

# Untitled

2024-12-16

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(lexicon)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(topicmodels)
library(tidytext)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(dplyr)
library(ggplot2)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)
movies <- read.csv("movie_plots.csv")
movies <- movies %>%
mutate(Genre = case_when(
  str_detect(Plot, "(?i)science|space|experiment|future|alien|
robot|technology|planet|human|new world|earth") ~ "Sci-Fi",
  str_detect(Plot, "(?i)love|romance|relationship|affair|
1
wedding|couple|meets|girl") ~ "Romance",
  str_detect(Plot, "(?i)war|battle|army|soldier|conflict|
military|enemy|tank|battlefield|dead") ~ "War",
  str_detect(Plot, "(?i)ghost|haunt|horror|fear|terror|
scary|supernatural|creepy") ~ "Horror",
  str_detect(Plot, "(?i)crime|detective|murder|investigate|
```

```
thriller|mafia|heist|mystery") ~ "Crime",
str_detect(Plot, "(?i)action|fight|fighting|adventure|hero|explosion|
battle|rescue") ~ "Action",
str_detect(Plot, "(?i)comedy|funny|humor|laugh|joke|
satire|parody") ~ "Comedy",
str_detect(Plot, "(?i)history|historical|biography|true story|
period drama|century|ancient") ~ "History",
str_detect(Plot, "(?i)fantasy|magic|myth|legend|superhero|
kingdom|evil") ~ "Fantasy",
str_detect(Plot, "(?i)western|cowboy|wild west|sheriff|ranch|
town|outlaw") ~ "Western",
str_detect(Plot, "(?i)documentary|docu|true events|reality|
biopic") ~ "Documentary",
str_detect(Plot, "(?i)sport|game|team|match|championship|
wrestling") ~ "Sport",
str_detect(Plot, "(?i)home|people|brother|daughter|brothers|
friend|wife|son|father|mother") ~ "Family",
TRUE ~ "Other"
))
genre_frequency <- movies %>%
count(Genre, name = "Frequency")
plot_words <- movies %>%
unnest_tokens(word, Plot) %>%
anti_join(get_stopwords()) %>%
count(Genre, word, sort = TRUE)
```

```
## Joining with `by = join_by(word)`
```

```
nested_data <- plot_words %>%
group_by(Genre) %>%
summarize(
Words = paste(unique(word), collapse = ", ")
) %>%
left_join(genre_frequency, by = "Genre")
view(nested_data)

dtm <- plot_words %>%
cast_dtm(Genre, word, n)
dtm
```

```
## <<DocumentTermMatrix (documents: 14, terms: 14842)>>
## Non-/sparse entries: 30812/176976
## Sparsity : 85%
## Maximal term length: 17
## Weighting : term frequency (tf)
```

```
k <- 20
lda_model <- LDA(dtm, k = k, control = list(seed = 999))

topics <- tidy(lda_model, matrix = "beta")

top_terms <- topics %>%
group_by(topic) %>%
slice_max(beta, n = 10) %>%
ungroup() %>%
```

```

arrange(topic, -beta)
print(top_terms)

```

```

## # A tibble: 205 x 3
##   topic term      beta
##   <int> <chr>    <dbl>
## 1     1 town    0.00552
## 2     1 one     0.00545
## 3     1 action 0.00474
## 4     1 life   0.00427
## 5     1 new    0.00387
## 6     1 fight  0.00378
## 7     1 man    0.00347
## 8     1 two    0.00339
## 9     1 world  0.00300
## 10    1 mcguire 0.00296
## # i 195 more rows

