

## web scraping

2023-09-25

```
# Url Links for web scraping
lista_url <-
c("https://www.federalreserve.gov/newsevents/pressreleases/monetary20230920a.htm",
  "https://www.federalreserve.gov/newsevents/pressreleases/monetary20230726a.htm",
  "https://www.federalreserve.gov/newsevents/pressreleases/monetary20230614a.htm",
  "https://www.federalreserve.gov/newsevents/pressreleases/monetary20230503a.htm")

##
$`https://www.federalreserve.gov/newsevents/pressreleases/monetary20230920a.htm`
## [1] "September 20, 2023"
## [2] "Recent indicators suggest that economic activity has been expanding at a
solid pace. Job gains have slowed in recent months but remain strong, and the
unemployment rate has remained low. Inflation remains elevated."
## [3] "The U.S. banking system is sound and resilient. Tighter credit conditions
for households and businesses are likely to weigh on economic activity, hiring,
and inflation. The extent of these effects remains uncertain. The Committee
remains highly attentive to inflation risks."
## [4] "The Committee seeks to achieve maximum employment and inflation at the
rate of 2 percent over the longer run. In support of these goals, the Committee
decided to maintain the target range for the federal funds rate at 5-1/4 to 5-1/2
percent. The Committee will continue to assess additional information and its
implications for monetary policy. In determining the extent of additional policy
firming that may be appropriate to return inflation to 2 percent over time, the
Committee will take into account the cumulative tightening of monetary policy, the
lags with which monetary policy affects economic activity and inflation, and
economic and financial developments. In addition, the Committee will continue
reducing its holdings of Treasury securities and agency debt and agency mortgage-
backed securities, as described in its previously announced plans. The Committee
is strongly committed to returning inflation to its 2 percent objective."
##
##
$`https://www.federalreserve.gov/newsevents/pressreleases/monetary20230726a.htm`
## [1] "July 26, 2023"
## [2] "Recent indicators suggest that economic activity has been expanding at a
moderate pace. Job gains have been robust in recent months, and the unemployment
rate has remained low. Inflation remains elevated."
## [3] "The U.S. banking system is sound and resilient. Tighter credit conditions
for households and businesses are likely to weigh on economic activity, hiring,
and inflation. The extent of these effects remains uncertain. The Committee
remains highly attentive to inflation risks."
## [4] "The Committee seeks to achieve maximum employment and inflation at the
rate of 2 percent over the longer run. In support of these goals, the Committee
```

decided to raise the target range for the federal funds rate to 5-1/4 to 5-1/2 percent. The Committee will continue to assess additional information and its implications for monetary policy. In determining the extent of additional policy firming that may be appropriate to return inflation to 2 percent over time, the Committee will take into account the cumulative tightening of monetary policy, the lags with which monetary policy affects economic activity and inflation, and economic and financial developments. In addition, the Committee will continue reducing its holdings of Treasury securities and agency debt and agency mortgage-backed securities, as described in its previously announced plans. The Committee is strongly committed to returning inflation to its 2 percent objective."

##

##

\$` <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230614a.htm>`

## [1] "June 14, 2023"

## [2] "Recent indicators suggest that economic activity has continued to expand at a modest pace. Job gains have been robust in recent months, and the unemployment rate has remained low. Inflation remains elevated."

## [3] "The U.S. banking system is sound and resilient. Tighter credit conditions for households and businesses are likely to weigh on economic activity, hiring, and inflation. The extent of these effects remains uncertain. The Committee remains highly attentive to inflation risks."

## [4] "The Committee seeks to achieve maximum employment and inflation at the rate of 2 percent over the longer run. In support of these goals, the Committee decided to maintain the target range for the federal funds rate at 5 to 5-1/4 percent. Holding the target range steady at this meeting allows the Committee to assess additional information and its implications for monetary policy. In determining the extent of additional policy firming that may be appropriate to return inflation to 2 percent over time, the Committee will take into account the cumulative tightening of monetary policy, the lags with which monetary policy affects economic activity and inflation, and economic and financial developments. In addition, the Committee will continue reducing its holdings of Treasury securities and agency debt and agency mortgage-backed securities, as described in its previously announced plans. The Committee is strongly committed to returning inflation to its 2 percent objective."

##

##

\$` <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230503a.htm>`

## [1] "May 03, 2023"

## [2] "Economic activity expanded at a modest pace in the first quarter. Job gains have been robust in recent months, and the unemployment rate has remained low. Inflation remains elevated."

## [3] "The U.S. banking system is sound and resilient. Tighter credit conditions for households and businesses are likely to weigh on economic activity, hiring, and inflation. The extent of these effects remains uncertain. The Committee remains highly attentive to inflation risks."

