Working with eye-tracking reading data in R

Loading and eye-balling a dataset

Daniela Palleschi

2023-04-12

Table of contents

Set-up

```r
knitr::opts_chunk$set(eval = T, # evaluate = T for REPRODUCIBLE analyses
                      echo = T, # 'print code chunk?'
                      message = F, # print messages?
                      error = T, # render even if errors encountered?
                      warning = F) # print warnings?
```

```r
library(here) # relative path
library(tidyverse) # tidy/transform
library(beepr) # beeps when code runs or fails
library(rbbt) # zotero plugin
```

```r
## play sound if error encountered
### from: https://sejohnston.com/2015/02/24/make-r-beep-when-r-markdown-finishes-or-when-i
options(error = function(){    # Beep on error
  beepr::beep(sound = "wilhelm")
  Sys.sleep(2) #
  }
 )
## and when knitting is complete
.Last <- function() {          # Beep on exiting session
  beepr::beep(sound = "ping")
  Sys.sleep(6) # allow to play for 6 seconds
  }
```

```r
# Create references.json file based on the citations in this script:
# 1. make sure you have 'bibliography: references/references.json' in the YAML
# 2. create a new folder called 'references'
# 3. run:
rbbt::bbt_update_bib("_et_dataset.qmd")
```

The Perfect Lifetime Effect

- the English Present Perfect (e.g., *has done*) (e.g., Comrie, 1976)
  - must be used in temporal contexts that *include the present*
    - *I have been sick* ***since last week***
    - **I have been sick* ***last year***

- The Lifetime Effect
  - a referent’s lifetime (dead/alive) constrains verb tense in certain circumstances (e.g., Mittwoch, 2008a)
    - *_Queen Elizabeth II_ **_is_** _the British monarch._
    - *_King Charles III_ **_was_** _the British monarch._
- the Perfect Lifetime Effect
  - the (English) Present Perfect cannot be used to describe events of a dead person (e.g., Mittwoch, 2008b)
    - *_Queen Elizabeth II_ **_has met_** _many politicians._
    - _King Charles III_ **_has met_** _many politicians._

Our first dataset

- referent-lifetime context
  - dead/alive
- critical sentence
  - Present Perfect/Simple Future
- binary naturalness judgement to end trial
  - accept/reject

Design description

- 2x2 mixed design
  - two 2-level factors (2x2 = 2-level x 2-level)
    - factor 1: lifetime (levels: dead, alive)
    - factor 2: tense (levels: PP, SF)

	alive	dead
PP	Eddie Redmayne…has won	Gene Kelly…*has won
SF	Eddie Redmayne…will win	Gene Kelly…*will win

- predictors/independent variables
  - lifetime
  - tense

- measure/dependent variables (verb region)
    - first-fixation time (milliseconds)
    - first-pass reading time (ms)
    - regression path duration (ms)
    - total reading time (ms)

Repeated measures design

- observations are repeated e.g., multiple data points per participant, and per item across participants
    - essentially, data are not independent
    - e.g., each participant will have their own reading speed, some items might be systematically less acceptable for some unforeseen reason, etc.

Working with the data

Day 1

1. load the data
2. inspect data
    - eyeball data structure
    - print summaries
    - plot data distributions

Day 2

3. tidy data
4. visualise data
5. communicate data

Day 3

6. analyse data
    - confirmatory (a priori)
    - exploratory (post-hoc)
7. report analyses

Install packages

```
install.packages("tidyverse")
install.packages("here")
```

- install
  - only do once
  - …or when you working on a new computer
  - …or after updating R
- might be a wise idea to create a script just for installing packages
  - can save time/energy when updating R

Load packages

```
library(tidyverse)
library(here)
```

- load packages
  - needed at the start of each session

Load dataset

```
df_lifetime <- readr::read_csv(here::here("data/data_lifetime_pilot.csv"))
```

- N.B., `readr::read_csv` can be read as "`read_csv()` function in the `readr` package"
  - i.e., `package::function()`
  - you only need to use this syntax if you haven't loaded the specific package yet (maybe because you only need it once), or if a function name is included in multiple packages (i.e., there's a discrepancy in what `read_csv` could be referring to)
  - why did I use it here?

`here` package

Using the `here` package, we can access files *relative* to where our .RProj is stored.
In 'olden times', we had to specify the file path with something like:

```
# load in data from an *absolute* file path
df_lifetime <- read_csv("Users/yournamehere/Documents/SoSe2023/ET_reading/data/data_life
```

Or, we'd set an *absolute* path as our working directory, to which all other file paths were *relative*

```
# set *absolute* path as working directory
setwd("Users/username/Documents/SoSe2023/ET_reading")

# load in data *relative* to our wd
df_lifetime <- read_csv("data/data_lifetime_pilot.csv")
```

This meant that if I sent my project folder to somebody else, they wouldn't be able to run my code because they would have to change the *absolute* file path to match their machine.