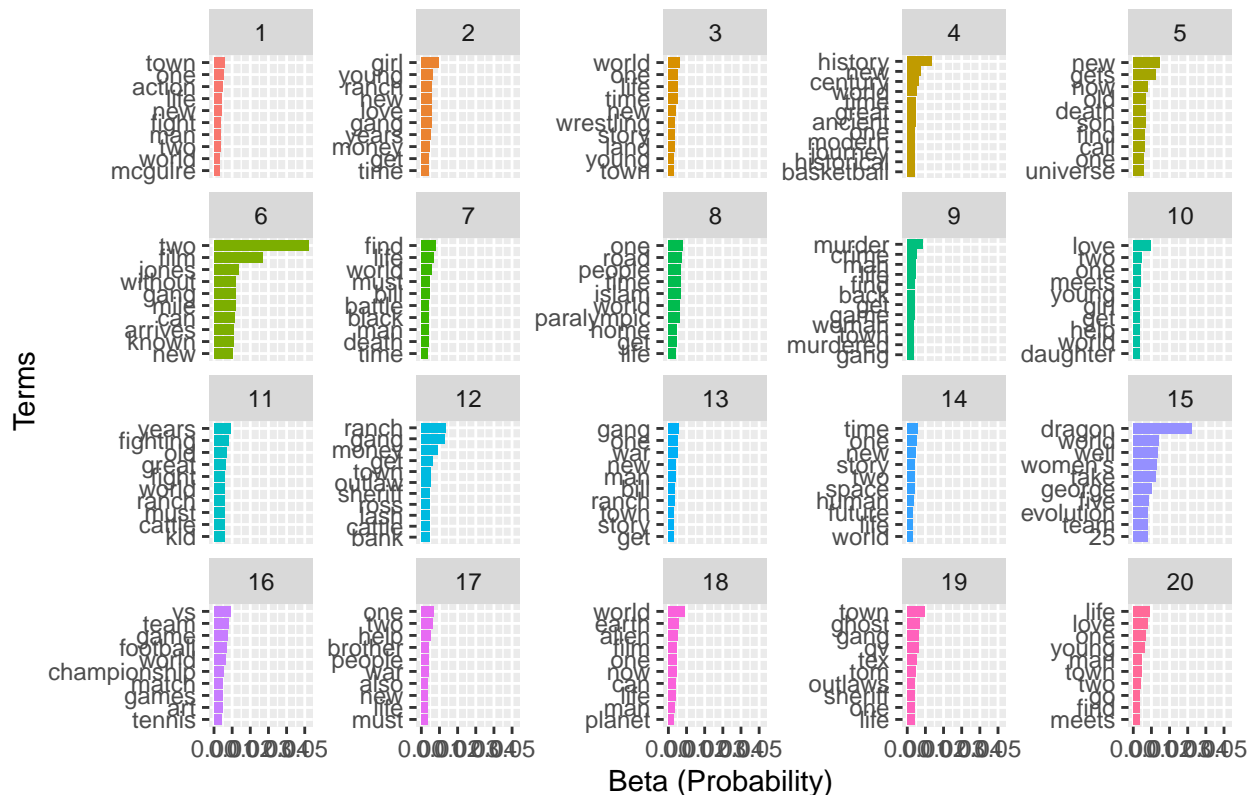
```

```

ggplot(top_terms, aes(x = reorder_within(term, beta, topic), y = beta,
fill = factor(topic))) +
geom_col(show.legend = FALSE) +
facet_wrap(~ topic, scales = "free_y") +
coord_flip() +
scale_x_reordered() +
labs(title = "Top 10 Terms for Each Topic", x = "Terms",
y = "Beta (Probability)")

```

Top 10 Terms for Each Topic



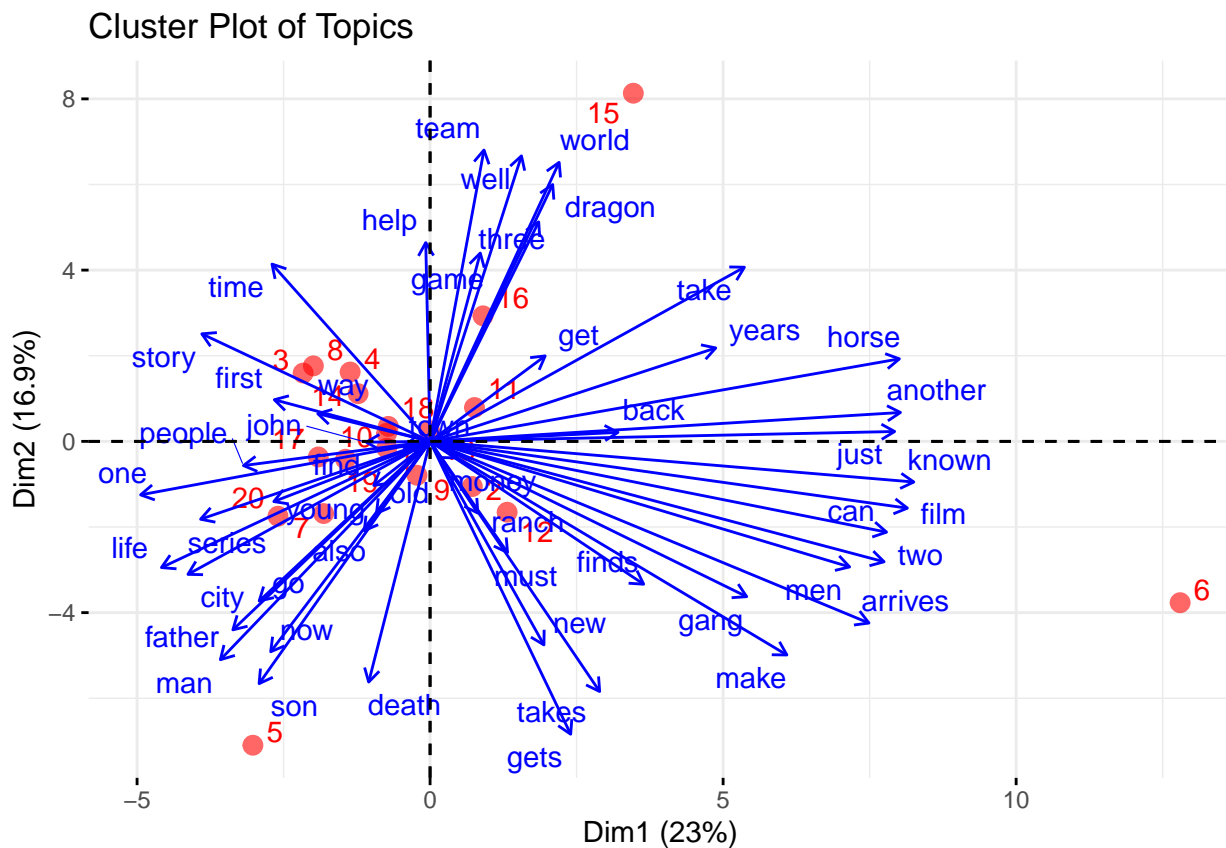
```

filtered_topics <- topics %>%
  group_by(term) %>%
  summarize(total_beta = sum(beta)) %>%
  top_n(50, total_beta)
topic_term_matrix <- topics %>%
  filter(term %in% filtered_topics$term) %>%
  spread(term, beta) %>%
  column_to_rownames(var = "topic")

pca_result <- prcomp(topic_term_matrix, scale. = TRUE)

fviz_pca_biplot(pca_result,
  repel = TRUE,
  col.var = "blue",
  col.ind = "red",
  pointsize = 3,
  alpha.ind = 0.6,
  title = "Cluster Plot of Topics")

```



```

doc_topic_matrix <- tidy(lda_model, matrix = "gamma")

genre_topic_distribution <- doc_topic_matrix %>%
  rename(Genre = document) %>%
  group_by(Genre, topic) %>%
  summarize(avg_gamma = mean(gamma, na.rm = TRUE), .groups = "drop")

