

# Tidy Tuesday

Week 35

AUTHOR  
Cristian T

PUBLISHED  
September 3, 2024

This week we are exploring the Power Rangers Franchise!

In 1993, five ordinary teenagers exploded on the pop-culture scene with the launch of Mighty Morphin Power Rangers. Together they broke down barriers. They defeated evil by demonstrating teamwork, inclusivity, and diversity to people of all ages. Today, this grand tradition continues as new Ranger teams and new generations of fans discover these essential values again.

The series, created by Haim Saban, has one of the most popular taglines in history, "It's Morphin Time!" The TV series "Mighty Morphin Power Rangers" (MMPR) launched on August 28, 1993. Power Rangers quickly became the #1 kids action brand and a global phenomenon. With its current 25th season, "Power Rangers Super Ninja Steel," the show is now the second-longest-running, non-soap-opera, scripted program on American TV (after "The Simpsons"). There are also over 830 episodes in its library. Currently, Power Rangers is seen in more than 150 markets around the world. It's also translated into numerous languages and is a favorite on many indispensable children's programming blocks around the world. Go Go Power Rangers on 8.28!

power\_rangers\_episodes.csv

variable	class	description
season_title	character	title of the overall season
episode_num	double	number of this episode within this season
episode_title	character	title of this episode
air_date	double	date on which this episode first aired in the U.S.
IMDB_rating	double	average rating among IMDB users
total_votes	double	total votes on IMDB
desc	character	free-text description of this episode

power\_rangers\_seasons.csv

variable	class	description
season_title	character	title of this season
season_num	double	season number
number_of_episodes	double	number of episodes in this season
air_date_first_ep	double	date on which the first episode in this season first aired in the U.S.

variable	class	description
air_date_last_ep	character	date on which the last episode in this season first aired in the U.S.
producer	character	the company that produced this season
IMDB_rating	double	average rating of this seasons among IMDB users

```
# Load the tidyuesday package
suppressMessages(library(tidyuesdayR)) # For accessing TidyTuesday datasets
suppressMessages(library(skimr)) # For summary and descriptive statistics
suppressMessages(library(tidyverse)) # For data manipulation and visualization
suppressMessages(library(dplyr)) # For data manipulation and transformation
suppressMessages(library(ggplot2)) # For data visualization
suppressMessages(library(RColorBrewer)) # For color palettes in visualizations
suppressMessages(library(mdsr)) # For spatial analysis
suppressMessages(library(tidytext)) # For text mining and analysis
suppressMessages(library(textstem)) # For lemmatization

# Load the current week's dataset
tuesdata <- tidyuesdayR::tt_load('2024-08-27')
```

Downloading file 1 of 2: `power\_rangers\_episodes.csv`  
 Downloading file 2 of 2: `power\_rangers\_seasons.csv`

```
power_rangers_episodes <- tuesdata$power_rangers_episodes
pwe <- power_rangers_episodes
power_rangers_seasons <- tuesdata$power_rangers_seasons
pws <- power_rangers_seasons

# Explore the structure of the dataset
str(pwe)
```

```
spc_tbl_ [921 × 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ season_title : chr [1:921] "Mighty Morphin (Season 1)" "Mighty Morphin (Season 1)" "Mighty
Morphin (Season 1)" "Mighty Morphin (Season 1)" ...
 $ episode_num  : num [1:921] 0 1 2 3 4 5 6 7 8 9 ...
 $ episode_title: chr [1:921] "The Lost Episode" "Day of the Dumpster" "High Five" "Teamwork"
...
 $ air_date      : Date[1:921], format: "1999-05-22" "1993-08-28" ...
 $ IMDB_rating   : num [1:921] 6.7 7.4 6.9 7.3 6.9 6.6 7.3 6.4 7 6 ...
 $ total_votes   : num [1:921] 113 687 564 546 535 516 531 510 504 494 ...
 $ desc          : chr [1:921] "Original version of the premiere episode." "Following the
accidental release of long-imprisoned evil space-sorceress Rita Repulsa, a benevolent sage
known "| __truncated__ "Rita plans to trap the Rangers in a time trap like the one she
trapped Zordon in and to do this she sends a fre"| __truncated__ "Trini and Kimberly set up a
petition to clean up pollution and there are plenty of people willing to sign but n"|
__truncated__ ...
- attr(*, "spec")=
.. cols(
```

```

.. season_title = col_character(),
.. episode_num = col_double(),
.. episode_title = col_character(),
.. air_date = col_date(format = ""),
.. IMDB_rating = col_double(),
.. total_votes = col_double(),
.. desc = col_character()
.. )
- attr(*, "problems")=<externalptr>

