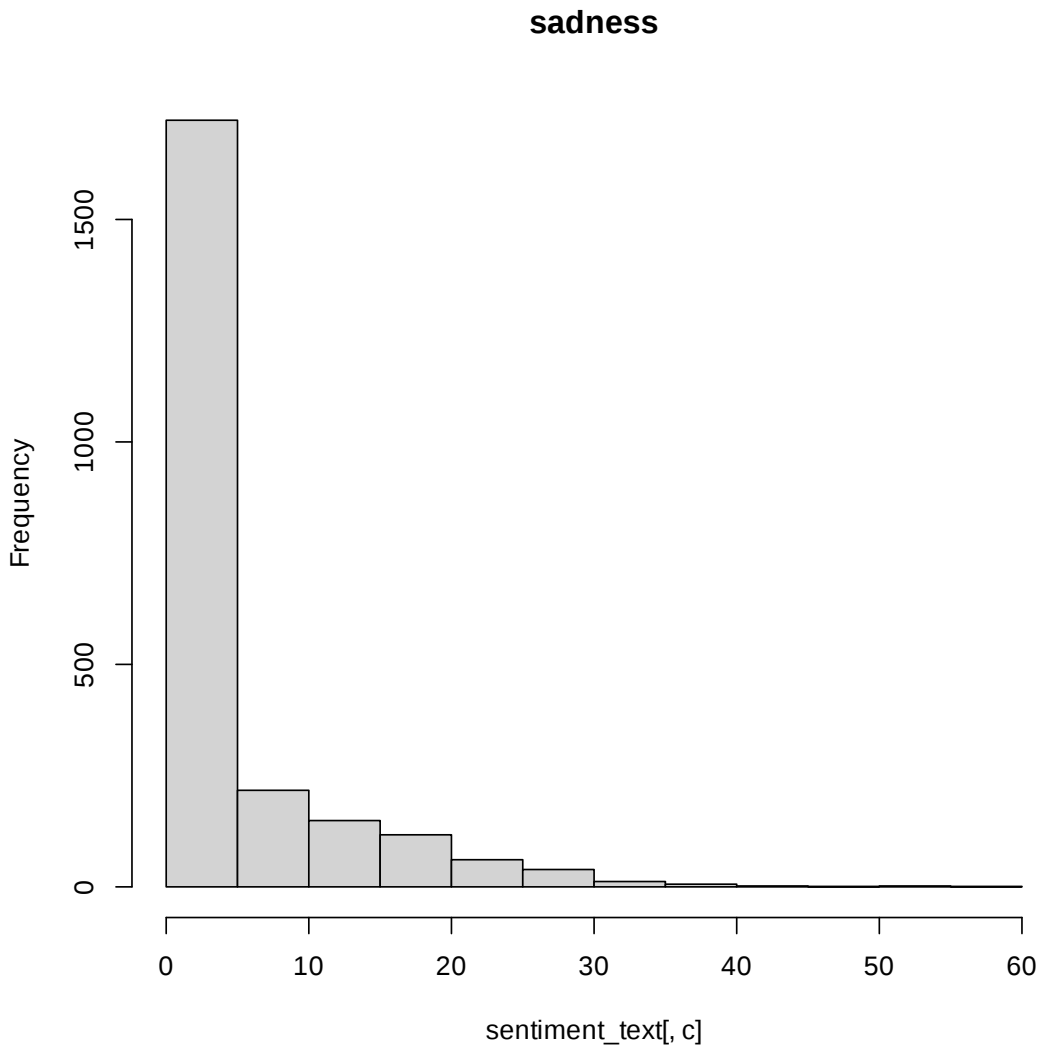
```
[1] "joy"
```


fear
Frequency
0
500
1000
1500
0
10
20
30
40
50
60
70
sentiment_text[, c]

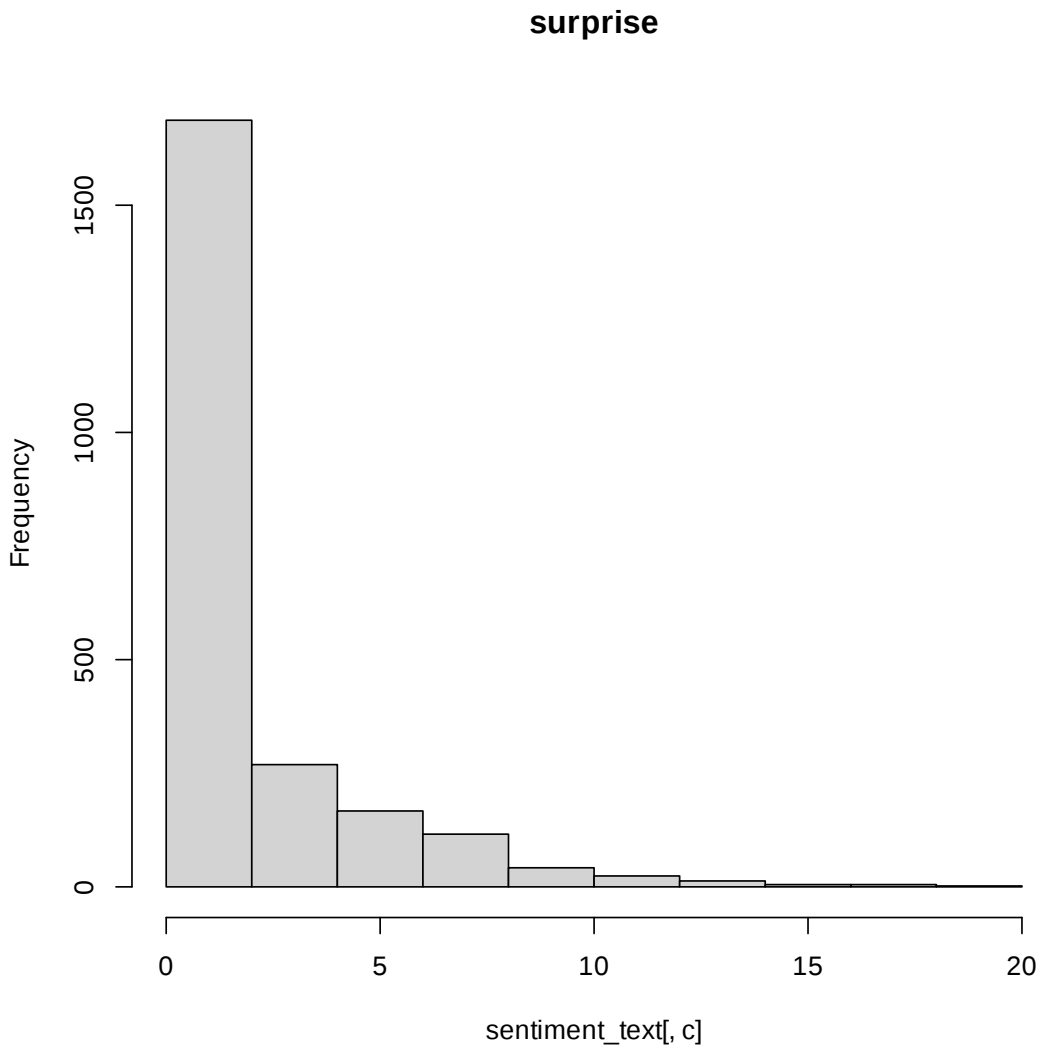
```
[1] "sadness"
```



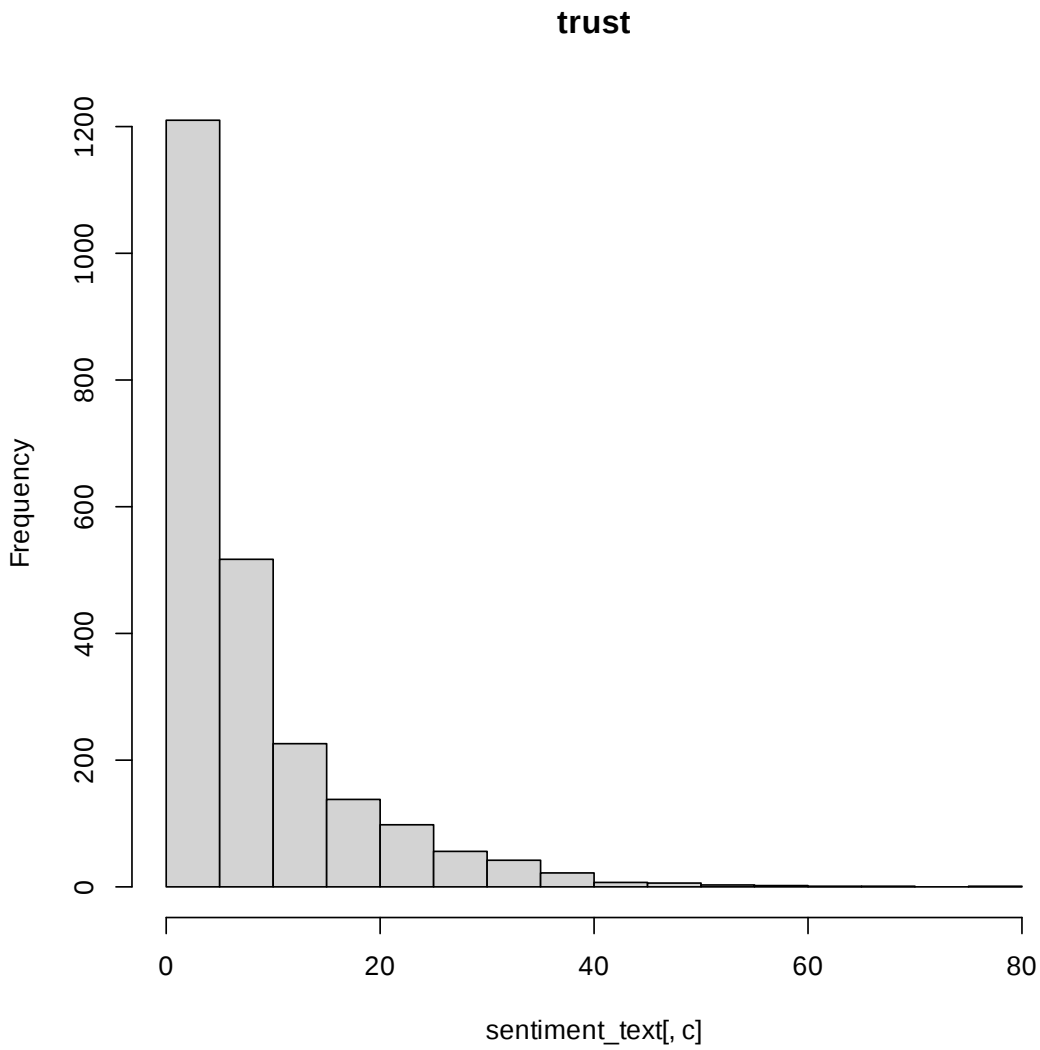

```
[1] "surprise"
```


sadness
Frequency
0
500
1000
1500
0
10
20
30
40
50
60
sentiment_text[, c]

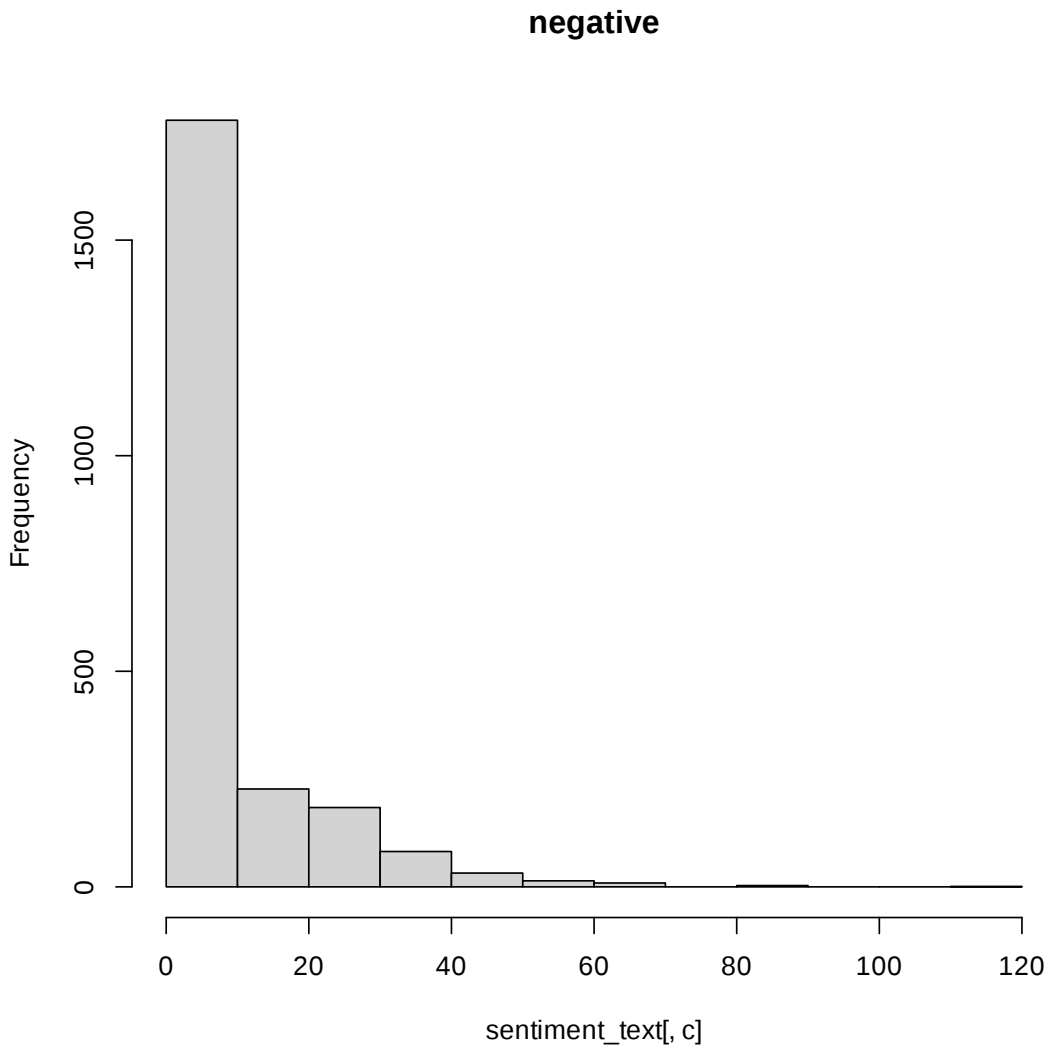
```
[1] "trust"
```

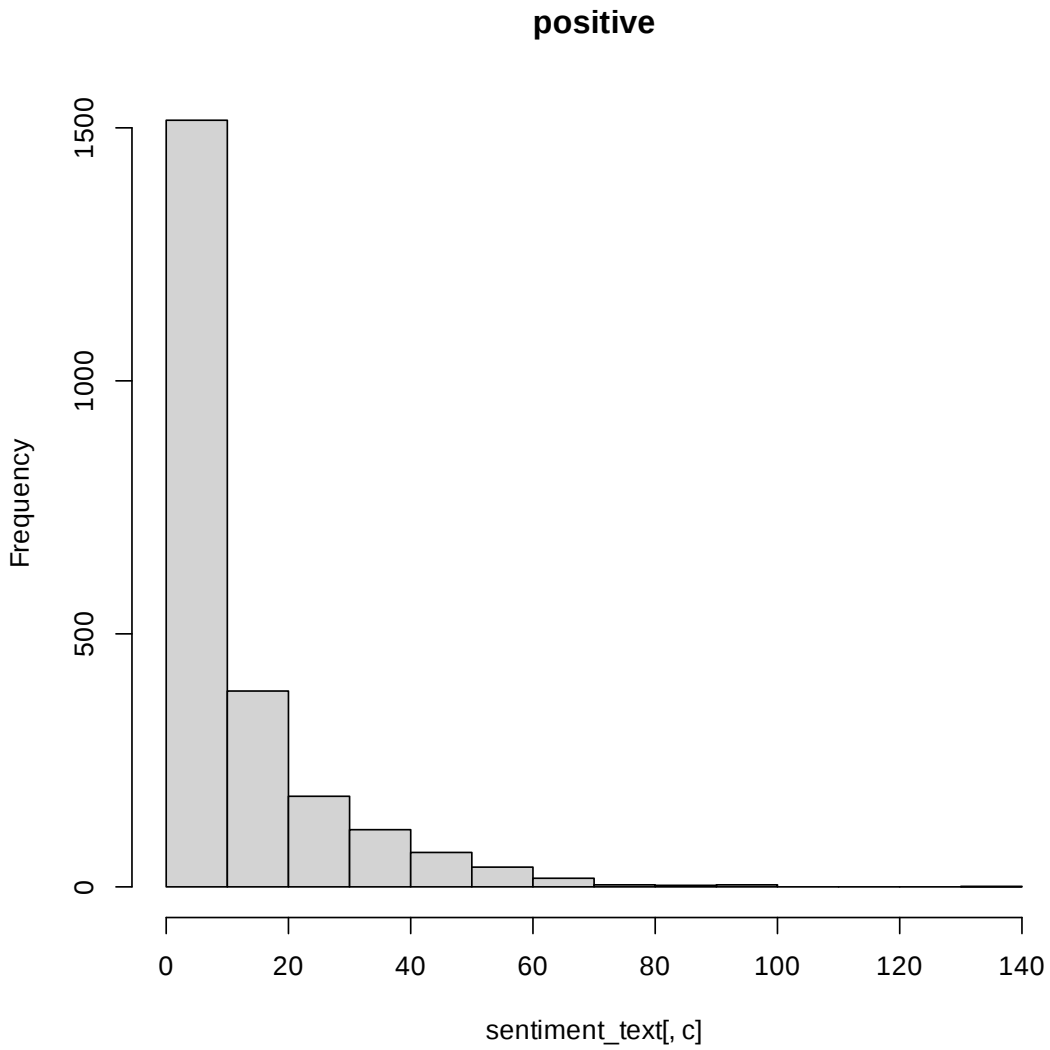

surprise
Frequency
0
500
1000
1500
0
5
10
15
20
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
200
400
600
800
1000
1200
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
500
1000
1500
0
20
40
60
80
100
120
sentiment_text[, c]



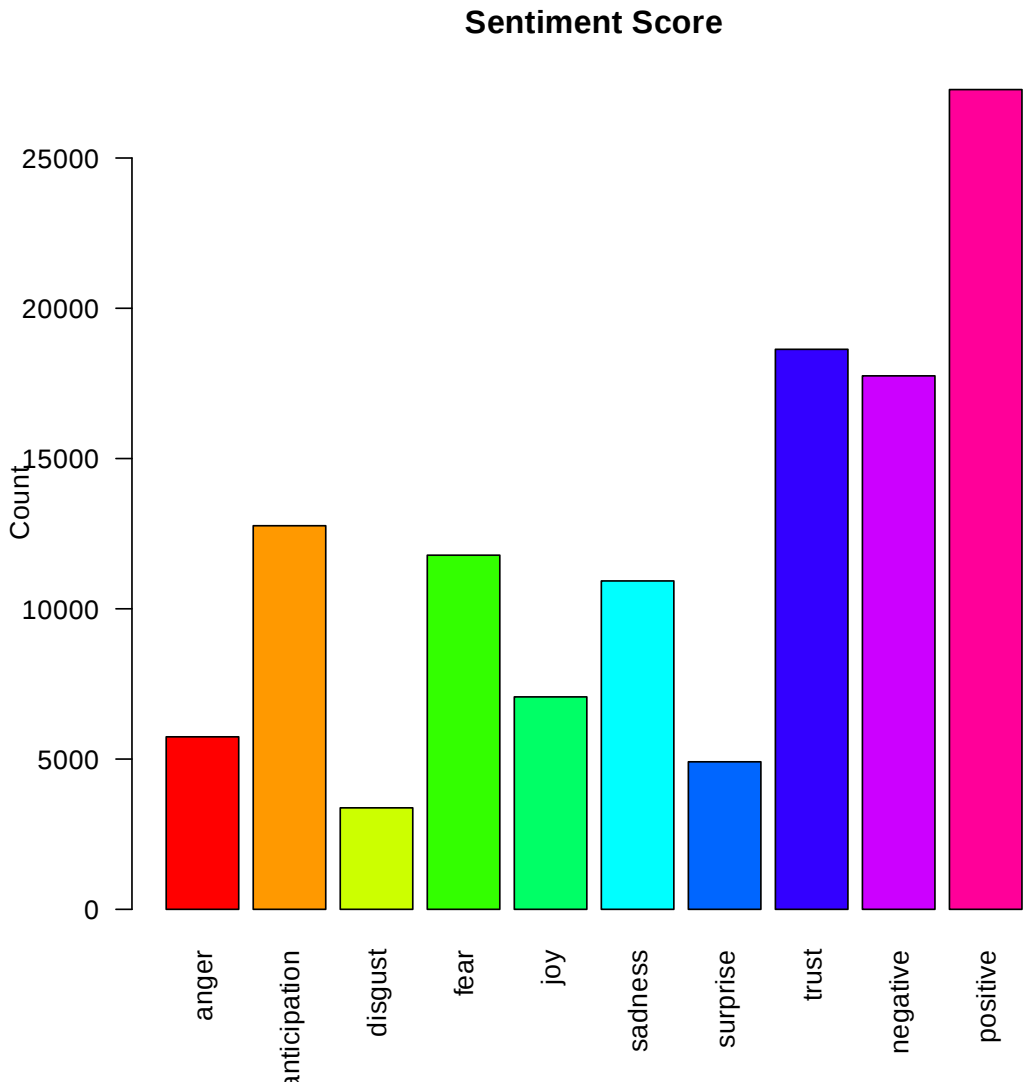
[98]:
```
#african average sentiment
average_african<- average_sentiment(african_sentiment)
summary(average_african)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-24.000   0.000   2.000   3.868   6.000  71.000
```

[57]:
```
#black barplot
barplot_function(clean.data.black$Text)
```

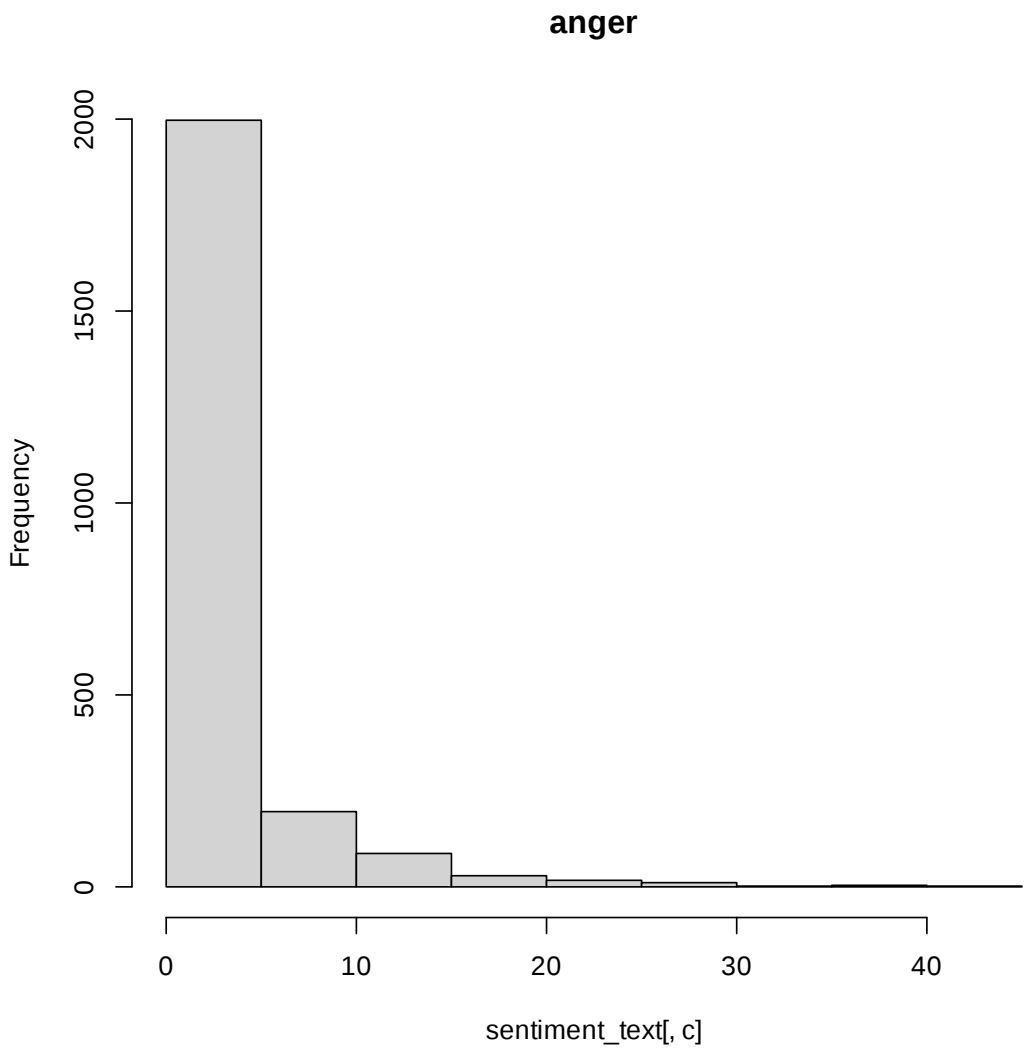
```
     anger          anticipation       disgust           fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.00   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000
```

```
Median : 0.000   Median : 3.000   Median : 0.00   Median : 2.000
Mean   : 2.448   Mean   : 5.443   Mean   : 1.44   Mean   : 5.025
3rd Qu.: 2.000   3rd Qu.: 7.000   3rd Qu.: 2.00   3rd Qu.: 6.000
Max.   :45.000   Max.   :48.000   Max.   :24.00   Max.   :67.000
     joy            sadness         surprise         trust
Min.   : 0.000   Min.   : 0.00   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000
Median : 1.000   Median : 2.00   Median : 1.000   Median : 5.000
Mean   : 3.014   Mean   : 4.66   Mean   : 2.093   Mean   : 7.946
3rd Qu.: 4.000   3rd Qu.: 6.00   3rd Qu.: 3.000   3rd Qu.:11.000
Max.   :34.000   Max.   :53.00   Max.   :25.000   Max.   :77.000
   negative         positive
Min.   :  0.00   Min.   :  0.00
1st Qu.:  0.00   1st Qu.:  0.00
Median :  2.00   Median :  7.00
Mean   :  7.57   Mean   : 11.63
3rd Qu.:  9.00   3rd Qu.: 15.00
Max.   :103.00   Max.   :116.00
```

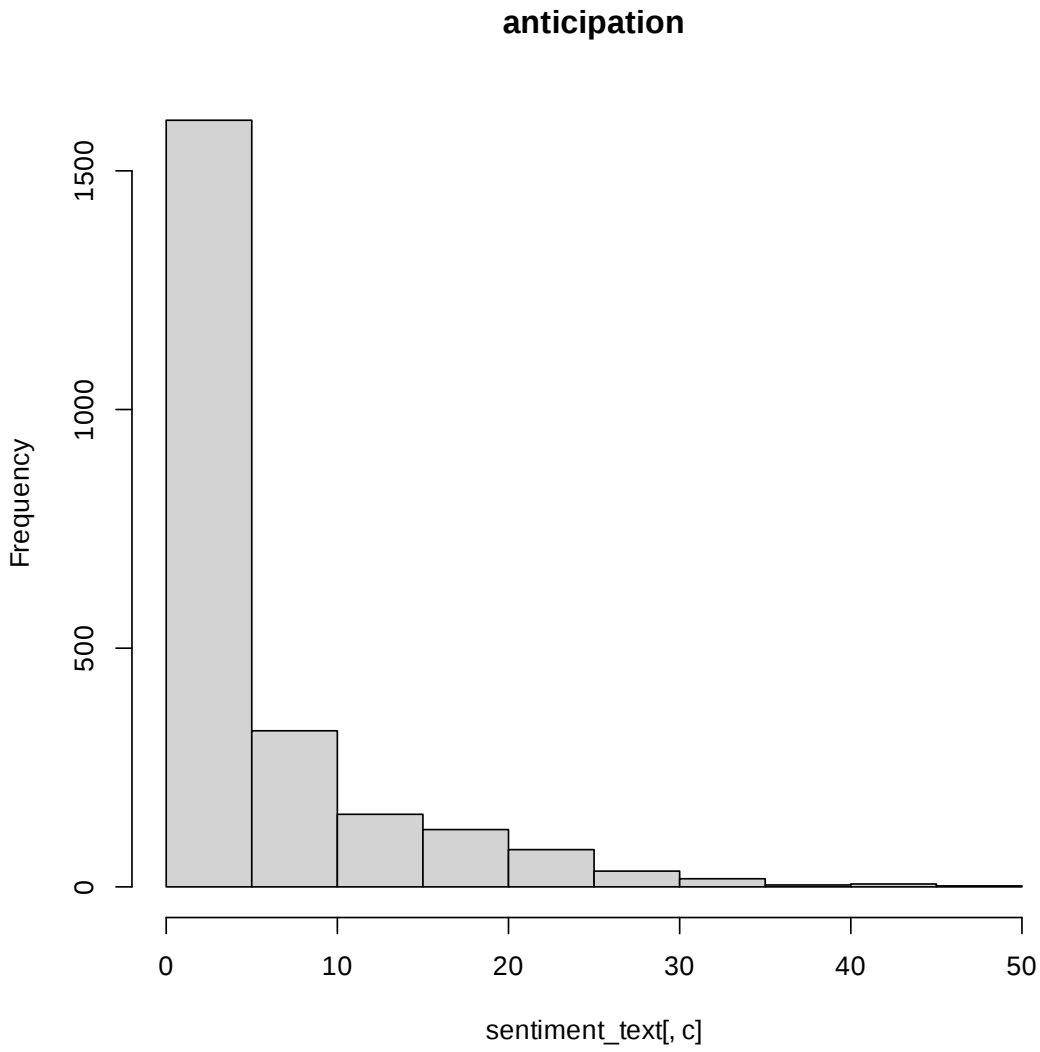


[82]:
```
#black sentiment histogram
sentiment_hist(black_sentiment)
```

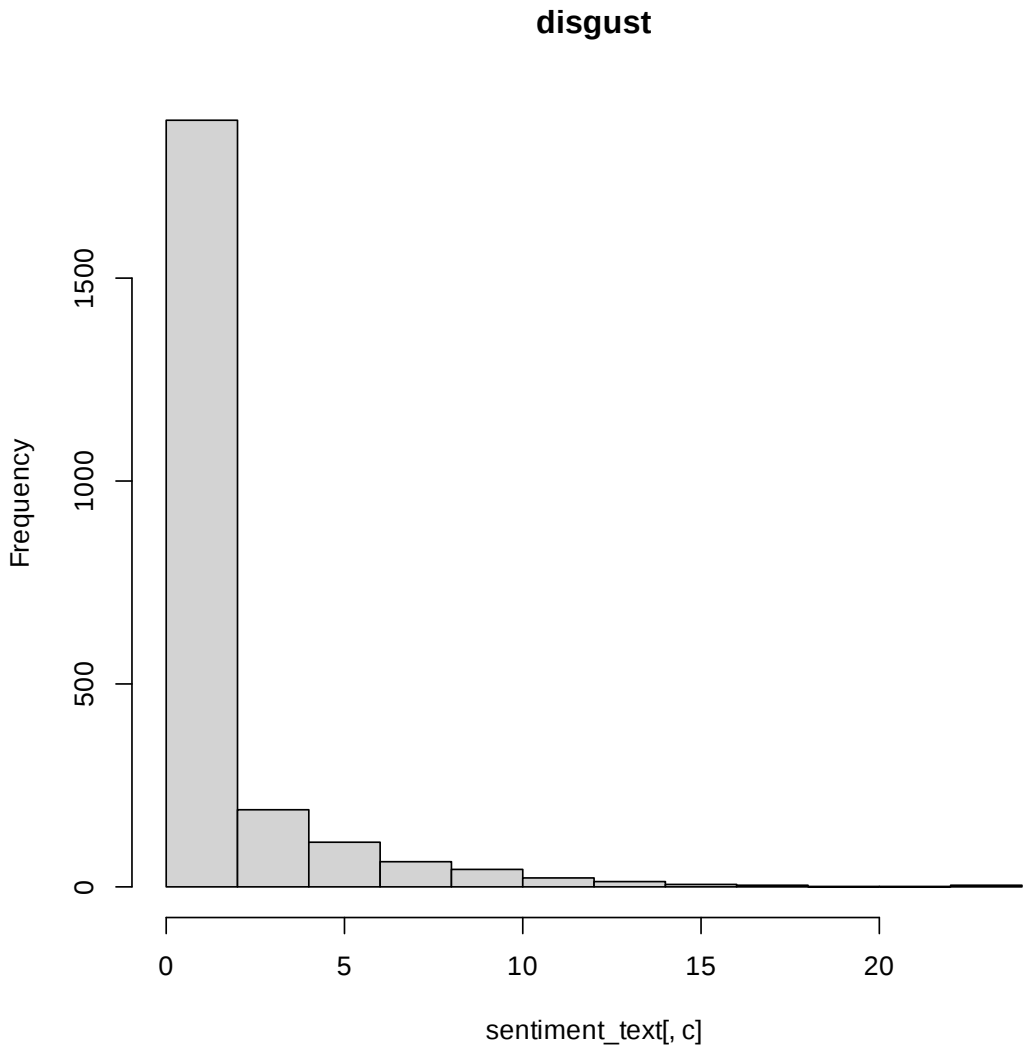
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
500
1000
1500
2000
0
10
20
30
40
sentiment_text[, c]

