**A**

## Metadata

| id | machine_name | date | ... | condition | sex | ... | p |
|---|---|---|---|---|---|---|---|
| xxx...xx\|x | machine_001 | 2016-09-01 | ... | A | M | ... | $p_1$ |
| xxx...xx\|y | machine_001 | 2016-09-01 | ... | B | M | ... | $p_2$ |
| xxx...xx\|z | machine_002 | 2016-09-03 | ... | A | F | ... | $p_3$ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ | ⋱ | ⋮ |
| n | $machine_n$ | $date_n$ | ... | $condition_n$ | $sex_n$ | ... | $p_n$ |

Platform fields | Experiment fields

+

## Data

| id | t | activity | ... | q |
|---|---|---|---|---|
| xxx...xx\|x | 1 | 1 | ⋱ | $q_{1,1}$ |
| xxx...xx\|x | 2 | 0 | ⋱ | $q_{1,2}$ |
| xxx...xx\|x | 3 | 0 | ⋱ | $q_{1,3}$ |
| xxx...xx\|x | ⋮ | ⋮ | ⋱ | ⋮ |
| xxx...xx\|y | ⋮ | ⋮ | ⋱ | ⋮ |
| xxx...xx\|z | 1 | 0 | ⋱ | $q_{3,1}$ |
| xxx...xx\|z | 2 | 2 | ⋱ | $q_{3,2}$ |
| xxx...xx\|z | 3 | 0 | ⋱ | $q_{3,3}$ |
| xxx...xx\|z | ⋮ | ⋮ | ⋱ | ⋮ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
| n | ⋮ | ⋮ | ⋱ | $q_{n,k_n}$ |

**B**

| | Metadata | Data |
|---|---|---|
| Select | **dt[CRITERIA, meta = TRUE]**<br><br>*# to subet the metadata only for males*<br>*> male_meta <- dt[sex == "M",*<br>*meta = TRUE]* | **dt[CRITERIA]**<br><br>*# to keep only data > 5s*<br>*> late_dt <- dt[t > 5]*<br><br>Note: metadata is updated when selection removes all data from one id. |
| Alter, create & delete (meta)variables | **dt[, X := value, meta = TRUE]**<br>*# to create a metavariable set to "wt"*<br>*> dt[, genotype := "wt", meta = TRUE]*<br>*# delete*<br>*> dt[, sex := NULL, meta = TRUE]* | **dt[, Y := value]**<br>*# to create t_2 (t - 1)*<br>*> dt[, t_2 := t - 1]*<br>*# to delete t*<br>*> dt[, t := NULL]*<br>Note: update data in place. No copy of dt in memory. |
| Expand metavariables as variables | **dt[xmv(X)]**<br><br>*# to select data with sex*<br>*> dt <- dt[xmv(sex) == "M"]*<br><br>*# to copy a metavariable as a variable*<br>*> dt[, s := xmv(sex)]* | |
| Aggregate & summary | **dt[, OPERATION, by = id]**<br><br># to compute mean activity, per individual<br>*> dt <- dt[,.(*<br>*mean_act = mean(activity)*<br>*), by = id]*<br><br>*# to count reads per id*<br>*> dt[, .N, by = id]* | **OPERATION** |
| Join data & metadata | **rejoin(dt)**<br><br>*# to reunite data and metadata*<br>*> full_table <- rejoin(dt)*<br><br>Note: used mostly after aggregation or preprocessing | **REJOIN** |