

## Step2: Do sentiment analysis at sentence level

```
spooky<-read.csv('../data/spooky.csv',as.is=T)
spooky.sentence<-spooky%>%
  mutate(sentiment = get_sentiment(text))
```

```
## Warning in split_warn(text.var, "sentiment_by", ...): Each time
## `sentiment_by` is run it has to do sentence boundary disambiguation when
## a raw `character` vector is passed to `text.var`. This may be costly of
## time and memory. It is highly recommended that the user first runs the raw
## `character` vector through the `get_sentences` function.
```

```
head(spooky.sentence)
```

```
##      id
## 1 id26305
## 2 id17569
## 3 id11008
## 4 id27763
## 5 id12958
## 6 id22965
##
## 1
## 2
## 3
## 4
## 5
## 6 A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so r
##   author sentiment
## 1    EAP  0.1561738
## 2    HPL -0.2004459
## 3    EAP  0.3250000
## 4    MWS  0.6002450
## 5    HPL -0.9083644
## 6    MWS  0.5982152
```

```
count(spooky.sentence, sentiment)
```

```
## # A tibble: 8,704 x 2
##   sentiment      n
##   <dbl> <int>
## 1   -2.42     1
## 2   -2.15     1
## 3   -1.92     1
## 4   -1.66     1
## 5   -1.65     1
## 6   -1.59     1
## 7   -1.56     1
## 8   -1.53     1
## 9   -1.48     1
## 10  -1.48     1
## # ... with 8,694 more rows
```

```
count(spooky.sentence, author, sentiment)
```

```
## # A tibble: 11,370 x 3
```

```
##      author sentiment      n
##      <chr>      <dbl> <int>
##  1 EAP          -2.42     1
##  2 EAP          -2.15     1
##  3 EAP          -1.66     1
##  4 EAP          -1.56     1
##  5 EAP          -1.45     1
##  6 EAP          -1.42     1
##  7 EAP          -1.40     1
##  8 EAP          -1.35     1
##  9 EAP          -1.33     1
## 10 EAP          -1.32     1
## # ... with 11,360 more rows
```

```
spooky.sentense.data<-spooky.sentense %>%
  mutate(sentiment_type = if_else(sentiment >0, "Positive", if_else(sentiment <0, "Negative", "Neutral")),
  select(sentiment, sentiment_type,text,author)
head(spooky.sentense.data)
```

```
##      sentiment sentiment_type
##  1  0.1561738      Positive
##  2 -0.2004459      Negative
##  3  0.3250000      Positive
##  4  0.6002450      Positive
##  5 -0.9083644      Negative
##  6  0.5982152      Positive
```

```
##
##  1
##  2
##  3
##  4
##  5
##  6 A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has so r
##      author
##  1      EAP
##  2      HPL
##  3      EAP
##  4      MWS
##  5      HPL
##  6      MWS
```

```
order.spooky.sentense<-spooky.sentense.data[order(spooky.sentense.data$sentiment),]
head(order.spooky.sentense,20)
```

```
##      sentiment sentiment_type
## 10780 -2.415748      Negative
##  8479 -2.154300      Negative
## 16438 -1.915133      Negative
##  5221 -1.663635      Negative
##  9109 -1.653757      Negative
## 17400 -1.588571      Negative
## 11608 -1.563142      Negative
## 12549 -1.531883      Negative
## 14021 -1.483405      Negative
## 15807 -1.482096      Negative
```