Inspect dataset

- there are several different things you can inspect
  - and different ways to accomplish those things
- the first thing I usually do is look at the column/variable names

`names()`

- the names in all caps are variables created during the experiment
  - i.e., they are our recorded *data*, mainly what we wanted to measure: dependent variables (DV)
  - also includes some information about the experiment set-up per participant
- the other names are variables from my stimuli lists
  - i.e., they mostly contain our independent variables (IV)/stimuli
- we typically want to see what effect our IVs had on any given DVs

- variable descriptions can be found on the Moodle: Data > Documentation

```r
names(df_lifetime)
```

```
 [1] "RECORDING_SESSION_LABEL"    "TRIAL_INDEX"
 [3] "EYE_USED"                   "IA_DWELL_TIME"
 [5] "IA_FIRST_FIXATION_DURATION" "IA_FIRST_RUN_DWELL_TIME"
 [7] "IA_FIXATION_COUNT"          "IA_ID"
 [9] "IA_LABEL"                   "IA_REGRESSION_IN"
[11] "IA_REGRESSION_IN_COUNT"     "IA_REGRESSION_OUT"
[13] "IA_REGRESSION_OUT_COUNT"    "IA_REGRESSION_PATH_DURATION"
[15] "KeyPress"                   "rt"
[17] "bio"                        "critical"
[19] "gender"                     "item_id"
[21] "list"                       "match"
[23] "condition"                  "name"
[25] "name_vital_status"          "tense"
[27] "type"                       "yes_press"
```

`rename()`

- the dependent variable names are pretty clunky, let's rename a few:
    - `RECORDING_SESSION_LABEL` corresponds to a single participant
    - `TRIAL_INDEX` logged the trial number
    - `EYE_USED` logged which eye was tracked

```r
df_lifetime <- df_lifetime %>%
  rename("px" = RECORDING_SESSION_LABEL,
         "trial" = TRIAL_INDEX,
         "eye" = EYE_USED)
```

Naming variables

Naming conventions

It's wise to keep variable and object names concise but informative

- all lowercase means fewer key strokes overall
- separate words with either periods or underscores, e.g., `trial.index` or `trial_index`

> - e.g., we called our dataset `df_lifetime` because it is a dataframe (`df`) with data from our lifetime experiment

Data structure

- datasets typically contain a lot of rows and columns
  - so we want to get a feel for how the data is structured

with base R

```r
head(df_lifetime)
```

```
# A tibble: 6 x 28
  px    trial eye   IA_DWELL_TIME IA_FIRST_FIXATION_DUR~1 IA_FIRST_RUN_DWELL_T~2
  <chr> <dbl> <chr>         <dbl>                   <dbl>                  <dbl>
1 px3       1 RIGHT             0                       0                      0
2 px3       2 RIGHT             0                       0                      0
3 px3       3 RIGHT             0                       0                      0
4 px3       3 RIGHT             0                       0                      0
5 px3       3 RIGHT             0                       0                      0
6 px3       3 RIGHT             0                       0                      0
# i abbreviated names: 1: IA_FIRST_FIXATION_DURATION,
#   2: IA_FIRST_RUN_DWELL_TIME
# i 22 more variables: IA_FIXATION_COUNT <dbl>, IA_ID <dbl>, IA_LABEL <chr>,
#   IA_REGRESSION_IN <dbl>, IA_REGRESSION_IN_COUNT <dbl>,
#   IA_REGRESSION_OUT <dbl>, IA_REGRESSION_OUT_COUNT <dbl>,
#   IA_REGRESSION_PATH_DURATION <dbl>, KeyPress <dbl>, rt <dbl>, bio <chr>,
#   critical <chr>, gender <chr>, item_id <dbl>, list <dbl>, match <chr>, ...
```

with the tidyverse pipe

```r
df_lifetime %>%
  head()
```

```
# A tibble: 6 x 28
  px    trial eye   IA_DWELL_TIME IA_FIRST_FIXATION_DUR~1 IA_FIRST_RUN_DWELL_T~2
  <chr> <dbl> <chr>         <dbl>                   <dbl>                  <dbl>
1 px3       1 RIGHT             0                       0                      0
```

```
2 px3       2 RIGHT             0                       0                      0
3 px3       3 RIGHT             0                       0                      0
4 px3       3 RIGHT             0                       0                      0
5 px3       3 RIGHT             0                       0                      0
6 px3       3 RIGHT             0                       0                      0
# i abbreviated names: 1: IA_FIRST_FIXATION_DURATION,
#   2: IA_FIRST_RUN_DWELL_TIME
# i 22 more variables: IA_FIXATION_COUNT <dbl>, IA_ID <dbl>, IA_LABEL <chr>,
#   IA_REGRESSION_IN <dbl>, IA_REGRESSION_IN_COUNT <dbl>,
#   IA_REGRESSION_OUT <dbl>, IA_REGRESSION_OUT_COUNT <dbl>,
#   IA_REGRESSION_PATH_DURATION <dbl>, KeyPress <dbl>, rt <dbl>, bio <chr>,
#   critical <chr>, gender <chr>, item_id <dbl>, list <dbl>, match <chr>, ...
```

