step2: Visualizing author topics

```
spooky_wrd_docs <- tidy(spooky_wrd_lda_6, matrix = "gamma")
head(spooky_wrd_docs)
```

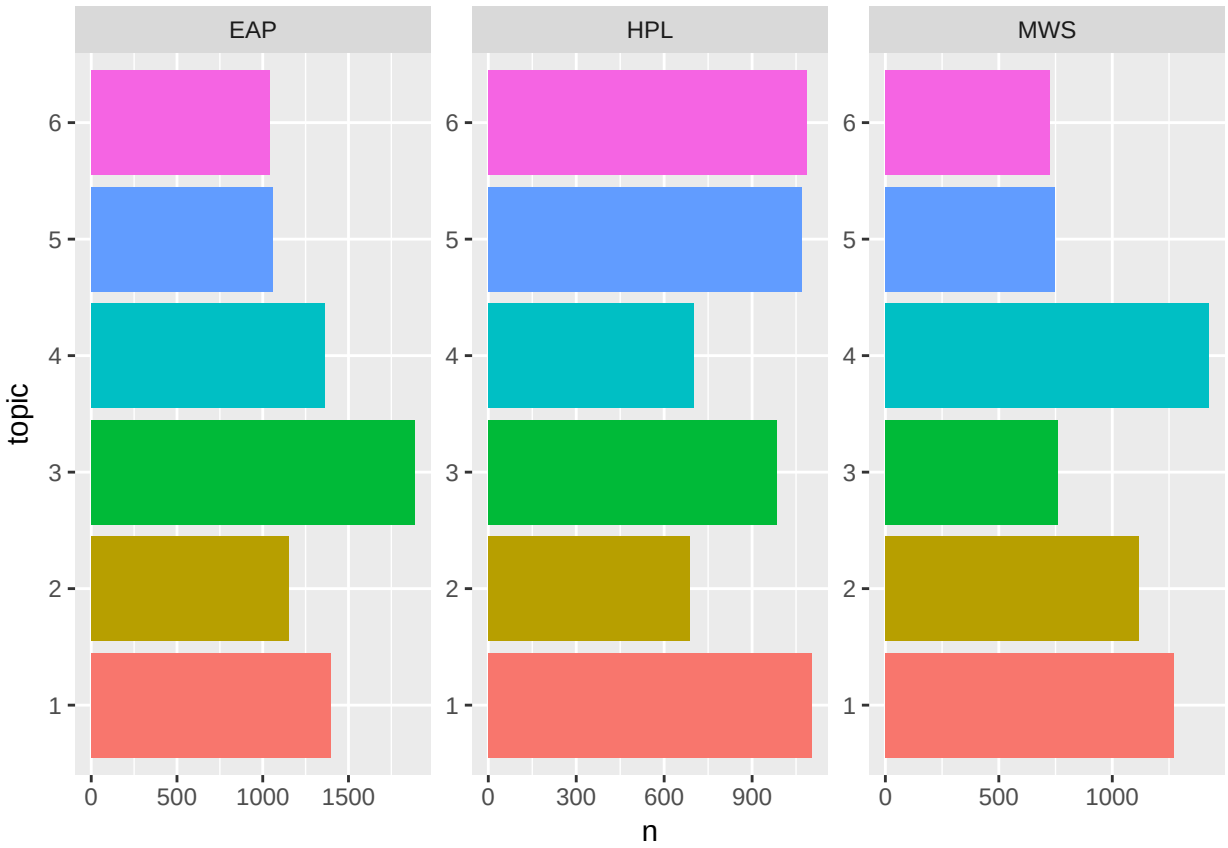
```
## # A tibble: 6 x 3
##   document topic gamma
##   <chr>    <int> <dbl>
## 1 id00001      1 0.168
## 2 id00002      1 0.166
## 3 id00003      1 0.167
## 4 id00004      1 0.170
## 5 id00005      1 0.168
## 6 id00006      1 0.165
```

```
author_topics <- left_join(spooky_wrd_docs, spooky, by = c("document" = "id"))
author_topics <- select(author_topics, -text)
author_topics$topic <- as.factor(author_topics$topic)
# Chooses the top topic per sentence
author_topics <- ungroup(top_n(group_by(author_topics, document), 1, gamma))

# Counts the number of sentences represented by each topic per author
author_topics <- ungroup(count(group_by(author_topics, author, topic)))
author_topics
```

```
## # A tibble: 18 x 3
##    author topic     n
##    <chr>  <fct> <int>
##  1 EAP    1      1395
##  2 EAP    2      1154
##  3 EAP    3      1887
##  4 EAP    4      1364
##  5 EAP    5      1059
##  6 EAP    6      1041
##  7 HPL    1      1104
##  8 HPL    2       687
##  9 HPL    3       983
## 10 HPL    4       702
## 11 HPL    5      1071
## 12 HPL    6      1088
## 13 MWS    1      1270
## 14 MWS    2      1116
## 15 MWS    3       762
## 16 MWS    4      1426
## 17 MWS    5       746
## 18 MWS    6       724
```

```
ggplot(author_topics) +
  geom_col(aes(topic, n, fill = factor(topic)), show.legend = FALSE) +
  facet_wrap(~ author, scales = "free", ncol = 4) +
  coord_flip()
```



From plot, we learn different author focus on diffrernt topics. And combine 5 top words for each topics, we can get theme for each author.

Section 5:Summary