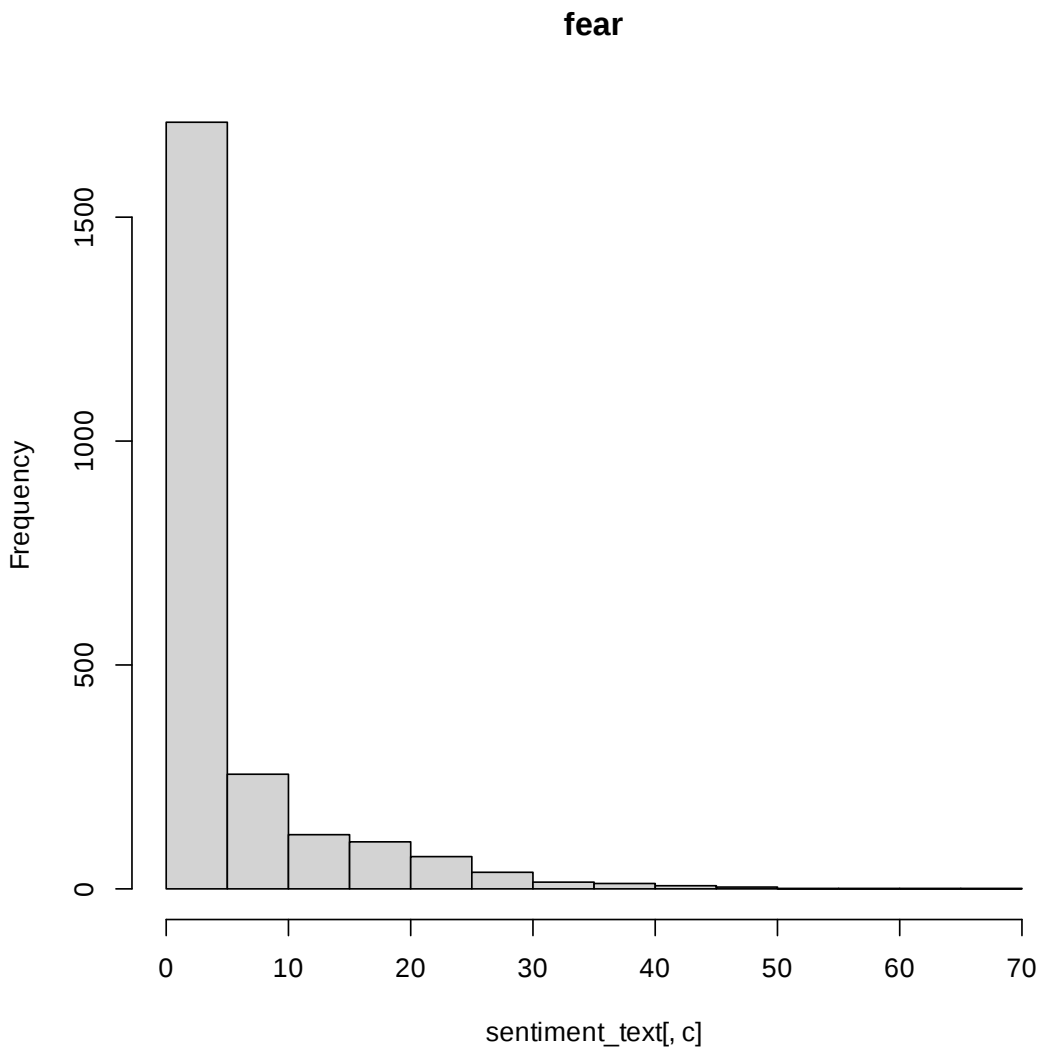
```
[1] "disgust"
```


anticipation
Frequency
0
500
1000
1500
0
10
20
30
40
50
sentiment_text[, c]

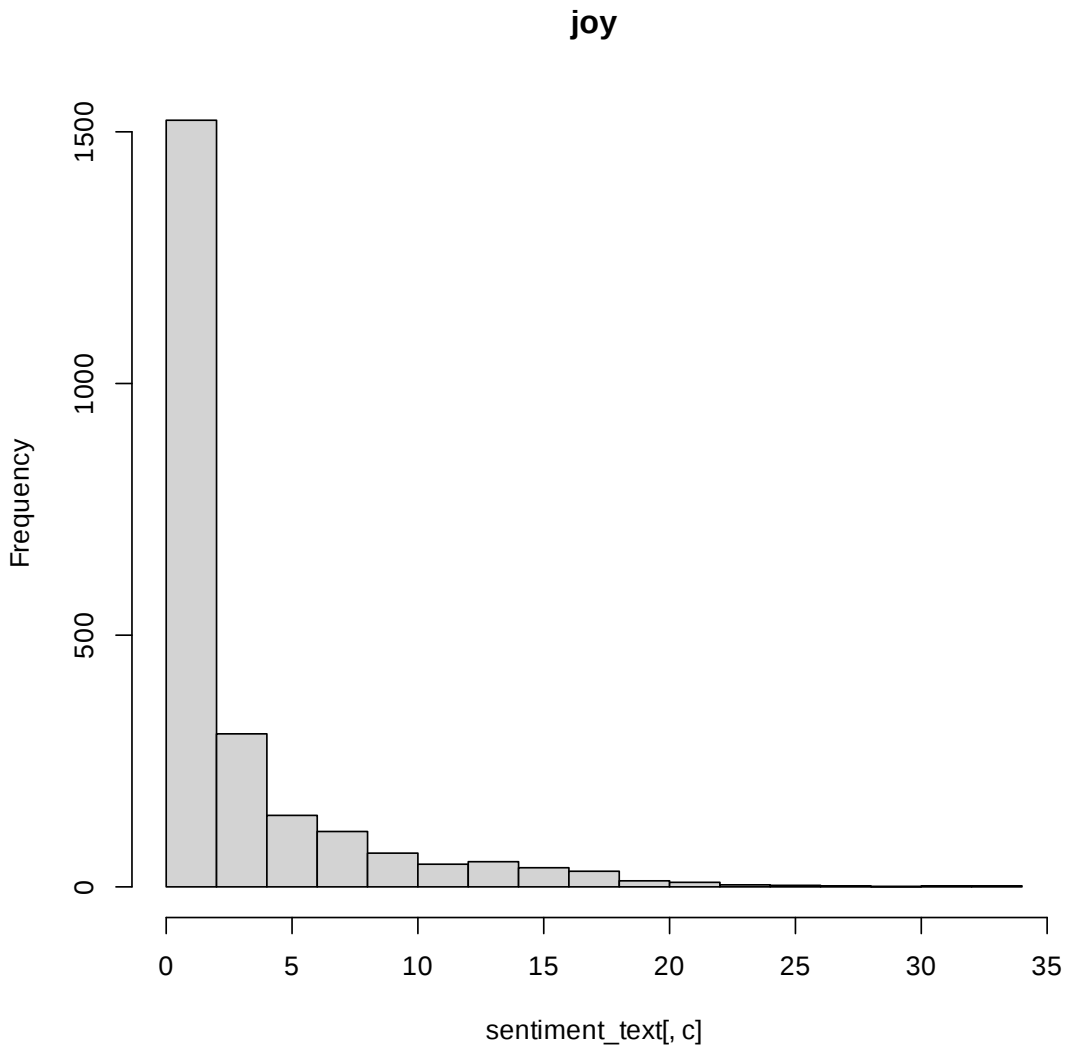
```
[1] "fear"
```


disgust
Frequency
0
500
1000
1500
0
5
10
15
20
sentiment_text[, c]

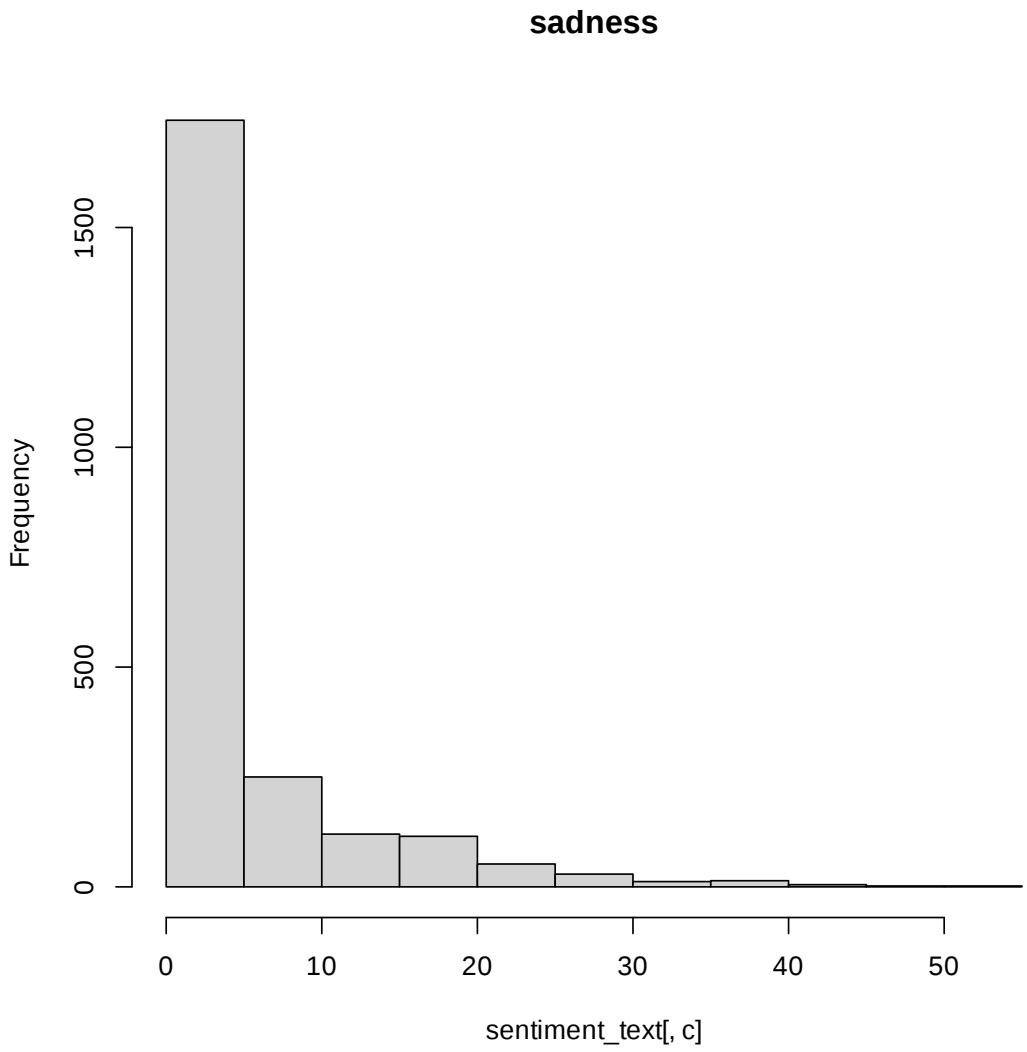
```
[1] "joy"
```


fear
Frequency
0
500
1000
1500
0
10
20
30
40
50
60
70
sentiment_text[, c]

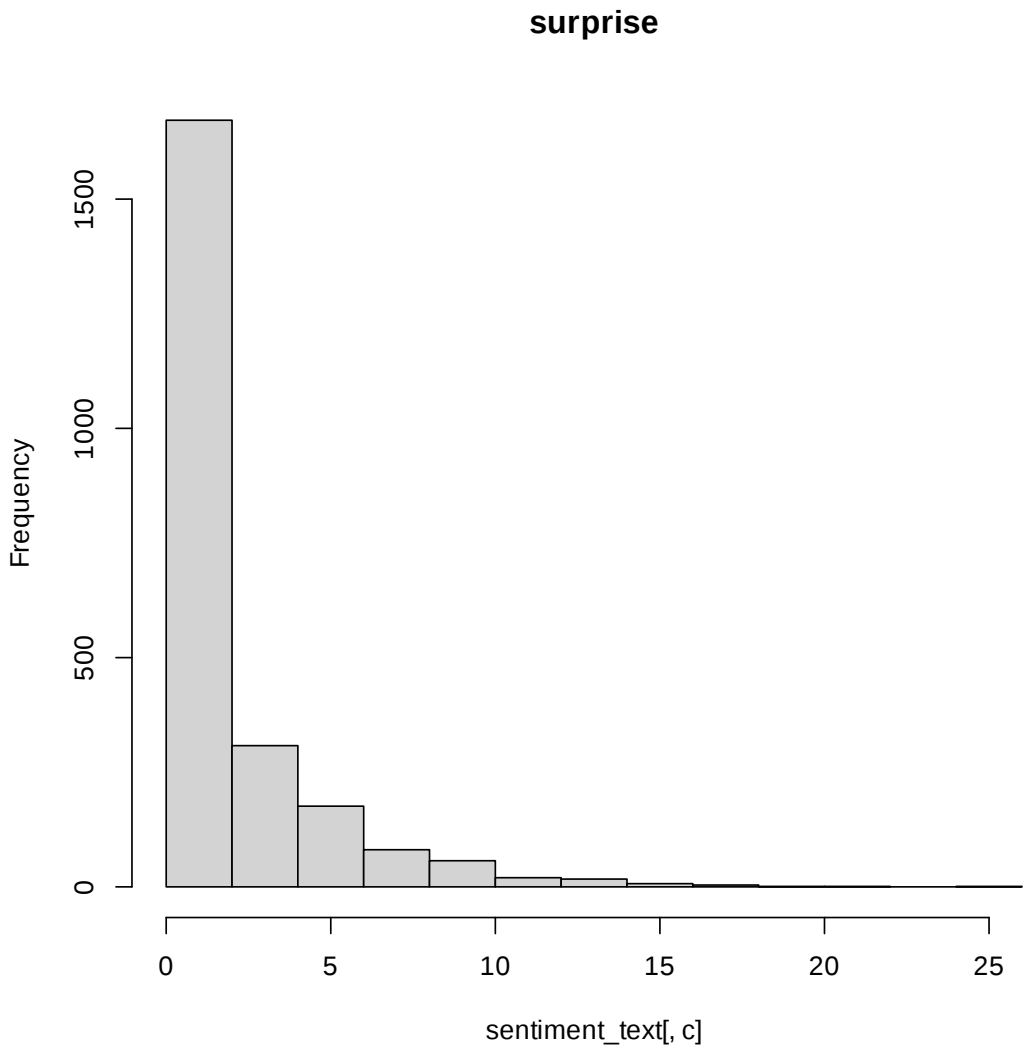
```
[1] "sadness"
```


joy
Frequency
0
500
1000
1500
0
5
10
15
20
25
30
35
sentiment_text[, c]

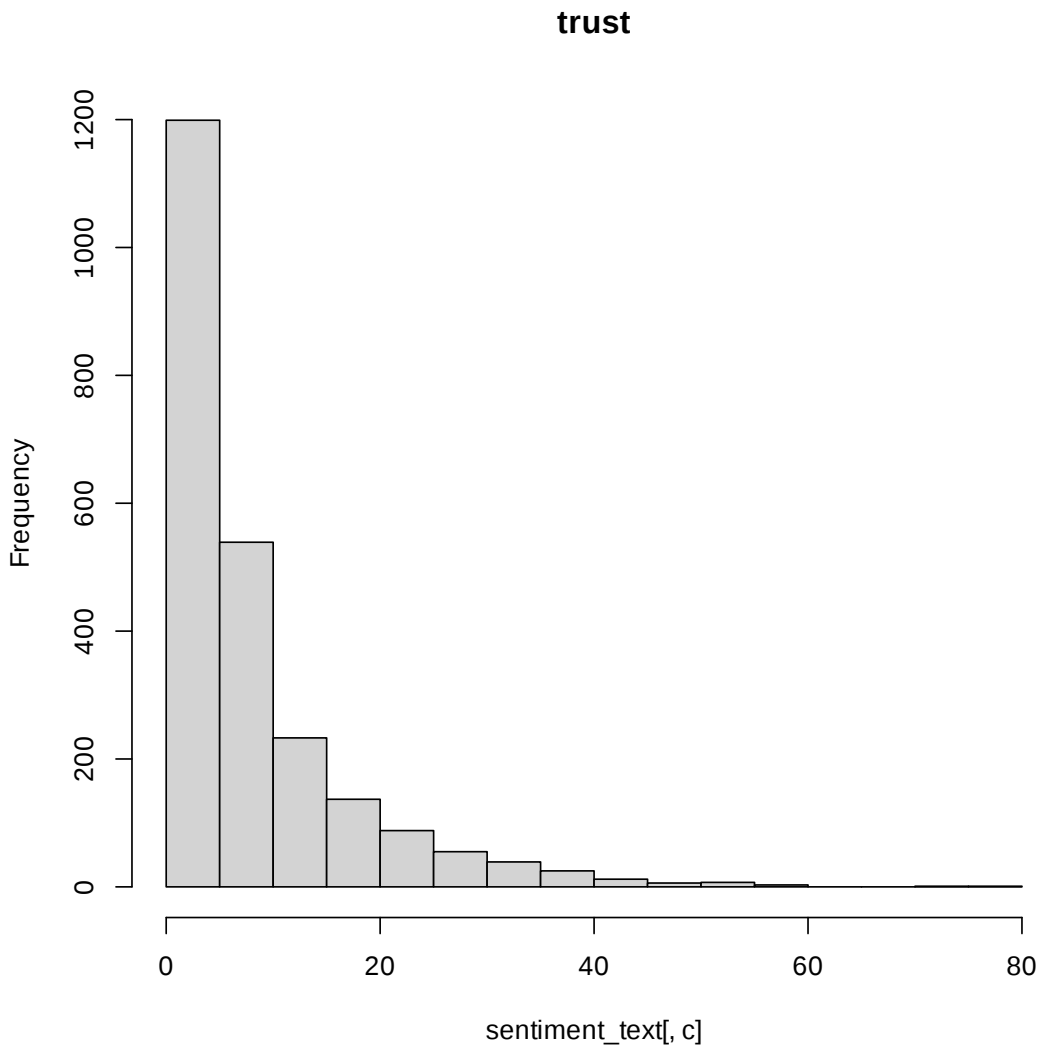
```
[1] "surprise"
```


sadness
Frequency
0
500
1000
1500
0
10
20
30
40
50
sentiment_text[, c]

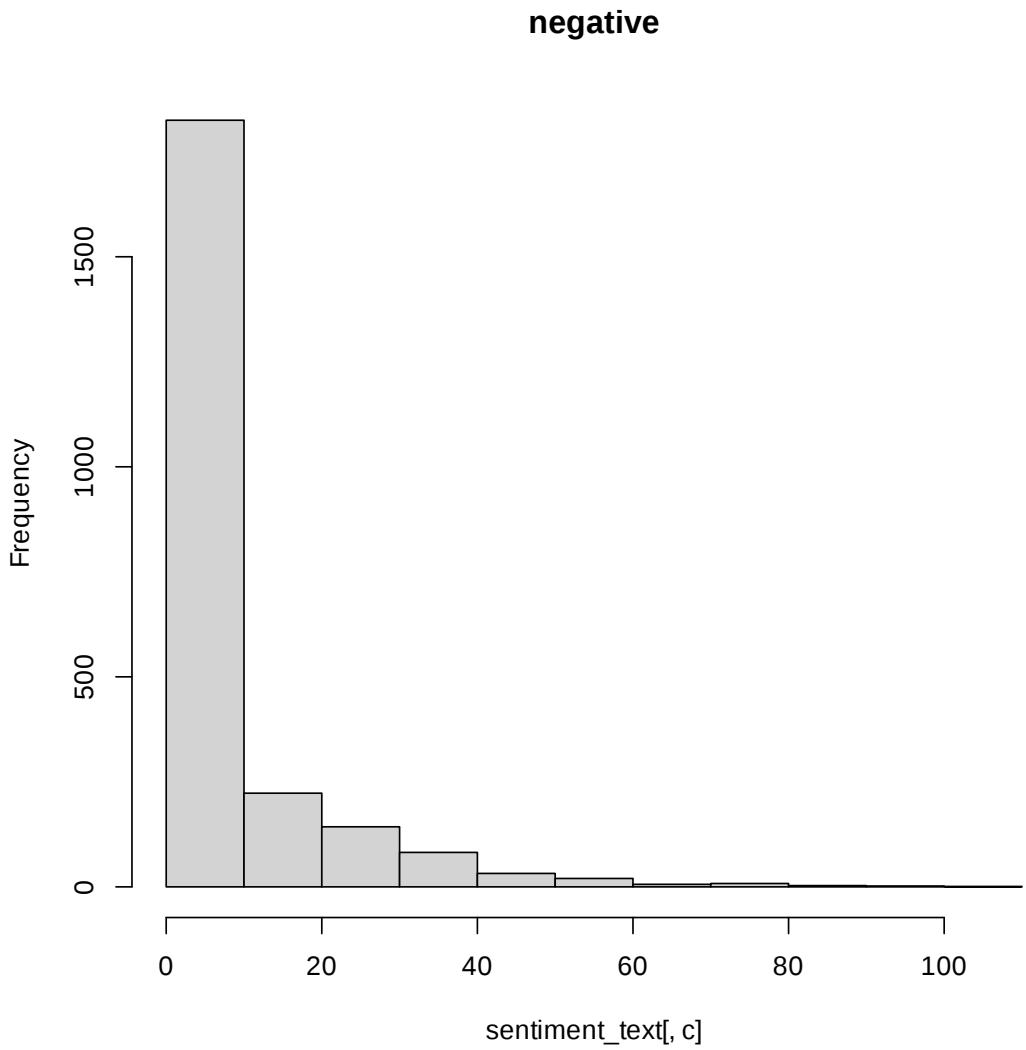
```
[1] "trust"
```

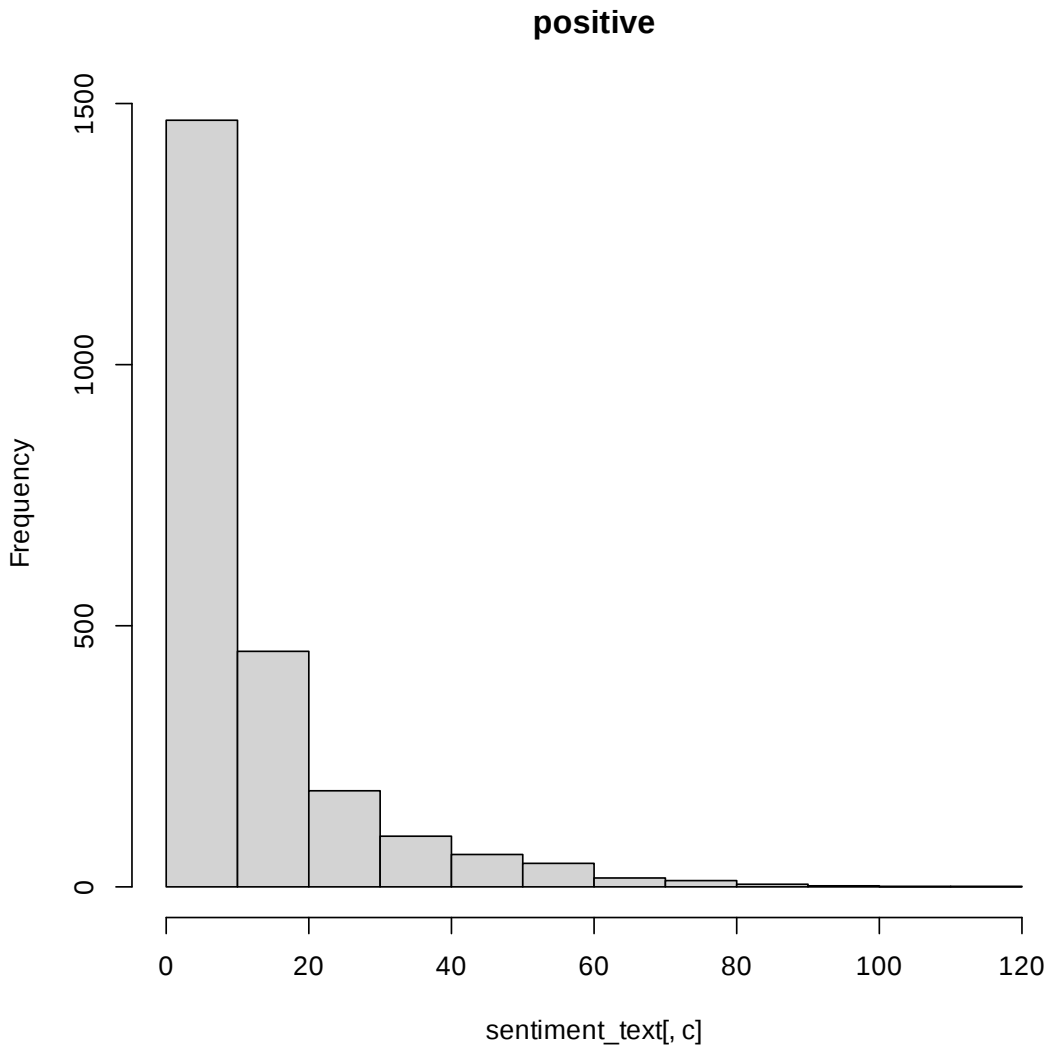

surprise
Frequency
0
500
1000
1500
0
5
10
15
20
25
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
200
400
600
800
1000
1200
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
500
1000
1500
0
20
40
60
80
100
sentiment_text[, c]



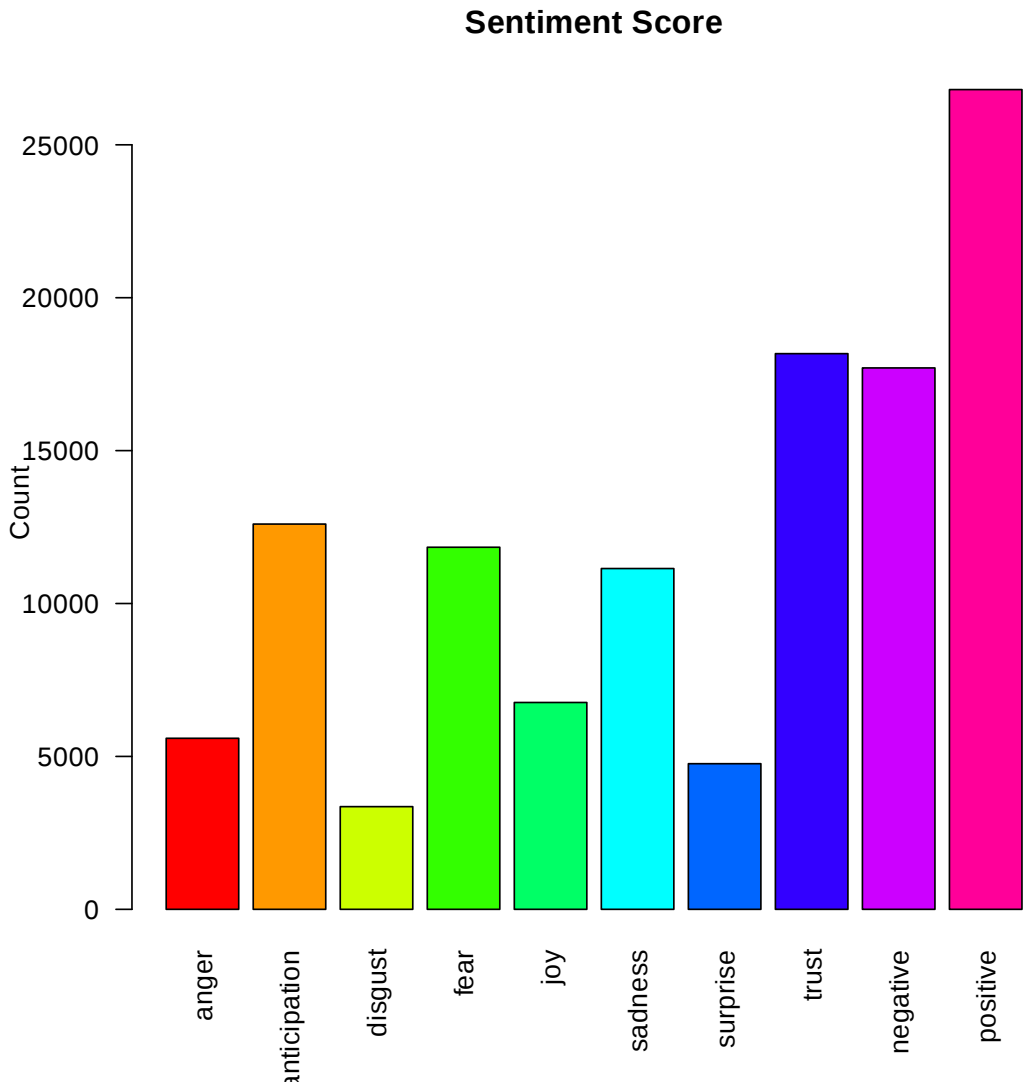
[84]:
```
#black average sentiment
average_black<- average_sentiment(black_sentiment)
summary(average_black)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-20.000   0.000   3.000   4.062   7.000  64.000
```