with the native R pipe (Ctrl/Cmd+Shift+M)

```
df_lifetime |>
  head()
```

```
# A tibble: 6 x 28
  px    trial eye   IA_DWELL_TIME IA_FIRST_FIXATION_DUR~1 IA_FIRST_RUN_DWELL_T~2
  <chr> <dbl> <chr>         <dbl>                   <dbl>                  <dbl>
1 px3       1 RIGHT             0                       0                      0
2 px3       2 RIGHT             0                       0                      0
3 px3       3 RIGHT             0                       0                      0
4 px3       3 RIGHT             0                       0                      0
5 px3       3 RIGHT             0                       0                      0
6 px3       3 RIGHT             0                       0                      0
# i abbreviated names: 1: IA_FIRST_FIXATION_DURATION,
#   2: IA_FIRST_RUN_DWELL_TIME
# i 22 more variables: IA_FIXATION_COUNT <dbl>, IA_ID <dbl>, IA_LABEL <chr>,
#   IA_REGRESSION_IN <dbl>, IA_REGRESSION_IN_COUNT <dbl>,
#   IA_REGRESSION_OUT <dbl>, IA_REGRESSION_OUT_COUNT <dbl>,
#   IA_REGRESSION_PATH_DURATION <dbl>, KeyPress <dbl>, rt <dbl>, bio <chr>,
#   critical <chr>, gender <chr>, item_id <dbl>, list <dbl>, match <chr>, ...
```

`head()` function

- *prints* the first 6 rows of your data
  - you can also specify the number of rows

```r
df_lifetime %>%
  head(n = 2)
```

```
# A tibble: 2 x 28
  px    trial eye   IA_DWELL_TIME IA_FIRST_FIXATION_DUR~1 IA_FIRST_RUN_DWELL_T~2
  <chr> <dbl> <chr>         <dbl>                   <dbl>                  <dbl>
1 px3       1 RIGHT             0                       0                      0
2 px3       2 RIGHT             0                       0                      0
# i abbreviated names: 1: IA_FIRST_FIXATION_DURATION,
#   2: IA_FIRST_RUN_DWELL_TIME
# i 22 more variables: IA_FIXATION_COUNT <dbl>, IA_ID <dbl>, IA_LABEL <chr>,
#   IA_REGRESSION_IN <dbl>, IA_REGRESSION_IN_COUNT <dbl>,
#   IA_REGRESSION_OUT <dbl>, IA_REGRESSION_OUT_COUNT <dbl>,
#   IA_REGRESSION_PATH_DURATION <dbl>, KeyPress <dbl>, rt <dbl>, bio <chr>,
#   critical <chr>, gender <chr>, item_id <dbl>, list <dbl>, match <chr>, ...
```

`head()` function task

Exercise: `head()`

1. print only 2 rows using whichever syntax you prefer
2. change `n = 2` to some other number and print
3. run `?head` in the `Console`
    - find the opposite function (i.e., prints last rows) in the function description?
4. run this function with `df_lifetime` as argument; how many rows does it print as default?
5. play with `n =` in this function to print some other number of rows

`tail()` function

- prints the last rows of a dataframe (or matrix, vector, table, or function)

```r
df_lifetime %>%
  tail()
```

```
# A tibble: 6 x 28
  px    trial eye   IA_DWELL_TIME IA_FIRST_FIXATION_DUR~1 IA_FIRST_RUN_DWELL_T~2
  <chr> <dbl> <chr>         <dbl>                   <dbl>                  <dbl>
```

```
1 px4     207 LEFT            509                     218                    509
2 px4     208 LEFT              0                       0                      0
3 px4     208 LEFT            317                     167                    317
4 px4     208 LEFT            162                     162                    162
5 px4     208 LEFT            139                     139                    139
6 px4     208 LEFT            280                     280                    280
# i abbreviated names: 1: IA_FIRST_FIXATION_DURATION,
#   2: IA_FIRST_RUN_DWELL_TIME
# i 22 more variables: IA_FIXATION_COUNT <dbl>, IA_ID <dbl>, IA_LABEL <chr>,
#   IA_REGRESSION_IN <dbl>, IA_REGRESSION_IN_COUNT <dbl>,
#   IA_REGRESSION_OUT <dbl>, IA_REGRESSION_OUT_COUNT <dbl>,
#   IA_REGRESSION_PATH_DURATION <dbl>, KeyPress <dbl>, rt <dbl>, bio <chr>,
#   critical <chr>, gender <chr>, item_id <dbl>, list <dbl>, match <chr>, ...
```

names()