```

```
skim(pws)
```

## Variable type: numeric

var	n	na	mean	sd	p0	p25	p50	p75	p100
season_num	28	0	14.50	8.23	1.0	7.75	14.5	21.25	28.0
number_of_episodes	28	0	34.00	11.13	20.0	22.00	32.5	40.75	60.0
IMDB_rating	28	0	6.36	0.56	5.1	6.00	6.5	6.73	7.3

```
str(pws)
```

```

spc_tbl_ [28 × 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ season_title      : chr [1:28] "Mighty Morphin (Season 1)" "Mighty Morphin (Season 2)"
"Mighty Morphin (Season 3)" "Zeo" ...
 $ season_num        : num [1:28] 1 2 3 4 5 6 7 8 9 10 ...
 $ number_of_episodes: num [1:28] 60 52 33 50 45 43 45 40 40 40 ...
 $ air_date_first_ep : Date[1:28], format: "1993-08-28" "1994-07-21" ...
 $ air_date_last_ep  : chr [1:28] "1994-05-23" "1995-05-20" "1995-11-27" "1996-11-27" ...
 $ producer          : chr [1:28] "Saban" "Saban" "Saban" "Saban" ...
 $ IMDB_rating       : num [1:28] 6.5 6.5 6.5 6.6 5.7 7.3 6.9 6.5 7.1 6.3 ...
- attr(*, "spec")=
.. cols(
..   season_title = col_character(),
..   season_num = col_double(),
..   number_of_episodes = col_double(),
..   air_date_first_ep = col_date(format = ""),
..   air_date_last_ep = col_character(),
..   producer = col_character(),
..   IMDB_rating = col_double()
.. )
- attr(*, "problems")=<externalptr>

```

```
skim(pws)
```

## Variable type: numeric

var	n	na	mean	sd	p0	p25	p50	p75	p100
season_num	28	0	14.50	8.23	1.0	7.75	14.5	21.25	28.0
number_of_episodes	28	0	34.00	11.13	20.0	22.00	32.5	40.75	60.0

var	n	na	mean	sd	p0	p25	p50	p75	p100
IMDB_rating	28	0	6.36	0.56	5.1	6.00	6.5	6.73	7.3

```
# View the first few rows of the dataset
head(pwe)
```

season_title	episode_num	episode_title
<chr>	<dbl>	<chr>
Mighty Morphin (Season 1)	0	The Lost Episode
Mighty Morphin (Season 1)	1	Day of the Dumpster
Mighty Morphin (Season 1)	2	High Five
Mighty Morphin (Season 1)	3	Teamwork
Mighty Morphin (Season 1)	4	A Pressing Engagement
Mighty Morphin (Season 1)	5	Different Drum

6 rows | 1-3 of 7 columns

```
head(pws)
```

season_title	season_num	number_of_episodes
<chr>	<dbl>	<dbl>
Mighty Morphin (Season 1)	1	60
Mighty Morphin (Season 2)	2	52
Mighty Morphin (Season 3)	3	33
Zeo	4	50
Turbo	5	45
In Space	6	43

6 rows | 1-3 of 7 columns

```
# Join the datasets on the common key (e.g., season_id)
combined_data <- pwe %>%
  inner_join(pws, by = "season_title")

combined_data %>% select("season_title", "season_num")
```

[illegible]

season_title	season_num
<chr>	<dbl>
Mighty Morphin (Season 1)	1
Mighty Morphin (Season 1)	1
1-10 of 921 rows	
Previous 1 2 3 4 5 6 ... 93 Next	

```
#tidytuesdayR::use_tidytemplate()
```

Clean the data

```
# Doesn't appear the data needs cleaned
```

Missing Values

DF doesn't have any missing data and column seem appropriate for what we are looking at