[58]:
```
#white barplot
barplot_function(clean.data.white$Text)
```

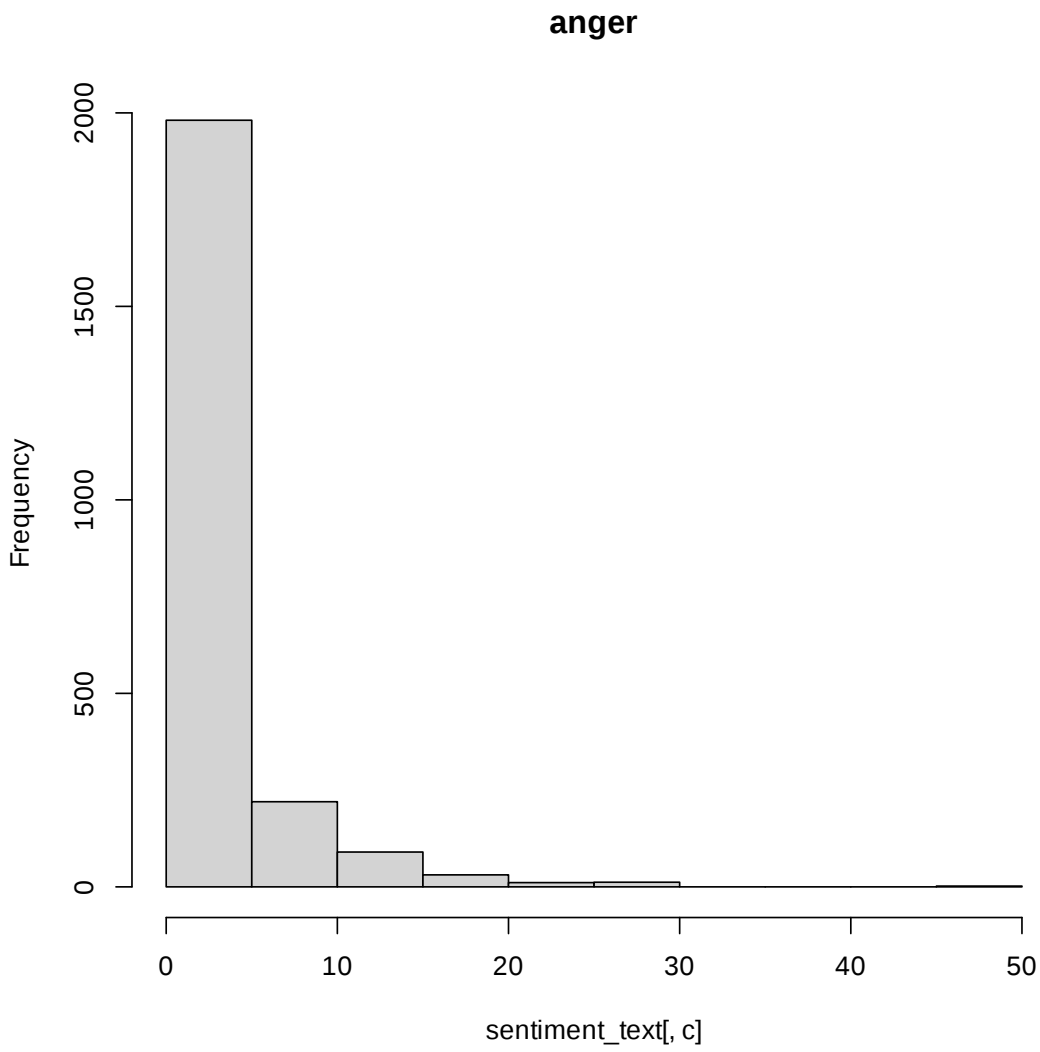
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 3.000   Median : 0.000   Median : 1.000
Mean   : 2.383   Mean   : 5.368   Mean   : 1.431   Mean   : 5.045
3rd Qu.: 3.000   3rd Qu.: 7.000   3rd Qu.: 2.000   3rd Qu.: 6.000
Max.   :49.000   Max.   :45.000   Max.   :30.000   Max.   :62.000
     joy            sadness          surprise         trust
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
Median : 1.000   Median : 1.000   Median : 1.000   Median : 5.000
Mean   : 2.882   Mean   : 4.748   Mean   : 2.029   Mean   : 7.742
3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 3.000   3rd Qu.:11.000
Max.   :29.000   Max.   :58.000   Max.   :20.000   Max.   :63.000
   negative          positive
Min.   :  0.000   Min.   :  0.00
1st Qu.:  0.000   1st Qu.:  0.00
Median :  2.000   Median :  7.00
Mean   :  7.543   Mean   : 11.42
3rd Qu.:  8.500   3rd Qu.: 15.00
Max.   :108.000   Max.   :101.00
```

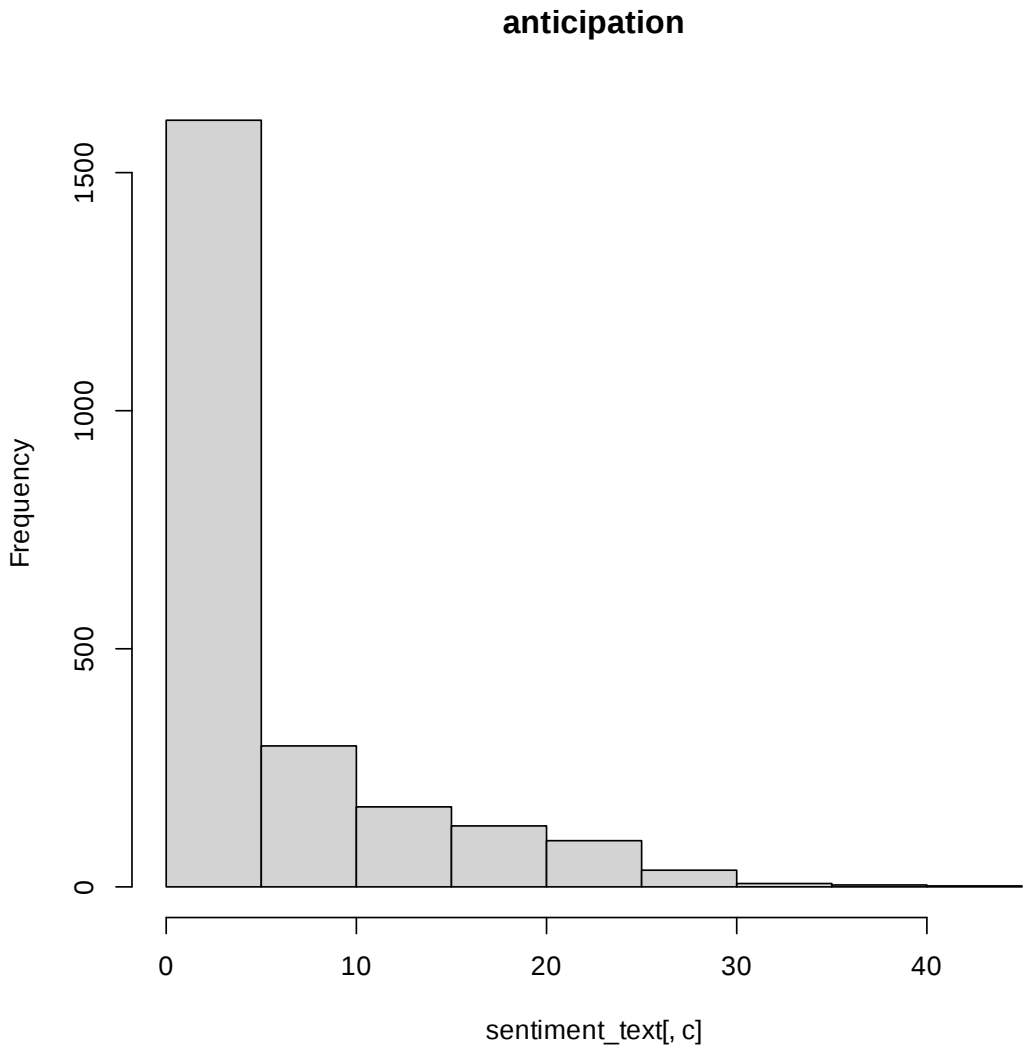


[85]:
```
#white sentiment histogram
sentiment_hist(white_sentiment)
```

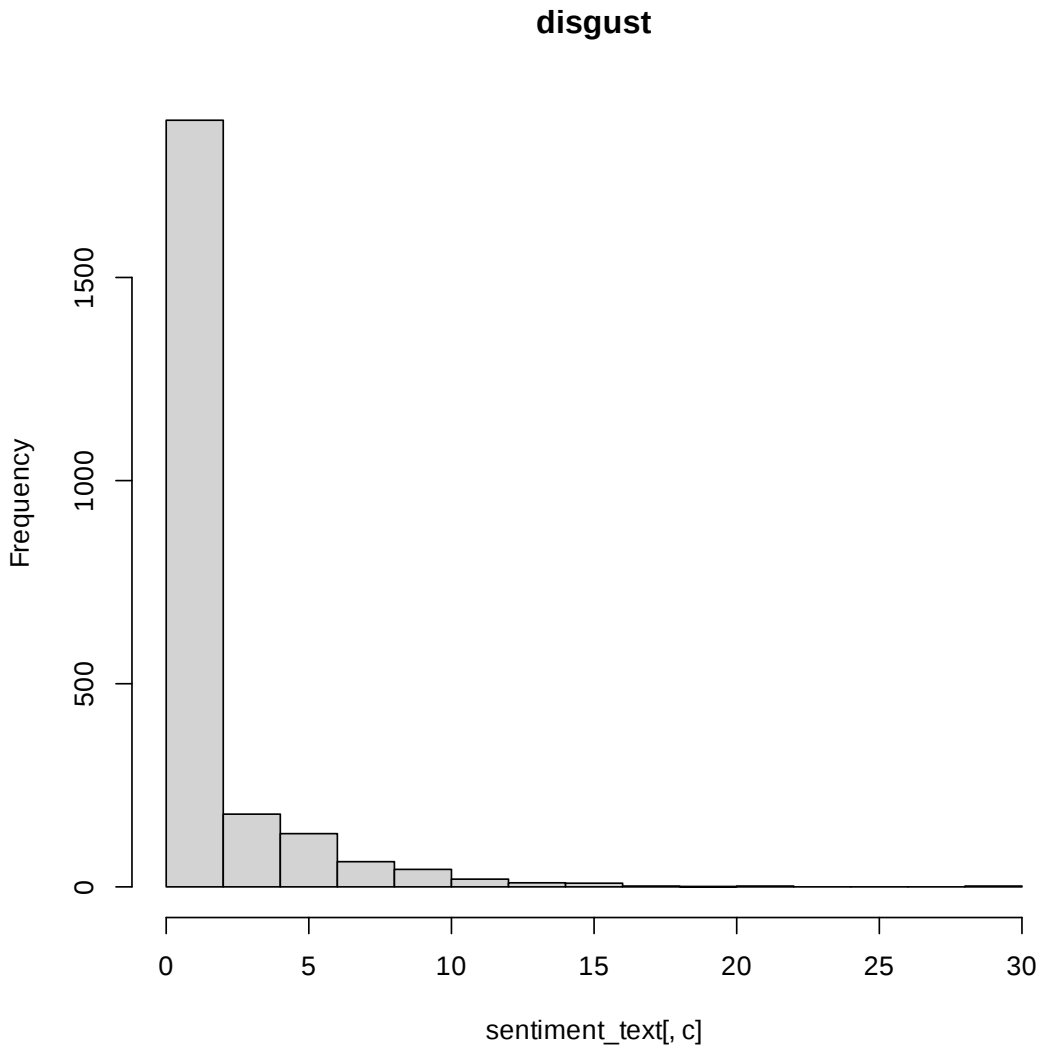
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
500
1000
1500
2000
0
10
20
30
40
50
sentiment_text[, c]

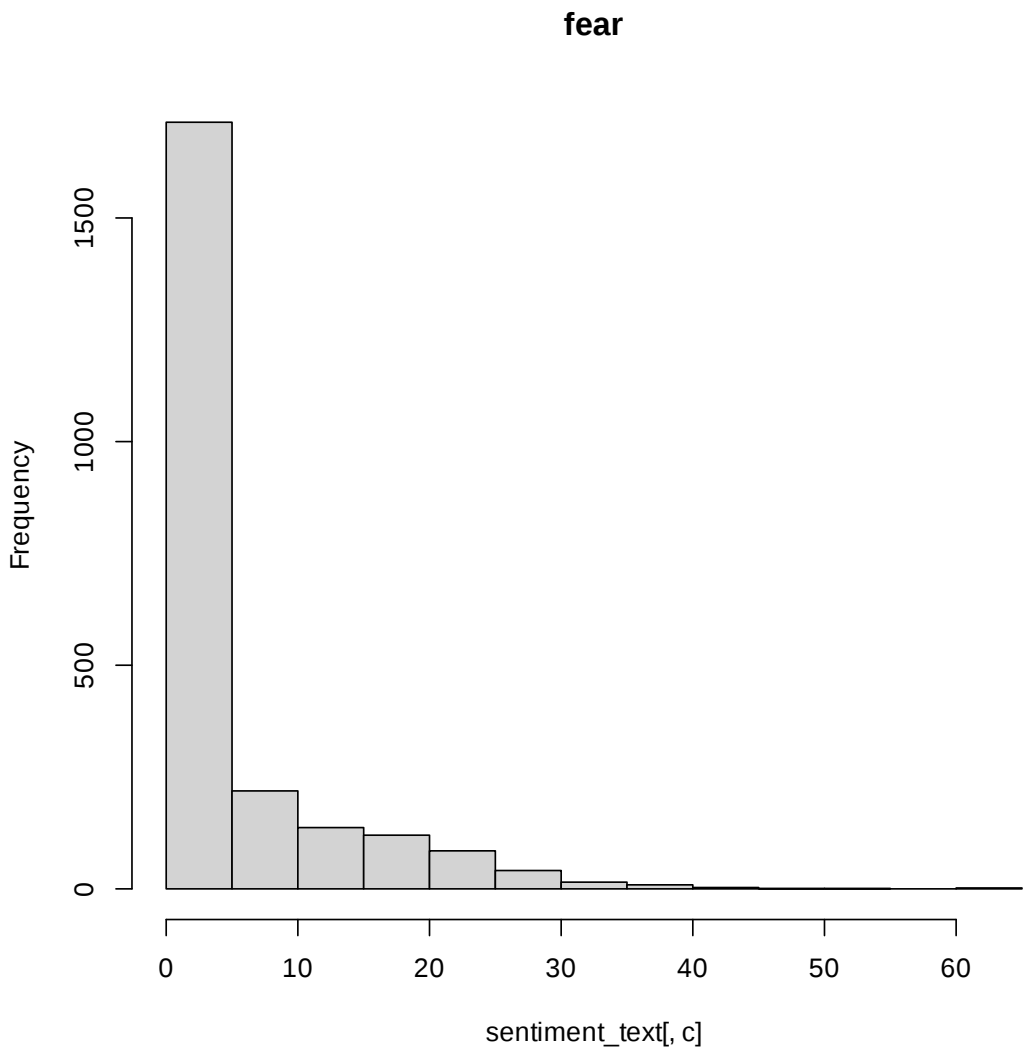
```
[1] "disgust"
```


anticipation
Frequency
0
500
1000
1500
0
10
20
30
40
sentiment_text[, c]

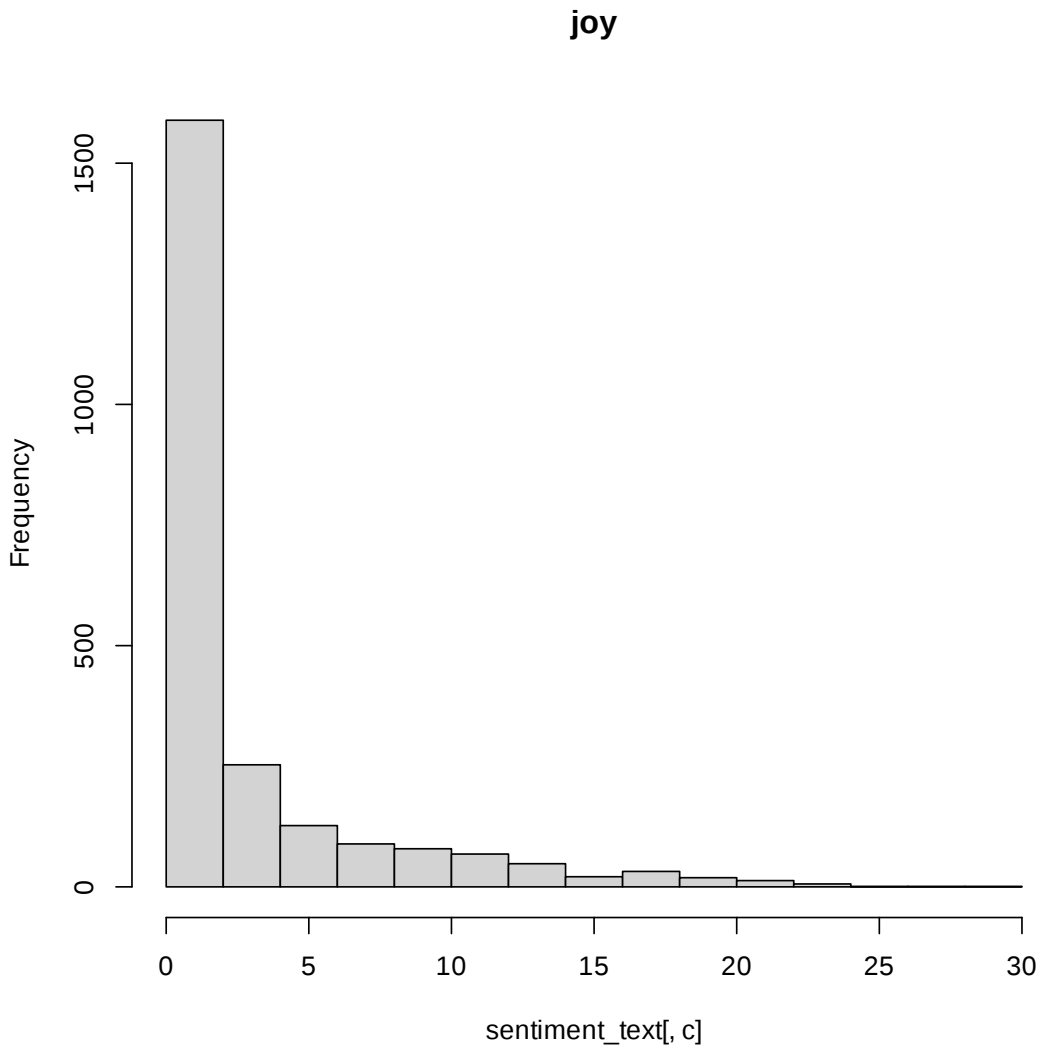
```
[1] "fear"
```


disgust
Frequency
0
500
1000
1500
0
5
10
15
20
25
30
sentiment_text[, c]

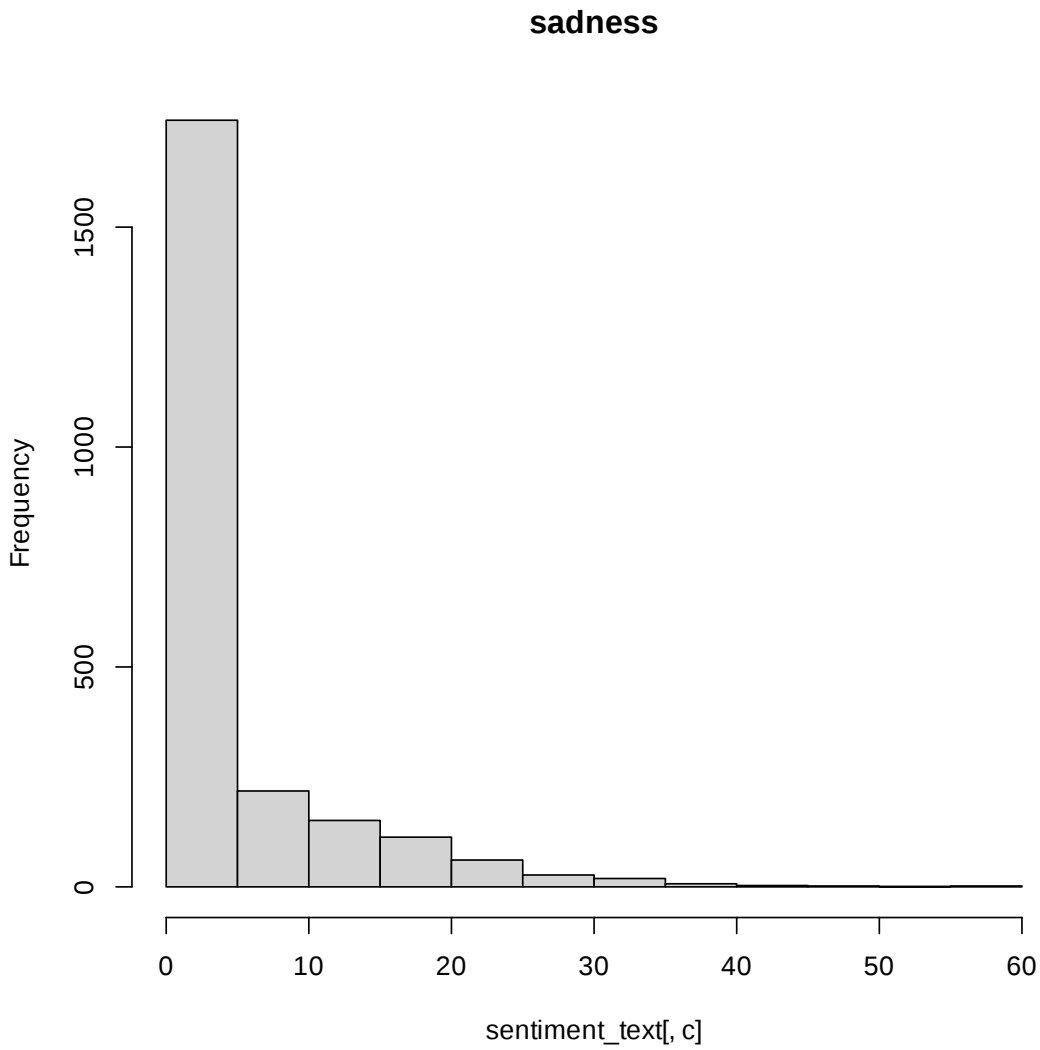
```
[1] "joy"
```


fear
Frequency
0
500
1000
1500
0
10
20
30
40
50
60
sentiment_text[, c]

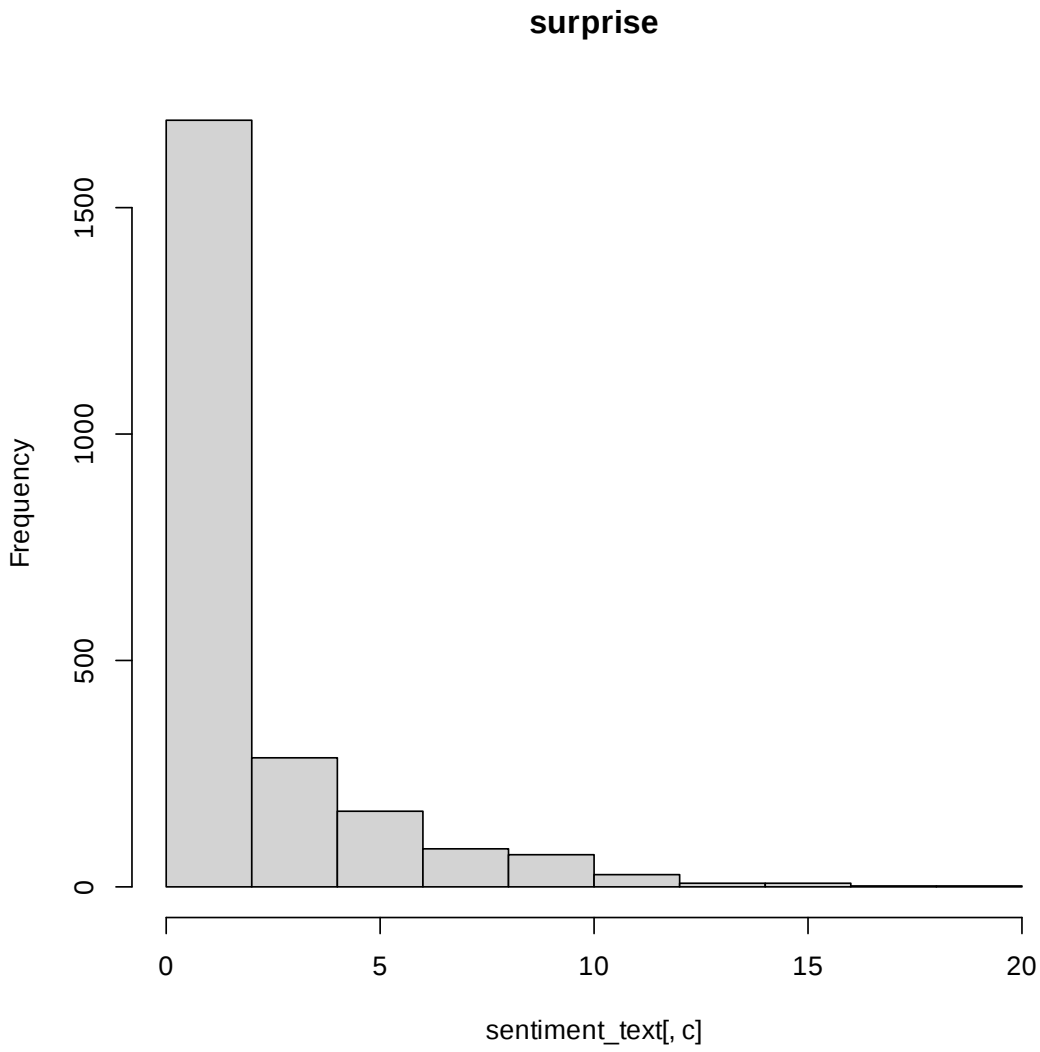
```
[1] "sadness"
```


joy
Frequency
0
500
1000
1500
0
5
10
15
20
25
30
sentiment_text[, c]

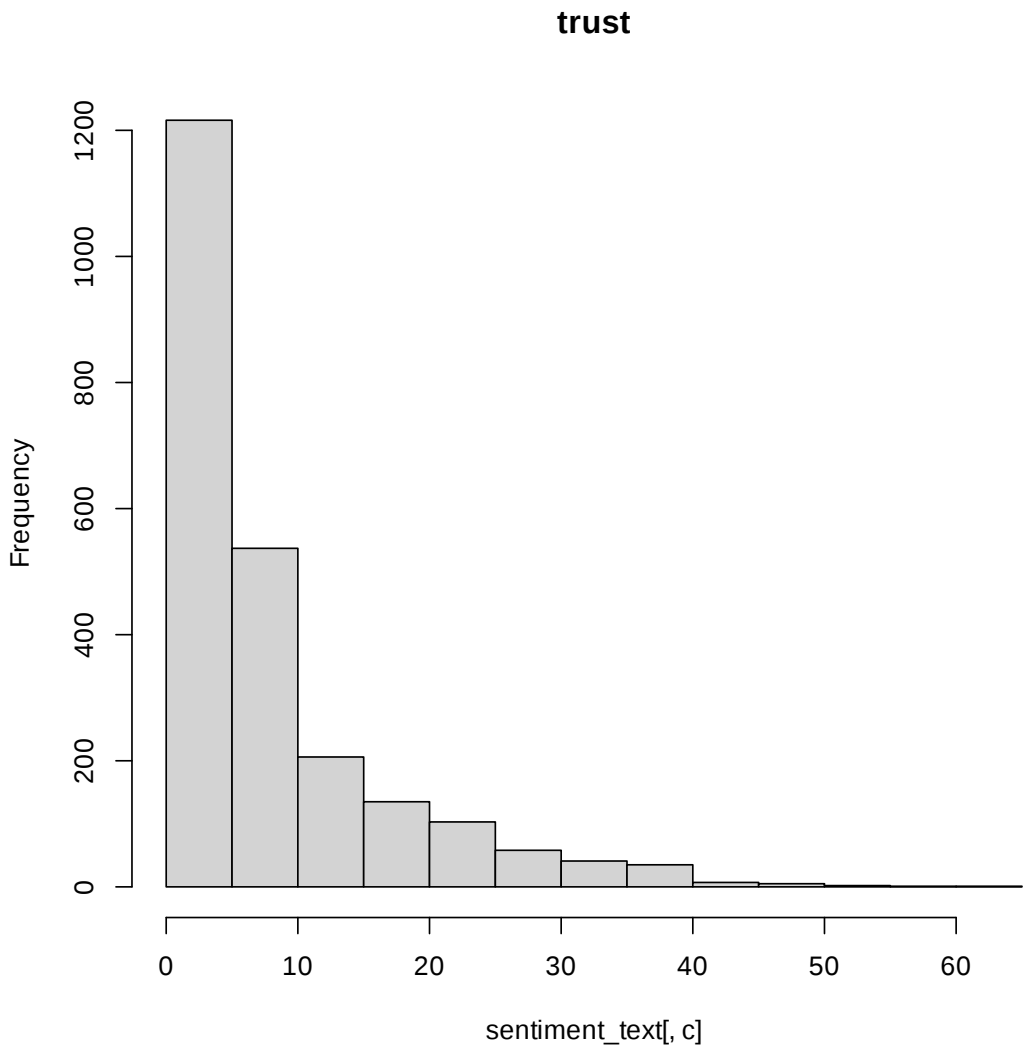
```
[1] "surprise"
```


sadness
Frequency
0
500
1000
1500
0
10
20
30
40
50
60
sentiment_text[, c]

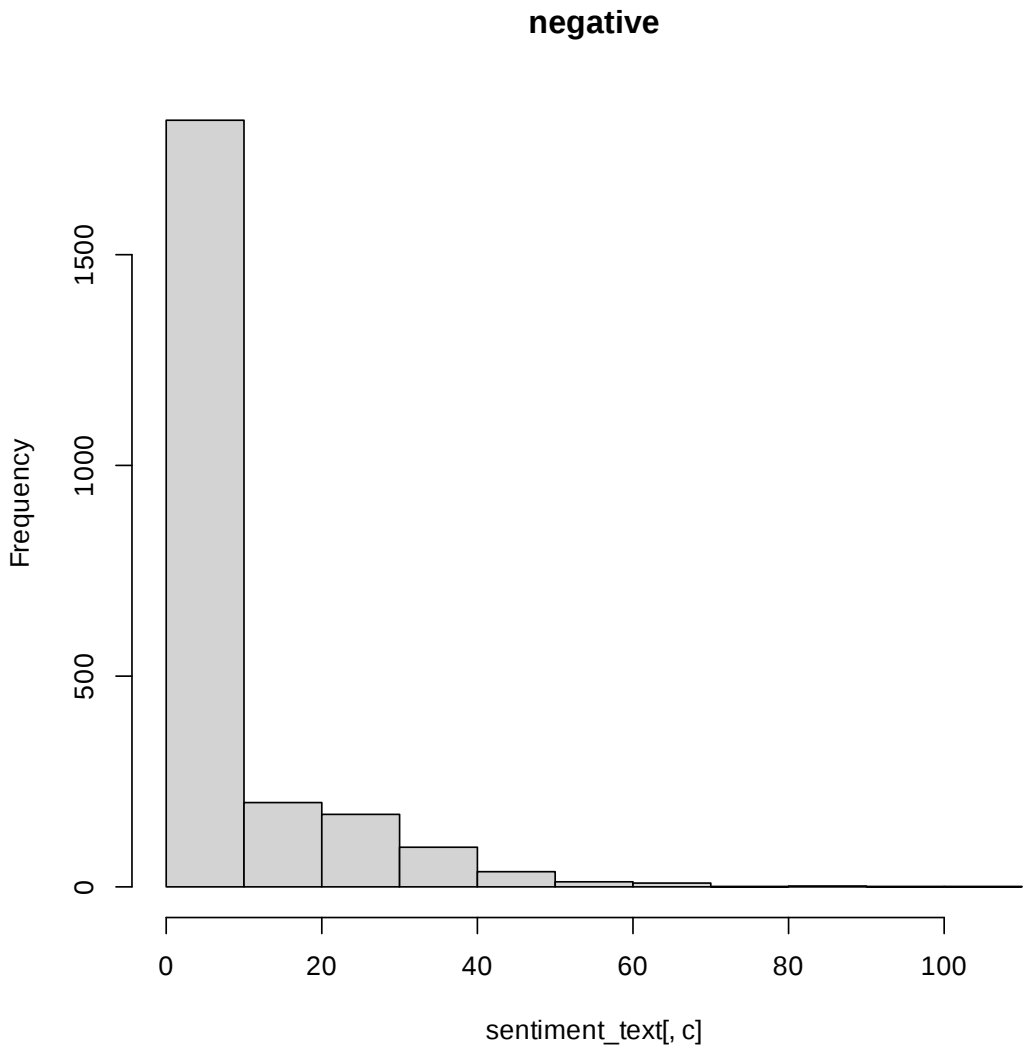
```
[1] "trust"
```

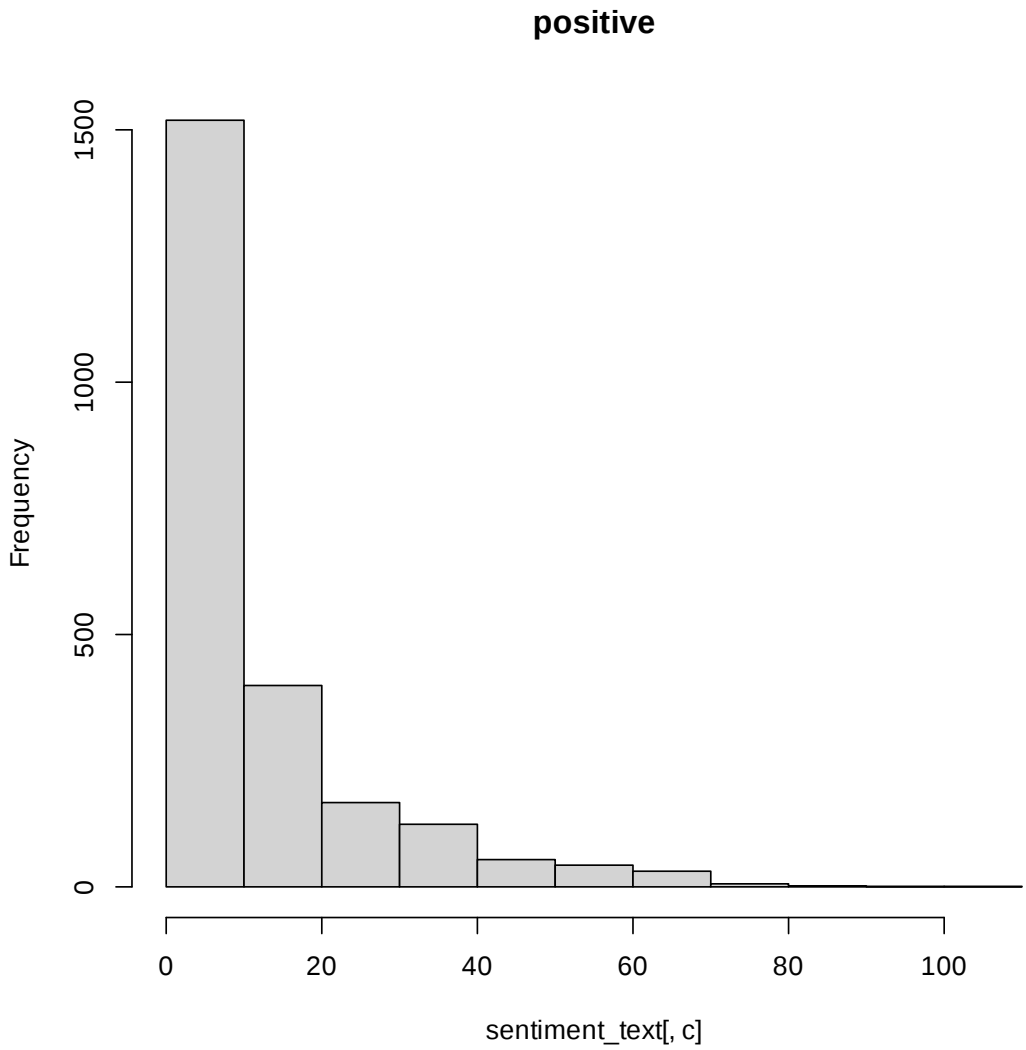

surprise
Frequency
0
500
1000
1500
0
5
10
15
20
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
200
400
600
800
1000
1200
0
10
20
30
40
50
60
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
500
1000
1500
0
20
40
60
80
100
sentiment_text[, c]



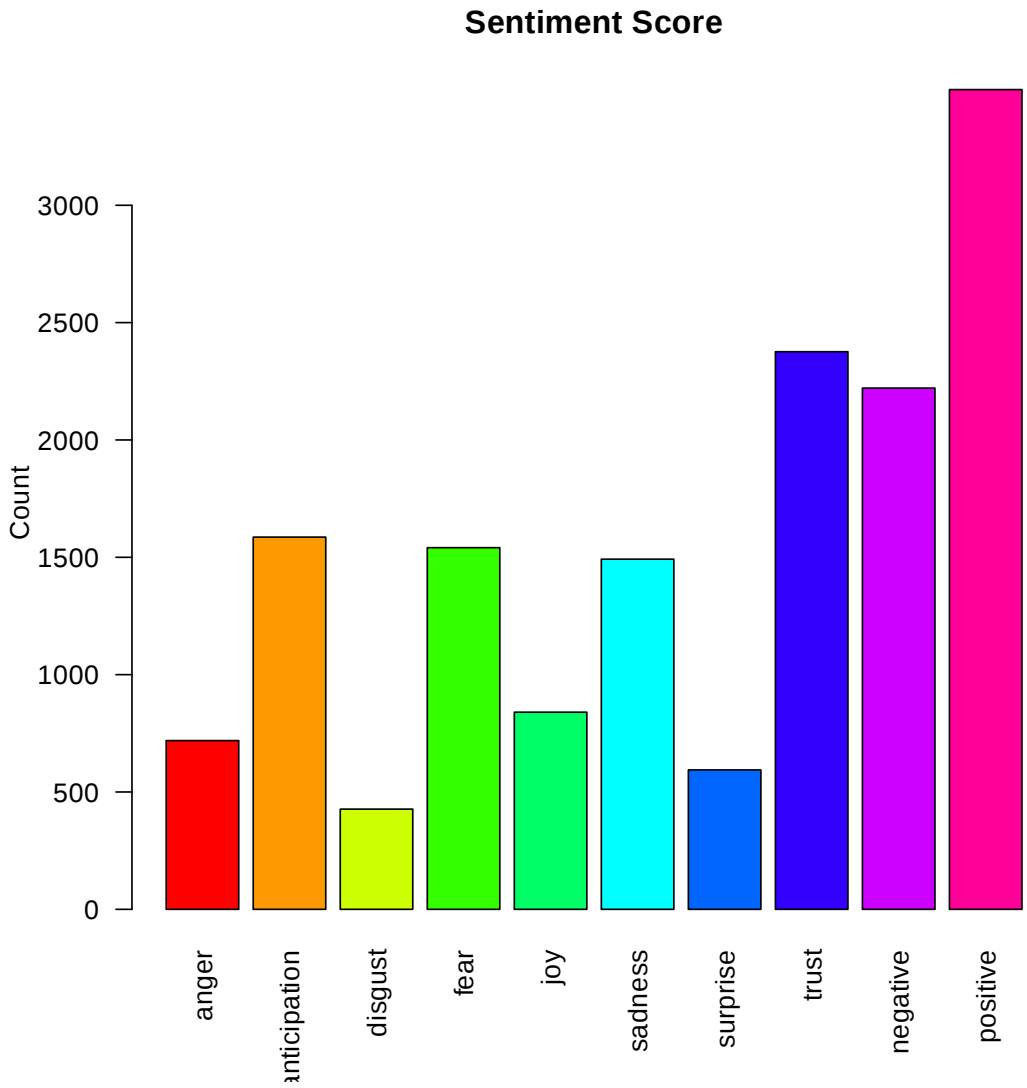
[86]:
```
#white average sentiment
average_white<- average_sentiment(white_sentiment)
summary(average_white)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-17.000   0.000   2.000   3.878   6.000  54.000
```

[59]:
```
#mixed barplot
barplot_function(clean.data.mixed$Text)
```

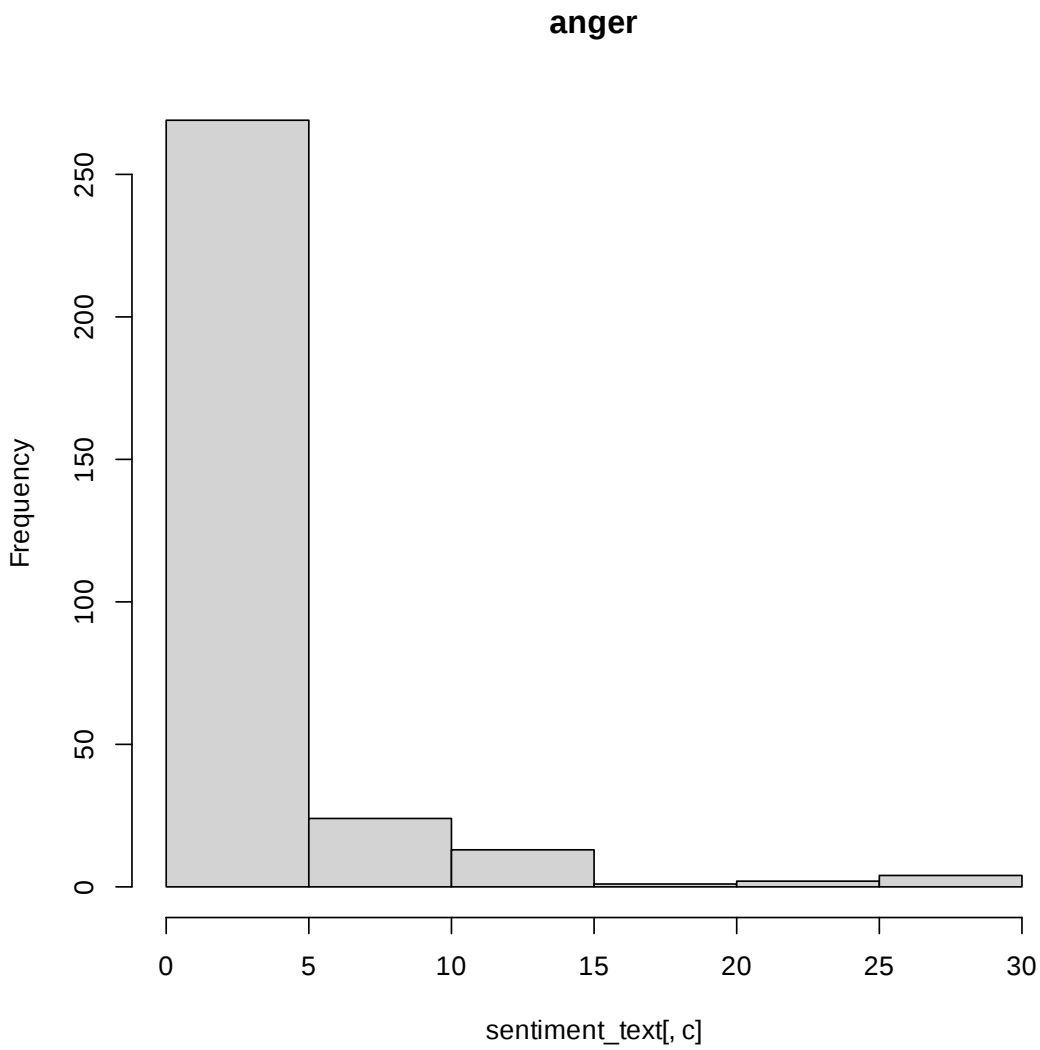
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 2.000   Median : 0.000   Median : 2.000
Mean   : 2.297   Mean   : 5.067   Mean   : 1.364   Mean   : 4.923
3rd Qu.: 2.000   3rd Qu.: 6.000   3rd Qu.: 1.000   3rd Qu.: 5.000
Max.   :30.000   Max.   :53.000   Max.   :25.000   Max.   :51.000
     joy            sadness         surprise         trust
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 1.000   1st Qu.: 0.000   1st Qu.: 2.000
Median : 1.000   Median : 2.000   Median : 1.000   Median : 5.000
Mean   : 2.684   Mean   : 4.767   Mean   : 1.898   Mean   : 7.591
3rd Qu.: 3.000   3rd Qu.: 5.000   3rd Qu.: 3.000   3rd Qu.:10.000
Max.   :34.000   Max.   :53.000   Max.   :17.000   Max.   :76.000
   negative         positive
Min.   : 0.000   Min.   :  0.00
1st Qu.: 0.000   1st Qu.:  2.00
Median : 3.000   Median :  6.00
Mean   : 7.096   Mean   : 11.16
3rd Qu.: 6.000   3rd Qu.: 14.00
Max.   :93.000   Max.   :113.00
```