- prints the column/variable names

```
df_lifetime %>%
  names()
```

```
 [1] "px"                          "trial"
 [3] "eye"                         "IA_DWELL_TIME"
 [5] "IA_FIRST_FIXATION_DURATION"  "IA_FIRST_RUN_DWELL_TIME"
 [7] "IA_FIXATION_COUNT"           "IA_ID"
 [9] "IA_LABEL"                    "IA_REGRESSION_IN"
[11] "IA_REGRESSION_IN_COUNT"      "IA_REGRESSION_OUT"
[13] "IA_REGRESSION_OUT_COUNT"     "IA_REGRESSION_PATH_DURATION"
[15] "KeyPress"                    "rt"
[17] "bio"                         "critical"
[19] "gender"                      "item_id"
[21] "list"                        "match"
[23] "condition"                   "name"
[25] "name_vital_status"           "tense"
[27] "type"                        "yes_press"
```

summary()

- prints a summary of each variable (column)

```r
df_lifetime %>%
  summary()
```

```
     px                trial           eye            IA_DWELL_TIME
Length:4431        Min.   :  1.0   Length:4431        Min.   :   0.0
Class :character   1st Qu.: 52.5   Class :character   1st Qu.:   0.0
Mode  :character   Median :104.0   Mode  :character   Median : 301.0
                   Mean   :105.0                      Mean   : 587.5
                   3rd Qu.:157.0                      3rd Qu.: 765.5
                   Max.   :208.0                      Max.   :8968.0
IA_FIRST_FIXATION_DURATION IA_FIRST_RUN_DWELL_TIME IA_FIXATION_COUNT
Min.   :  0.0              Min.   :   0.0          Min.   : 0.000
1st Qu.:  0.0              1st Qu.:   0.0          1st Qu.: 0.000
Median :161.0              Median : 245.0          Median : 2.000
Mean   :139.4              Mean   : 507.9          Mean   : 2.714
3rd Qu.:202.5              3rd Qu.: 586.0          3rd Qu.: 4.000
Max.   :775.0              Max.   :8968.0          Max.   :35.000
    IA_ID         IA_LABEL         IA_REGRESSION_IN  IA_REGRESSION_IN_COUNT
Min.   :1.000   Length:4431        Min.   :0.00000   Min.   :0.0000
1st Qu.:1.000   Class :character   1st Qu.:0.00000   1st Qu.:0.0000
Median :2.000   Mode  :character   Median :0.00000   Median :0.0000
Mean   :2.681                      Mean   :0.09817   Mean   :0.1318
3rd Qu.:4.000                      3rd Qu.:0.00000   3rd Qu.:0.0000
Max.   :6.000                      Max.   :1.00000   Max.   :5.0000
IA_REGRESSION_OUT IA_REGRESSION_OUT_COUNT IA_REGRESSION_PATH_DURATION
Min.   :0.00000   Min.   :0.00000         Min.   :    0.0
1st Qu.:0.00000   1st Qu.:0.00000         1st Qu.:    0.0
Median :0.00000   Median :0.00000         Median :  282.0
Mean   :0.08147   Mean   :0.09185         Mean   :  595.6
3rd Qu.:0.00000   3rd Qu.:0.00000         3rd Qu.:  747.0
Max.   :1.00000   Max.   :7.00000         Max.   :10242.0
   KeyPress           rt            bio              critical
Min.   :4.000   Min.   :  533   Length:4431        Length:4431
1st Qu.:4.000   1st Qu.: 1332   Class :character   Class :character
Median :4.000   Median : 1890   Mode  :character   Mode  :character
Mean   :4.496   Mean   : 2467
3rd Qu.:5.000   3rd Qu.: 2910
Max.   :5.000   Max.   :15654
   gender             item_id            list          match
Length:4431        Min.   :  1.00   Min.   :14.00   Length:4431
Class :character   1st Qu.: 26.00   1st Qu.:15.00   Class :character
Mode  :character   Median : 51.00   Median :25.00   Mode  :character
```

```
                    Mean   : 64.16   Mean   :29.45
                    3rd Qu.: 78.50   3rd Qu.:35.00
                    Max.   :208.00   Max.   :45.00
  condition             name           name_vital_status     tense
 Length:4431        Length:4431        Length:4431        Length:4431
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character



     type             yes_press
 Length:4431        Min.   :4.000
 Class :character   1st Qu.:4.000
 Mode  :character   Median :4.000
                    Mean   :4.499
                    3rd Qu.:5.000
                    Max.   :5.000
```

Exercise

Take some time to explore the dataset.

- double click on the dataset name in the Environment pane to view it like a spreadsheet
- look at the names, can you figure out what they represent?

class types

- there are difference classes of data that R can read
    - the function `class()` takes as its argument an object or number

```r
df_lifetime$rt %>%
  class()
```

```
[1] "numeric"
```

Selecting a column

```r
# with column index
df_lifetime[2] %>% summary()
```

```
    trial
Min.   :  1.0
1st Qu.: 52.5
Median :104.0
Mean   :105.0
3rd Qu.:157.0
Max.   :208.0
```

```r
# with column name
df_lifetime[,"trial"] %>% summary()
```

```
    trial
Min.   :  1.0
1st Qu.: 52.5
Median :104.0
Mean   :105.0
3rd Qu.:157.0
Max.   :208.0
```

```r
# with data$column_name
df_lifetime$trial %>% summary()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0    52.5   104.0   105.0   157.0   208.0
```

```r
# with the tidyverse: select()
df_lifetime %>%
  select(trial) %>%
  summary()
```

```
    trial
Min.   :  1.0
1st Qu.: 52.5
```

```
Median :104.0
Mean   :105.0
3rd Qu.:157.0
Max.   :208.0
```

