Introduction to data.table

Bay Area R Users Group

12 May 2015

Matt Dowle

In one slide

```
DT[type=="books", sum(sales)]
```

filter and do by group

```
DT[type=="books", sum(sales), by=country]
```

Why inside [...] ?

- **R's lazy evaluation enables optimizations of i, j and by together**
- **Fast development**
- **Flexible – do anything in j**
- **Fast to read and maintain in production**

PRC

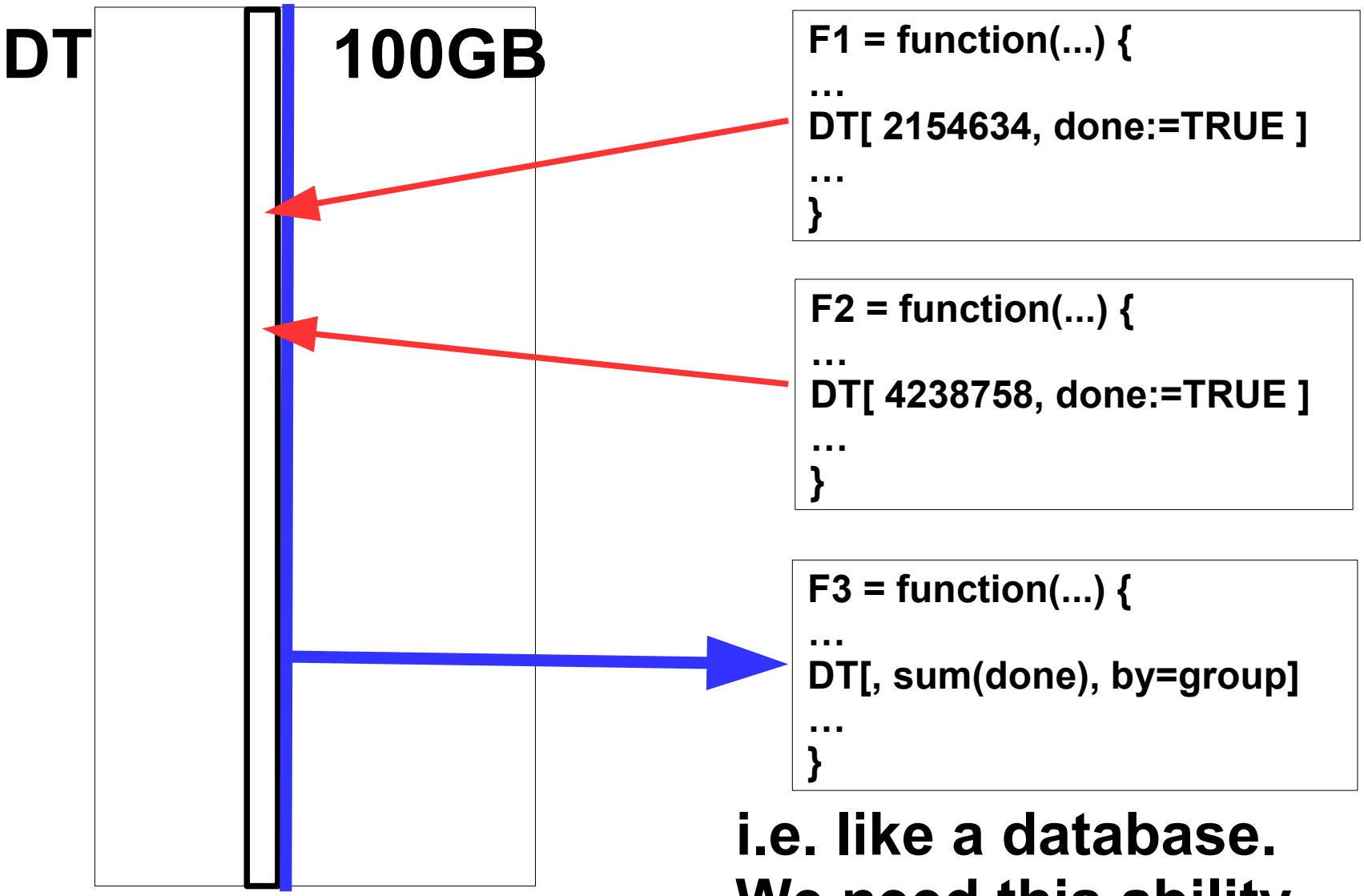
id	date	price
SBRY	20080501	380.50
SBRY	20080502	391.50
SBRY	20080506	389.00
VOD	20080501	159.30
VOD	20080502	163.30
VOD	20080506	160.80

setkey(PRC, id, date)