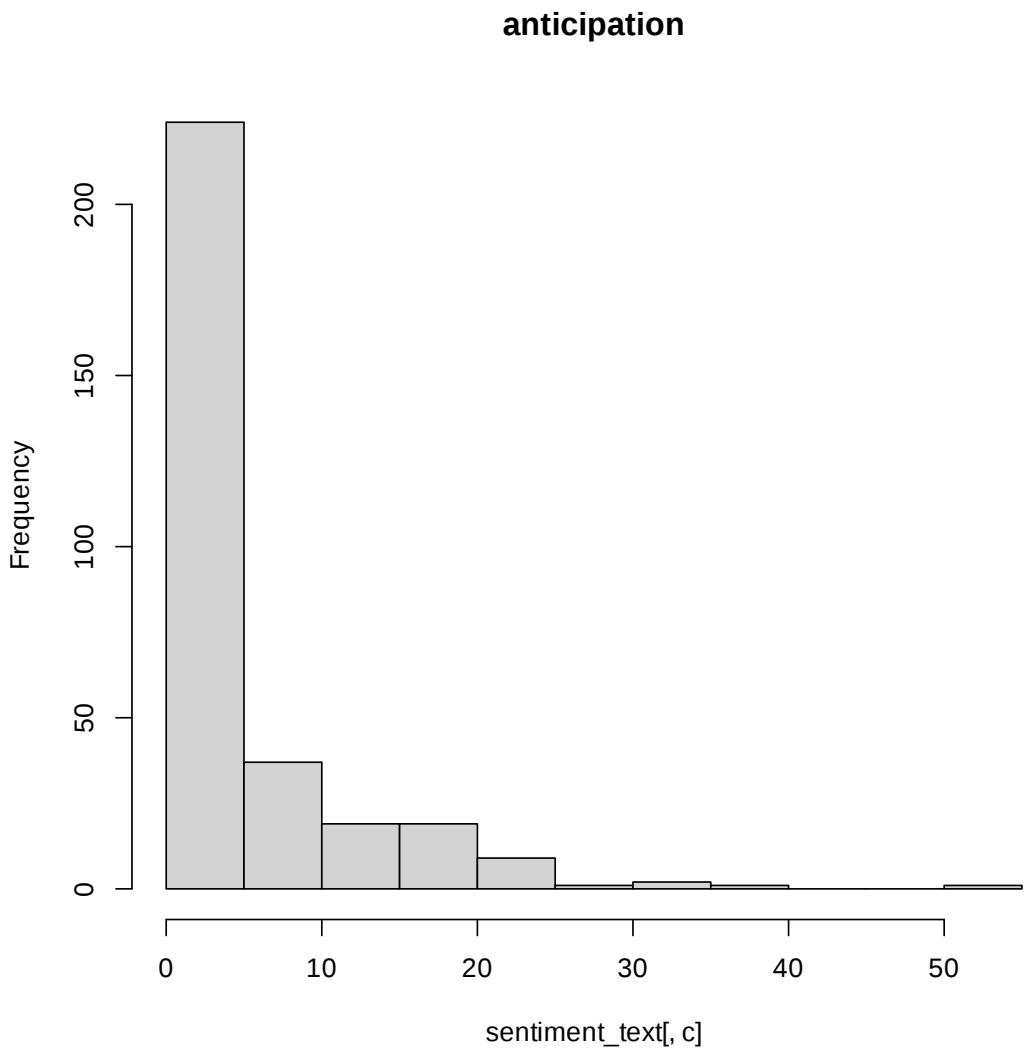


[87]:
```
#mixed sentiment histogram
sentiment_hist(mixed_sentiment)
```

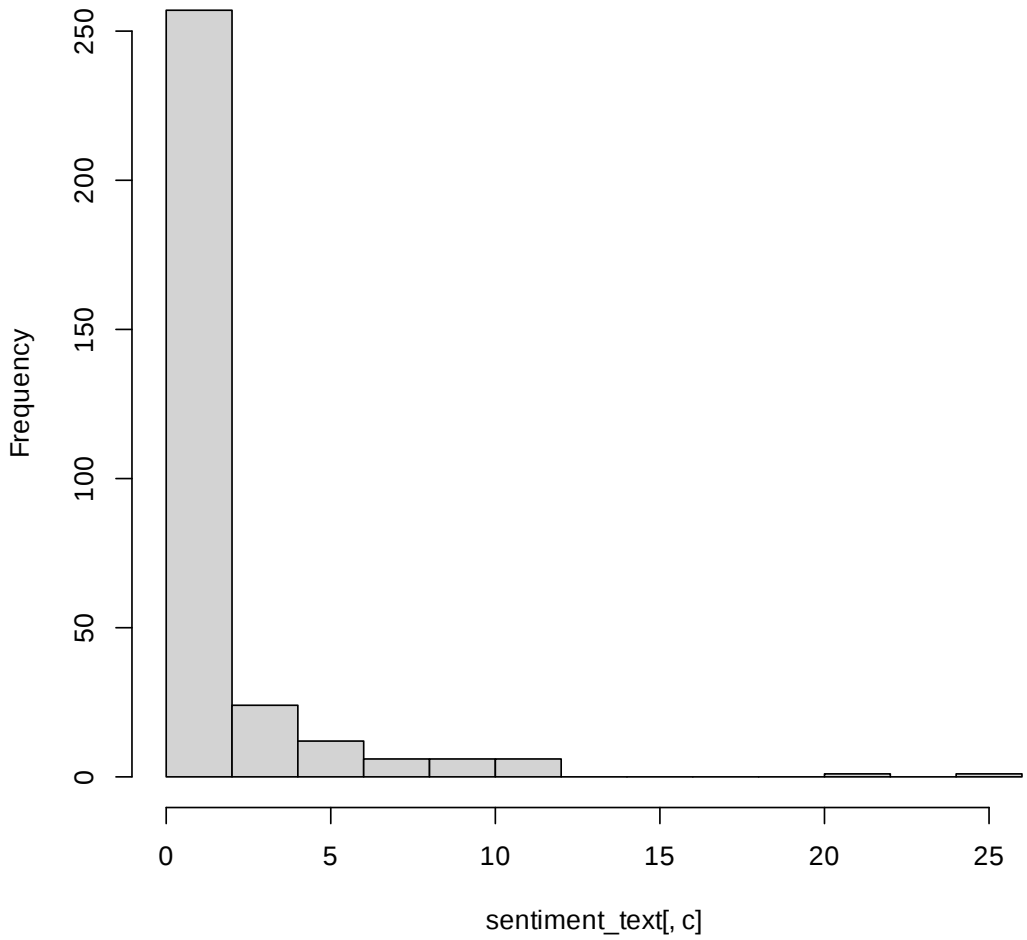
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
50
100
150
200
250
0
5
10
15
20
25
30
sentiment_text[, c]

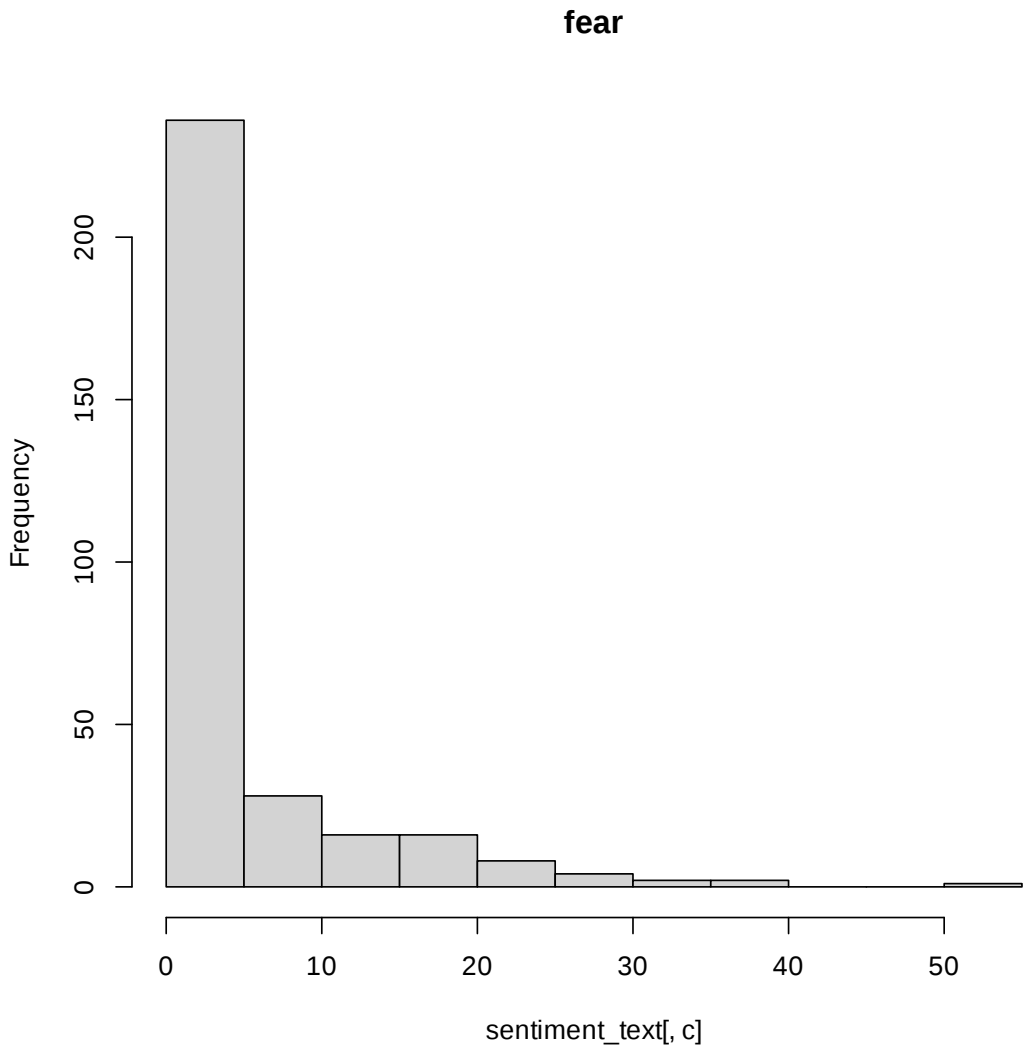
```
[1] "disgust"
```


anticipation
Frequency
0
50
100
150
200
0
10
20
30
40
50
sentiment_text[, c]

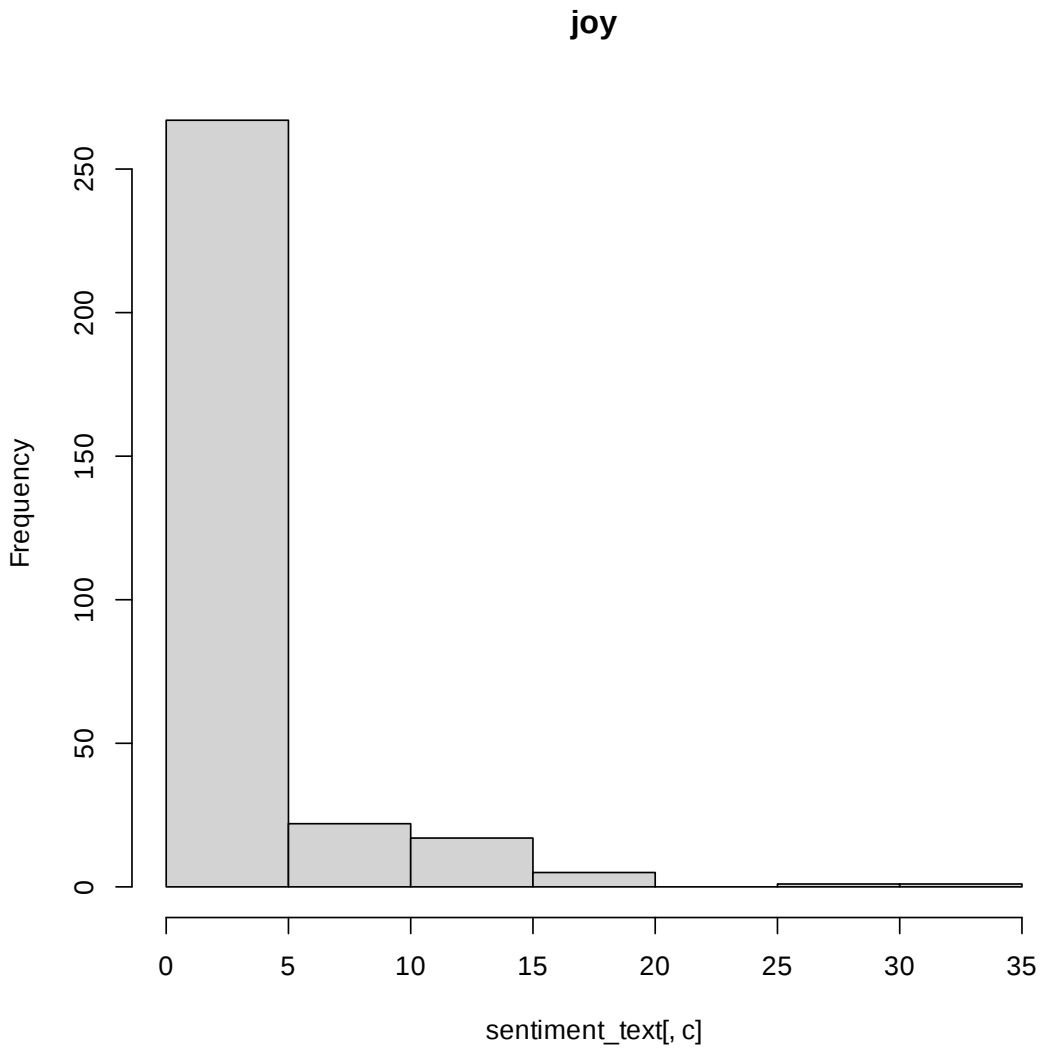
```
[1] "fear"
```


disgust
Frequency
0
50
100
150
200
250
0
5
10
15
20
25
sentiment_text[, c]

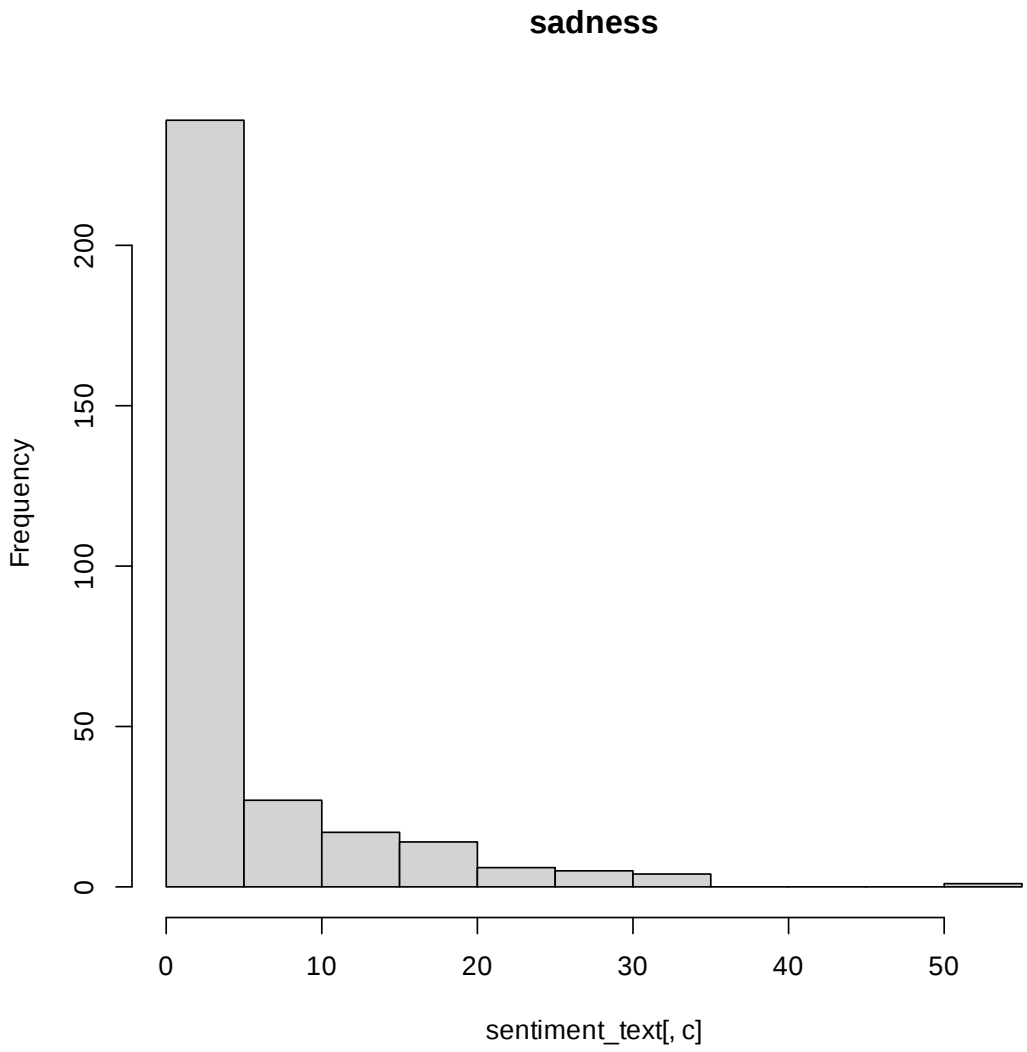
```
[1] "joy"
```


fear
Frequency
0
50
100
150
200
0
10
20
30
40
50
sentiment_text[, c]

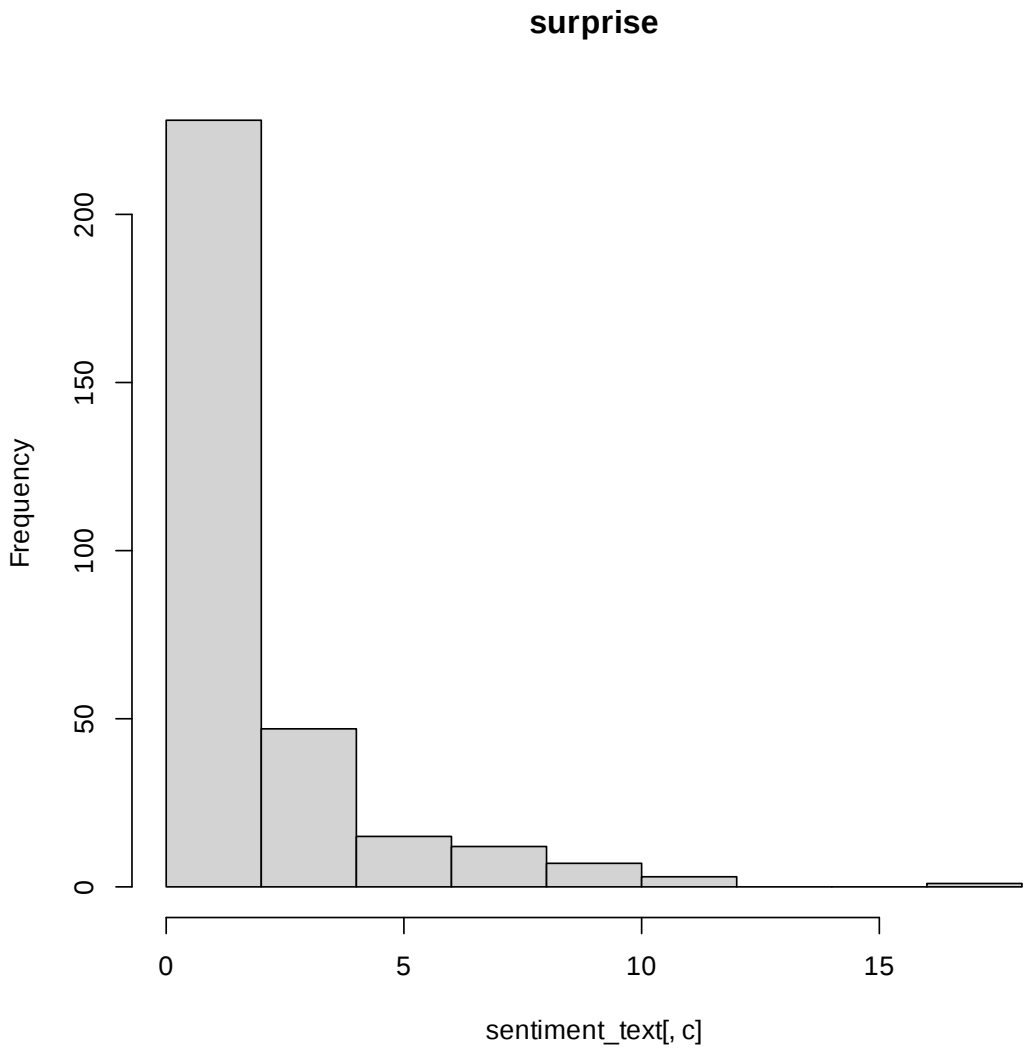
```
[1] "sadness"
```


joy
Frequency
0
50
100
150
200
250
0
5
10
15
20
25
30
35
sentiment_text[, c]

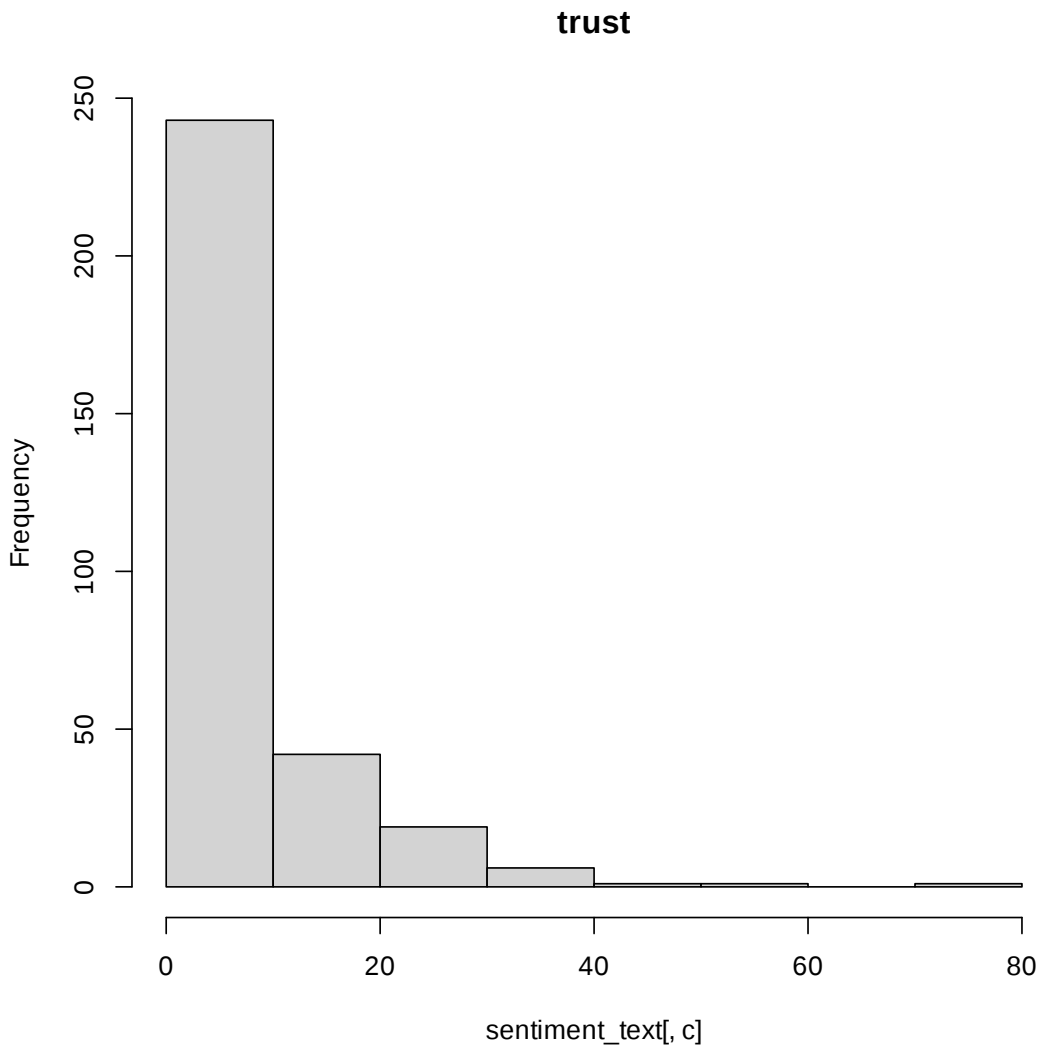
```
[1] "surprise"
```


sadness
Frequency
0
50
100
150
200
0
10
20
30
40
50
sentiment_text[, c]

