

Sentiment Analysis

EDP 618 Week 8

Dr. Abhik Roy

Getting Prepped

Opening a Script Setting the working directory

1. Open up RStudio
2. Go to `File > New File > R Script`
3. Go to `File > Save As` and save the R Script in the same folder as the `csv` file. Name it whatever you want (e.g. **Week 7 R Walkthrough**)
4. Run the following command in your RStudio console

```
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
```

Loading Packages

Please load up the following packages by placing these at the top of your script

```
library(tidyverse)  
library(tidytext)  
library(textclean)
```

You may also want to put

```
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
```

below the packages so its there

Getting Data

We will be working with scripts from the first three seasons of the show *Rick and Morty*. Run the following to load the data

```
rickmorty <- read_csv("RickAndMortyScripts.csv")

##

Rows: 1905 Columns: 6
## [36m [39m Rendering ]8;;file:///Users/skynet/Documents/WVU/Teaching/GitHub.nosync/edp618/static/slides
##

## [36m [39m Rendering ]8;;file:///Users/skynet/Documents/WVU/Teaching/GitHub.nosync/edp618/static/slides

— Column specification —————
## [36m [39m Rendering ]8;;file:///Users/skynet/Documents/WVU/Teaching/GitHub.nosync/edp618/static/slides

Delimiter: ","
## chr (3): episode name, name, line
## dbl (3): index, season no., episode no.
##

## [36m [39m Rendering ]8;;file:///Users/skynet/Documents/WVU/Teaching/GitHub.nosync/edp618/static/slides
##

##
```

Assessing Data

We can take a look at the first ten rows of the data by running

```
head(rickmorty)
```

```
## # A tibble: 6 × 6
##   index `season no.` `episode no.` `episode name` name line
##   <dbl>      <dbl>      <dbl> <chr>      <chr> <chr>
## 1      0          1          1 Pilot      Rick  Morty! You gotta come on. Jus'... you go...
## 2      1          1          1 Pilot      Morty What, Rick? What's going on?
## 3      2          1          1 Pilot      Rick  I got a surprise for you, Morty.
## 4      3          1          1 Pilot      Morty It's the middle of the night. What are y...
## 5      4          1          1 Pilot      Rick  Come on, I got a surprise for you. Come...
## 6      5          1          1 Pilot      Morty Ow! Ow! You're tugging me too hard!
```

We can also take a look at the types of columns by running

```
glimpse(rickmorty)
```

```
## Rows: 1,905
## Columns: 6
## $ index      <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20...
## $ `season no.` <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ `episode no.` <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ `episode name` <chr> "Pilot", "Pilot", "Pilot", "Pilot", "Pilot", "Pilot", "Pilot", "Pilot", ...
## $ name        <chr> "Rick", "Morty", "Rick", "Morty", "Rick", "Morty", "Rick", "Rick", "Mort...
## $ line        <chr> "Morty! You gotta come on. Jus'... you gotta come with me.", "What, Rick..."
```

Wrangling Terms

Selecting Needed Columns

```
rickmortality_selected <-  
  rickmortality %>%  
  select(index, line)
```

```
rickmortality_selected
```

```
## # A tibble: 1,905 × 2  
##   index line  
##   <dbl> <chr>  
## 1      0 Morty! You gotta come on. Jus'... you gotta come with me.  
## 2      1 What, Rick? What's going on?  
## 3      2 I got a surprise for you, Morty.  
## 4      3 It's the middle of the night. What are you talking about?  
## 5      4 Come on, I got a surprise for you. Come on, hurry up.  
## 6      5 Ow! Ow! You're tugging me too hard!  
## 7      6 We gotta go, gotta get outta here, come on. Got a surprise for you Morty.  
## 8      7 What do you think of this... flying vehicle, Morty? I built it outta stuff I found in t...  
## 9      8 Yeah, Rick... I-it's great. Is this the surprise?  
## 10     9 Morty. I had to... I had to do it. I had— I had to— I had to make a bomb, Morty. I had ...  
## # ... with 1,895 more rows
```