## [4] "The Committee seeks to achieve maximum employment and inflation at the rate of 2 percent over the longer run. In support of these goals, the Committee decided to raise the target range for the federal funds rate to 5 to 5-1/4 percent. The Committee will closely monitor incoming information and assess the implications for monetary policy. In determining the extent to which additional policy firming may be appropriate to return inflation to 2 percent over time, the

Committee will take into account the cumulative tightening of monetary policy, the lags with which monetary policy affects economic activity and inflation, and economic and financial developments. In addition, the Committee will continue reducing its holdings of Treasury securities and agency debt and agency mortgage-backed securities, as described in its previously announced plans. The Committee is strongly committed to returning inflation to its 2 percent objective."

*# The 'tm' package (text mining) can be used to clean up the text. First, text should be converted to 'Corpus', or a list of the text we would like to use.*

```
corpus <- Corpus(VectorSource(dati_siti))
```

```
corpus <- tm_map(corpus, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(corpus, content_transformer(tolower)):
```

```
## transformation drops documents
```

```
corpus <- tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
```

```
## documents
```

```
corpus <- tm_map(corpus, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removeNumbers): transformation drops
```

```
## documents
```

```
corpus <- tm_map(corpus, removeWords, stopwords("english"))
```

```
## Warning in tm_map.SimpleCorpus(corpus, removeWords, stopwords("english")):
```

```
## transformation drops documents
```

```
inspect(corpus)
```

```
## <<SimpleCorpus>>
```

```
## Metadata: corpus specific: 1, document level (indexed): 0
```

```
## Content: documents: 4
```

```
##
```

```
## [1] cseptember recent indicators suggest economic activity expanding  
solid pace job gains slowed recent months remain strong unemployment rate  
remained low inflation remains elevated us banking system sound resilient  
tighter credit conditions households businesses likely weigh economic  
activity hiring inflation extent effects remains uncertain committee remains  
highly attentive inflation risks \n committee seeks achieve maximum employment  
inflation rate percent longer run support goals committee decided  
maintain target range federal funds rate percent committee will continue  
assess additional information implications monetary policy determining extent  
additional policy firming may appropriate return inflation percent time  
committee will take account cumulative tightening monetary policy lags  
monetary policy affects economic activity inflation economic financial  
developments addition committee will continue reducing holdings treasury  
securities agency debt agency mortgagebacked securities described previously  
announced plans committee strongly committed returning inflation percent  
objective\n
```

```
## [2] cjuly recent indicators suggest economic activity expanding moderate
```

pace job gains robust recent months unemployment rate remained low inflation remains elevated us banking system sound resilient tighter credit conditions households businesses likely weigh economic activity hiring inflation extent effects remains uncertain committee remains highly attentive inflation risks \n committee seeks achieve maximum employment inflation rate percent longer run support goals committee decided raise target range federal funds rate percent committee will continue assess additional information implications monetary policy determining extent additional policy firming may appropriate return inflation percent time committee will take account cumulative tightening monetary policy lags monetary policy affects economic activity inflation economic financial developments addition committee will continue reducing holdings treasury securities agency debt agency mortgagebacked securities described previously announced plans committee strongly committed returning inflation percent objective\n

```
## [3] cJune recent indicators suggest economic activity continued expand modest pace job gains robust recent months unemployment rate remained low inflation remains elevated us banking system sound resilient tighter credit conditions households businesses likely weigh economic activity hiring inflation extent effects remains uncertain committee remains highly attentive inflation risks \n committee seeks achieve maximum employment inflation rate percent longer run support goals committee decided maintain target range federal funds rate percent holding target range steady meeting allows committee assess additional information implications monetary policy determining extent additional policy firming may appropriate return inflation percent time committee will take account cumulative tightening monetary policy lags monetary policy affects economic activity inflation economic financial developments addition committee will continue reducing holdings treasury securities agency debt agency mortgagebacked securities described previously announced plans committee strongly committed returning inflation percent objective\n
```

```
## [4] cMay economic activity expanded modest pace first quarter job gains robust recent months unemployment rate remained low inflation remains elevated us banking system sound resilient tighter credit conditions households businesses likely weigh economic activity hiring inflation extent effects remains uncertain committee remains highly attentive inflation risks committee seeks achieve maximum employment inflation rate percent longer run support goals committee decided raise target range federal funds rate percent committee will closely monitor incoming information assess implications monetary policy determining extent additional policy firming may appropriate return inflation percent time committee will take account cumulative tightening monetary policy lags monetary policy affects economic activity inflation economic financial developments addition committee will continue reducing holdings treasury securities agency debt agency mortgagebacked securities described previously announced plans committee strongly committed returning inflation percent objective\n
```