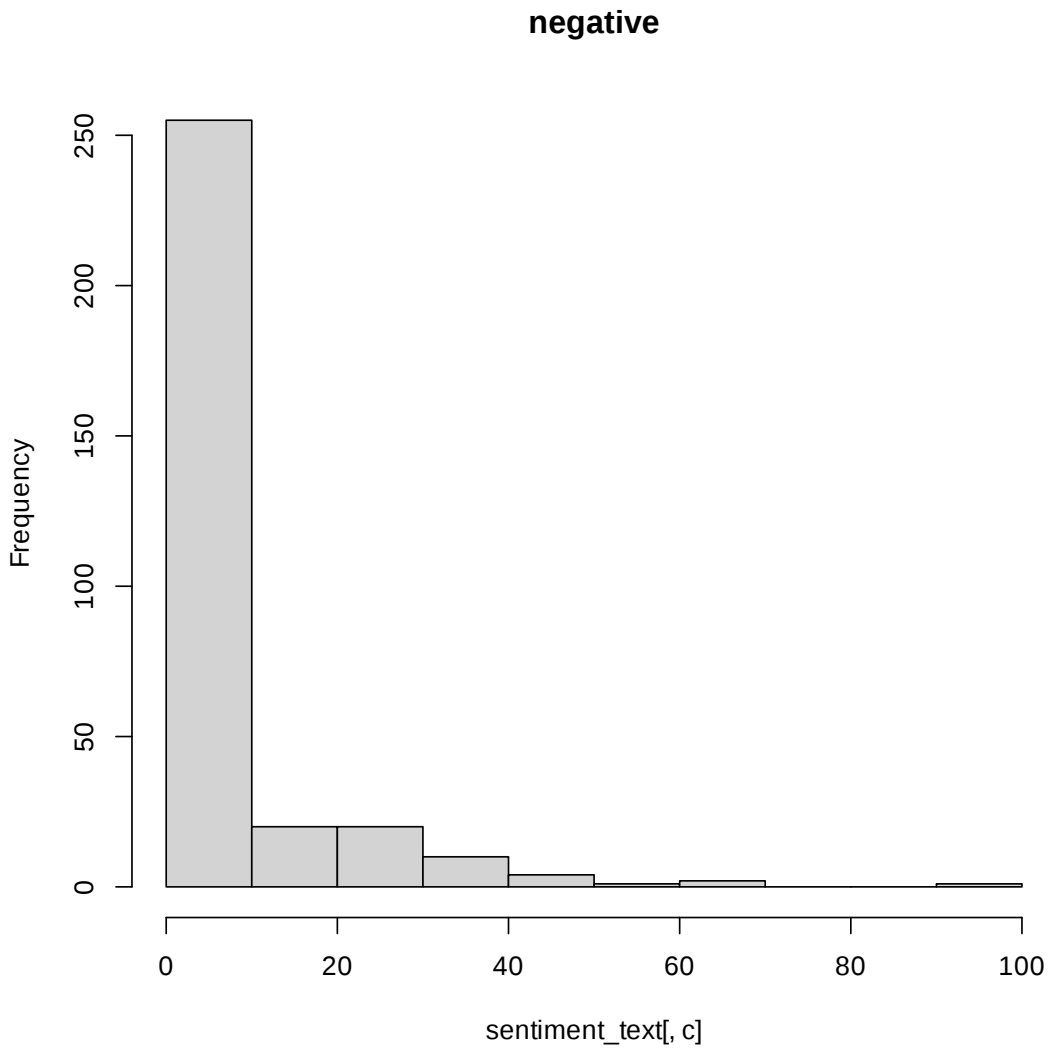
```
[1] "trust"
```

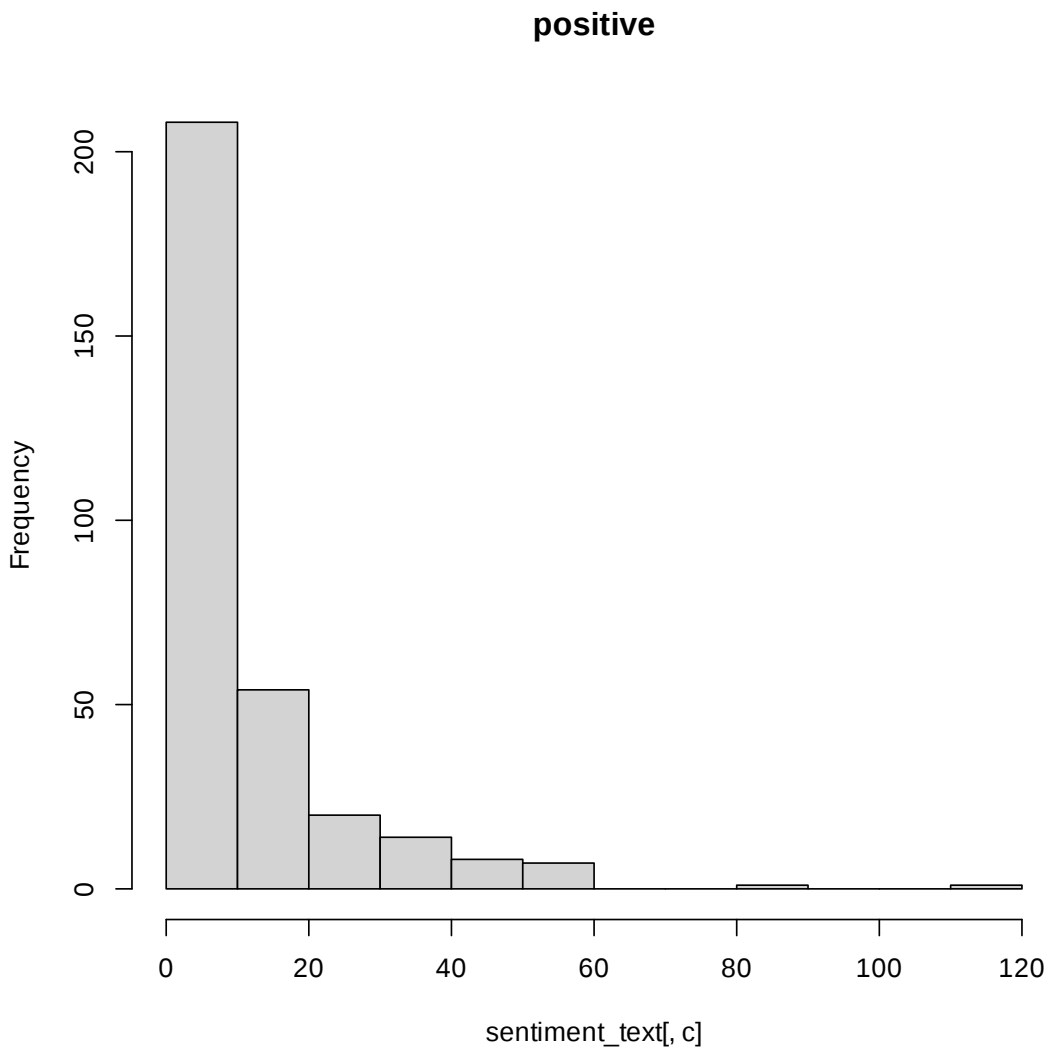



```
[1] "negative"
```


trust
Frequency
0
50
100
150
200
250
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
50
100
150
200
250
0
20
40
60
80
100
sentiment_text[, c]



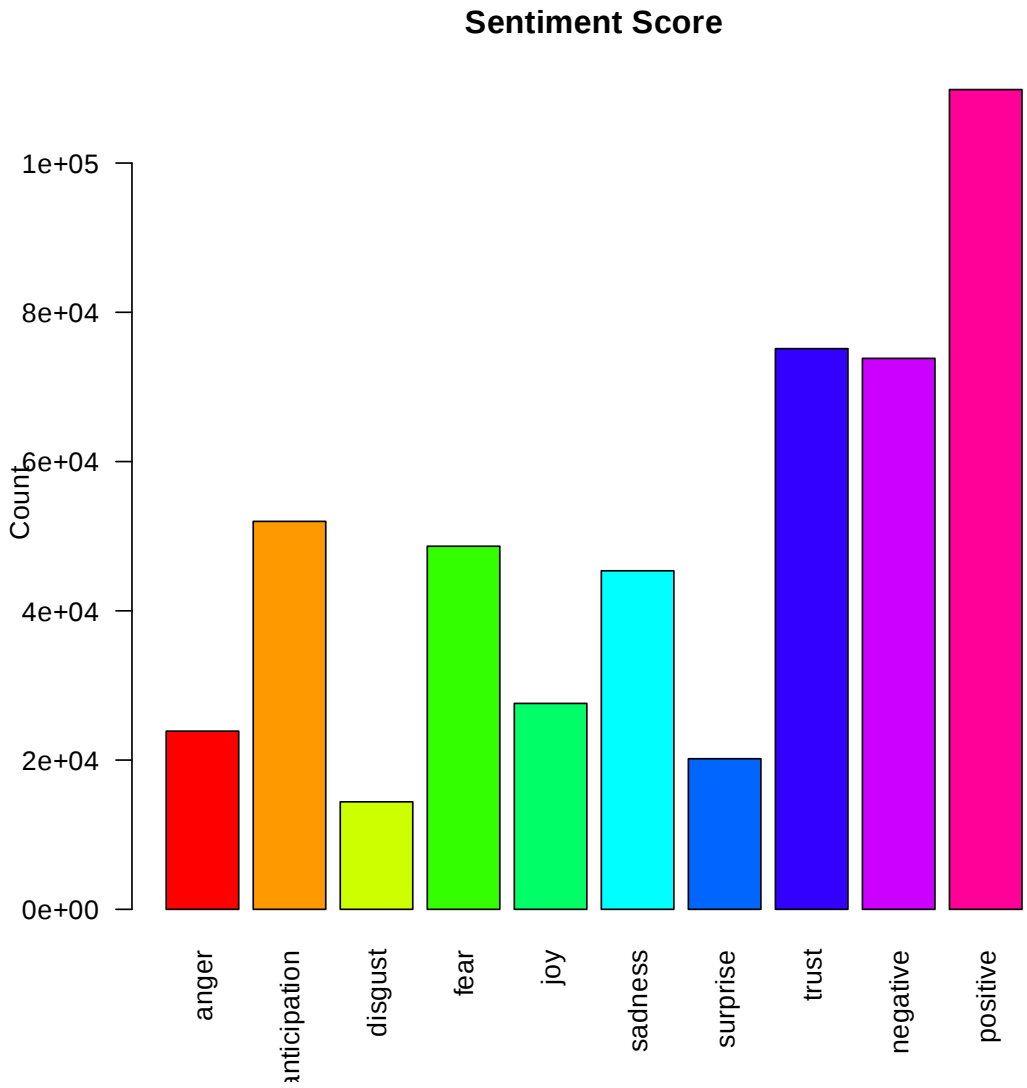
[88]:
```
#mixed average sentiment
average_mixed<- average_sentiment(mixed_sentiment)
summary(average_mixed)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-15.000   0.000   3.000   4.064   6.000  40.000
```

