

Text analysis and regular expressions

Nicholas Horton (nhorton@amherst.edu)

May 19, 2022

Note: you'll need to:

1. install the DickinsonPoems package and
2. run `get_sentiments("afinn")` interactively

to be able to knit this.

Warmup #1

How often does the word “science” appear in Emily Dickinson’s poetry?

```
all_poems <- list_poems() %>%
  map(get_poem) %>%
  unlist() %>%
  enframe(value = "words") %>%
  unnest_tokens(word, words)
head(all_poems)
```

```
## # A tibble: 6 x 2
##   name word
##   <int> <chr>
## 1     1 i
## 2     3 success
## 3     5 published
## 4     5 in
## 5     5 a
## 6     5 masque
```

```
all_poems %>%
  filter(stringr::str_detect(word, "science"))
```

```
## # A tibble: 4 x 2
##   name word
##   <int> <chr>
## 1   511 conscience
## 2  4618 science
## 3  6549 sciences
## 4  7849 science
```

Warmup #2

Classify Emily Dickinson’s poem *The Lonely House* as either positive or negative using the AFINN lexicon. Does this match with your own interpretation of the poem?

```
poem <- get_poem("gutenberg1.txt014")
poem
```

```

## [1] "XV. "
## [2] " "
## [3] "THE LONELY HOUSE. "
## [4] " "
## [5] "I know some lonely houses off the road "
## [6] "A robber 'd like the look of, -- "
## [7] "Wooden barred, "
## [8] "And windows hanging low, "
## [9] "Inviting to "
## [10] "A portico, "
## [11] "Where two could creep: "
## [12] "One hand the tools, "
## [13] "The other peep "
## [14] "To make sure all's asleep. "
## [15] "Old-fashioned eyes, "
## [16] "Not easy to surprise! "
## [17] " "
## [18] "How orderly the kitchen 'd look by night, "
## [19] "With just a clock, -- "
## [20] "But they could gag the tick, "
## [21] "And mice won't bark; "
## [22] "And so the walls don't tell, "
## [23] "None will. "
## [24] " "
## [25] "A pair of spectacles ajar just stir -- "
## [26] "An almanac's aware. "
## [27] "Was it the mat winked, "
## [28] "Or a nervous star? "
## [29] "The moon slides down the stair "
## [30] "To see who's there. "
## [31] " "
## [32] "There's plunder, -- where? "
## [33] "Tankard, or spoon, "
## [34] "Earring, or stone, "
## [35] "A watch, some ancient brooch "
## [36] "To match the grandmamma, "
## [37] "Staid sleeping there. "
## [38] " "
## [39] "Day rattles, too, "
## [40] "Stealth's slow; "
## [41] "The sun has got as far "
## [42] "As the third sycamore. "
## [43] "Screams chanticleer, "
## [44] "\"Who's there?\" "
## [45] "And echoes, trains away, "
## [46] "Sneer -- \"Where?\" "
## [47] "While the old couple, just astir, "
## [48] "Fancy the sunrise left the door ajar! "
## [49] " "
## [50] " "
## [51] " "
## [52] " "

```

```
# need to first run `get_sentiments("afinn")` interactively.
poem_sentiments <- data.frame(poem) %>%
  unnest_tokens(word, poem) %>%
  inner_join(get_sentiments("afinn"))
```

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

```
sum(poem_sentiments$value) # negative value
```

```
## [1] -8
```

Homework

Use the collection of Emily Dickinson poems to identify patterns or insights. This is a very open ended assignment: you may pick a few poems or consider all of them.