

# Introduction to data.table

20190927

Jan Gorecki

Why R?, Warsaw, Poland

# workshop agenda

- hands-on exercises: type code rather than just watch and listen
- syntax similarities and differences: `data.table` and `data.frame`
- using Arun Srinivasan workshops materials from useR!2017  
[arunsrinivasan/user2017-data.table-tutorial](https://arunsrinivasan.github.io/user2017-data.table-tutorial/)
- extras:
  - chaining of `data.table` queries
  - reference semantics
  - subset of data: `.SD`
  - R function argument matching
  - LEFT / RIGHT outer join

# data

```
DF1 = data.frame(id = c(rep(1L,4L), 2L, 2L, 1L, 2L, 1L),  
                 code = c("c","b","c","c","a","a","b","a","c"),  
                 valA = 1:9, valB = 10:18,  
                 stringsAsFactors = FALSE)
```

```
DF2 = data.frame(id = c(3L, 1L, 1L, 2L, 3L),  
                 code = c("b","a","c","c","d"),  
                 mul = 5:1, stringsAsFactors = FALSE)
```

```
library(data.table)  
DT1 = as.data.table(DF1)  
DT2 = as.data.table(DF2)
```

# exercises

1-3. subset rows and columns

4-5. update/add columns

6-7. aggregate

8-9. window aggregate function

10-11. join

12-13. update/add columns on join

14. aggregate on join

# live coding session

From now on we will move to hands-on session.

Script is available at:

<https://jangorecki.gitlab.io/r-talks-why2019/intro.R>