

Tidy Tuesday Office

Poonam Shah

3/21/2020

```
# Get the Office rating data
office_ratings <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/
tidytuesday/master/data/2020/2020-03-17/
office_ratings.csv')

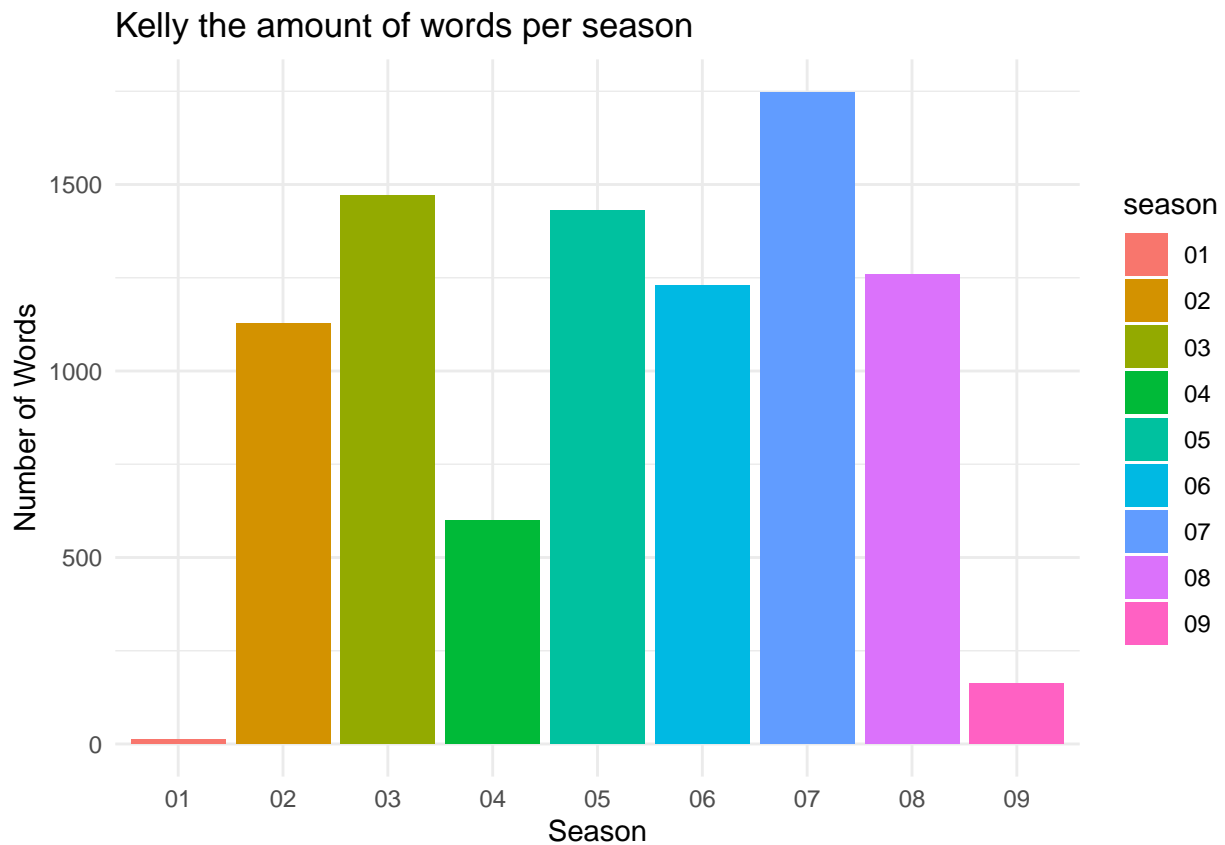
#Get the Schrute data
mydata <- schrute::theoffice
#convert a dataframe with a text column to be one-token-per-row
mydata %>%
  tidytext::unnest_tokens(words,text) ->Schrute_tidytext

#Kelly the amount of words per season
Schrute_tidytext %>%
  #filter the character by Kelly
  filter(character=="Kelly") %>%
  #deleted the unused columns
  select(-index,-episode,-episode_name,-director,
        -writer,-imdb_rating,-air_date,-total_votes, -text_w_direction)%>%
  group_by(season,character) %>%
  count() -> Mindy_Project

## filter: removed 561,415 rows (98%), 9,035 rows remaining
## select: dropped 9 variables (index, episode, episode_name, director, writer, ...)
## group_by: 2 grouping variables (season, character)
## count: now 9 rows and 3 columns, 2 group variables remaining (season, character)

#Making the plot
Mindy_Project %>%
  ggplot(aes(x=season, y=n, fill=season)) +
  geom_bar(stat = "identity") +
  #Adding a title and cleaning up the axis
  labs(x="Season", y="Number of Words",
        title="Kelly the amount of words per season") +
  scale_color_brewer(palette = "Set1") -> Mindy_Project_plot

print(Mindy_Project_plot)
```



```
#Saving the plot as a picture  
ggsave("Mindy_Project_plot.png", Mindy_Project_plot, height = 4, width = 6)
```

This week TidyTuesday was using a new package called tidytext. It is a package to help in text mining dataset and books. I used with package to see how amopunt words the character Kelly said in every season. The tidytext package was great in The Office dataset. I have only scatched the surface of the tidytext package I would love to practice more in the future for other tidytuesday.