Socio-Informatics 348
Practical 6

Submission Instructions

- Submit your completed practical as `studentnumber.qmd` on SocSciLearn.

- Submissions are checked for completeness, not correctness.

- At least 80% of exercises must be attempted to receive 1% towards AF assessment.

- Attendance of at least one practical session per week is required to earn the 1% for that week's practical.

Deadline

Friday 26 September, 17:00 (submit on SocSciLearn)

Chapters Covered:

- R4DS: Chapters 15, 16 and 17

Exercises

Section 1: `babynames` dataset from `babynames` package

1. Which baby name has the most vowels? (Hint: use a regular expression.)

2. Which baby name has the highest proportion of vowels? (Hint: what is the denominator?)

Section 2: `words` from `stringr` package

3. Given the corpus of common words in `stringr::words`, create separate regular expressions that find all words that:

    - Start with "y".
    - Don't start with "y".
    - End with "x".
    - Are exactly three letters long (without using `str_length()`).
    - Contain a vowel-consonant pair.
    - Contain at least two vowel-consonant pairs in a row.
    - Only consist of repeated vowel-consonant pairs.

Section 3: Dates and Times with `lubridate`

4. What happens if you parse a string that contains invalid dates?

   ```
   ymd(c("2010-10-10", "bananas"))
   ```

5. What does the `tzone` argument to `today()` do? Why is it important?

6. For each of the following date-times, show how you'd parse it using:

   (a) A `readr` column specification

   (b) A `lubridate` function

   ```
   d1 <- "January 1, 2010"
   d2 <- "2015-Mar-07"
   d3 <- "06-Jun-2017"
   d4 <- c("August 19 (2015)", "July 1 (2015)")
   d5 <- "12/30/14" # Dec 30, 2014
   ```