`character` class

- contain *strings*: collection of characters (i.e., text)
- there's no grouping in character variables
  - each value is considered 'unique' and assumed to not be repeated
- we usually aren't interested in character class variables
  - unless e.g., we have unique values per row (e.g., if a participant gave a free-text answer)
  - or perhaps we have stored some stimuli sentences
    - although this would arguably be better as a 'category', since there should be multiple trials across participants that contain the same sentences

`numeric` class

- variables with numeric values, usually some variable we'd want to compute summaries on, e.g., means
- sometimes we don't want numbers to be stored as numeric class, however
  - this is the case for our variables `yes_press` and `KeyPress` (with `4` or `5`)
- the same is true for our variable `item_id`, which ranges from 1:120
  - the numbers are just unique codes for our stimuli, the difference between `item 1` and `item 2` has nothing to do with the difference between the numbers `1` and `2`

`factor` class

- we typically want *grouping* variables to be `factor` class
  - factors contain ***categorical*** data
  - any number that could be replaced with some other label should be a factor
- region of interest (ROI) = 1:7
  - but we want to know how many observations per region, the number is not informative
  - ROI could alternatively be coded as, e.g., "adverb", "pronoun", "verb", "spillover"

`factor` class

- let’s change `df_lifetime$yes_press` to `factor`
  - using the `mutate()` verb from `dplyr`
  - and `as_factor()` from `forcats`

```r
# change yes_press to factor
df_lifetime %>%
  mutate(yes_press = as_factor(yes_press)) %>%
  summary()
```

```
      px                trial            eye            IA_DWELL_TIME 
 Length:4431        Min.   :  1.0   Length:4431        Min.   :   0.0  
 Class  :character  1st Qu.: 52.5   Class  :character  1st Qu.:   0.0  
 Mode   :character  Median :104.0   Mode   :character  Median : 301.0  
                    Mean   :105.0                      Mean   : 587.5  
                    3rd Qu.:157.0                      3rd Qu.: 765.5  
                    Max.   :208.0                      Max.   :8968.0  
 IA_FIRST_FIXATION_DURATION IA_FIRST_RUN_DWELL_TIME IA_FIXATION_COUNT
 Min.   :  0.0              Min.   :   0.0          Min.   : 0.000   
 1st Qu.:  0.0              1st Qu.:   0.0          1st Qu.: 0.000   
 Median :161.0              Median : 245.0          Median : 2.000   
 Mean   :139.4              Mean   : 507.9          Mean   : 2.714   
 3rd Qu.:202.5              3rd Qu.: 586.0          3rd Qu.: 4.000   
 Max.   :775.0              Max.   :8968.0          Max.   :35.000   
     IA_ID         IA_LABEL         IA_REGRESSION_IN  IA_REGRESSION_IN_COUNT
 Min.   :1.000   Length:4431        Min.   :0.00000   Min.   :0.0000        
 1st Qu.:1.000   Class  :character  1st Qu.:0.00000   1st Qu.:0.0000        
 Median :2.000   Mode   :character  Median :0.00000   Median :0.0000        
 Mean   :2.681                      Mean   :0.09817   Mean   :0.1318        
 3rd Qu.:4.000                      3rd Qu.:0.00000   3rd Qu.:0.0000        
 Max.   :6.000                      Max.   :1.00000   Max.   :5.0000        
 IA_REGRESSION_OUT IA_REGRESSION_OUT_COUNT IA_REGRESSION_PATH_DURATION
 Min.   :0.00000   Min.   :0.00000         Min.   :    0.0             
 1st Qu.:0.00000   1st Qu.:0.00000         1st Qu.:    0.0             
 Median :0.00000   Median :0.00000         Median :  282.0             
 Mean   :0.08147   Mean   :0.09185         Mean   :  595.6             
 3rd Qu.:0.00000   3rd Qu.:0.00000         3rd Qu.:  747.0             
 Max.   :1.00000   Max.   :7.00000         Max.   :10242.0             
    KeyPress           rt            bio              critical        
 Min.   :4.000   Min.   :  533   Length:4431        Length:4431       
```

```
1st Qu.:4.000   1st Qu.: 1332   Class :character   Class :character
Median :4.000   Median : 1890   Mode  :character   Mode  :character
Mean   :4.496   Mean   : 2467
3rd Qu.:5.000   3rd Qu.: 2910
Max.   :5.000   Max.   :15654
   gender             item_id            list          match
Length:4431        Min.   :  1.00   Min.   :14.00   Length:4431
Class :character   1st Qu.: 26.00   1st Qu.:15.00   Class :character
Mode  :character   Median : 51.00   Median :25.00   Mode  :character
                   Mean   : 64.16   Mean   :29.45
                   3rd Qu.: 78.50   3rd Qu.:35.00
                   Max.   :208.00   Max.   :45.00
 condition             name           name_vital_status     tense
Length:4431        Length:4431        Length:4431        Length:4431
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character


    type           yes_press
Length:4431        4:2218
Class :character   5:2213
Mode  :character
```