[60]:
```
#eu barplot
barplot_function(clean.data.eu$Text)
```

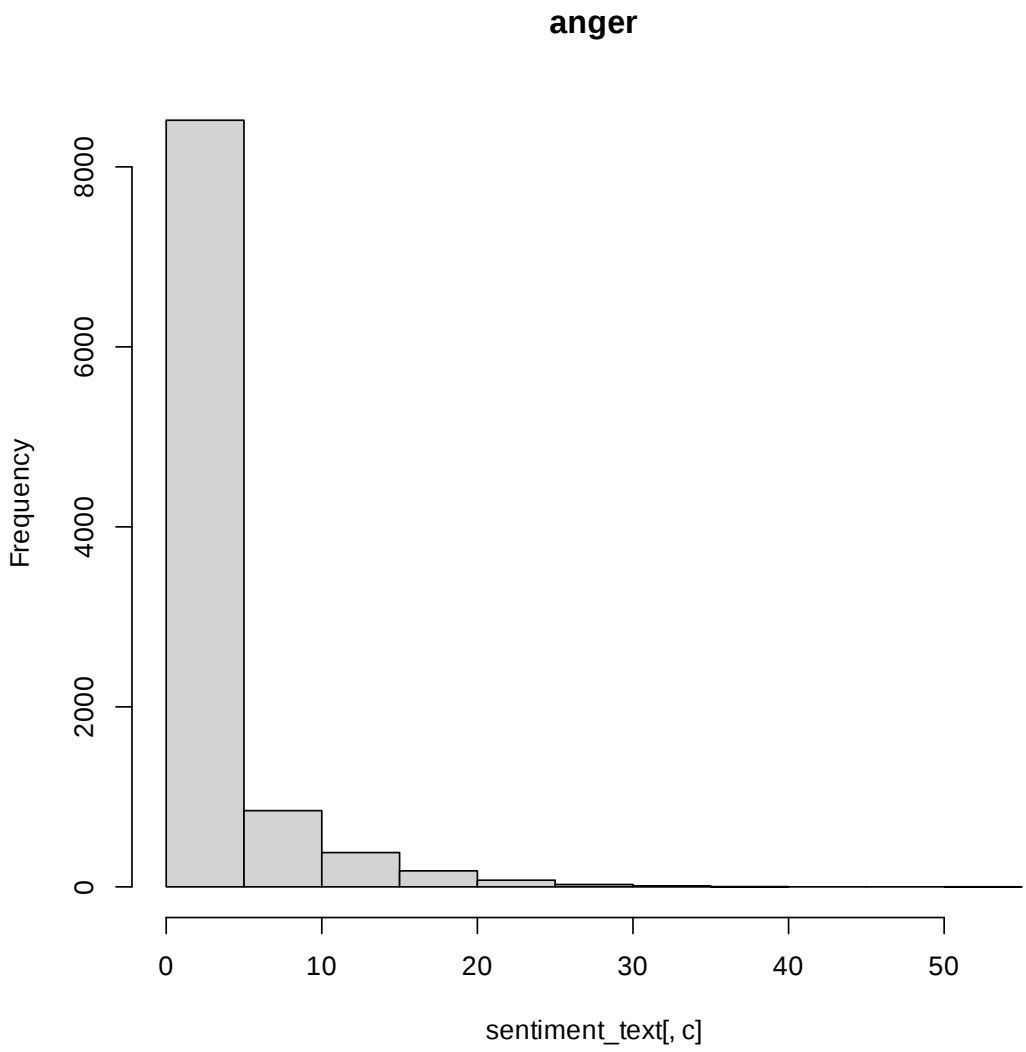
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 2.000   Median : 0.000   Median : 1.000
Mean   : 2.378   Mean   : 5.178   Mean   : 1.435   Mean   : 4.847
3rd Qu.: 2.000   3rd Qu.: 7.000   3rd Qu.: 1.000   3rd Qu.: 6.000
Max.   :51.000   Max.   :46.000   Max.   :24.000   Max.   :63.000
     joy            sadness         surprise         trust
Min.   : 0.000   Min.   : 0.000   Min.   : 0.00   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000
Median : 1.000   Median : 1.000   Median : 1.00   Median : 5.000
Mean   : 2.749   Mean   : 4.519   Mean   : 2.01   Mean   : 7.484
3rd Qu.: 4.000   3rd Qu.: 5.000   3rd Qu.: 3.00   3rd Qu.:11.000
Max.   :38.000   Max.   :53.000   Max.   :23.00   Max.   :76.000
   negative         positive
Min.   : 0.000   Min.   :  0.00
1st Qu.: 0.000   1st Qu.:  0.00
Median : 2.000   Median :  6.00
Mean   : 7.354   Mean   : 10.94
3rd Qu.: 8.000   3rd Qu.: 15.00
Max.   :98.000   Max.   :120.00
```

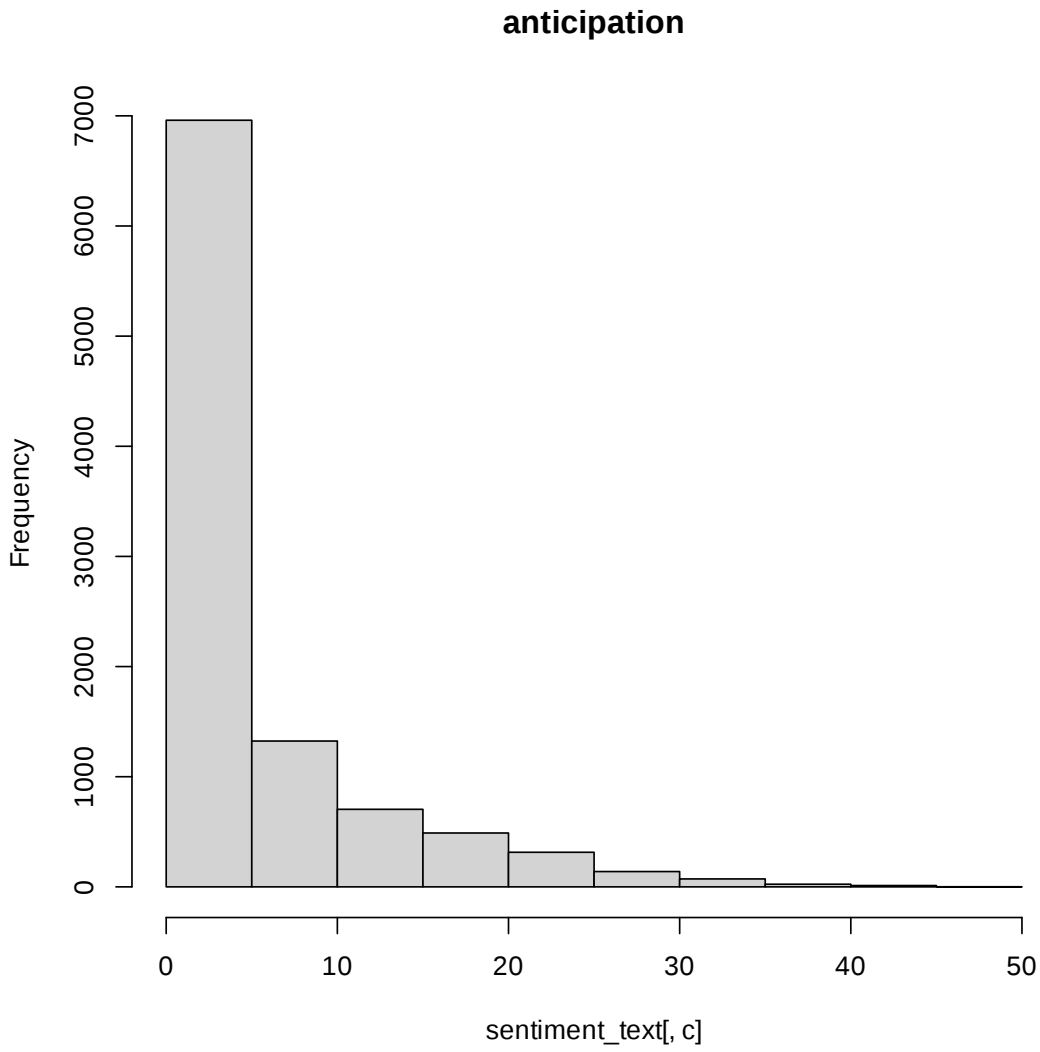


[89]:
```
#eu sentiment histogram
sentiment_hist(eu_sentiment)
```

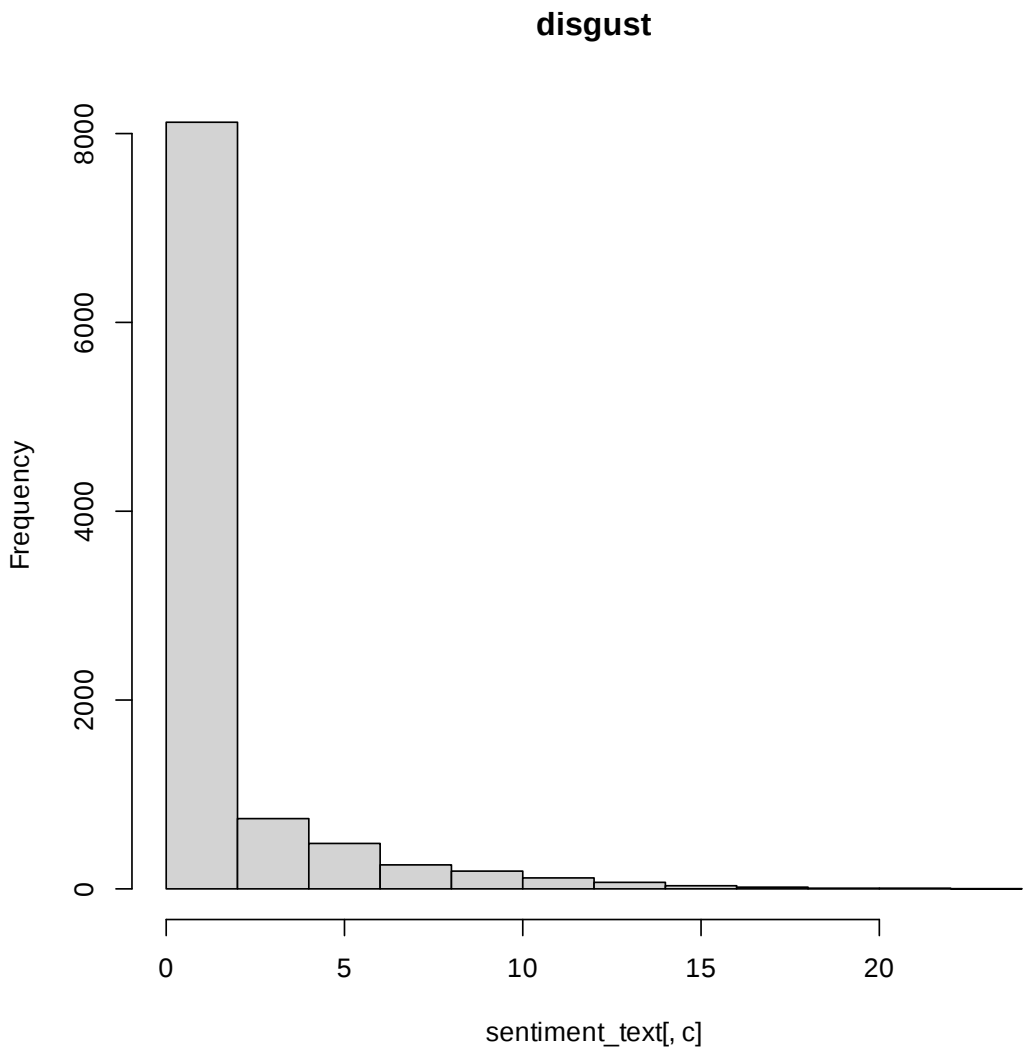
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
2000
4000
6000
8000
0
10
20
30
40
50
sentiment_text[, c]

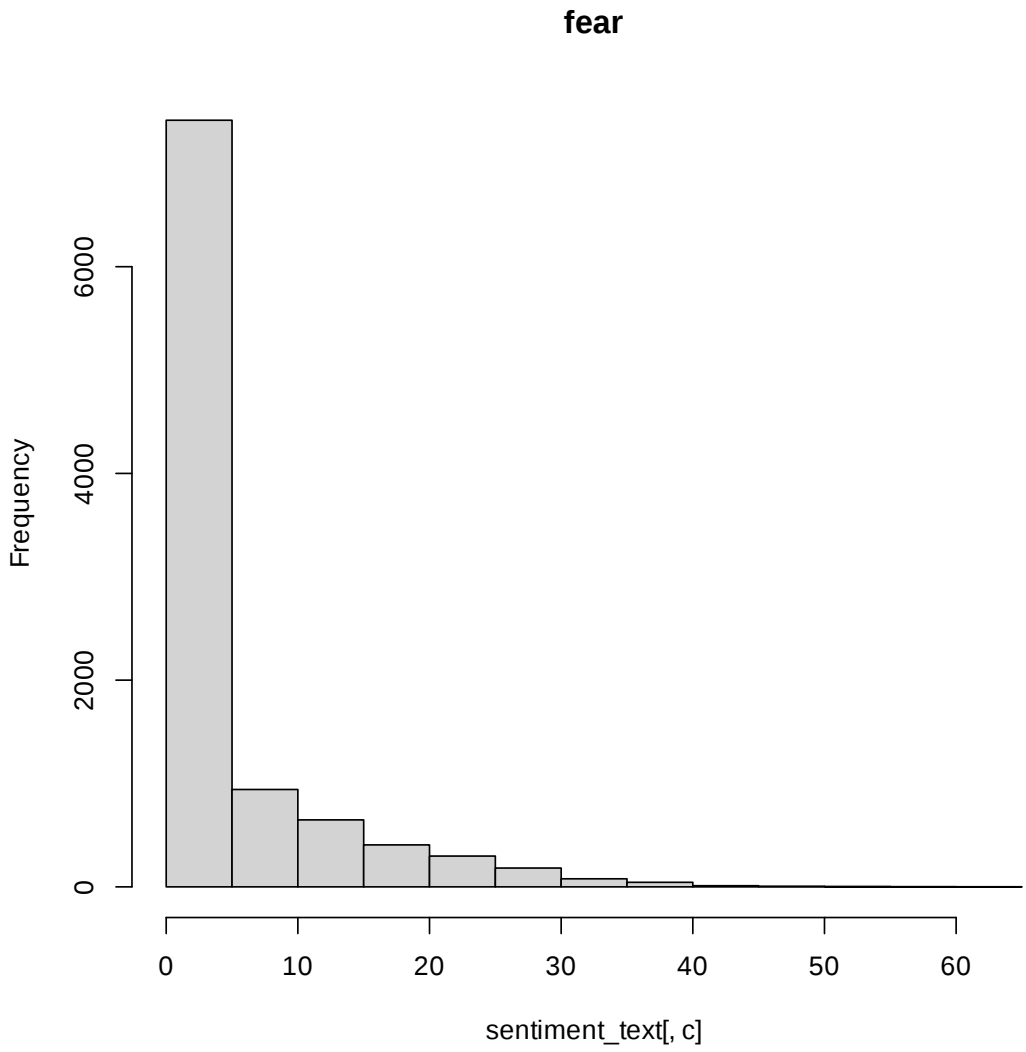
```
[1] "disgust"
```


anticipation
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0
10
20
30
40
50
sentiment_text[, c]

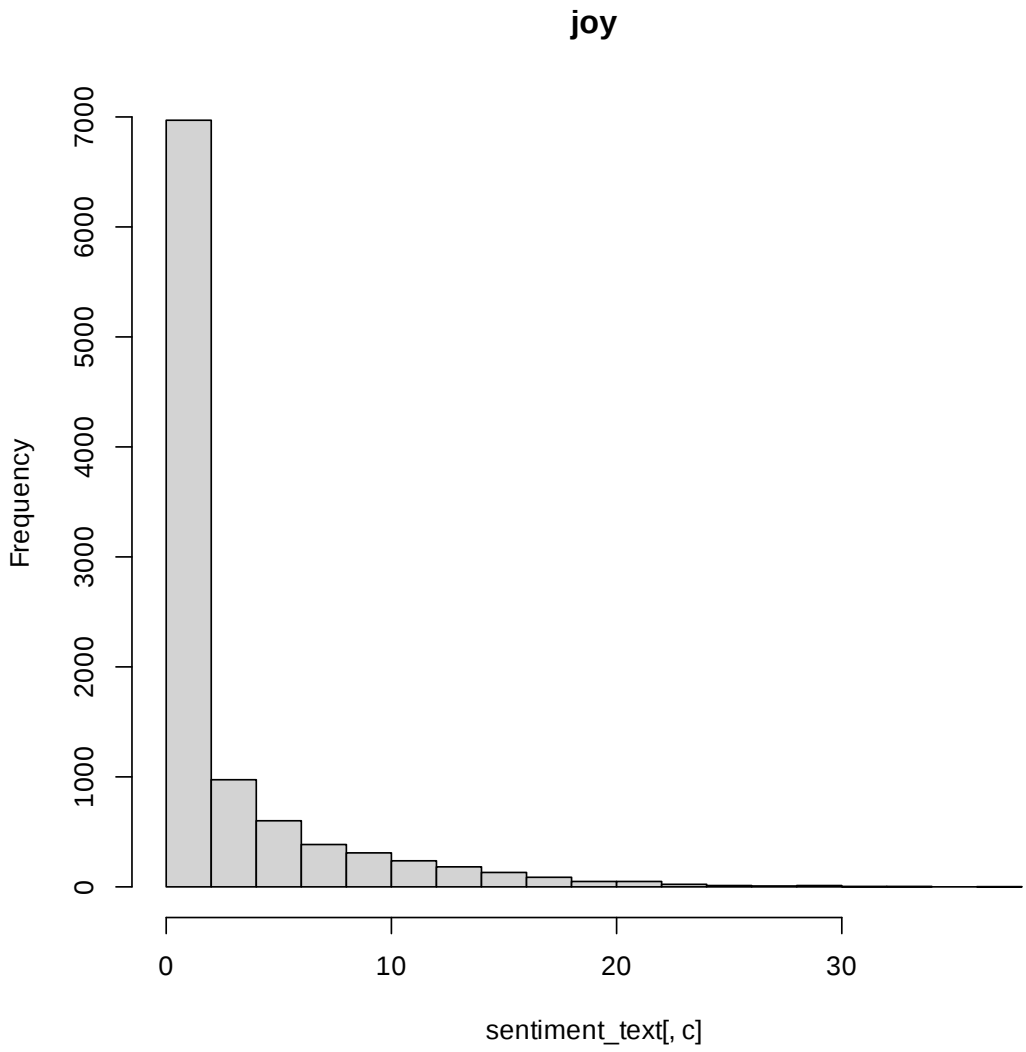
```
[1] "fear"
```


disgust
Frequency
0
2000
4000
6000
8000
0
5
10
15
20
sentiment_text[, c]

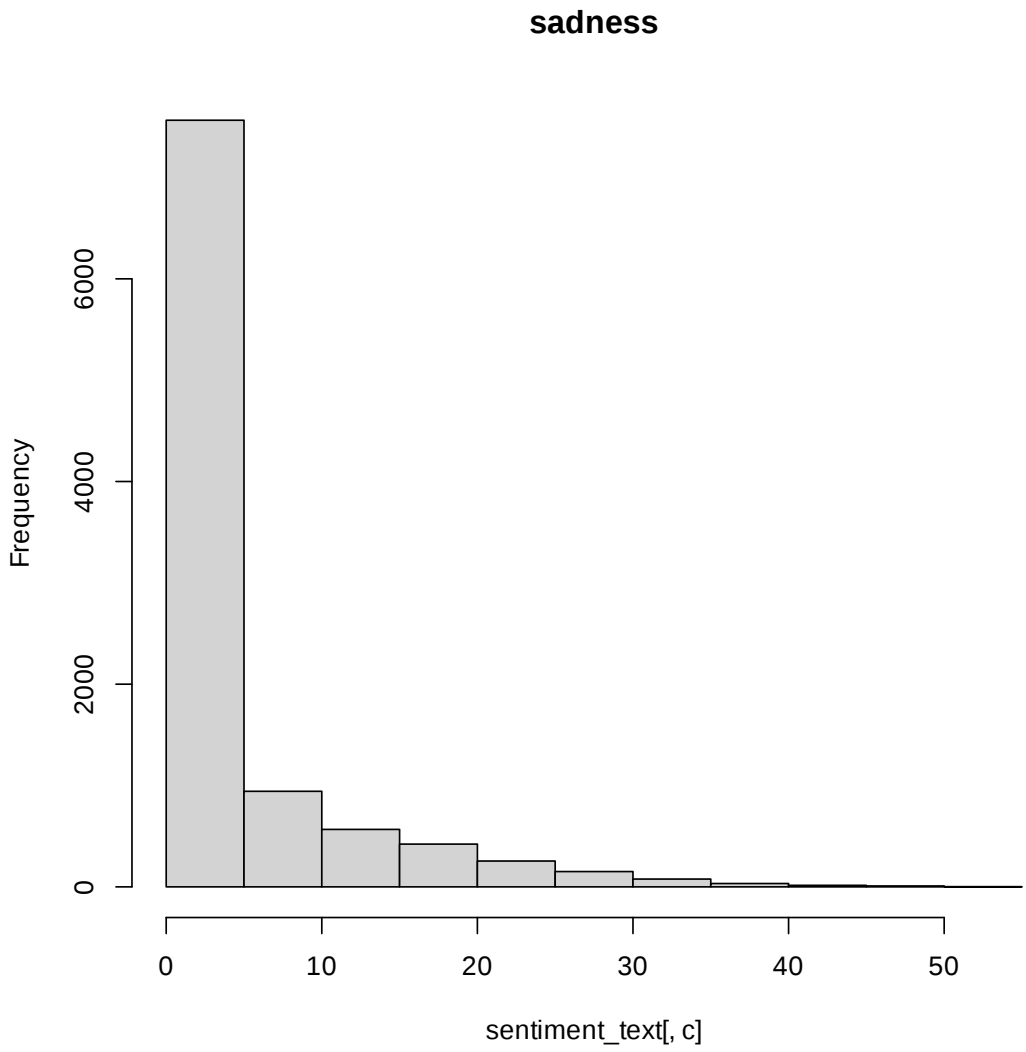
```
[1] "joy"
```


fear
Frequency
0
2000
4000
6000
0
10
20
30
40
50
60
sentiment_text[, c]

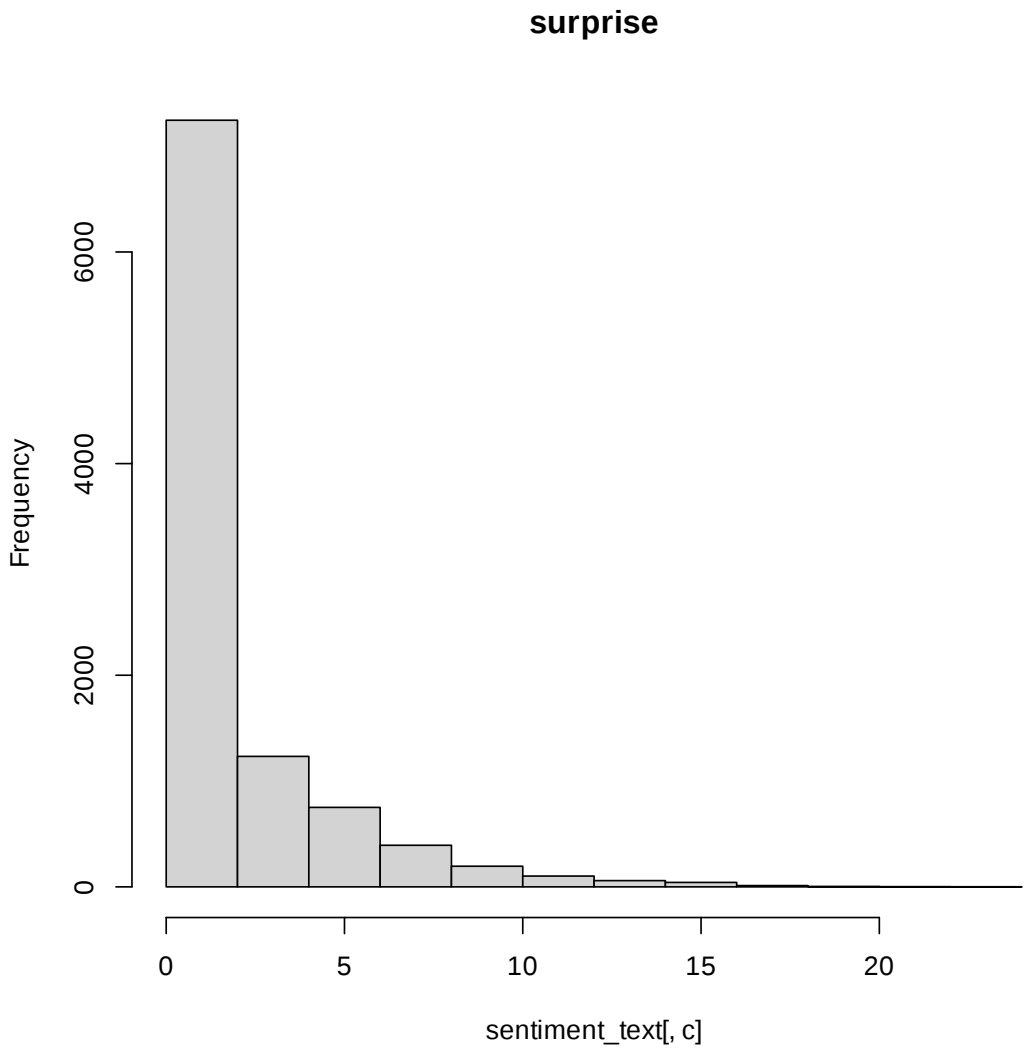
```
[1] "sadness"
```


joy
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0
10
20
30
sentiment_text[, c]

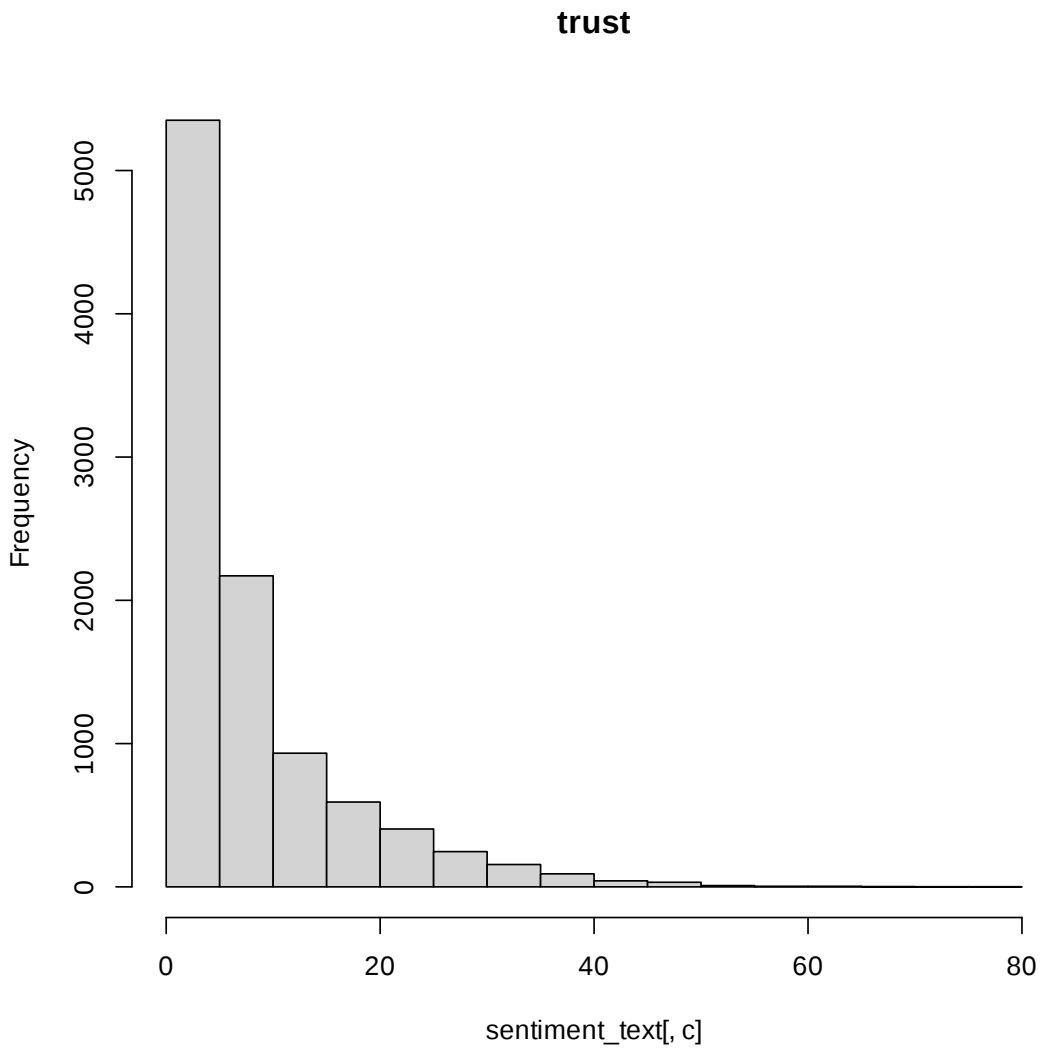
```
[1] "surprise"
```


sadness
Frequency
0
2000
4000
6000
0
10
20
30
40
50
sentiment_text[, c]

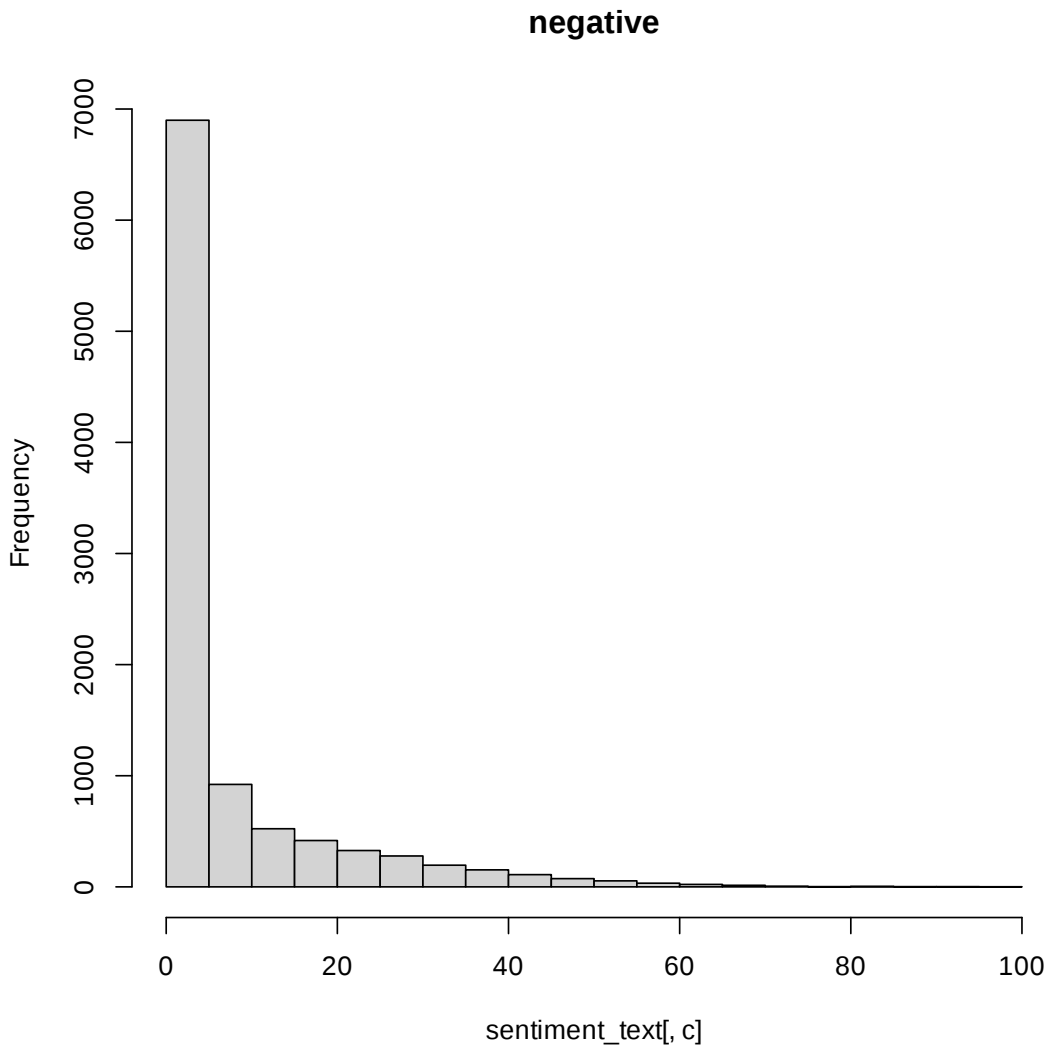
```
[1] "trust"
```

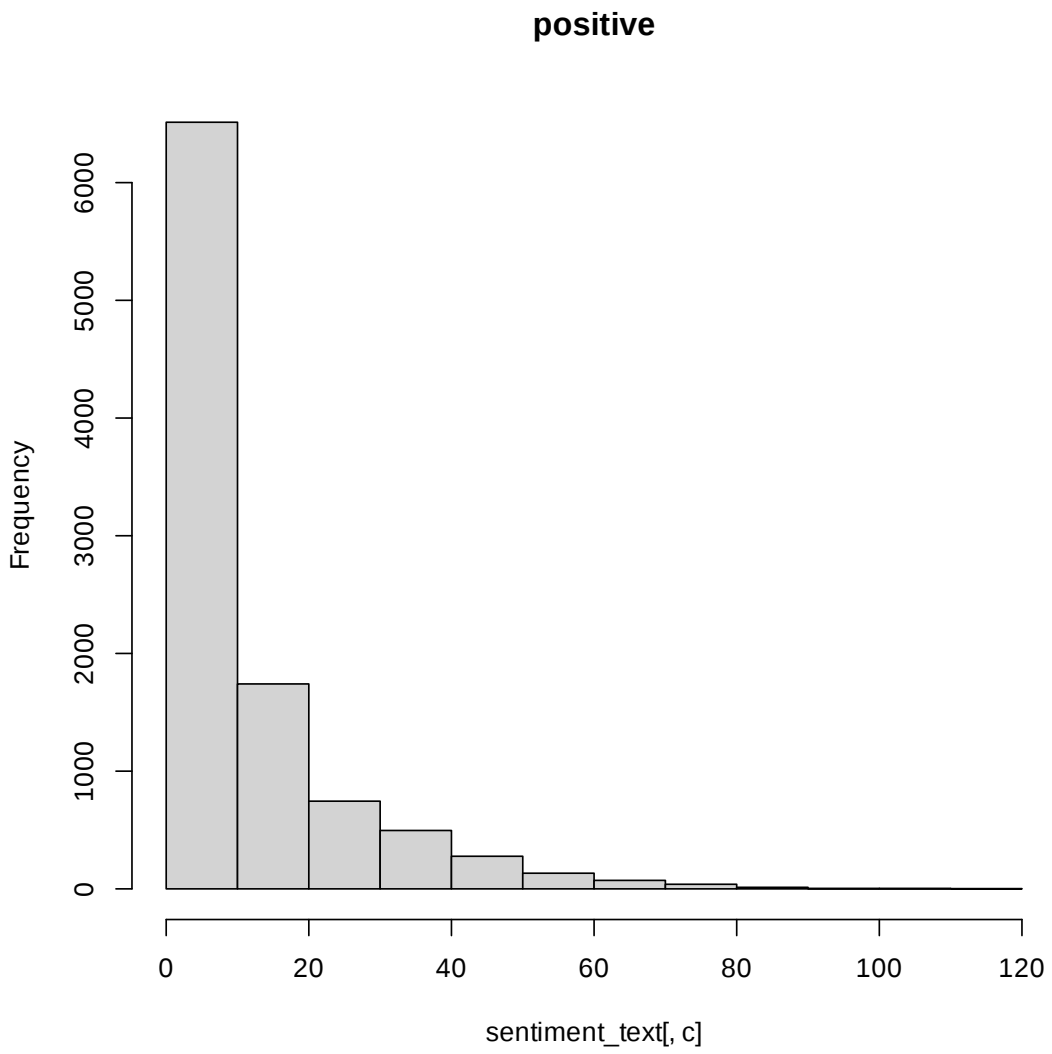

surprise
Frequency
0
2000
4000
6000
0
5
10
15
20
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
1000
2000
3000
4000
5000
0
20
40
60
80
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0
20
40
60
80
100
sentiment_text[, c]



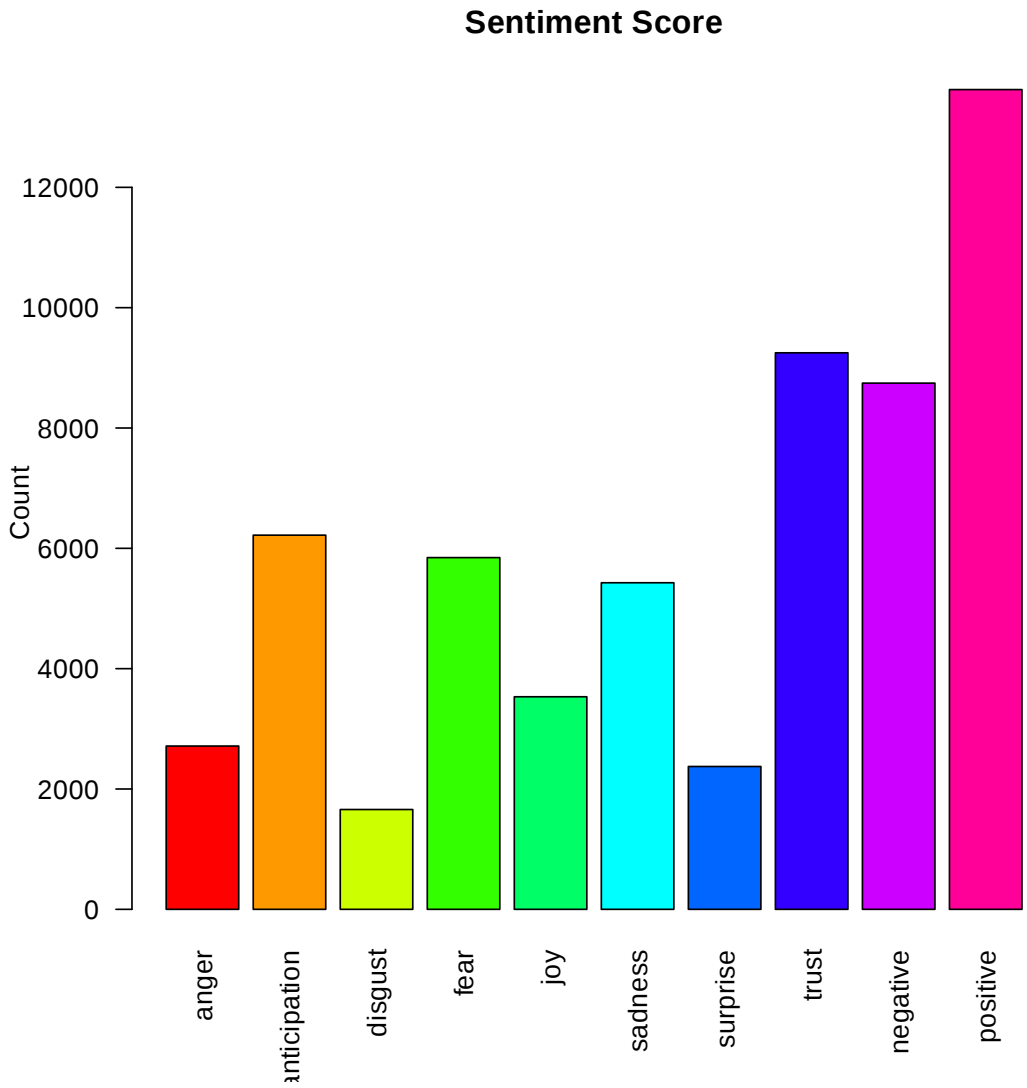
[90]:
```
#eu average sentiment
average_eu<- average_sentiment(eu_sentiment)
summary(average_eu)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-22.000   0.000   2.000   3.588   6.000  54.000
```

[61]:
```
#caribbean barplot
barplot_function(clean.data.caribbean$Text)
```

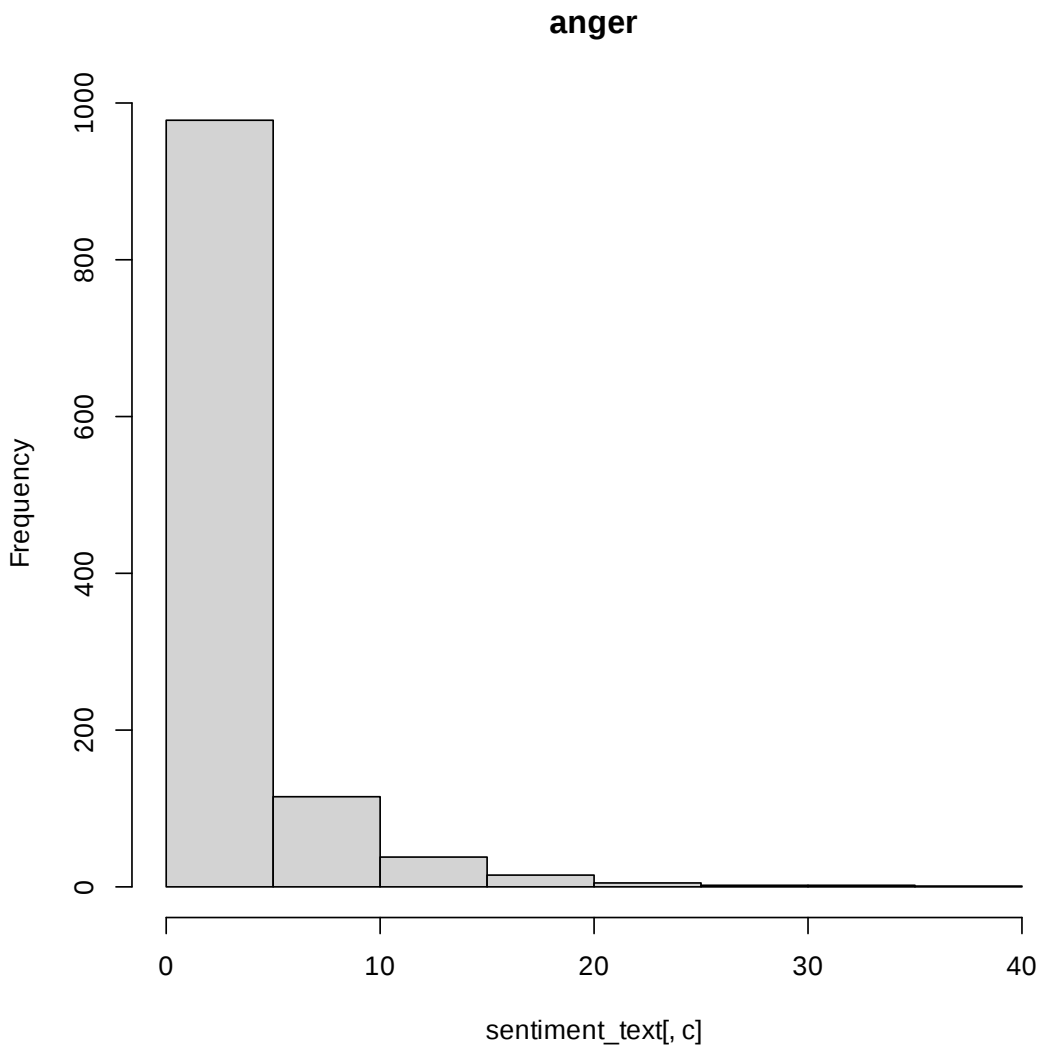
```
     anger          anticipation      disgust           fear
 Min.   : 0.000   Min.   : 0.00   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000
```