Text Analysis

Plot 1 Top Words

```
# Convert text to lowercase
pwe$desc <- tolower(pwe$desc)

# Count the words excluding stopwords
pwe_text_analysis <- pwe %>%
  unnest_tokens(word, desc) %>%
  anti_join(stop_words, by = "word") %>%
  count(word, sort = TRUE)

# Print Tidy Text
pwe_text_analysis
```

word	n
<chr>	<int>
rangers	720
power	182
ranger	161
monster	158
battle	86
time	82
tommy	79
earth	78
team	75
evil	74
1-10 of 4,498 rows	
Previous 1 2 3 4 5 6 ... 450 Next	

```
# Get the top 20 words
top_20_words <- pwe_text_analysis %>%
  top_n(20, n)

suppressMessages(library(treemapify))

# Graph the top 20 words
ggplot(top_20_words, aes(area = n, fill = n, label = word)) +
  geom_treemap() +
  geom_treemap_text(color = "white", place = "centre", grow = TRUE) +
  scale_fill_gradientn(colors = pals::alphabet(20)) +
  labs(title = "Top 20 Words Found in Power Rangers Episode Descriptions",
       x = NULL,
       y = "Count") +
  theme_minimal(base_size = 14) + # Set base font size for readability
  theme(axis.text.y = element_text(size = 8),
        axis.text.x = element_text(size = 10),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.position = "none") +
  scale_y_continuous(labels = NULL)
```

## Top 20 Words Found in Power Rangers Episode Descriptions



A “bigram” is a pair of consecutive words that appear together in a text. It is a fundamental concept in n-gram analysis, where “n” represents the number of words in the sequence. By generating and analyzing bigrams from the top words data frame, we can uncover the relationships between key terms within the text. Visualizing these bigrams as a network graph allows us to see how words are connected, revealing prominent themes and recurring concepts within the episodes.

```
suppressMessages(library(textdata))
suppressMessages(library(igraph))
suppressMessages(library(ggraph))
suppressMessages(library(tidygraph))

# Generates bigrams from the episode descriptions and counts their occurrences
pwe_bigrams <- pwe %>%
  unnest_tokens(bigram, desc, token = "ngrams", n = 2) %>%
  count(bigram, sort = TRUE)

# Separates each bigram into individual words for further processing
bigrams_separated <- pwe_bigrams %>%
  separate(bigram, c("word1", "word2"), sep = " ")

# Removes stop words
bigrams_filtered <- bigrams_separated %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)

# Prepare the top_words data
top_words <- bigrams_filtered %>%
  arrange(desc(n)) %>%
  slice_head(n = 100)

# Create nodes data frame (unique words from bigrams)
nodes <- top_words %>%
  select(word1, word2) %>%
  pivot_longer(cols = everything(), names_to = "type", values_to = "name") %>%
  distinct(name) %>%
  mutate(id = row_number())

# Defines the edges (relationships) between words on bigram frequencies.
edges <- top_words %>%
  inner_join(nodes, by = c("word1" = "name")) %>%
  rename(from = id) %>%
  inner_join(nodes, by = c("word2" = "name")) %>%
  rename(to = id) %>%
  select(from, to, n)

# We created a network graph of the word relationships. The nodes data frame is made

# Create tbl_graph
graph <- tbl_graph(nodes = nodes, edges = edges, directed = TRUE)

# The tbl_graph() function creates a graph using the nodes and edges data frames.

# The layout for the graph is set to "fr" (Fruchterman-Reingold) arranges the nodes i
```





## Plot 3 NRC

The NRC Emotion Lexicon is a powerful tool designed to analyze and categorize the emotional content of text. It associates words with eight basic emotions—anger, fear, anticipation, trust, surprise, sadness, joy, and disgust—and two overarching sentiments: positive and negative. This lexicon helps us understand the emotional underpinnings of various texts by mapping the words to specific emotional and sentiment categories.