multiple arguments in a verb

- we can also change multiple columns at once:

```
# change ROI & label to factor
df_lifetime %>%
  mutate(KeyPress = as_factor(KeyPress),
         item_id = as_factor(item_id)) %>%
  summary()
```

```
     px                trial           eye             IA_DWELL_TIME
Length:4431        Min.   :  1.0   Length:4431        Min.   :   0.0
Class :character   1st Qu.: 52.5   Class :character   1st Qu.:   0.0
Mode  :character   Median :104.0   Mode  :character   Median : 301.0
```

```
                   Mean   :105.0                      Mean   : 587.5
                   3rd Qu.:157.0                      3rd Qu.: 765.5
                   Max.   :208.0                      Max.   :8968.0

IA_FIRST_FIXATION_DURATION IA_FIRST_RUN_DWELL_TIME IA_FIXATION_COUNT
Min.   :  0.0              Min.   :   0.0          Min.   : 0.000
1st Qu.:  0.0              1st Qu.:   0.0          1st Qu.: 0.000
Median :161.0              Median : 245.0          Median : 2.000
Mean   :139.4              Mean   : 507.9          Mean   : 2.714
3rd Qu.:202.5              3rd Qu.: 586.0          3rd Qu.: 4.000
Max.   :775.0              Max.   :8968.0          Max.   :35.000

    IA_ID         IA_LABEL         IA_REGRESSION_IN  IA_REGRESSION_IN_COUNT
Min.   :1.000   Length:4431        Min.   :0.00000   Min.   :0.0000
1st Qu.:1.000   Class :character   1st Qu.:0.00000   1st Qu.:0.0000
Median :2.000   Mode  :character   Median :0.00000   Median :0.0000
Mean   :2.681                      Mean   :0.09817   Mean   :0.1318
3rd Qu.:4.000                      3rd Qu.:0.00000   3rd Qu.:0.0000
Max.   :6.000                      Max.   :1.00000   Max.   :5.0000

IA_REGRESSION_OUT IA_REGRESSION_OUT_COUNT IA_REGRESSION_PATH_DURATION KeyPress
Min.   :0.00000   Min.   :0.00000         Min.   :    0.0             4:2234
1st Qu.:0.00000   1st Qu.:0.00000         1st Qu.:    0.0             5:2197
Median :0.00000   Median :0.00000         Median :  282.0
Mean   :0.08147   Mean   :0.09185         Mean   :  595.6
3rd Qu.:0.00000   3rd Qu.:0.00000         3rd Qu.:  747.0
Max.   :1.00000   Max.   :7.00000         Max.   :10242.0

      rt            bio              critical            gender
Min.   :  533   Length:4431        Length:4431        Length:4431
1st Qu.: 1332   Class :character   Class :character   Class :character
Median : 1890   Mode  :character   Mode  :character   Mode  :character
Mean   : 2467
3rd Qu.: 2910
Max.   :15654

   item_id          list          match            condition
2      :  48   Min.   :14.00   Length:4431        Length:4431
7      :  48   1st Qu.:15.00   Class :character   Class :character
8      :  48   Median :25.00   Mode  :character   Mode  :character
9      :  48   Mean   :29.45
10     :  48   3rd Qu.:35.00
12     :  48   Max.   :45.00
```

```
(Other):4143
    name           name_vital_status     tense              type
Length:4431        Length:4431        Length:4431        Length:4431
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
  yes_press
Min.   :4.000
1st Qu.:4.000
Median :4.000
Mean   :4.499
3rd Qu.:5.000
Max.   :5.000
```

Pop quiz

1. Which class *should* the following variables be (`numeric`, `factor`, or `character`)?:
    - participant ID
    - trial number
    - first-pass reading time
    - regression path duration
    - regressions in
    - context sentence
    - lifetime
    - tense
    - celebrity name
2. change them to these class types, and print a summary
3. save and render the document