Getting Rid of Common Terms

```
tidy_script <-  
  rickmarty_selected %>%  
  unnest_tokens(word, line) %>%  
  anti_join(stop_words)
```

```
## Joining, by = "word"
```

```
tidy_script
```

```
## # A tibble: 8,513 × 2  
##   index word  
##   <dbl> <chr>  
## 1      0 morty  
## 2      0 gotta  
## 3      0 jus  
## 4      0 gotta  
## 5      1 rick  
## 6      1 what's  
## 7      2 surprise  
## 8      2 morty  
## 9      3 middle  
## 10     3 night  
## # ... with 8,503 more rows
```

```
tidy_script %>%
```

```
count(word, sort = TRUE)
```

```
## # A tibble: 3,072 × 2
```

```
##   word      n
```

```
##   <chr> <int>
```

```
## 1 morty   334
```

```
## 2 rick    169
```

```
## 3 gonna   113
```

```
## 4 time     89
```

```
## 5 yeah     78
```

```
## 6 uh       53
```

```
## 7 whoa     53
```

```
## 8 jerry    50
```

```
## 9 god      48
```

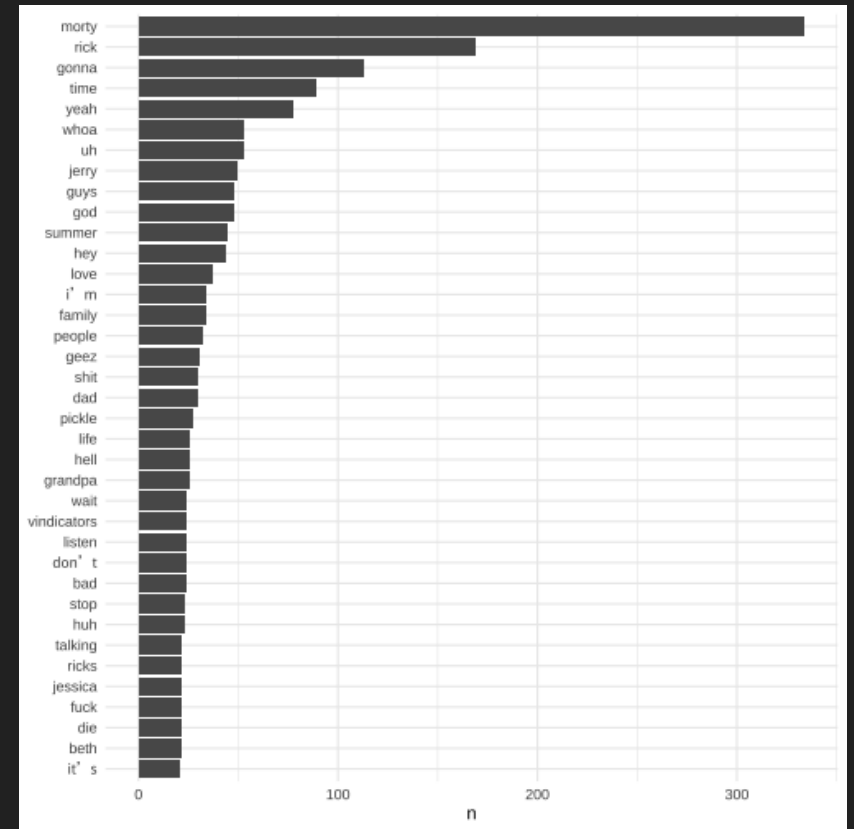
```
## 10 guys    48
```

```
## # ... with 3,062 more rows
```

```

tidy_script %>%
  count(word, sort = TRUE) %>%
  filter(n > 20) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(n, word)) +
  geom_col() +
  labs(y = NULL) +
  theme_minimal()

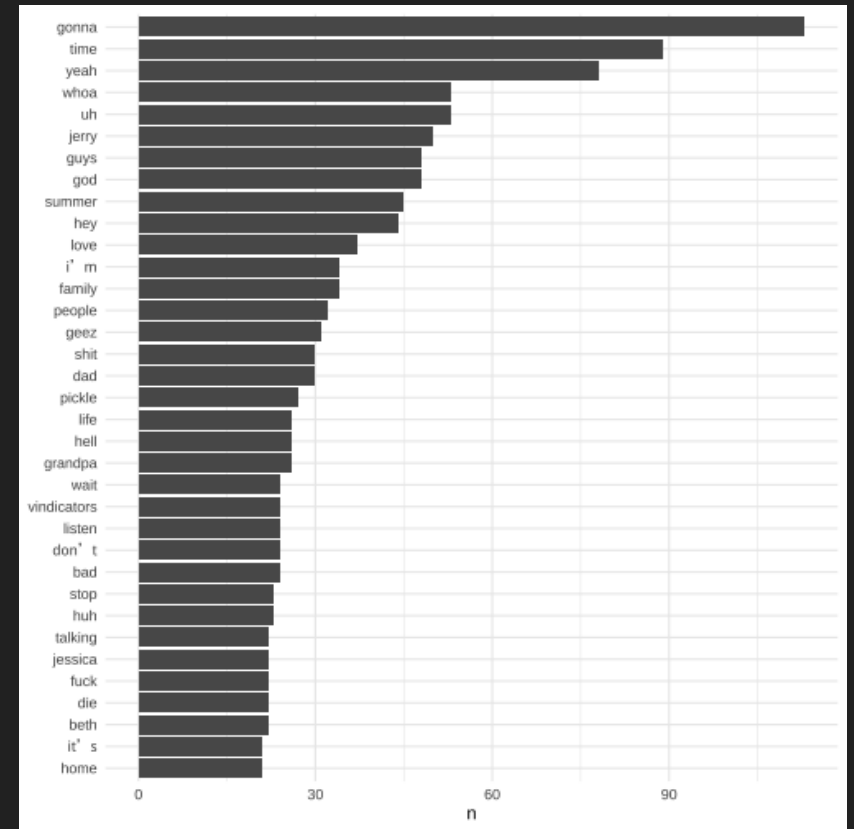
```



```
rickmorty_selected %>%  
mutate(line = str_remove_all(line, "Rick")) %>%  
mutate(line = str_remove_all(line, "Morty")) %>%  
mutate(line = replace_contraction(line)) %>%  
unnest_tokens(word, line) %>%  
anti_join(stop_words) %>%  
count(word, sort = TRUE)
```

```
## Joining, by = "word"  
  
## # A tibble: 3,056 × 2  
##   word      n  
##   <chr> <int>  
## 1 gonna    113  
## 2 time      89  
## 3 yeah     78  
## 4 uh       53  
## 5 whoa     53  
## 6 jerry    50  
## 7 god      48  
## 8 guys     48  
## 9 summer   45  
## 10 hey     44  
## # ... with 3,046 more rows
```

```
rickandmorty_filtered %>%  
  filter(n > 20) %>%  
  mutate(word = reorder(word, n)) %>%  
  ggplot(aes(n, word)) +  
  geom_col() +  
  labs(y = NULL) +  
  theme_minimal()
```