When applied to TV show descriptions, such as those for “Power Rangers,” the NRC Emotion Lexicon provides insights into the emotional landscape of the content. By examining the frequency and distribution of different emotions, we can uncover how various feelings are portrayed and how they contribute to the overall narrative.

```
# Download NRC lexicon
nrc_lexicon <- get_sentiments("nrc")

# Unnest tokens to get individual words
pwe_words <- pwe %>%
  unnest_tokens(word, desc)

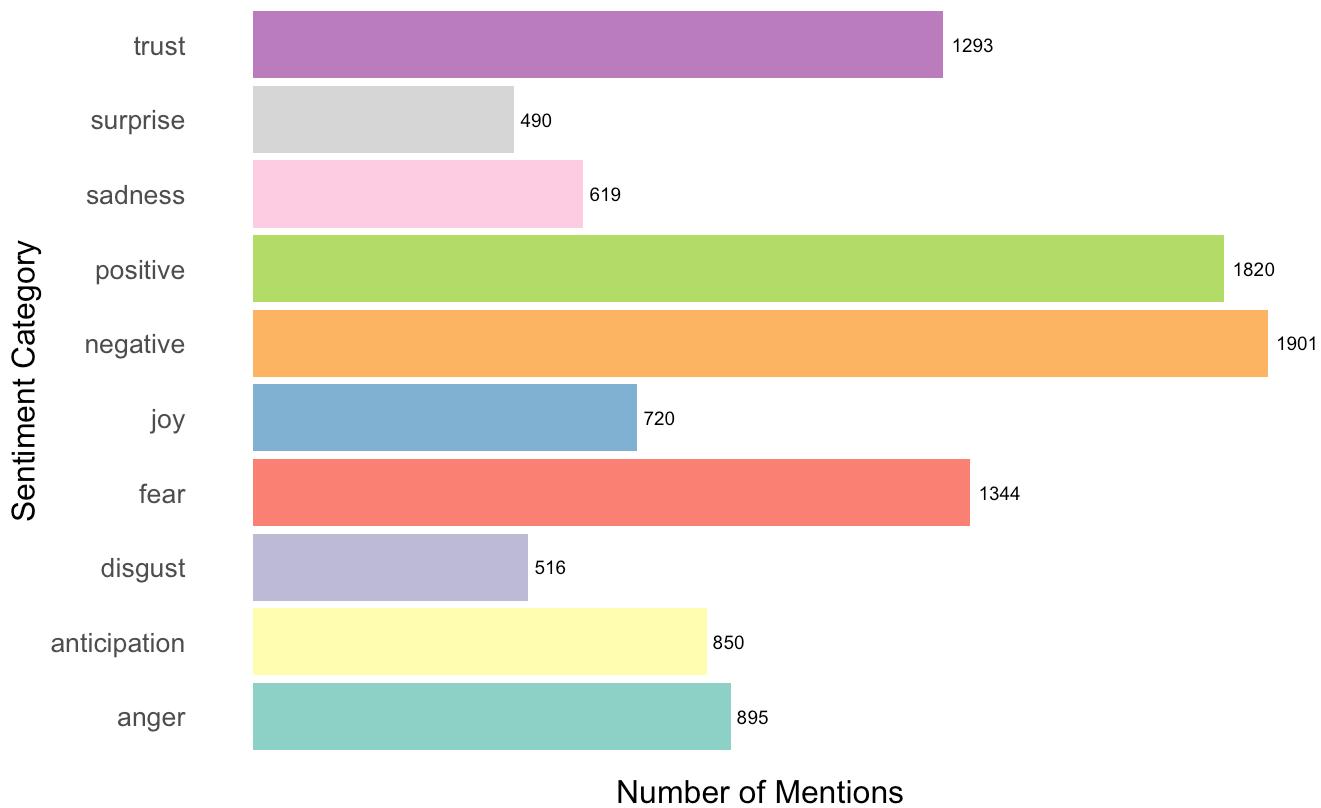
# Join with NRC lexicon to get sentiment scores
pwe_sentiments <- pwe_words %>%
  inner_join(nrc_lexicon, by = "word")

# Count the sentiment scores
sentiment_counts <- pwe_sentiments %>%
  count(sentiment)

# Plot the sentiment counts
ggplot(sentiment_counts, aes(x = sentiment, y = n, fill = sentiment)) +
  geom_bar(stat = "identity") + # Create the bar plot
  geom_text(aes(label = n), hjust = -0.2, size = 2.5) + # Add totals at the end of t
  scale_fill_manual(values = brewer.pal(10, name = "Set3")) + # Use color-blind-frie
  coord_flip() + # Flip coordinates
  labs(
    title = "Sentiment Analysis of Power Rangers Episodes",
    subtitle = "Using NRC Emotion Lexicon",
    x = "Sentiment Category",
    y = "Number of Mentions",
  ) +
  theme_minimal(base_size = 14) + # Set base font size for readability
  theme(axis.text.y = element_text(size = 10),
        axis.text.x = element_text(size = 10),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.position = "none") +
  scale_y_continuous(labels = NULL)
```