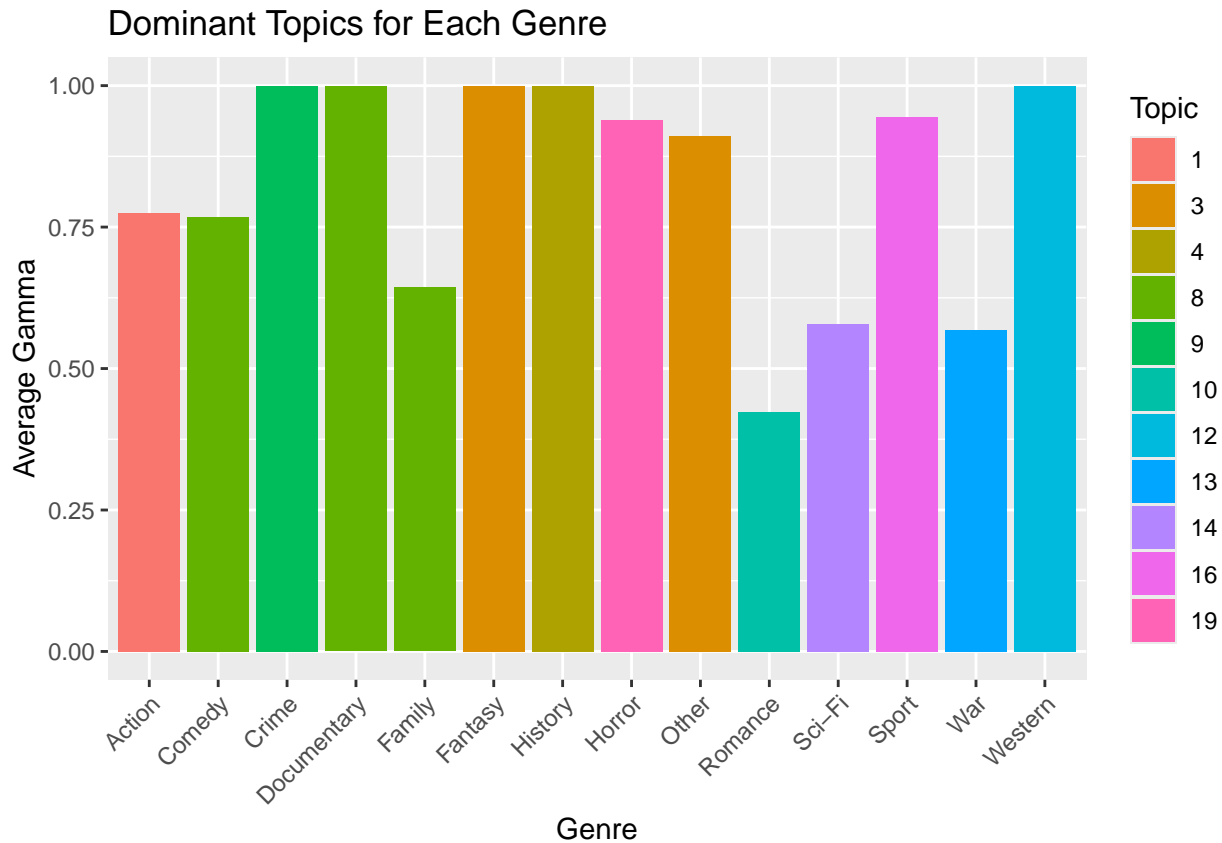
dominant_topics <- genre_topic_distribution %>%

```

```
group_by(Genre) %>%
slice_max(avg_gamma, n = 1) %>%
ungroup()
print(dominant_topics)
```

```
## # A tibble: 14 x 3
##   Genre      topic avg_gamma
##   <chr>      <int>   <dbl>
## 1 Action         1     0.775
## 2 Comedy         8     0.769
## 3 Crime          9     1.00
## 4 Documentary    8     0.998
## 5 Family         8     0.643
## 6 Fantasy        3     1.00
## 7 History        4     1.00
## 8 Horror        19     0.939
## 9 Other          3     0.910
## 10 Romance       10     0.423
## 11 Sci-Fi        14     0.579
## 12 Sport         16     0.945
## 13 War           13     0.567
## 14 Western       12     1.00
```

```
ggplot(dominant_topics, aes(x = Genre, y = avg_gamma, fill = factor(topic))) +
geom_col() +
labs(title = "Dominant Topics for Each Genre",
x = "Genre",
y = "Average Gamma",
fill = "Topic") +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
palette <- brewer.pal(8, "Dark2")

k <- 14
for (i in 1:k) {
  topic_words <- topics %>%
  filter(topic == i) %>%
  arrange(desc(beta))

  wordcloud(words = topic_words$term,
    freq = topic_words$beta,
    max.words = 100,
    random.order = FALSE,
    colors = palette)
}
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## freedom could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## discover could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## california could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## behind could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## people could not be fit on page. It will not be plotted.
```

```

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : away
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : horse
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : owner
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## scroll could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : jack
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : white
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## brother could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## championship could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : money
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : finds
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## saloon could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : gene
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : way
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : never
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## gunfighter could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : lives
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : round
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : brick
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : time
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : truck
## could not be fit on page. It will not be plotted.