Sentiment Analysis

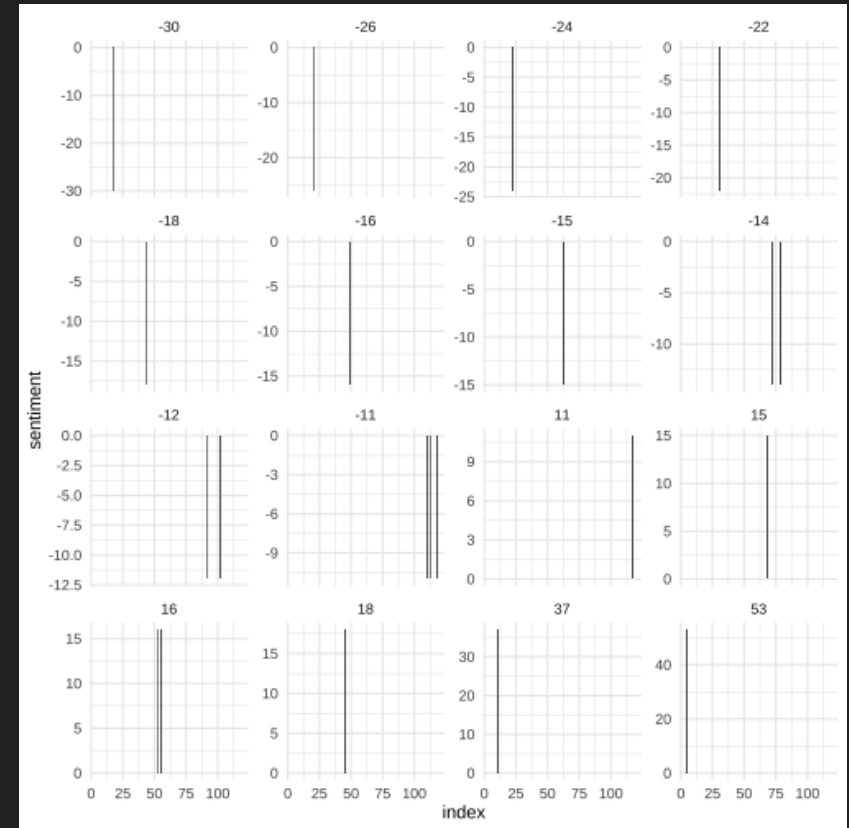
- is used to determine whether a given text contains negative, positive, or neutral emotions
- employs *Natural Language Processing* - computer program to understand human language as it is spoken and written

```
rickandmorty_filtered %>%
  rowid_to_column(var = "index") %>%
  inner_join(get_sentiments("bing")) %>%
  pivot_wider(names_from = sentiment,
              values_from = n,
              values_fill = 0) %>%
  mutate(sentiment = positive - negative)
```

```
## Joining, by = "word"
```

```
## # A tibble: 22 × 5
##   index word    positive negative sentiment
##   <int> <chr>      <int>      <int>      <int>
## 1     5 whoa         53         0         53
## 2    11 love         37         0         37
## 3    17 shit          0        30        -30
## 4    20 hell          0        26        -26
## 5    22 bad           0        24        -24
## 6    30 die           0        22        -22
## 7    31 fuck          0        22        -22
## 8    43 crap          0        18        -18
## 9    45 pretty       18         0         18
## 10   49 bitch         0        16        -16
## # ... with 12 more rows
```

```
ggplot(rickandmorty_bing,
      aes(index, sentiment)) +
  geom_bar(stat = "identity",
          show.legend = FALSE) +
  theme_minimal() +
  facet_wrap(~sentiment, scales = "free_y")
```




```
rickandmorty_filtered %>%  
  rowid_to_column(var = "index") %>%  
  inner_join(get_sentiments("afinn"))
```

