In-class Assignment 11

Andrew Shao (NetID: as13381)

Question 1 (1 pt): Use `str_length()` and `str_sub()` to extract the middle character from each string in the vector `x <- c("a", "abc", "abcd", "abcde", "abcdef")`. What will you do if a string has an even number of characters? (Hint: Base R function `floor()` or `ceiling()` may be useful here.)

Answer:

```
x <- c("a", "abc", "abcd", "abcde", "abcdef")

str_sub(x, ceiling(str_length(x) / 2), ceiling(str_length(x) / 2))
```

```
## [1] "a" "b" "b" "c" "c"
```

Question 2 (1 pt): `words` is a vector with each element being a word, which is available after you run `library(tidyverse)`. From the vector `words`, find the words that have the highest number of vowels?

Answer:

```
words[sapply(str_extract_all(words, '[aeiou]'), length) ==
↪  max(sapply(str_extract_all(words, '[aeiou]'), length))]
```

```
## [1] "appropriate" "associate"   "available"   "colleague"   "encourage"
## [6] "experience"  "individual"  "television"
```

Question 3 (1 pt): `sentences` is a vector with each element being a sentence, which is available after you run `library(tidyverse)`. Extract the first word from each sentence in `sentences`. Use `%>% head(100)` to output the result.

Answer:

```
str_split(sentences, ' ', simplify = T)[,1] %>% head(100)
```

```
##   [1] "The"      "Glue"     "It's"     "These"    "Rice"     "The"
##   [7] "The"      "The"      "Four"     "A"        "The"      "A"
##  [13] "The"      "Kick"     "Help"     "A"        "Smoky"    "The"
##  [19] "The"      "The"      "The"      "The"      "Press"    "The"
##  [25] "The"      "Two"      "Her"      "The"      "It"       "Read"
##  [31] "Hoist"    "Take"     "Note"     "Wipe"     "Mend"     "The"
##  [37] "The"      "The"      "The"      "What"     "A"        "The"
```

```
##  [43] "Sickness" "The"      "The"      "Lift"     "The"      "Hop"
##  [49] "The"      "Mesh"     "The"      "The"      "Adding"   "The"
##  [55] "A"        "The"      "March"    "A"        "Place"    "Both"
##  [61] "We"       "Use"      "He"       "The"      "A"        "Cars"
##  [67] "The"      "This"     "The"      "Those"    "A"        "The"
##  [73] "The"      "The"      "The"      "A"        "The"      "The"
##  [79] "The"      "The"      "The"      "See"      "There"    "The"
##  [85] "The"      "The"      "Cut"      "Men"      "Always"   "He"
##  [91] "The"      "A"        "A"        "The"      "The"      "Bail"
##  [97] "The"      "A"        "Ten"      "The"
```

Question 4 (1 pt): In the vector `sentences`, find all words that contain an apostrophe. In each found word, separate out the pieces before and after the apostrophe. (Hint: See 14.4.3 Grouped Matches in the textbook.)

Answer:

```
str_split(sentences, ' ', simplify = T)[str_split(sentences, ' ', simplify = T) %>%
↪  str_detect('\'')] %>%
  str_split('\'', simplify = T)
```

```
##       [,1]       [,2]
##  [1,] "It"       "s"
##  [2,] "Let"      "s"
##  [3,] "It"       "s"
##  [4,] "child"    "s"
##  [5,] "don"      "t"
##  [6,] "don"      "t"
##  [7,] "store"    "s"
##  [8,] "don"      "t"
##  [9,] "don"      "t"
## [10,] "don"      "t"
## [11,] "man"      "s"
## [12,] "don"      "t"
## [13,] "workman"  "s"
## [14,] "sun"      "s"
## [15,] "queen"    "s"
## [16,] "pirate"   "s"
## [17,] "neighbor" "s"
## [18,] "king"     "s"
```

Question 5 (1 pt): What are the five most common words in the vector `sentences`? (Hint: You may use `boundary("word")` with an appropriate function. Use the function `unlist()` to convert a list to a vector. The function `count()` may be useful here.)

Answer:

```
sort(table(str_to_lower(unlist(str_split(sentences, boundary('word'))))), decreasing =
↪  T)[1:5]
```

```
##
```

```
## the   a  of  to and
## 744 213 132 123 118
```