*# Making text document*

```
tdm <- TermDocumentMatrix(corpus)
```

*#Term Frequencies*

```
frequenze <- row_sums(as.matrix(tdm))
tabella_frequenze <- data.frame(parola = names(frequenze), frequenza = frequenze)
tabella_frequenze <- tabella_frequenze[order(-tabella_frequenze$frequenza), ]
```

```
head(tabella_frequenze,15)
```

```
##          parola frequenza
## committee committee      28
## inflation inflation      28
## economic economic       16
## percent percent        16
## policy policy          16
## activity activity       12
## monetary monetary      12
## rate rate             12
## remains remains       12
## will will            11
## agency agency         8
## extent extent         8
## securities securities   8
## additional additional   7
## recent recent         7
```

```
term_freq <- frequenze[order(-frequenze)]
head(term_freq)
```

```
## committee inflation economic percent policy activity
##          28          28          16          16          16          12
```

```
# WordCloud
```

```
word.counts<-as.matrix(TermDocumentMatrix(corpus))
word.freq<-sort(rowSums(word.counts), decreasing=TRUE)
head(word.freq)##what are the top words?
```

```
## committee inflation economic percent policy activity
##          28          28          16          16          16          12
```

```
set.seed(32) #be sure to set the seed if you want to reproduce the same again
```

```
wordcloud(words=names(word.freq), freq=word.freq, scale=c(4,.3),max.words = 100,
          random.order = TRUE, color=wes_palette("Royal1"))
```



inflation rate percent longer run support goals committee decided  
maintain target range federal funds rate percent committee will continue  
assess additional information implications monetary policy determining extent  
additional policy firming may appropriate return inflation percent time  
committee will take account cumulative tightening monetary policy lags  
monetary policy affects economic activity inflation economic financial  
developments addition committee will continue reducing holdings treasury  
securities agency debt agency mortgagebacked securities described previously  
announced plans committee strongly committed returning inflation percent  
objective\n"

## [2] "c|july recent indicators suggest economic activity expanding  
moderate pace job gains robust recent months unemployment rate remained low  
inflation remains elevated us banking system sound resilient tighter credit  
conditions households businesses likely weigh economic activity hiring  
inflation extent effects remains uncertain committee remains highly attentive  
inflation risks \n committee seeks achieve maximum employment inflation rate  
percent longer run support goals committee decided raise target range  
federal funds rate percent committee will continue assess additional  
information implications monetary policy determining extent additional  
policy firming may appropriate return inflation percent time committee will  
take account cumulative tightening monetary policy lags monetary policy  
affects economic activity inflation economic financial developments addition  
committee will continue reducing holdings treasury securities agency debt  
agency mortgagebacked securities described previously announced plans  
committee strongly committed returning inflation percent objective\n"

## [3] "c|june recent indicators suggest economic activity continued expand  
modest pace job gains robust recent months unemployment rate remained low  
inflation remains elevated us banking system sound resilient tighter credit  
conditions households businesses likely weigh economic activity hiring  
inflation extent effects remains uncertain committee remains highly attentive  
inflation risks \n committee seeks achieve maximum employment inflation rate  
percent longer run support goals committee decided maintain target range  
federal funds rate percent holding target range steady meeting allows  
committee assess additional information implications monetary policy  
determining extent additional policy firming may appropriate return inflation  
percent time committee will take account cumulative tightening monetary  
policy lags monetary policy affects economic activity inflation economic  
financial developments addition committee will continue reducing holdings  
treasury securities agency debt agency mortgagebacked securities described  
previously announced plans committee strongly committed returning inflation  
percent objective\n"