```
Median : 0.000   Median : 3.00   Median : 0.000   Median : 2.000
Mean   : 2.348   Mean   : 5.38   Mean   : 1.435   Mean   : 5.057
3rd Qu.: 3.000   3rd Qu.: 8.00   3rd Qu.: 2.000   3rd Qu.: 6.000
Max.   :40.000   Max.   :43.00   Max.   :26.000   Max.   :52.000
     joy            sadness         surprise         trust
Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
Median : 1.000   Median : 2.000   Median : 1.000   Median : 5.500
Mean   : 3.056   Mean   : 4.696   Mean   : 2.054   Mean   : 8.002
3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 3.000   3rd Qu.:11.000
Max.   :34.000   Max.   :47.000   Max.   :18.000   Max.   :62.000
   negative         positive
Min.   : 0.000   Min.   :  0.00
1st Qu.: 0.000   1st Qu.:  0.00
Median : 3.000   Median :  8.00
Mean   : 7.566   Mean   : 11.79
3rd Qu.:10.000   3rd Qu.: 16.00
Max.   :87.000   Max.   :104.00
```

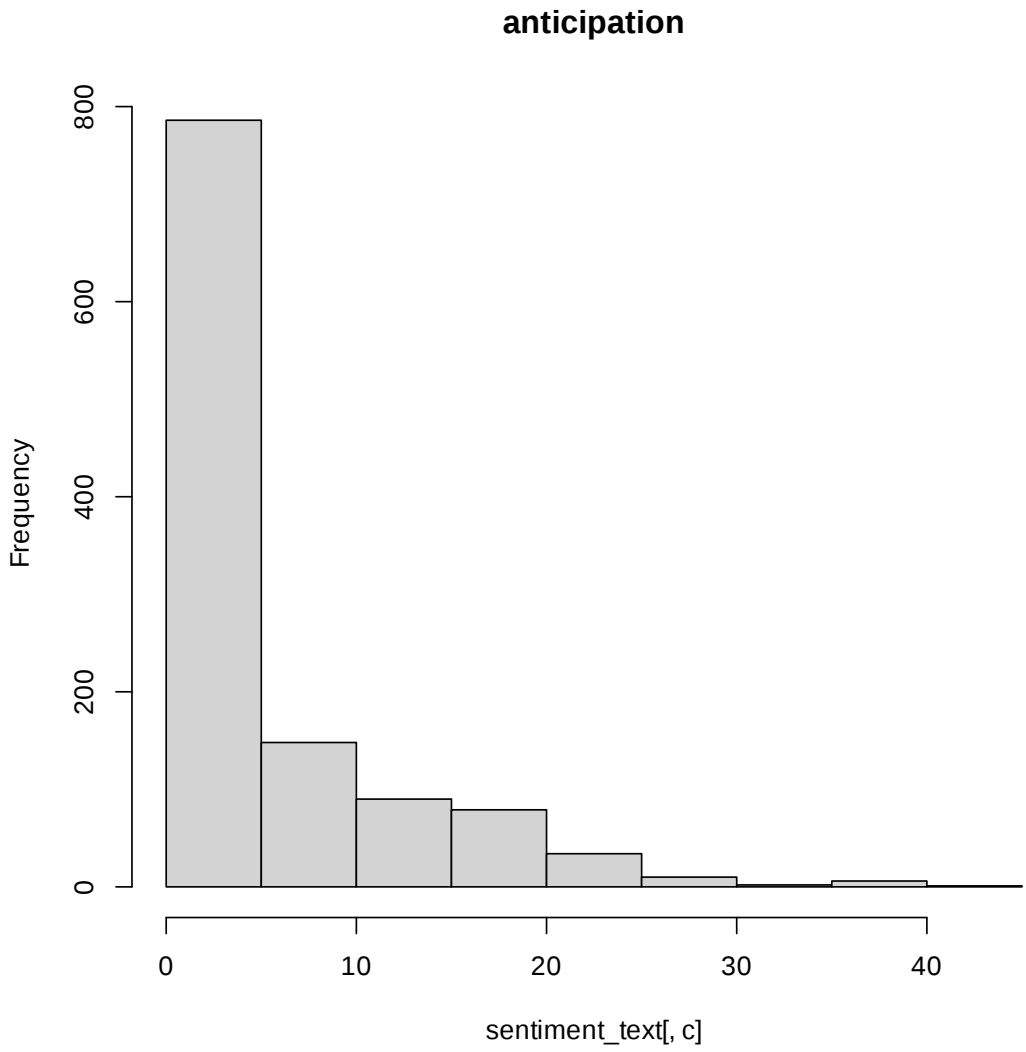


[92]:
```
#caribbean sentiment histogram
sentiment_hist(caribbean_sentiment)
```

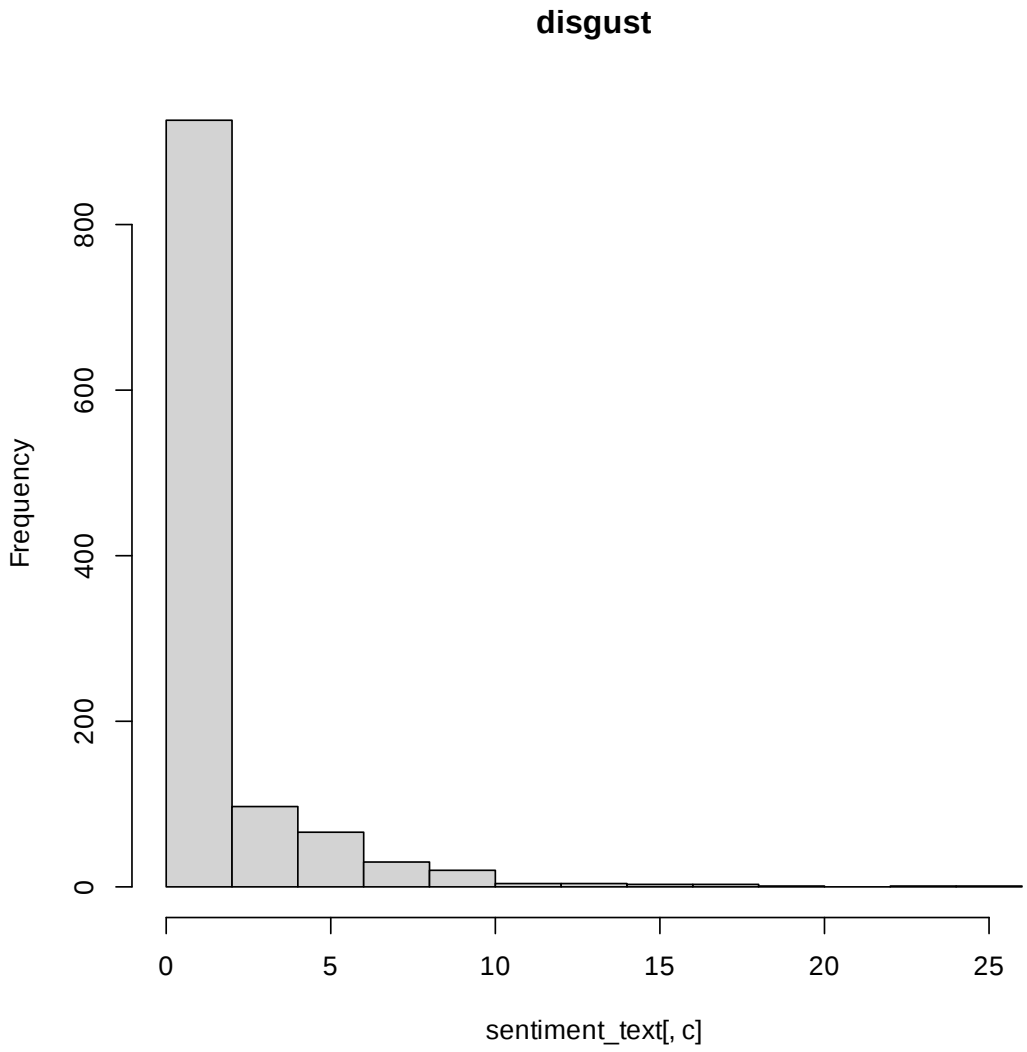
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
200
400
600
800
1000
0
10
20
30
40
sentiment_text[, c]

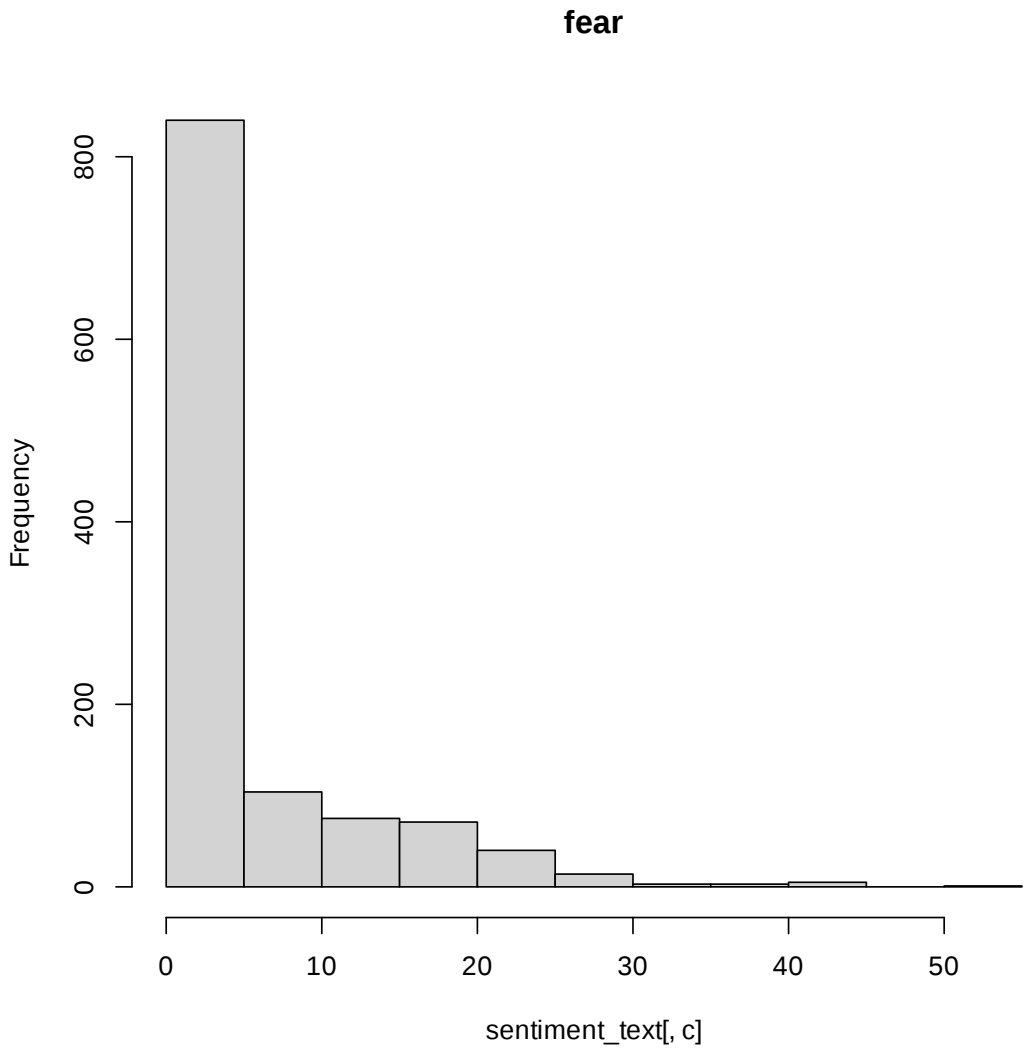
```
[1] "disgust"
```


anticipation
Frequency
0
200
400
600
800
0
10
20
30
40
sentiment_text[, c]

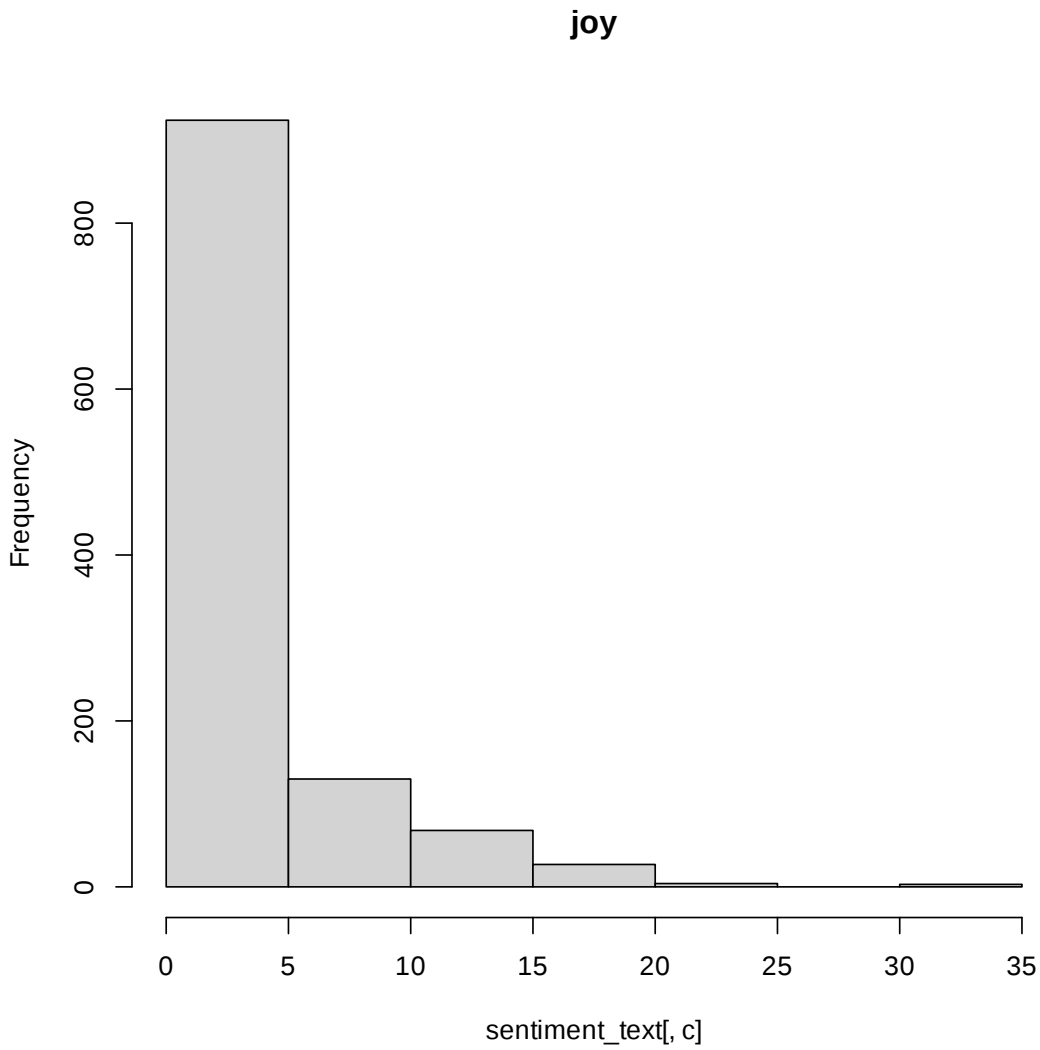
```
[1] "fear"
```


disgust
Frequency
0
200
400
600
800
0
5
10
15
20
25
sentiment_text[, c]

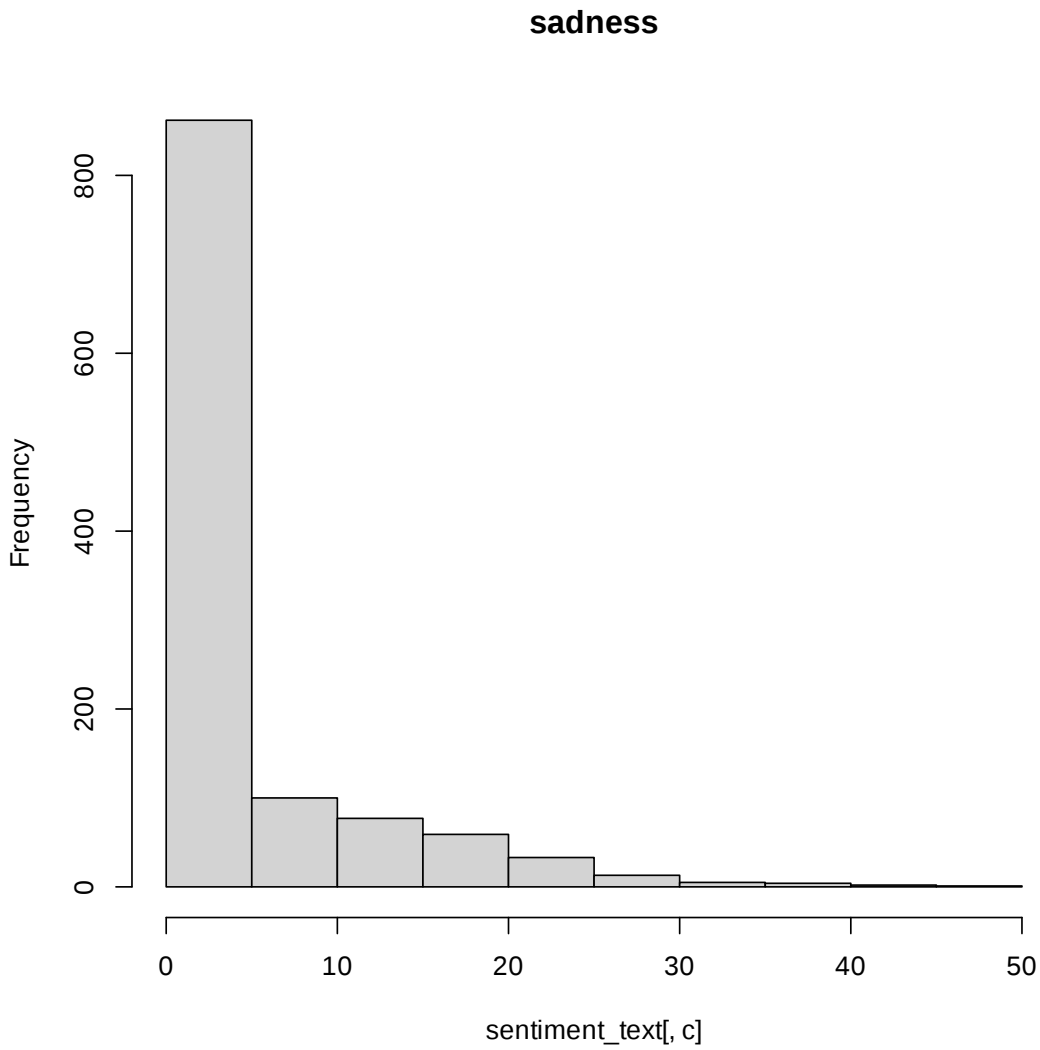
```
[1] "joy"
```


fear
Frequency
0
200
400
600
800
0
10
20
30
40
50
sentiment_text[, c]

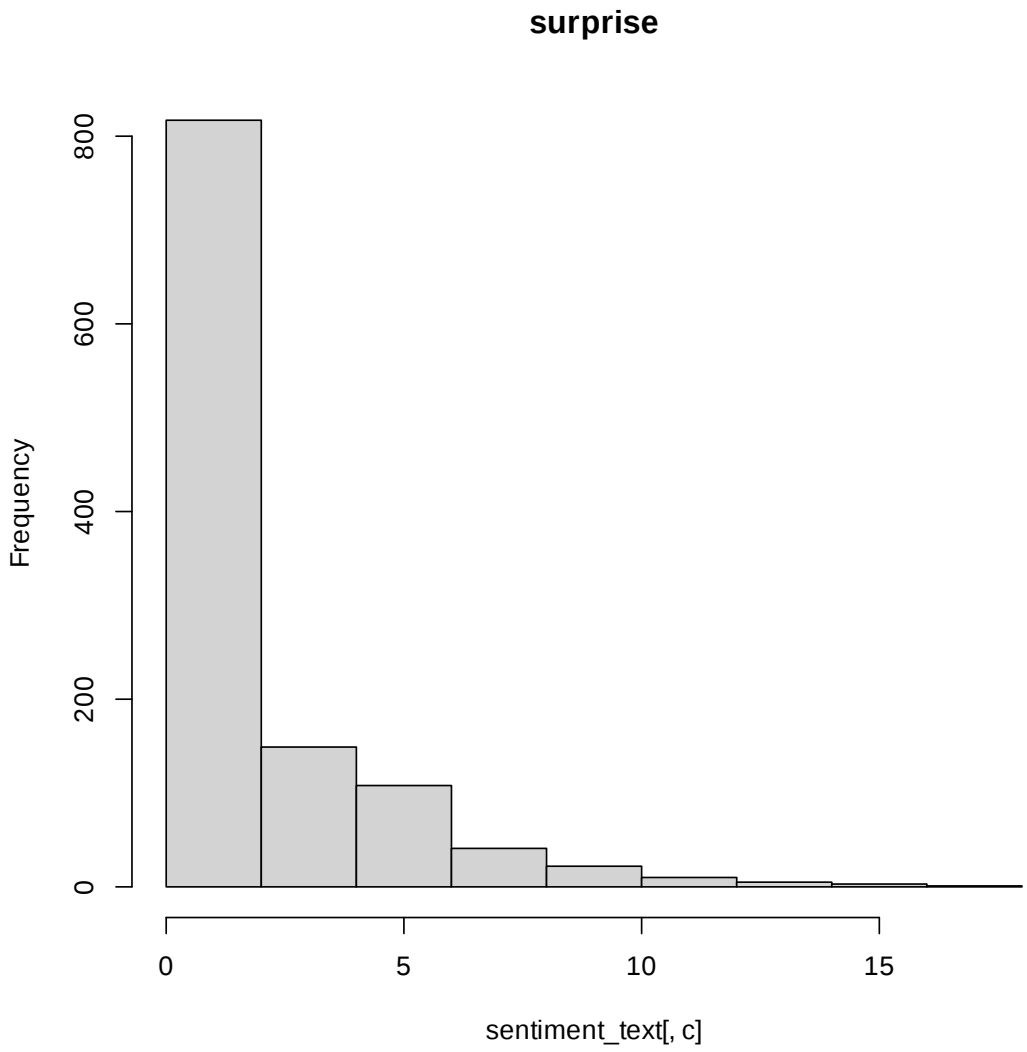
```
[1] "sadness"
```


joy
Frequency
0
200
400
600
800
0
5
10
15
20
25
30
35
sentiment_text[, c]

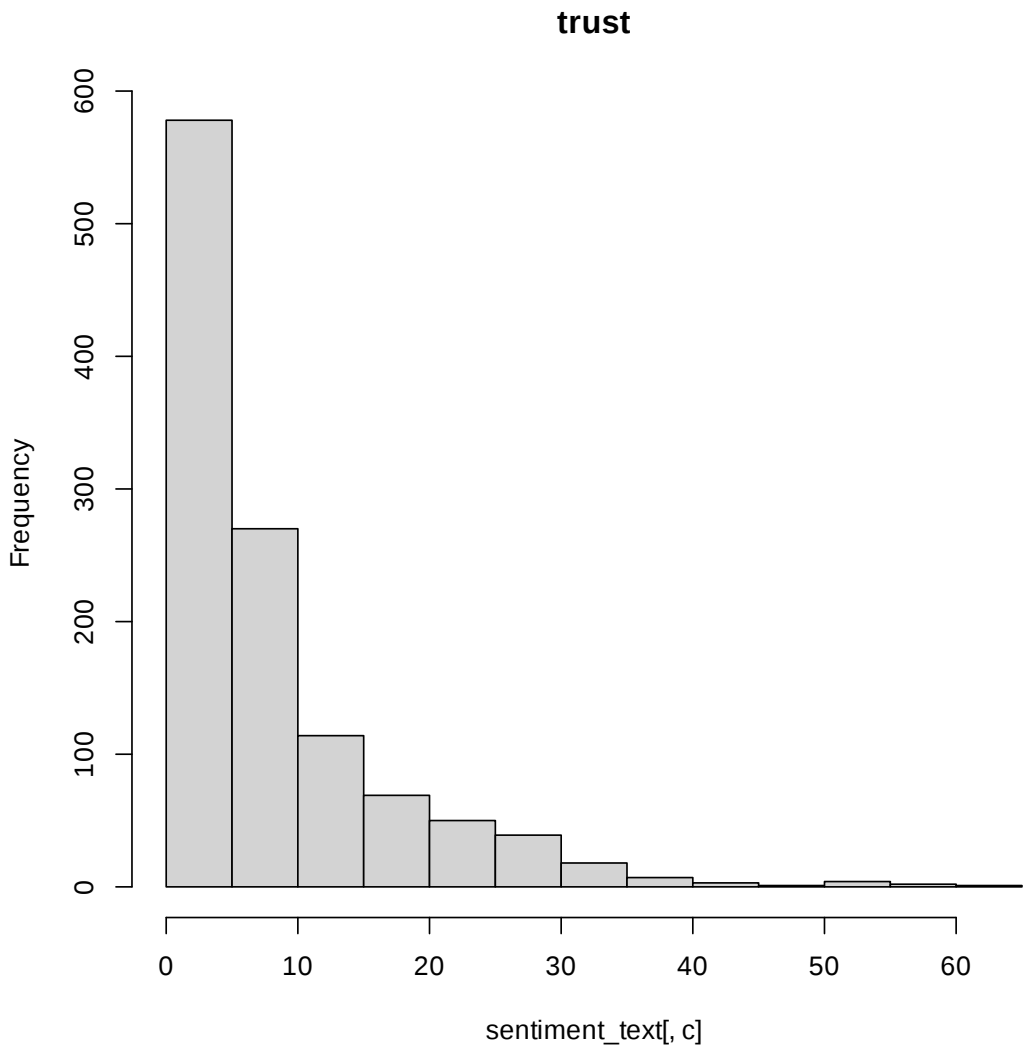
```
[1] "surprise"
```


sadness
Frequency
0
200
400
600
800
0
10
20
30
40
50
sentiment_text[, c]

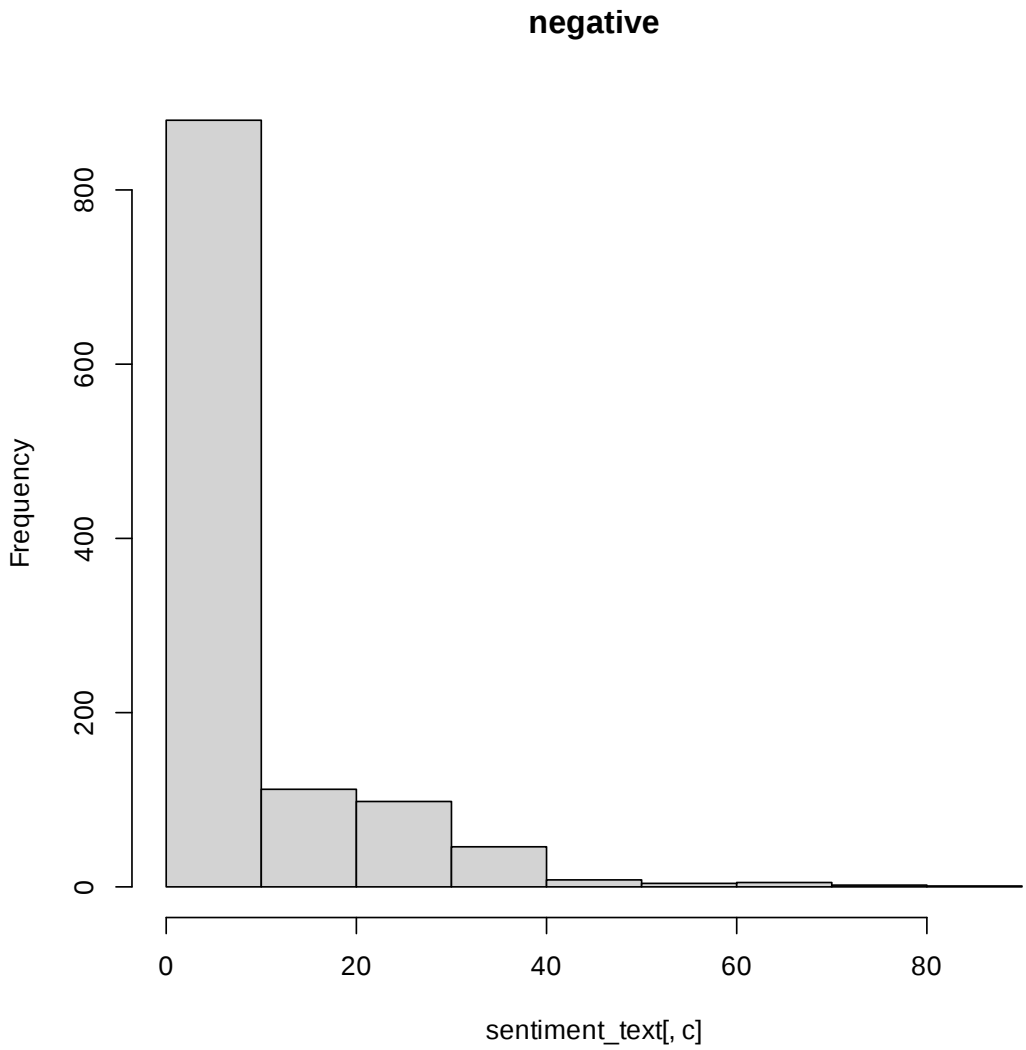
```
[1] "trust"
```

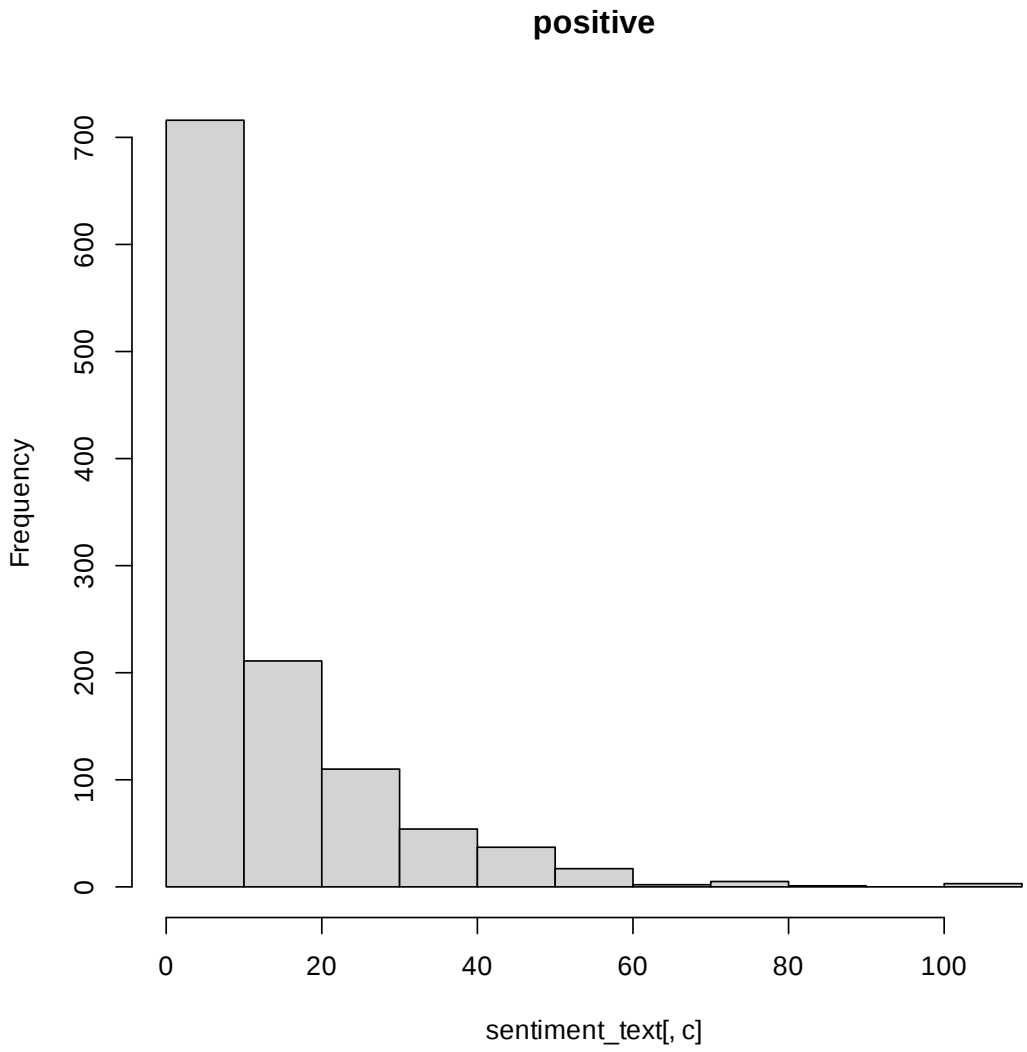

surprise
Frequency
0
200
400
600
800
0
5
10
15
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
100
200
300
400
500
600
0
10
20
30
40
50
60
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
200
400
600
800
0
20
40
60
80
sentiment_text[, c]



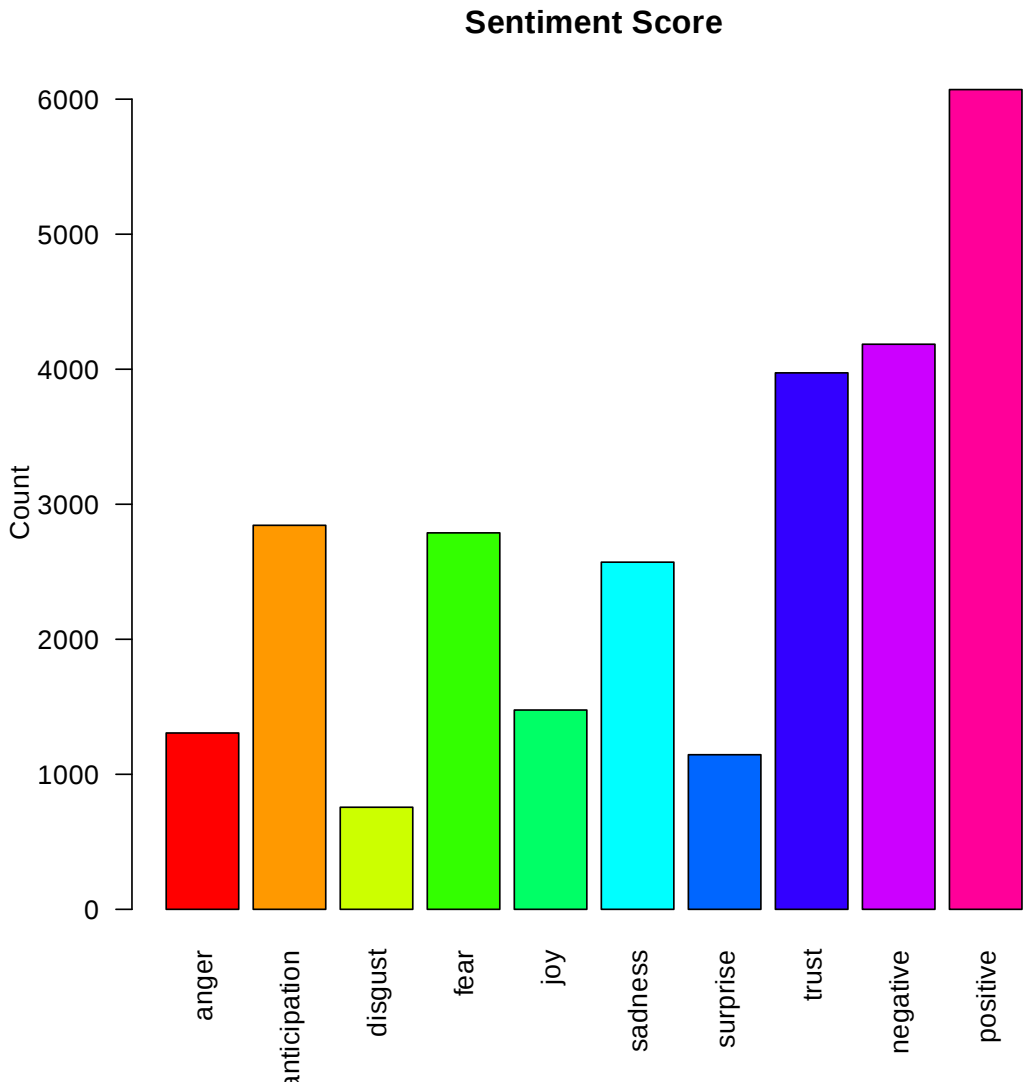
[93]:
```
#caribbean average sentiment
average_caribbean<- average_sentiment(caribbean_sentiment)
summary(average_caribbean)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-20.000   0.000   3.000   4.221   7.000  80.000
```

[62]:
```
#asian barplot
barplot_function(clean.data.asian$Text)
```

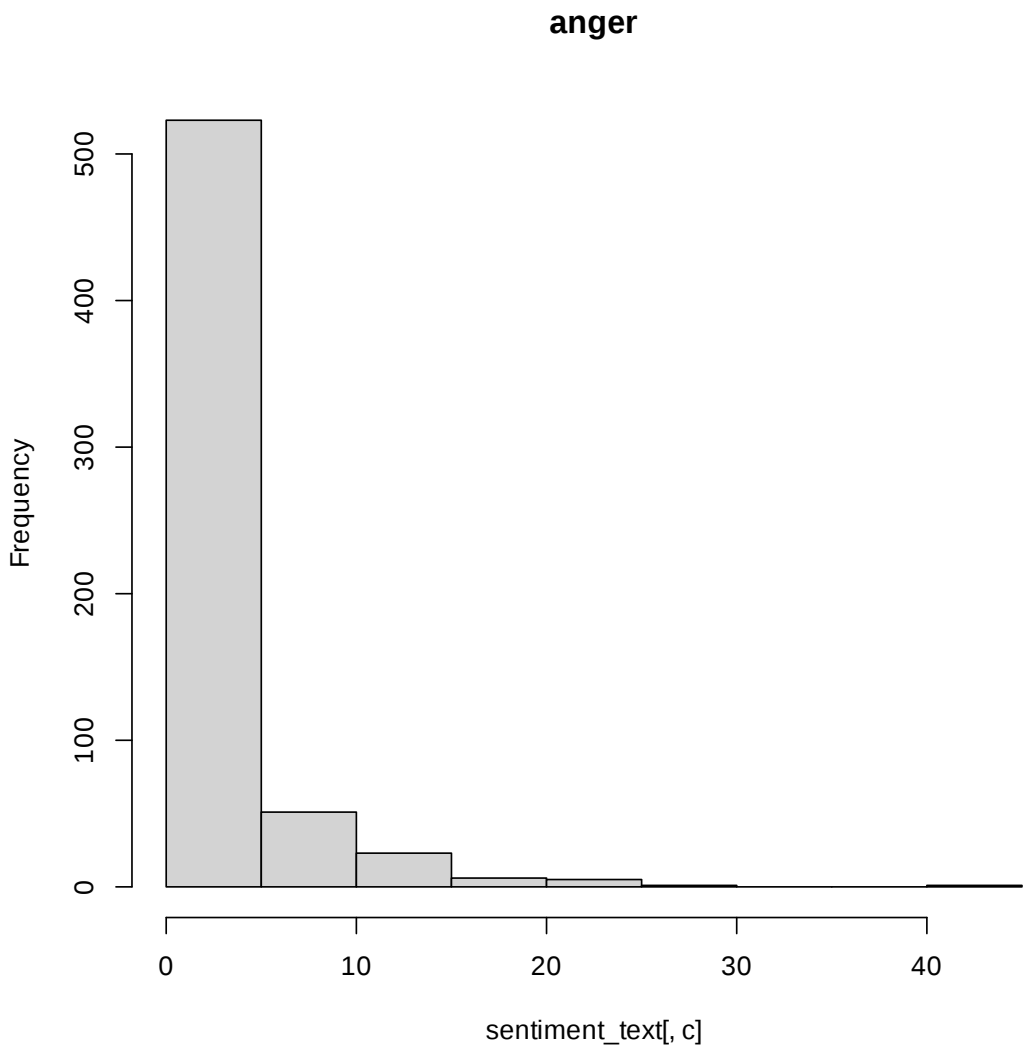
```
     anger          anticipation       disgust            fear
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.00
 1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.00
```