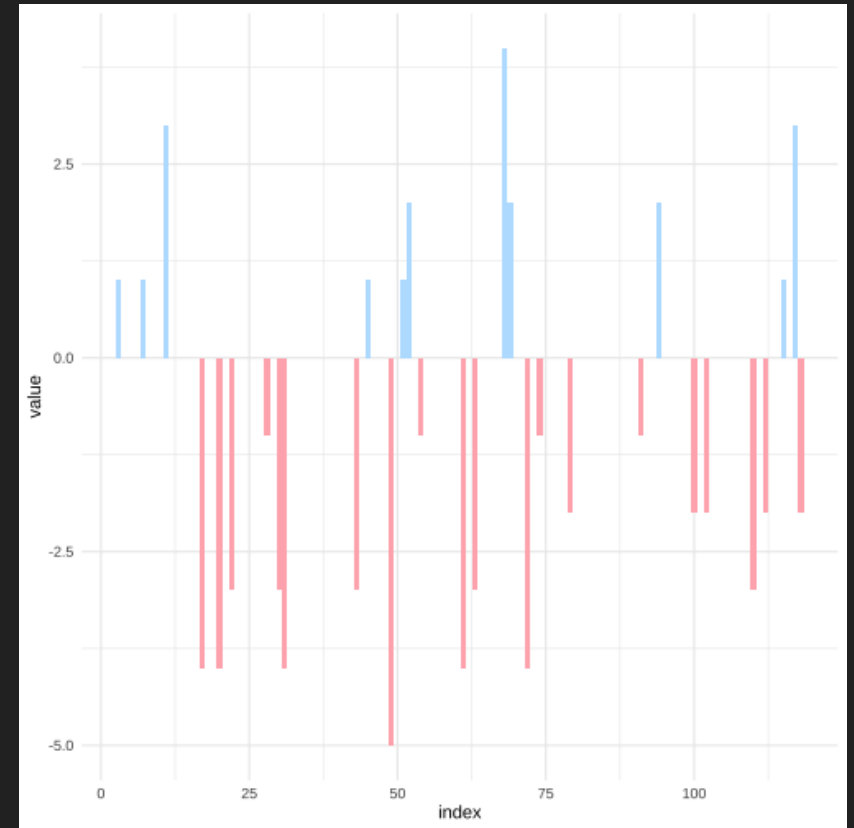
```
## Joining, by = "word"  
  
## # A tibble: 31 × 4  
##   index word      n value  
##   <int> <chr> <int> <dbl>  
## 1     3 yeah     78      1  
## 2     7 god      48      1  
## 3    11 love     37      3  
## 4    17 shit     30     -4  
## 5    20 hell     26     -4  
## 6    22 bad      24     -3  
## 7    28 stop     23     -1  
## 8    30 die      22     -3  
## 9    31 fuck      22     -4  
## 10   43 crap      18     -3  
## # ... with 21 more rows
```

```
rickandmorty_filtered %>%
  rowid_to_column(var = "index") %>%
  inner_join(get_sentiments("afinn")) %>%
  mutate(sentiment = if_else(value > 0,
                             "positive",
                             "negative",
                             "NA"))
```

```
## Joining, by = "word"
```

```
## # A tibble: 31 × 5
##   index word      n value sentiment
##   <int> <chr> <int> <dbl> <chr>
## 1     3 yeah     78      1 positive
## 2     7 god      48      1 positive
## 3    11 love     37      3 positive
## 4    17 shit     30     -4 negative
## 5    20 hell     26     -4 negative
## 6    22 bad      24     -3 negative
## 7    28 stop     23     -1 negative
## 8    30 die      22     -3 negative
## 9    31 fuck     22     -4 negative
## 10   43 crap     18     -3 negative
## # ... with 21 more rows
```

```
ggplot(rickandmorty_afinn_posneg,  
       aes(index, value, fill = sentiment)) +  
  geom_bar(stat = "identity",  
           show.legend = FALSE) +  
  theme_minimal() +  
  scale_fill_manual(values = c("#ffb3ba", "#bae1ff"))
```