## 4932	-1.467582	Negative	
## 17764	-1.451766	Negative	
## 7140	-1.449639	Negative	
## 4907	-1.446667	Negative	
## 3933	-1.431829	Negative	
## 6675	-1.421725	Negative	
## 2815	-1.412500	Negative	
## 6269	-1.401715	Negative	
## 11881	-1.398000	Negative	
## 17127	-1.380261	Negative	
##			
## 10780			
## 8479			
## 16438			
## 5221			
## 9109			
## 17400			
## 11608			Pest spirits, plagu
## 12549			
## 14021			
## 15807			
## 4932			
## 17764			
## 7140			
## 4907			
## 3933			
## 6675	Mimes, in the form of God on high, Mutter and mumble low, And hither and thither fly; Mere pup		
## 2815			
## 6269			
## 11881			
## 17127			
##	author		
## 10780	EAP		
## 8479	EAP		
## 16438	MWS		
## 5221	EAP		
## 9109	HPL		
## 17400	MWS		
## 11608	EAP		
## 12549	HPL		
## 14021	HPL		
## 15807	HPL		
## 4932	HPL		
## 17764	MWS		
## 7140	HPL		
## 4907	EAP		
## 3933	HPL		
## 6675	EAP		
## 2815	MWS		
## 6269	EAP		
## 11881	HPL		
## 17127	HPL		

```
tail(order.spooky.sentence,20)
```

```
##      sentiment sentiment_type
## 1574    1.340218      Positive
## 16695   1.353791      Positive
## 18754   1.385363      Positive
## 3749    1.404374      Positive
## 11395   1.449772      Positive
## 11332   1.471375      Positive
## 16648   1.545220      Positive
## 17664   1.563239      Positive
## 6275    1.573033      Positive
## 17393   1.574701      Positive
## 16415   1.583669      Positive
## 6633    1.606865      Positive
## 12508   1.612000      Positive
## 18641   1.627278      Positive
## 11173   1.640488      Positive
## 12144   1.665600      Positive
## 17384   1.675193      Positive
## 14045   1.787973      Positive
## 5728    1.889643      Positive
## 18108   2.142120      Positive
```

```
##
```

```
## 1574
## 16695
## 18754
## 3749
## 11395
## 11332
## 16648
## 17664
## 6275
## 17393
## 16415
## 6633
## 12508
## 18641
## 11173
## 12144
## 17384
## 14045
## 5728
```

```
## 18108 Oh no I will become wise I will study my own heart and there discovering as I may the spring o
```

```
##      author
```

```
## 1574    MWS
## 16695    EAP
## 18754    MWS
## 3749    MWS
## 11395    MWS
## 11332    EAP
## 16648    EAP
## 17664    MWS
## 6275    HPL
```

```
## 17393    HPL
## 16415    MWS
## 6633     EAP
## 12508    EAP
## 18641    HPL
## 11173    MWS
## 12144    MWS
## 17384    EAP
## 14045    MWS
## 5728     EAP
## 18108    MWS
```

```
positive.rate<-sum(spooky.sentence.data$sentiment_type=='Positive')/nrow(spooky.sentence.data)
positive.rate
```

```
## [1] 0.4305634
```

```
count.whole.table<-count(spooky.sentence.data%>%group_by(author))
as.integer(count.whole.table[count.whole.table$author=='EAP'],$n)
```

```
## [1] 7900
```

```
as.integer(count.whole.table[count.whole.table$author=='HPL'],$n)
```

```
## [1] 5635
```

```
as.integer(count.whole.table[count.whole.table$author=='MWS'],$n)
```

```
## [1] 6044
```

```
count.table<-count(spooky.sentence.data%>%group_by(sentiment_type, author))
frequency.EAP<-count.table[count.table$author=='EAP'],$n/as.integer(count.whole.table[count.whole.table$author=='EAP'],$n)
frequency.HPL<-count.table[count.table$author=='HPL'],$n/as.integer(count.whole.table[count.whole.table$author=='HPL'],$n)
frequency.MWS<-count.table[count.table$author=='MWS'],$n/as.integer(count.whole.table[count.whole.table$author=='MWS'],$n)
n<-c(frequency.MWS,frequency.HPL,frequency.EAP)
author<-c('MWS','MWS','MWS','HPL','HPL','HPL','EAP','EAP','EAP')
sentiment_type<-c('Negative','Negative','Negative','Neutral','Neutral','Neutral','Positive','Positive','Positive')
frequency.table<-as.data.frame(cbind(sentiment_type,author,n))
ggplot(frequency.table)+geom_col(aes(sentiment_type, n, fill = sentiment_type)) +
  facet_wrap(~ author) +
  coord_flip() +
  theme(legend.position = "none")
```