```
Median : 0.000   Median : 2.000   Median : 0.000   Median : 1.00
Mean   : 2.141   Mean   : 4.662   Mean   : 1.239   Mean   : 4.57
3rd Qu.: 2.000   3rd Qu.: 6.000   3rd Qu.: 1.000   3rd Qu.: 5.00
Max.   :41.000   Max.   :52.000   Max.   :28.000   Max.   :63.00
     joy           sadness          surprise          trust
Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
Median : 0.00   Median : 1.000   Median : 1.000   Median : 4.000
Mean   : 2.42   Mean   : 4.215   Mean   : 1.877   Mean   : 6.513
3rd Qu.: 3.00   3rd Qu.: 5.000   3rd Qu.: 3.000   3rd Qu.: 9.000
Max.   :41.00   Max.   :66.000   Max.   :25.000   Max.   :66.000
   negative          positive
Min.   :  0.000   Min.   :  0.000
1st Qu.:  0.000   1st Qu.:  0.000
Median :  1.000   Median :  6.000
Mean   :  6.861   Mean   :  9.952
3rd Qu.:  7.000   3rd Qu.: 14.000
Max.   :107.000   Max.   :120.000
```

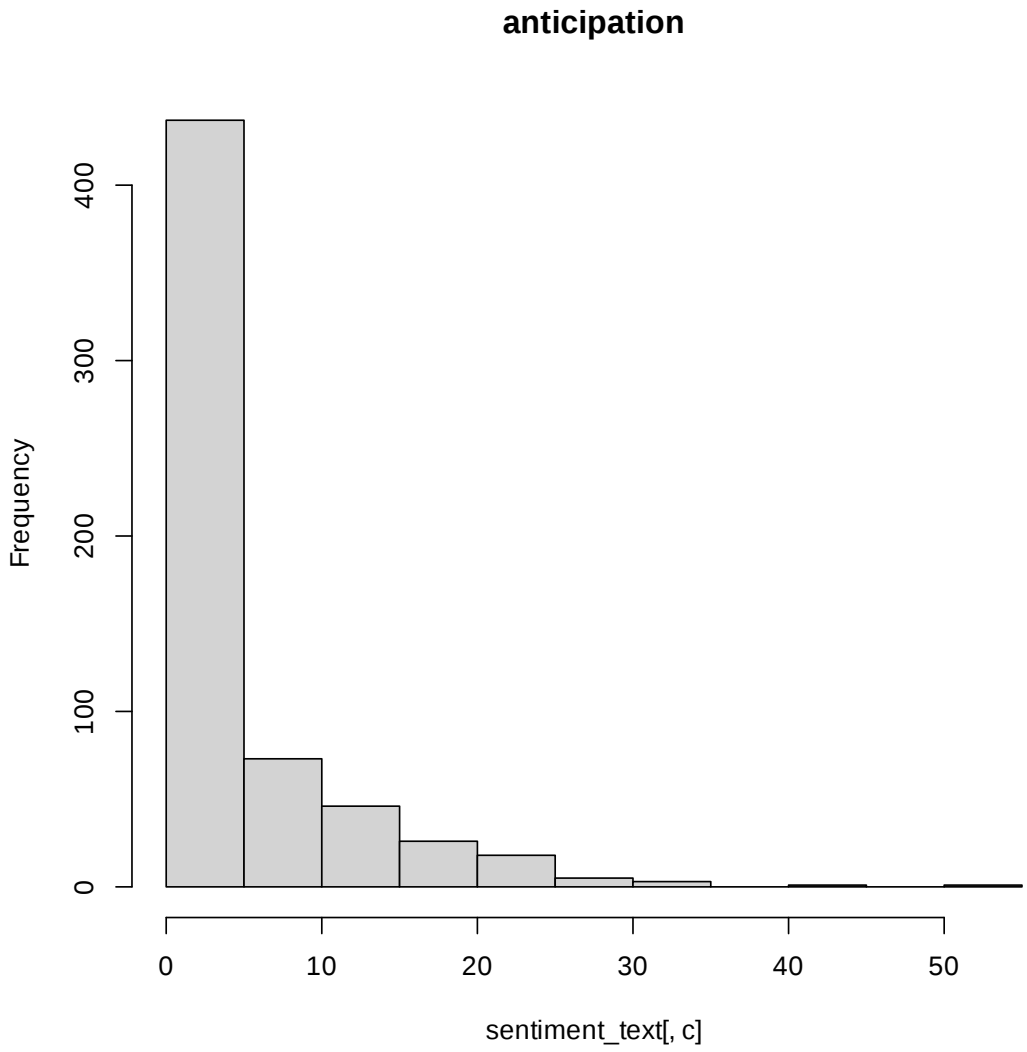


[91]:
```
#asian sentiment histogram
sentiment_hist(asian_sentiment)
```

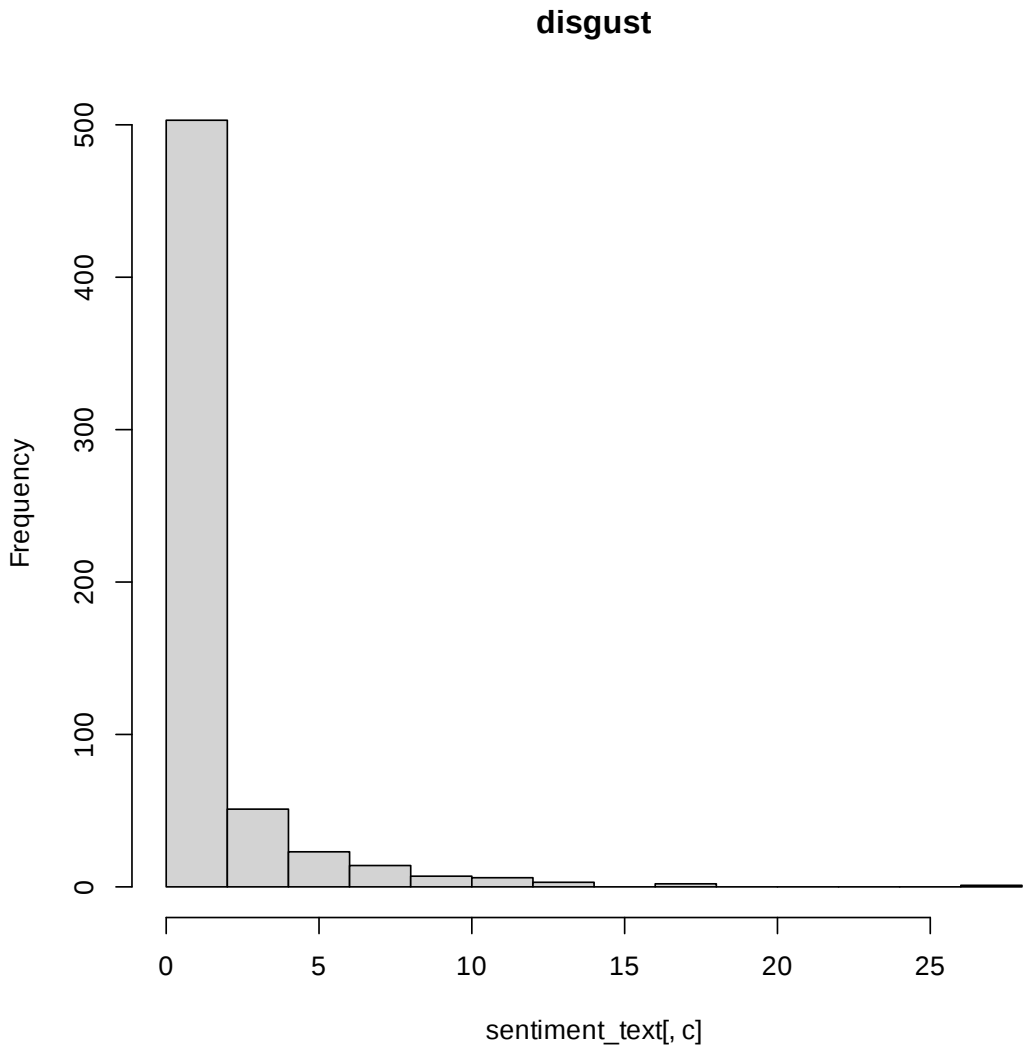
```
[1] "anger"
[1] "anticipation"
```


anger
Frequency
0
100
200
300
400
500
0
10
20
30
40
sentiment_text[, c]

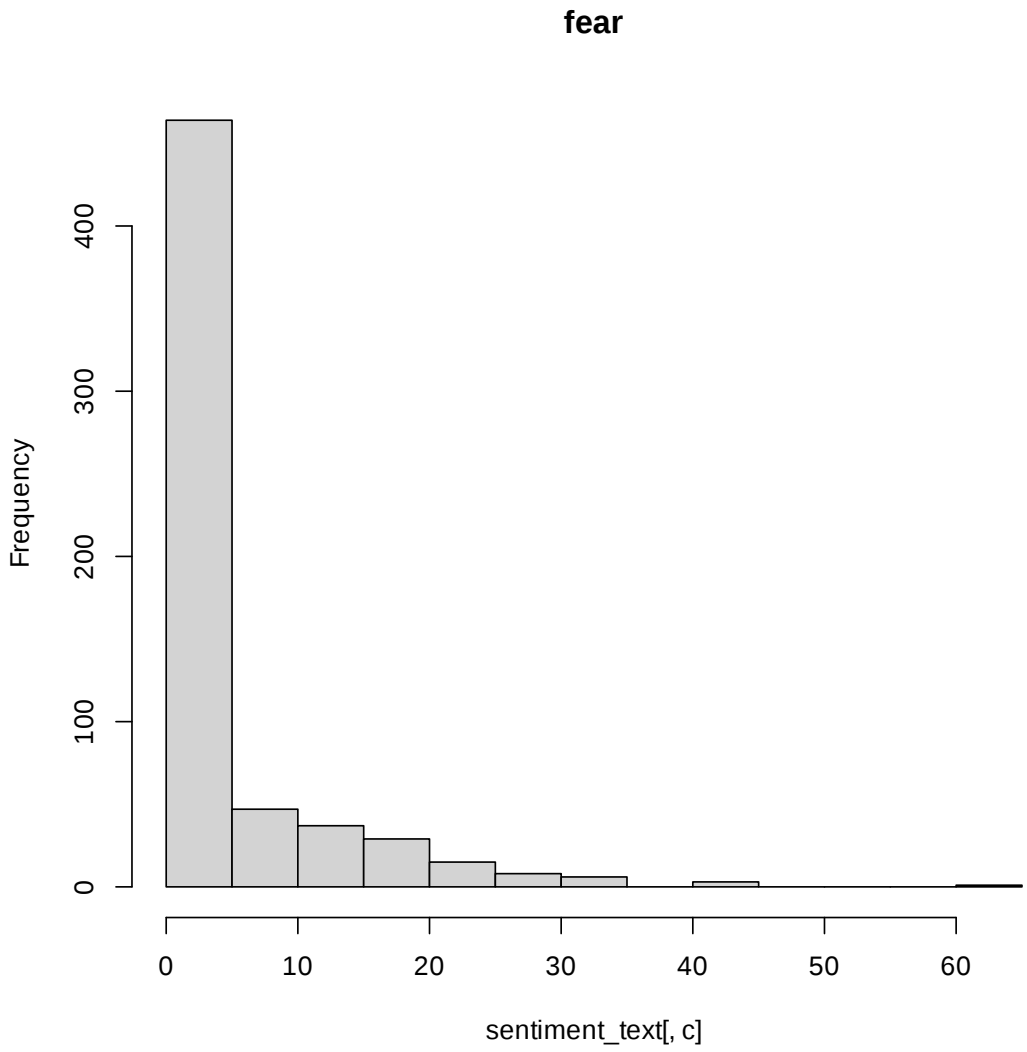
```
[1] "disgust"
```


anticipation
Frequency
0
100
200
300
400
0
10
20
30
40
50
sentiment_text[, c]

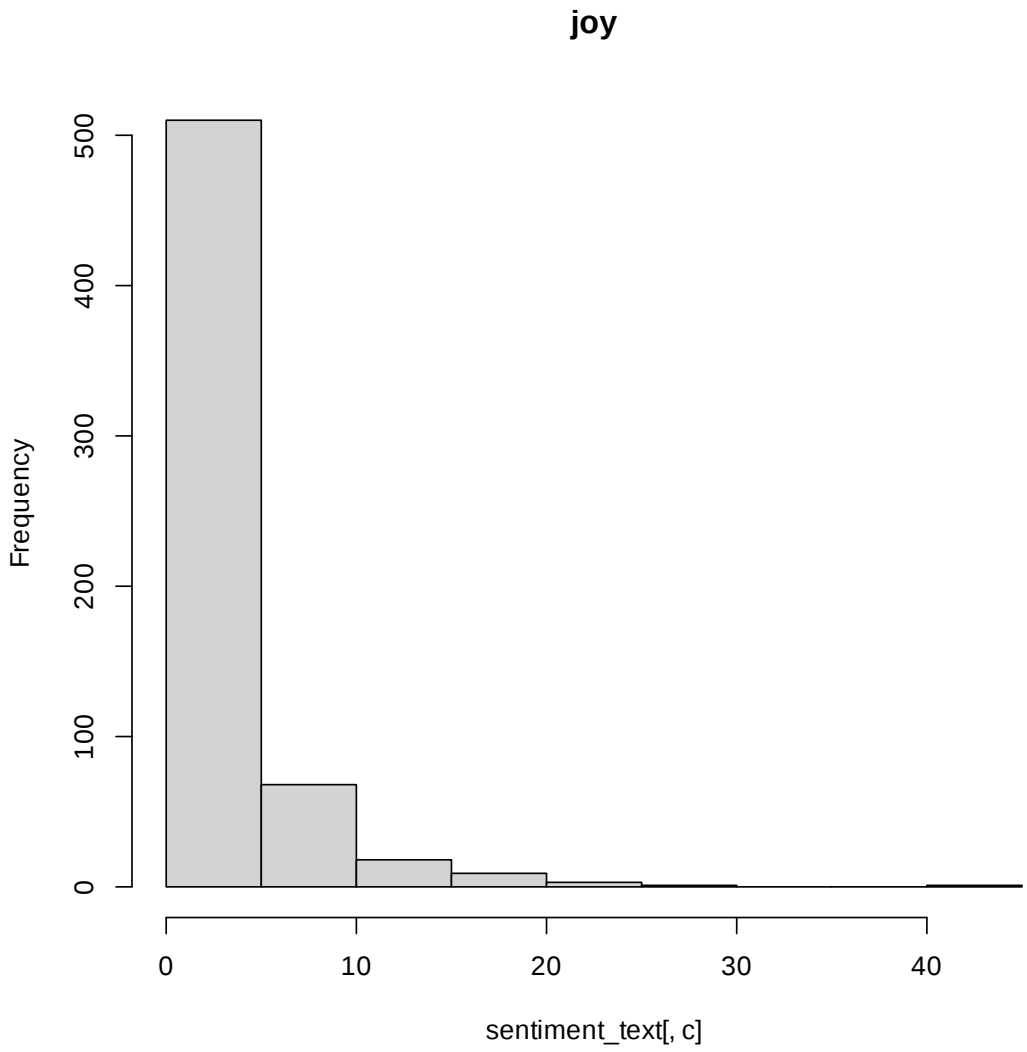
```
[1] "fear"
```


disgust
Frequency
0
100
200
300
400
500
0
5
10
15
20
25
sentiment_text[, c]

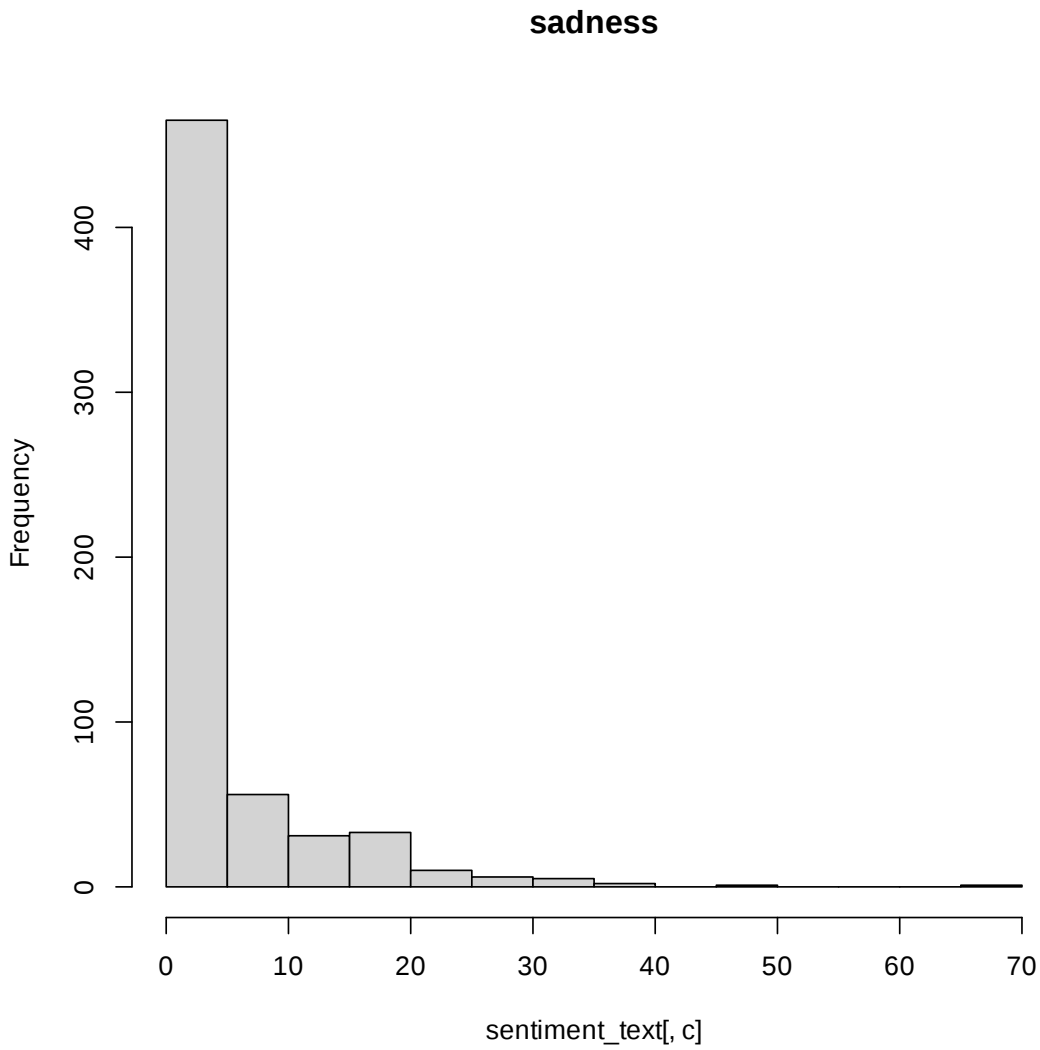
```
[1] "joy"
```


fear
Frequency
0
100
200
300
400
0
10
20
30
40
50
60
sentiment_text[, c]

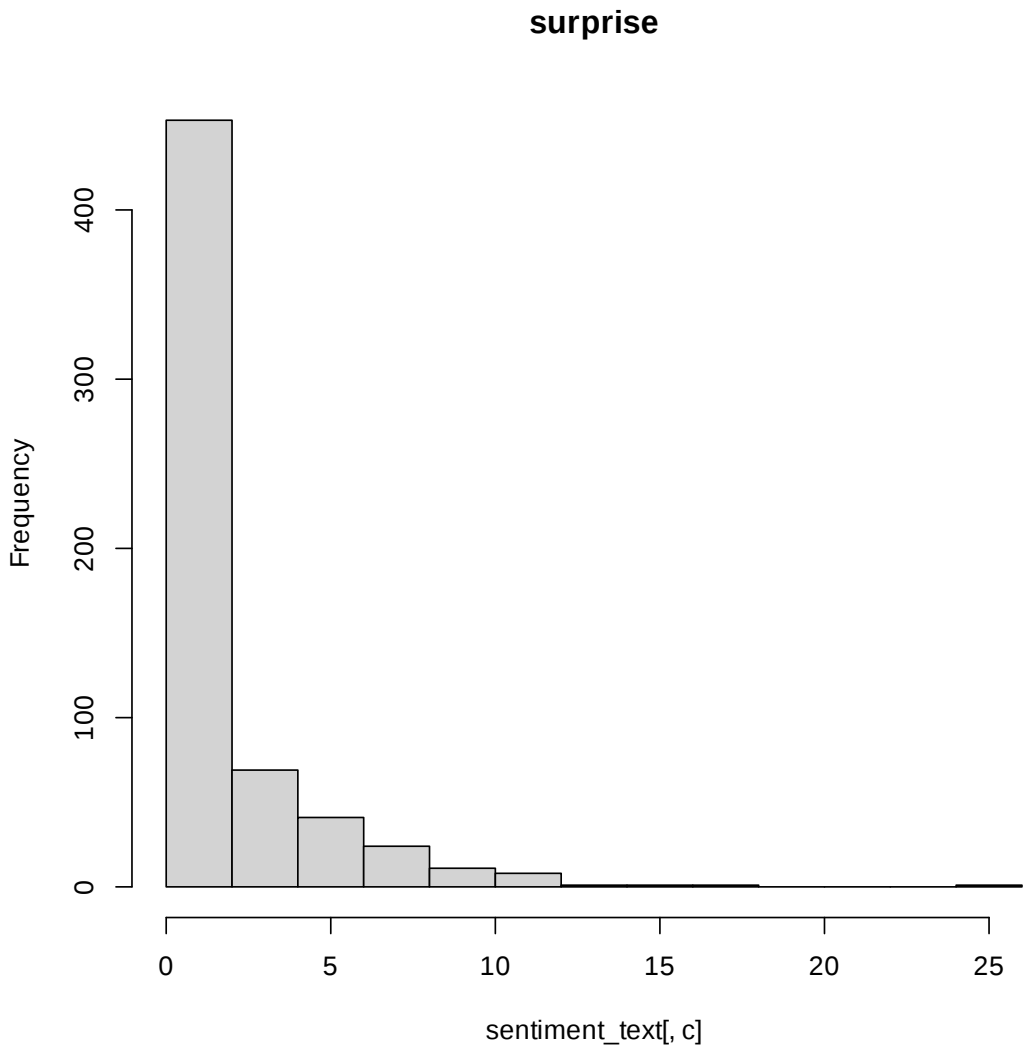
```
[1] "sadness"
```


joy
Frequency
0
100
200
300
400
500
0
10
20
30
40
sentiment_text[, c]

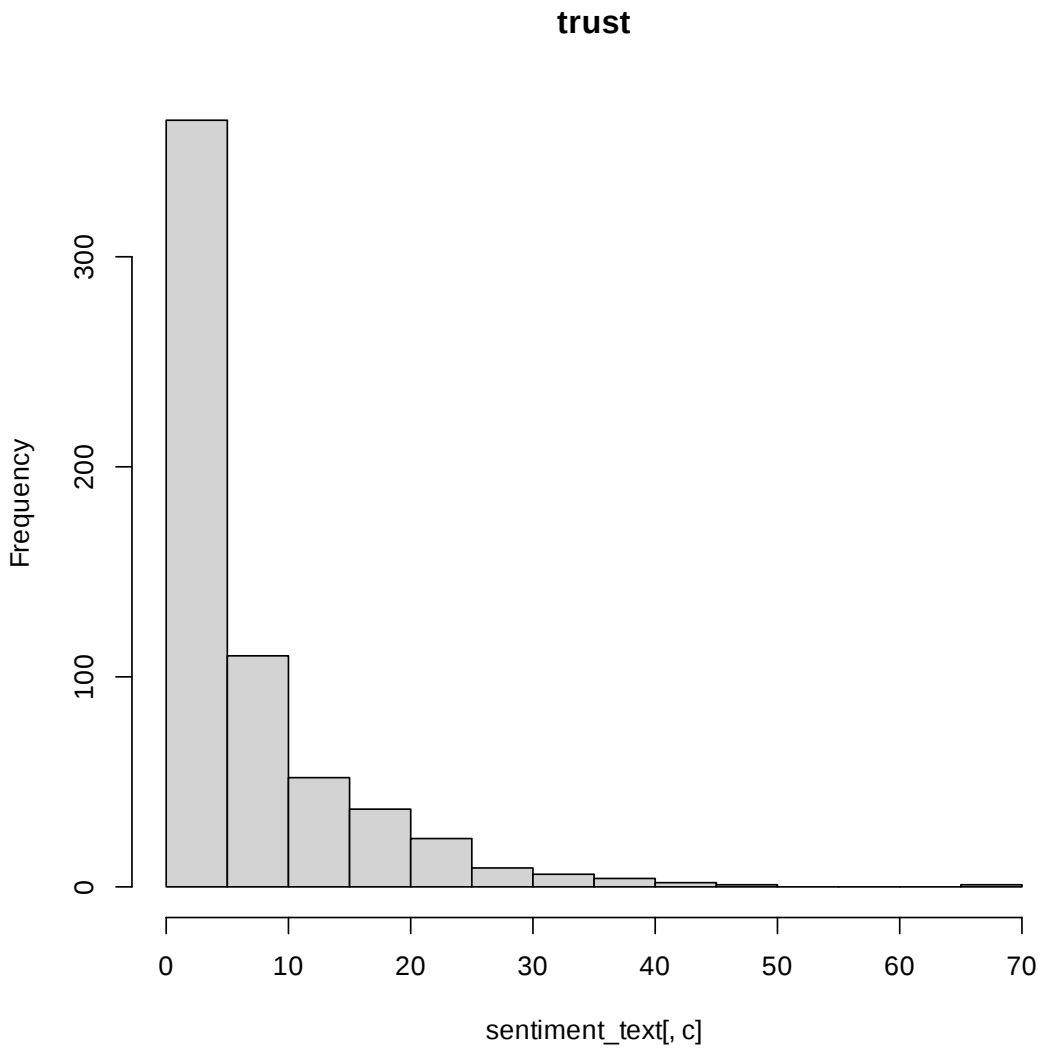
```
[1] "surprise"
```


sadness
Frequency
0
100
200
300
400
0
10
20
30
40
50
60
70
sentiment_text[, c]

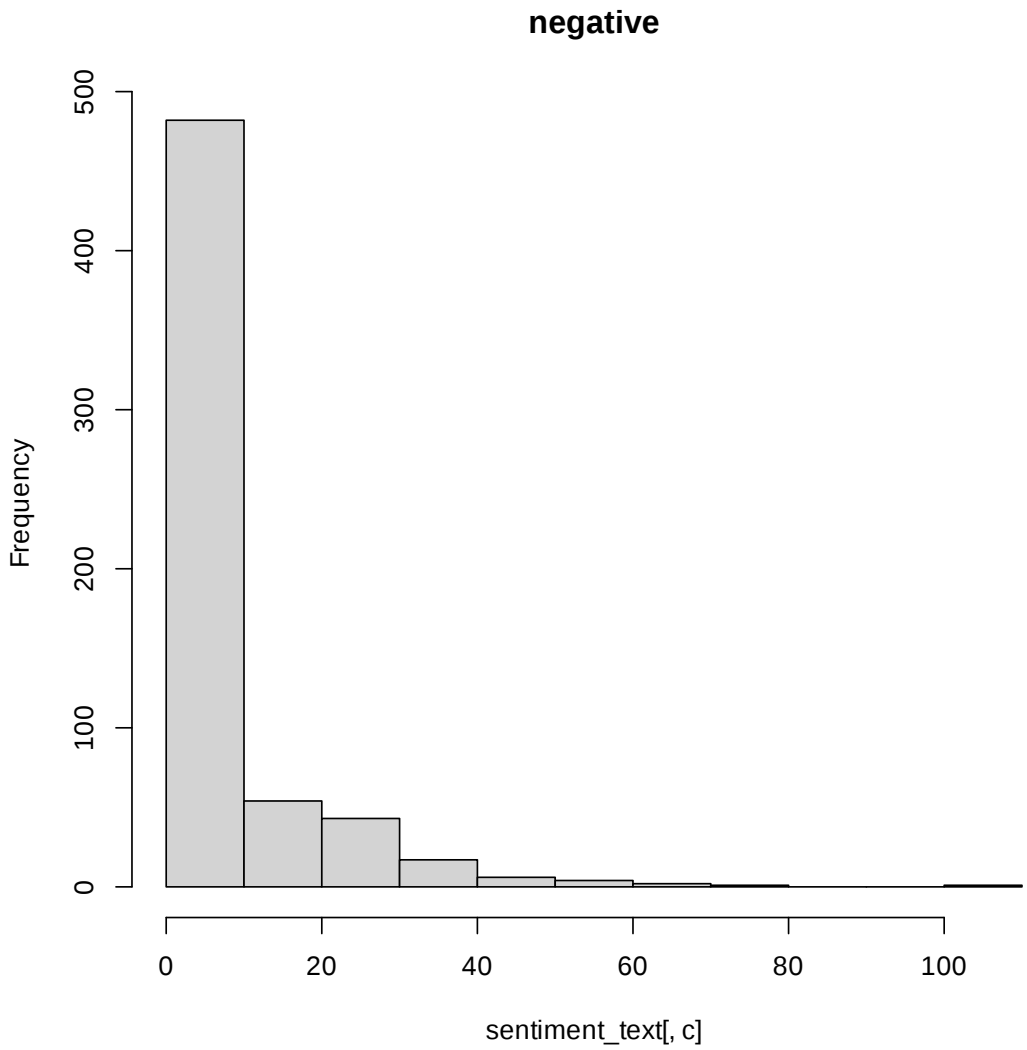
```
[1] "trust"
```

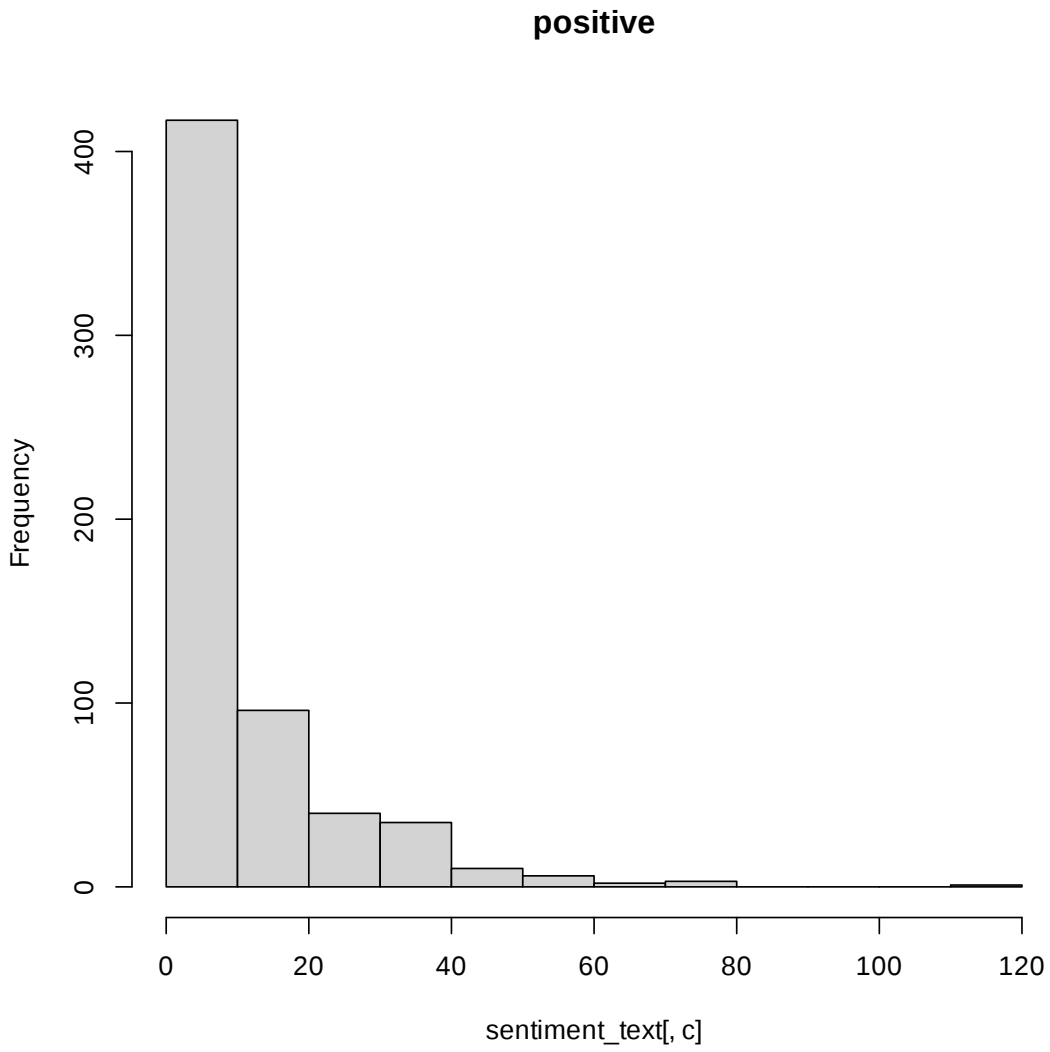

surprise
Frequency
0
100
200
300
400
0
5
10
15
20
25
sentiment_text[, c]

```
[1] "negative"
```


trust
Frequency
0
100
200
300
0
10
20
30
40
50
60
70
sentiment_text[, c]

```
[1] "positive"
```


negative
Frequency
0
100
200
300
400
500
0
20
40
60
80
100
sentiment_text[, c]



[94]:
```
#asian average sentiment
average_asian<- average_sentiment(asian_sentiment)
summary(average_asian)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-16.000   0.000   1.000   3.092   6.000  23.000
```

[73]:
```
#eunw average sentiment
average_eunw<- average_sentiment(eunw_sentiment)
summary(average_eunw)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-22.000   0.000   2.000   3.588   6.000  54.000
```

[76]:
```
#caribbean average sentiment
average_caribbean<- average_sentiment(caribbean_sentiment)
summary(average_caribbean)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-20.000   0.000   3.000   4.221   7.000  80.000
```

[66]:
```
#asian average sentiment
average_asian<- average_sentiment(asian_sentiment)
asian_summary<-summary(average_asian)
```