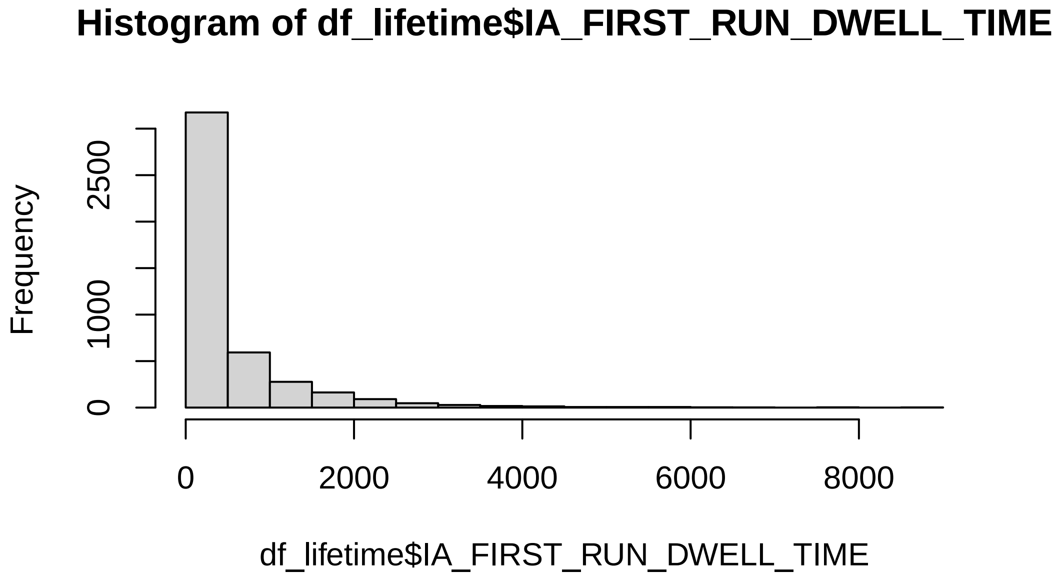
Plot the data

- at this stage we want to explore the data
    - distribution
        - peaks, spread

- boundaries

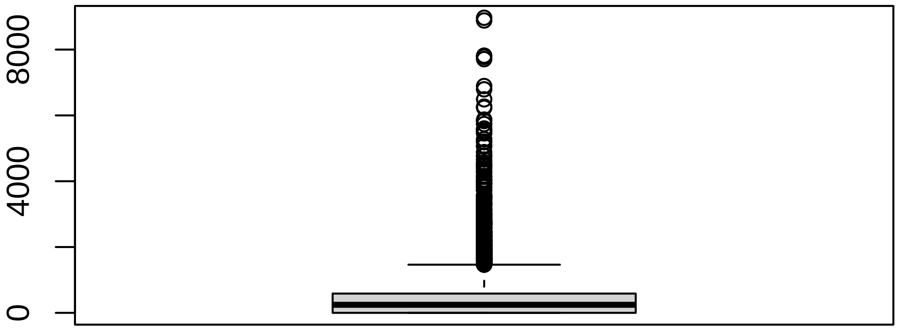
Histogram

```
hist(df_lifetime$IA_FIRST_RUN_DWELL_TIME)
```



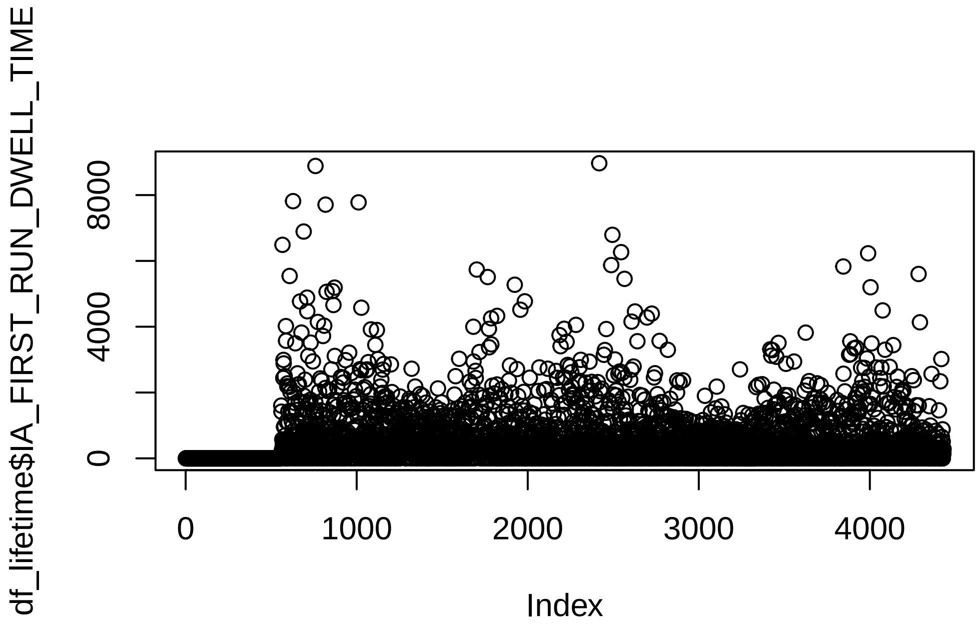
Boxplot

```
boxplot(df_lifetime$IA_FIRST_RUN_DWELL_TIME)
```