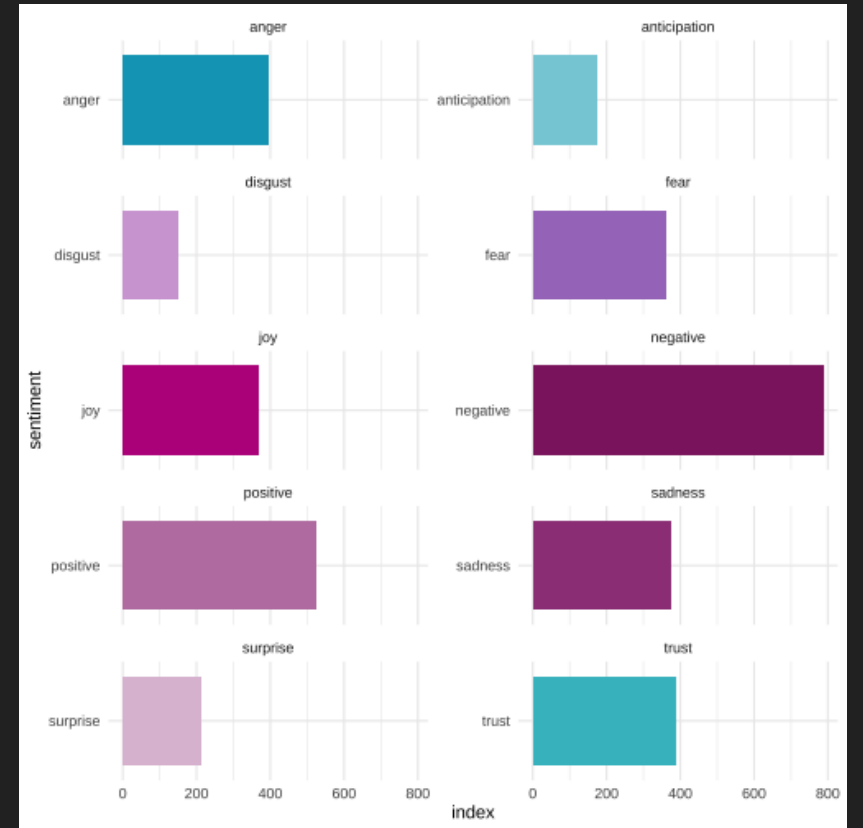
```
rickandmorty_filtered %>%  
  rowid_to_column(var = "index") %>%  
  inner_join(get_sentiments("nrc"))
```

```
## Joining, by = "word"  
  
## # A tibble: 72 × 4  
##   index word      n sentiment  
##   <int> <chr> <int> <chr>  
## 1     2 time      89 anticipation  
## 2     7 god       48 anticipation  
## 3     7 god       48 fear  
## 4     7 god       48 joy  
## 5     7 god       48 positive  
## 6     7 god       48 trust  
## 7    11 love      37 joy  
## 8    11 love      37 positive  
## 9    17 shit      30 anger  
## 10   17 shit      30 disgust  
## # ... with 62 more rows
```

```

ggplot(rickandmorty_nrc,
      aes(index, sentiment,
          fill = sentiment)) +
  geom_bar(stat = "identity",
          show.legend = FALSE) +
  theme_minimal() +
  facet_wrap(~sentiment, scales = "free_y",
            nrow = 5, ncol = 2) +
  scale_fill_manual(values = c("#05A4C0", "#85CEDA",
                                "#D2A7D8", "#A67BC5",
                                "#BB1C8B", "#8D266E",
                                "#BE82AF", "#9D4387",
                                "#DEC0D7", "#40BDC8",
                                "#80D3DB", "#BFE9ED"))

```



That's It!

Any questions?



This work is licensed under a
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License