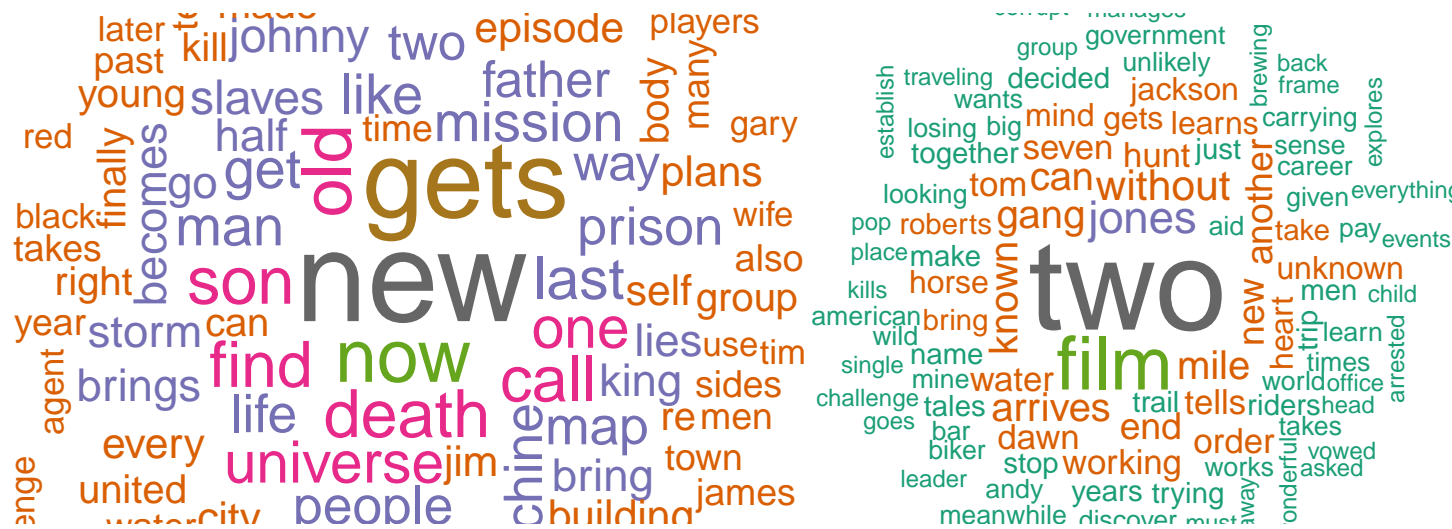
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : stock
## could not be fit on page. It will not be plotted.

```









```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## discovered could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## indian could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## america could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## streets could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## justice could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## forced could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : great  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : west  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## attempt could not be fit on page. It will not be plotted.
```



```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## training could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## members could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : gang  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : thing  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## around could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## behind could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : story  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : way  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : boy  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : watch  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## documentary could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## swimming could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## trials could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : waves  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## saloon could not be fit on page. It will not be plotted.
```

```

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : lion
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## homesteaders could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : terry
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## muhammed could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : joan
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## interviews could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : dam
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## explores could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## season could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : cup
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## sometimes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : young
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## scenes could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## looking could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## government could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## person could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : santa
## could not be fit on page. It will not be plotted.

```