# Sentiment Analysis of Power Rangers Episodes

## Using NRC Emotion Lexicon



The "negative" sentiment category has the highest count (1,901), closely followed by "positive" (1,820). Despite a higher count of "negative" mentions, the "positive" sentiment is nearly as frequent (1,820 vs. 1,901), indicating that while the show deals with conflicts, it balances this with uplifting or positive moments. This balance could reflect the moral lessons or inspirational themes typical in children's programming, where good triumphs over evil. However, specific negative emotions such as "fear" (1,344) and "anger" (895) are also prevalent, suggesting that the Mighty Morphin Power Rangers frequently deals with challenging situations, conflicts, or threats, which might contribute to the suspense and tension while also experiencing elements of danger or threat, likely due to the show's premise involving battles against villains and monsters that is typical of an action-adventure series. We also see that "Disgust" (516) and "sadness" (619) also appear, indicating that the series occasionally explores themes of loss, disappointment, or moral disdain.

On the other hand, the "trust" sentiment (1,293) has a significant presence, which might reflect themes of teamwork, friendship, and reliability among the characters. Trust is likely an important recurring element in the narrative, given the show's focus on a team of heroes. "Anticipation" (850) also ranks highly, suggesting that the storyline frequently builds suspense or excitement about future events. This is consistent with the show's episodic nature, where each episode often ends with a cliffhanger or foreshadows future conflicts. "Joy" (720) and "surprise" (490) are less frequent than other emotions but still significant, suggesting moments of happiness, celebration, and unexpected events.

### Plot 4 LDA with lemmatize

Latent Dirichlet Allocation (LDA) is a statistical model for topic modeling. We attempt to identify topics that occur in a collection of documents. We are treating each document as a season to see if anything stands out. LDA uses a beta measurement that shows the weight (probability) of each term within a particular topic.

