Module: Getting Started

April 24, 2021

Contents

GOAL: By the end of this module, you will have set up your computing environment with the appropriate software versions. Additionally, you would also be able to read and view the curated data files using `R`.

CODE FOR THIS MODULE: `module-getting-started.R`.

1 Set Up of Software

Let's load the two packages we will use throughout: `dplyr` for data manipulation and `lme4` for estimating generalized linear mixed models (GLMMs) with longitudinal data.

```r
library(dplyr)
library(lme4)
```

We note that we are using the following combination of versions of software for Module 1: `R 4.0.4`, `dplyr 1.0.5`, `lme4 1.1-25`. The code examples presented in the current and succeeding modules in this workshop may not display output as expected when using older versions of these software. In particular, examples reading csv files using versions of `R` prior to `4.0` or using versions of `lme4` prior to `1.0` will not result in output identical to what will be shown in this module.

Specific versions of `R` itself can be downloaded from the following URLs, depending on your operating system:

- Windows

  - https://cran.r-project.org/bin/windows/base/ for newer versions of `R`
  - https://cran.r-project.org/bin/windows/base/old/ for older versions of `R`

- MACOSX

  - https://cran.r-project.org/bin/macosx/base/ for newer versions of `R`
  - https://cran.r-project.org/bin/macosx/old/ for older versions of `R`

CAUTION: Check the particular version of `R` you are working with prior to starting any analysis. If you are not working with the required version, you may switch versions through the `Tools -> Global Options -> General` tab if you are using R Studio.

You may already have some version of the `dplyr` or `lme4` package installed. However, if you would like to use the specific versions used in this workshop, one of the simplest ways is to first uninstall any existing version you may have using the `remove.packages` function from the `utils` package, and then install specific versions of `R` packages through the `install_version()` function of the `devtools` package, like so:

```r
utils::remove.packages("dplyr")
utils::remove.packages("lme4")

devtools::install_version("dplyr", version = "1.0.5", repos = "http://cran.us.r-project.org")
devtools::install_version("lme4", version = "1.1-25", repos = "http://cran.us.r-project.org")
```

As check, you may run `sessionInfo()`. Displayed below is the combination of software versions used for the modules in this workshop.

```r
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices datasets  utils     methods   base
##
## other attached packages:
## [1] lme4_1.1-25  Matrix_1.3-2 dplyr_1.0.5
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6         pillar_1.5.1      compiler_4.0.4    formatR_1.8
##  [5] nloptr_1.2.2.2     tools_4.0.4       boot_1.3-27       digest_0.6.27
##  [9] statmod_1.4.35     evaluate_0.14     lifecycle_1.0.0   tibble_3.1.0
## [13] nlme_3.1-152       lattice_0.20-41   pkgconfig_2.0.3   rlang_0.4.10
## [17] yaml_2.2.1         xfun_0.22         stringr_1.4.0     knitr_1.31
## [21] generics_0.1.0     vctrs_0.3.7       grid_4.0.4        tidyselect_1.1.0
## [25] glue_1.4.2         R6_2.5.0          fansi_0.4.2       rmarkdown_2.7
## [29] minqa_1.2.4        purrr_0.3.4       magrittr_2.0.1    ellipsis_0.3.1
## [33] htmltools_0.5.1.1  splines_4.0.4     MASS_7.3-53.1     renv_0.13.1
## [37] utf8_1.2.1         stringi_1.5.3     crayon_1.4.1
```

2 Set Up of Locations to Store Data Files and Output

We specify the location of input files and output files in the variables `path_pns_input_data` and `path_pns_output_data`, respectively.

```
path_pns_input_data <- "C:/Users/jamieyap/Desktop/input_data"
path_pns_output_data <- "C:/Users/jamieyap/Desktop/output_data"
```

3 Reading in Data Files

Let's now read in the data files we will be using in Module 1. The following are noteworthy:

- These data files should be within the location specified in `path_pns_input_data`.
- We set the value of `na.strings` to be a blank character, which tells `R` to treat such characters in a csv file as missing values. This is necessary since, by default, `R` treats cells coded as `NA` as missing values, i.e., by default, `na.strings = "NA"`, but all missing values in the curated data files are coded as blanks.

```
dat_big_merged <- read.csv(file.path(path_pns_input_data, "merged.csv"), header = TRUE,
    na.strings = "")
dat_quit_dates <- read.csv(file.path(path_pns_input_data, "quit_dates_final.csv"),
    header = TRUE, na.strings = "")
ema_item_names <- read.csv(file.path(path_pns_input_data, "ema_item_names.csv"),
    header = TRUE, na.strings = "")
```

As a check for whether we have been able to read the files successfully, let's view the first few rows of `ema_item_names`.

```
head(ema_item_names, n = 10)
```

```
##    is_postquit_assessment_type        assessment_type      name_codebook
## 1                            1 Post-Quit Already Slipped  Consume_PastTense1
## 2                            1 Post-Quit Already Slipped  Consume_PastTense2
## 3                            1 Post-Quit Already Slipped          LocationSm
## 4                            1 Post-Quit Already Slipped         SS_PastTense1
## 5                            1 Post-Quit Already Slipped         SS_PastTense2
## 6                            1 Post-Quit Already Slipped         SS_PastTense3
## 7                            1 Post-Quit Already Slipped         SS_PastTense4
## 8                            1 Post-Quit Already Slipped AbSelfEff_PastTense
## 9                            1 Post-Quit Already Slipped   Affect_PastTense1
## 10                           1 Post-Quit Already Slipped  Affect_PastTense10
##                         name_new
## 1    postquit_alreadyslipped_item_1
## 2    postquit_alreadyslipped_item_2
## 3    postquit_alreadyslipped_item_3
## 4    postquit_alreadyslipped_item_4
## 5    postquit_alreadyslipped_item_5
## 6    postquit_alreadyslipped_item_6
## 7    postquit_alreadyslipped_item_7
## 8    postquit_alreadyslipped_item_8
## 9    postquit_alreadyslipped_item_9
## 10 postquit_alreadyslipped_item_10
```