## [4] "c|m|ay economic activity expanded modest pace first quarter job gains  
robust recent months unemployment rate remained low inflation remains elevated  
us banking system sound resilient tighter credit conditions households  
businesses likely weigh economic activity hiring inflation extent effects  
remains uncertain committee remains highly attentive inflation risks committee  
seeks achieve maximum employment inflation rate percent longer run  
support goals committee decided raise target range federal funds rate  
percent committee will closely monitor incoming information assess implications  
monetary policy determining extent additional policy firming may appropriate  
return inflation percent time committee will take account cumulative

```
tightening monetary policy lags monetary policy affects economic activity
inflation economic financial developments addition committee will continue
reducing holdings treasury securities agency debt agency mortgagebacked
securities described previously announced plans committee strongly committed
returning inflation percent objective\n"
```

```
# tokenize
```

```
tokens <- tibble(text = testi) %>% unnest_tokens(word, text)
```

```
tokens %>%
```

```
  inner_join(get_sentiments("bing")) %>% # pull out only sentiment words
```

```
  count(sentiment) %>% # count the # of positive & negative words
```

```
  spread(sentiment, n, fill = 0) %>% # made data wide rather than narrow
```

```
  mutate(sentiment = positive - negative) # # of positive words - # of negative words
```

```
## Joining, by = "word"
```

```
## # A tibble: 1 x 3
```

```
##   negative positive sentiment
```

```
##   <dbl>    <dbl>    <dbl>
```

```
## 1      17      28      11
```

```
# get words already get_sentiments("bing")
```

```
(sentiment <- tokens %>%
```

```
  inner_join(get_sentiments("bing")) %>%
```

```
  count(word, sentiment, sort = TRUE))
```

```
## Joining, by = "word"
```

```
## # A tibble: 15 x 3
```

```
##   word      sentiment    n
```

```
##   <chr>    <chr>    <int>
```

```
## 1 appropriate positive    4
```

```
## 2 attentive  positive    4
```

```
## 3 debt      negative    4
```

```
## 4 gains     positive    4
```

```
## 5 lags      negative    4
```

```
## 6 resilient positive    4
```

```
## 7 risks     negative    4
```

```
## 8 support   positive    4
```

```
## 9 uncertain negative    4
```

```
## 10 robust   positive    3
```

```
## 11 modest   positive    2
```

```
## 12 slowed   negative    1
```

```
## 13 solid    positive    1
```

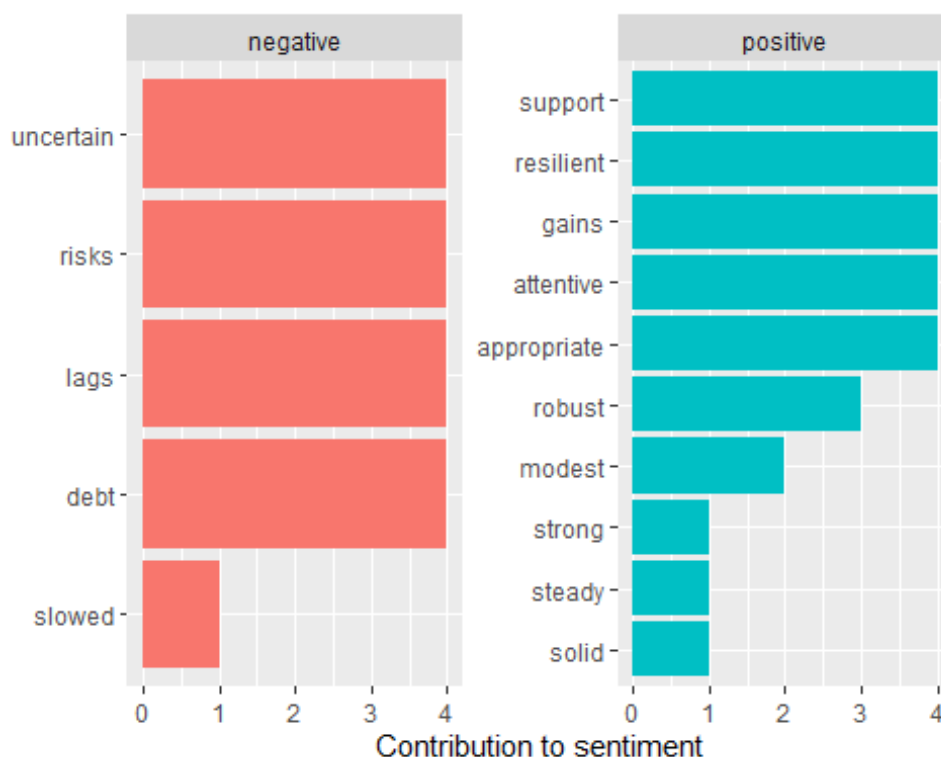
```
## 14 steady   positive    1
```

```
## 15 strong   positive    1
```



```
sentiment %>%
  group_by(sentiment) %>%
  top_n(20) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(x = "Contribution to sentiment",
       y = NULL)
```

## Selecting by n



```
vector.nrc <- get_nrc_sentiment(testi)
df.nrc <- data.frame(t(vector.nrc))
td_new <- data.frame( rowSums(df.nrc))

td_new %>% rename(count=rowSums.df.nrc.) -> td_new

td_new <- cbind( "sentiment" = rownames(td_new), td_new )
rownames(td_new) <- NULL
td_new2 <- td_new[1:10,]

quickplot(data = td_new2 , x = sentiment, fill = sentiment,
          weight= count, geom = 'bar', ylab = "count")+ theme(legend.position =
"none")
```

