

Practicing Regular Expressions

Math 241, Week 8

```
#Load tidyverse
library(tidyverse)
library(mdsr)
```

Problem 1 (Easy):

Use the `Macbeth_raw` data from the `mdsr` package to answer the following questions:

- a. Find all of the adjectives in Macbeth that end in more or less

```
data(Macbeth_raw)
lines <- strsplit(Macbeth_raw, "\r\n")[[1]]
words <- strsplit(paste(lines, collapse = " "), " ")[[1]]
grep("[a-z]+(more|less)$", words, value = TRUE)
```

```
## [1] "harmless"    "merciless"   "careless"    "peerless"    "sightless"
## [6] "sightless"   "boneless"    "measureless" "bless"        "bless"
## [11] "Thriftless"  "dauntless"   "fruitless"    "reckless"     "restless"
## [16] "Unless"      "confineless" "Boundless"    "stanchless"
```

- b. Find all of the lines containing the stage direction Exit or Exeunt.

```
lines <- strsplit(Macbeth_raw, "\r\n")[[1]]
grep("(Exit|Exeunt)", lines, value = TRUE)
```

```
## [1] "      Hover through the fog and filthy air.                Exeunt."
## [2] "                                Exit Sergeant, attended."
## [3] "                                Exeunt."
## [4] "  MACBETH. Till then, enough. Come, friends.              Exeunt."
## [5] "  Which the eye fears, when it is done, to see.            Exit."
## [6] "  It is a peerless kinsman.                                Flourish. Exeunt."
## [7] "  He brings great news.                                    Exit Messenger."
## [8] "  Leave all the rest to me.                                Exeunt."
## [9] "  By your leave, hostess.                                  Exeunt."
## [10] "                                Exeunt."
## [11] "                                Exeunt Banquo. and Fleance."
## [12] "  She strike upon the bell. Get thee to bed.              Exit Servant."
## [13] "  That summons thee to heaven, or to hell.                  Exit."
## [14] "  For it must seem their guilt.                            Exit. Knocking within."
## [15] "                                Exeunt."
## [16] "  For 'tis my limited service.                             Exit."
```

```

## [17] "                               Exeunt Macbeth and Lennox."
## [18] "                   Exeunt all but Malcolm and Donalbain."
## [19] "                               Exeunt."
## [20] "                               Exeunt."
## [21] "    Farewell.                               Exit Banquo."
## [22] "                               Exeunt all but Macbeth and an Attendant."
## [23] "    MACBETH. Bring them before us.           Exit Attendant."
## [24] "                               Exit Attendant."
## [25] "                               Exeunt Murtherers."
## [26] "    If it find heaven, must find it out tonight.       Exit."
## [27] "    SERVANT. Madam, I will.                   Exit."
## [28] "    So, prithee, go with me.                 Exeunt."
## [29] "                               Exeunt."
## [30] "                               Exit Murtherer."
## [31] "    Shall be the maws of kites.               Exit Ghost."
## [32] "    Unreal mockery, hence!                   Exit Ghost."
## [33] "                               Exeunt all but Macbeth and Lady Macbeth."
## [34] "    We are yet but young in deed.             Exeunt."
## [35] "    Sits in a foggy cloud and stays for me.       Exit."
## [36] "                               Exeunt."
## [37] "                               Exeunt."
## [38] "    Come, bring me where they are.           Exeunt."
## [39] "    I take my leave at once.                 Exit."
## [40] "    I dare abide no longer.                 Exit."
## [41] "                               Exit Lady Macduff, crying \"Murther!\""
## [42] "                               Exeunt Murtherers, following her."
## [43] "    MALCOLM. I thank you, Doctor.             Exit Doctor."
## [44] "    The night is long that never finds the day.     Exeunt."
## [45] "Exit."
## [46] "                               Exeunt."
## [47] "    Make we our march towards Birnam.         Exeunt marching."
## [48] "    MACBETH. Take thy face hence.             Exit Servant."
## [49] "    Profit again should hardly draw me here.     Exeunt."
## [50] "                               Exeunt Marching."
## [51] "    SEYTON. It is the cry of women, my good lord.   Exit."
## [52] "    At least we'll die with harness on our back.   Exeunt."
## [53] "                               Exeunt."
## [54] "    Brandish'd by man that's of a woman born.     Exit."
## [55] "    And more I beg not.                       Exit. Alarums."
## [56] "                               Exeunt. Alarum."
## [57] "                               Exeunt fighting. Alarums."
## [58] "                               Flourish. Exeunt."

```

Problem 2 (Medium):

- a. Use regular expressions to determine the number of speaking lines in the Complete Works of William Shakespeare. Here, we care only about how many times a character speaks—not what they say or for how long they speak.

```

shakespeare_url <- "https://www.gutenberg.org/files/100/100-0.txt"
shakespeare <- RCurl::getURL(shakespeare_url) %>% str_split("\r\n") %>%
  pluck(1)
length(shakespeare)

```

```
## [1] 196019
```

```
speaker_lines_pattern <- "[A-Z ]+\\." # I want any number of uppercase characters followed by a period
speaker_lines <- grep(speaker_lines_pattern,shakespeare)
length(speaker_lines)
```

```
## [1] 33590
```

- b. Make a bar chart displaying the top 100 characters with the greatest number of lines. Hint you may want to use either the `stringr::str_extract` or `strsplit` function here.

```
speakers <- str_extract(shakespeare, speaker_lines_pattern)
speaker_df <- speakers %>%
  data.frame(speaker = .) %>%
  filter(!is.na(speaker)) %>%
  group_by(speaker) %>%
  summarize(N = n()) %>%
  arrange(desc(N)) %>% head(100)

ggplot(data = speaker_df, aes(x = reorder(speaker, N), y = N)) +
  geom_bar(stat = "identity") +
  coord_flip()
```