Scatterplot

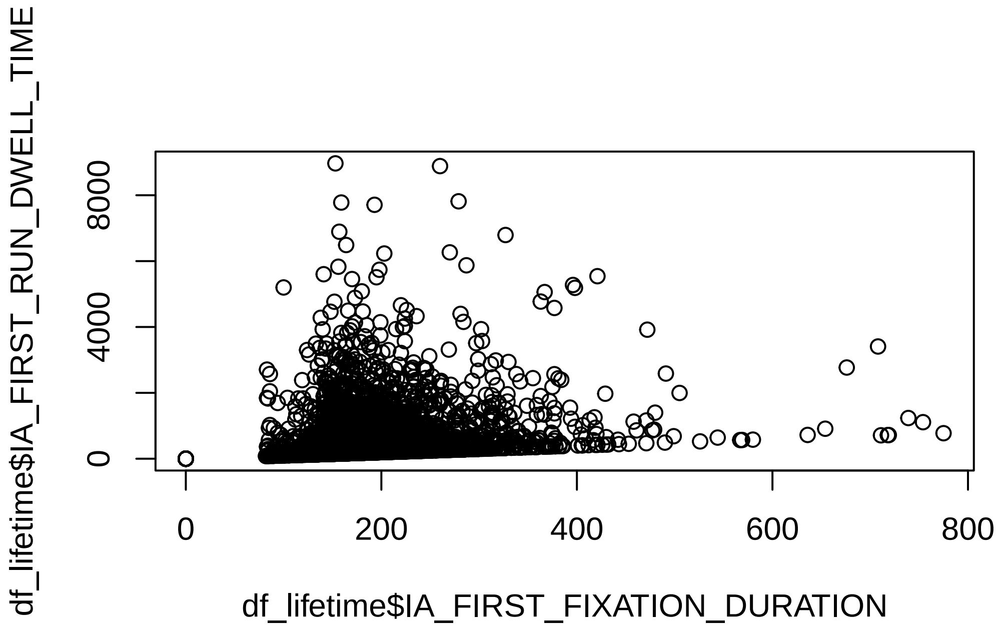
```
plot(df_lifetime$IA_FIRST_RUN_DWELL_TIME)
```



Plotting two variables

Scatterplot

```
plot(df_lifetime$IA_FIRST_FIXATION_DURATION, df_lifetime$IA_FIRST_RUN_DWELL_TIME)
```



Exercise

In your Quarto document:

1. create a heading 'Data exploration'

- briefly describe the data

2. For each of our depenent variables:

- create a subheading
- calculate the mean and standard deviation of the variable (`mean()`, `sd()`) + create a boxplot of the variable

3. Render the document often to make sure it runs
4. Upload the source file (day1-nachname_vorname.qmd) to Moodle
5. download the source file below yours in the list to the same folder, and try to run it

- does it run?

print options

- each code chunk can have different print options:
    - `eval = FALSE`: do not evaluate this chunk
    - `include = FALSE` evaluate this chunk but don't show it or its results
    - `echo = FALSE` print this chunk code
    - `message = FALSE`/`warning = false` don't print warnings or messages
    - `error = TRUE` continue rendering document even if there's an error
        - do not use `error = TRUE` for final versions! You want to make sure things work as they should

~~~
```{r, eval = T, echo = T, results = "asis", warning}
code here
```
~~~

or

~~~
```{r}
#| eval: false
code here
```
~~~

Session Info

```
sessionInfo()
```

```
R version 4.2.3 (2023-03-15)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Ventura 13.2.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] rbbt_0.0.0.9000 beepr_1.3       lubridate_1.9.2 forcats_1.0.0
 [5] stringr_1.5.0   dplyr_1.1.1     purrr_1.0.1     readr_2.1.4
 [9] tidyr_1.3.0     tibble_3.2.1    ggplot2_3.4.2   tidyverse_2.0.0
[13] here_1.0.1

loaded via a namespace (and not attached):
 [1] pillar_1.9.0     compiler_4.2.3   tools_4.2.3      bit_4.0.5
 [5] digest_0.6.31    timechange_0.2.0 jsonlite_1.8.4   evaluate_0.20
 [9] lifecycle_1.0.3  gtable_0.3.3     pkgconfig_2.0.3  rlang_1.1.0
[13] cli_3.6.1        rstudioapi_0.14  parallel_4.2.3   curl_5.0.0
[17] yaml_2.3.7       xfun_0.38        fastmap_1.1.1    httr_1.4.5
[21] withr_2.5.0      knitr_1.42       fs_1.6.1         generics_0.1.3
[25] vctrs_0.6.1      hms_1.1.3        bit64_4.0.5      rprojroot_2.0.3
[29] grid_4.2.3       tidyselect_1.2.0 glue_1.6.2       R6_2.5.1
[33] fansi_1.0.4      vroom_1.6.1      rmarkdown_2.21   tzdb_0.3.0
[37] magrittr_2.0.3   scales_1.2.1     htmltools_0.5.5  colorspace_2.1-0
[41] utf8_1.2.3       stringi_1.7.12   munsell_0.5.0    crayon_1.5.2
[45] audio_0.1-10
```

References

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Mittwoch, A. (2008a). Tenses for the living and the dead. *Theoretical and Crosslinguistic Approaches to the Semantics of Aspect*, *110*, 167.

Mittwoch, A. (2008b). The English Resultative perfect and its relationship to the Experiential perfect and the simple past tense. *Linguistics and Philosophy*, *31*(3), 323–351. https://doi.org/10.1007/s10988-008-9037-y