```

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : ned
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## become could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : night
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : men
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## protect could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## estranged could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## continues could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : save
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : death
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## gunfight could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## making could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## gunfighter could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : quest
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : come
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## begins could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## masque could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : films
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : home
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : see
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## reputation could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : king
## could not be fit on page. It will not be plotted.

```

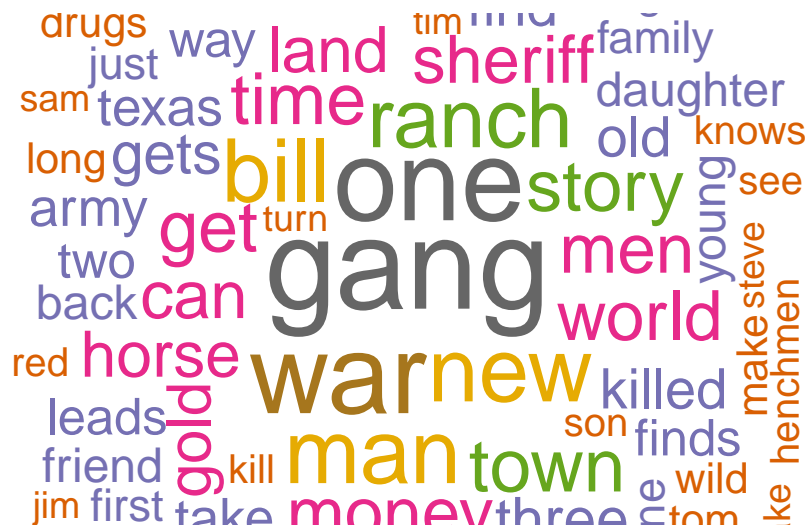
```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## homesteaders could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : judge  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## cowboy could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## framed could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## kobblestone could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## legend could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## saunders could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## together could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : night  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : must  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : mary  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## behind could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : based  
## could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## johnny could not be fit on page. It will not be plotted.  
  
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :  
## government could not be fit on page. It will not be plotted.
```



```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : great
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## attack could not be fit on page. It will not be plotted.
```



```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : star
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## mankind could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : named
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : james
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## government could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : every
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## attack could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : moon
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : small
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : may
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## mysterious could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : ghost
## could not be fit on page. It will not be plotted.
```

```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, :
## escape could not be fit on page. It will not be plotted.
```

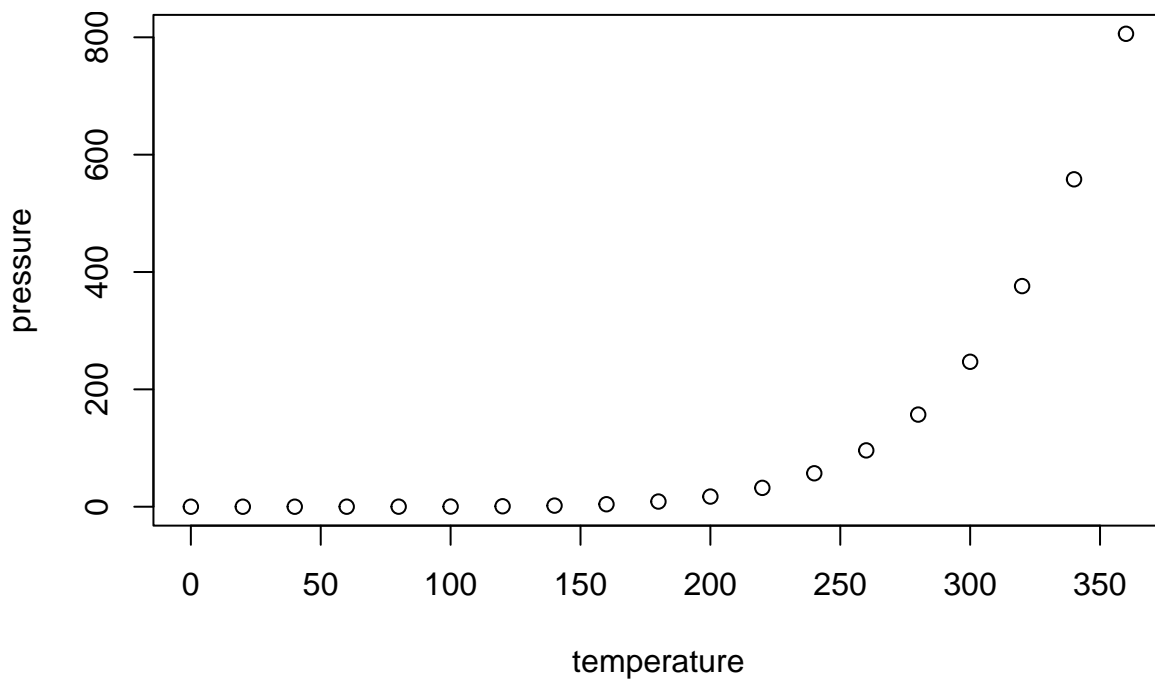


```
## Warning in wordcloud(words = topic_words$term, freq = topic_words$beta, : bring
## could not be fit on page. It will not be plotted.
```



## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.