The beta value represents how strongly a term is associated with the topic. A higher value indicates that the term is more important or prevalent within that topic.

```
suppressMessages(library(topicmodels)) # For topic modeling analysis
# Combine episode title by season
season_title <- combined_data %>%
  group_by(season_num) %>%
  summarize(episode_title = paste(episode_title, collapse = " "))

# Unnest tokens to get individual words and remove stop words
season_clean <- season_title %>%
  unnest_tokens(word, episode_title) %>%
  anti_join(stop_words, by = "word")

# Lemmatize the words to their base form
season_clean$word <- lemmatize_words(season_clean$word)

# Create a document-term matrix grouped by season
# Convert the count into a Document-Term Matrix (DTM), where rows represent documents
season_dtm <- season_clean %>%
  count(season_num, word) %>%
  cast_dtm(season_num, word, n)

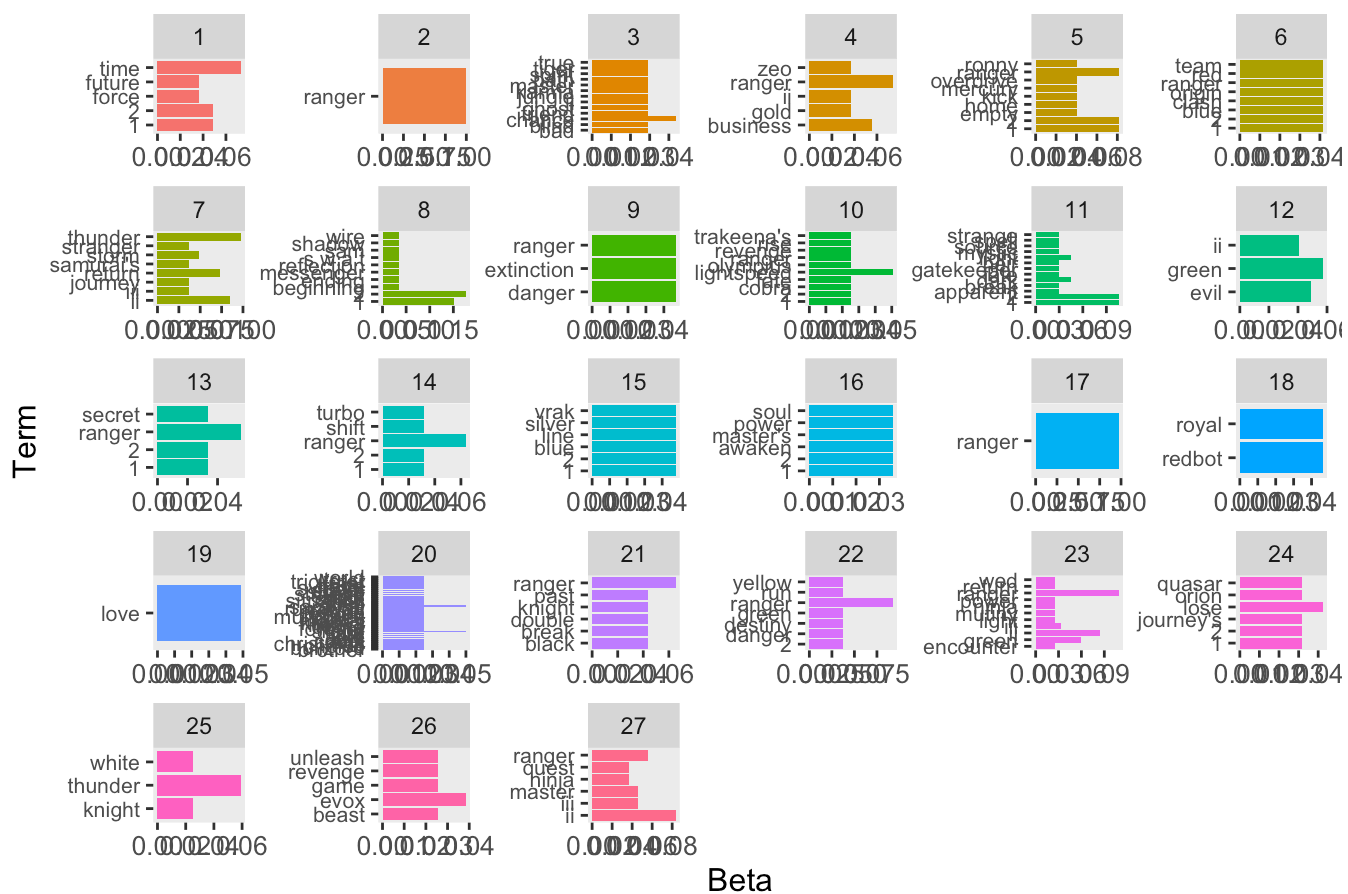
# Fit the LDA model
lda_model <- LDA(season_dtm, k = 27, control = list(seed = 1842))

# Extract the top terms in each topic
lda_topics <- tidy(lda_model, matrix = "beta")

# Get the top terms for each topic
# Beta is the term-topic matrix and represents the probability of each word belonging
# We are filtering on a beta value to limit the amount of topics displayed
top_terms <- lda_topics %>%
  group_by(topic) %>%
  top_n(5, beta) %>%
  ungroup() %>%
  filter(beta >= 0.0245) %>%
  arrange(topic, -beta)

# Plot the top terms in each topic
ggplot(top_terms, aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  scale_color_gradientn(colors = brewer.pal(25, "Set3")) + # Use a color gradient
  coord_flip() +
  labs(title = "Top Words from LDA Analysis Across Seasons",
       x = "Term",
       y = "Beta") +
  theme(axis.text.y = element_text(size = 8),
        axis.text.x = element_text(size = 10),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

# Top Words from LDA Analysis Across Seasons



We see the term "ranger" appear prominently across several seasons, which we would expect since we previously saw a color associated with a ranger, indicating that central characters are the focus of the episodes, which could also tie different narratives or storylines together. Suggesting the ranger character faces challenges, enemies, or adventures. We also see that we have many topics (e.g., seasons 2, 7, 9, 18) that involve terms associated with battles, conflicts, or confrontations, suggesting that a major overarching theme across the TV series could be action, adventure, and struggle against various threats. We also see elements (e.g., seasons 1, 3, 7) reference natural elements ("thunder," "storm," "wild") and supernatural or fate-related concepts ("ghost," "karma," "fate"). The Mighty Morphin Power Rangers appear to include both natural and supernatural elements, potentially blending genres like fantasy, science fiction, and adventure. We also get a glimpse in season 3, which suggests themes related to personal growth, ethical decisions, and interpersonal relationships. As our intro indicated, "Together they broke down barriers. They defeated evil by demonstrating teamwork, inclusivity, and diversity to people of all ages." These themes indicate subplots and secondary themes where characters undergo moral or personal challenges in addition to the main action.