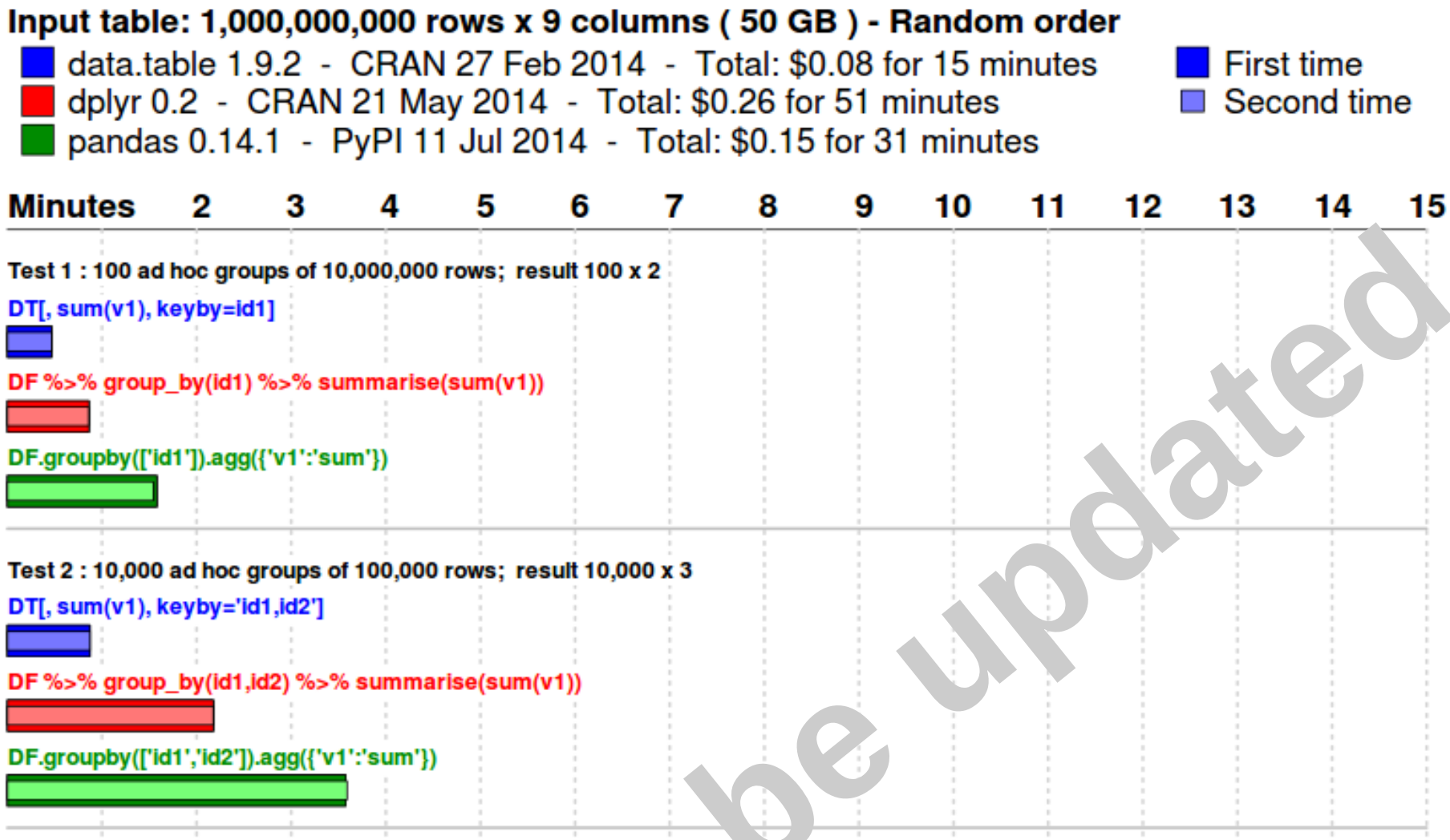
	Result
1. PRC[.(“SBRY”)]	rows 1:3
2. PRC[.(“SBRY”,20080502),price]	391.50
3. PRC[.(“SBRY”,20080505),price]	NA
4. PRC[.(“SBRY”,20080505),price,roll=TRUE]	391.50
5. PRC[.(“SBRY”,20080601),price,roll=TRUE]	389.00
6. PRC[.(“SBRY”,20080601),price,roll=TRUE,rollends=FALSE]	NA
7. PRC[.(“SBRY”,20080601),price,roll=20]	NA
8. PRC[.(“SBRY”,20080601),price,roll=40]	389.00

Client server demo

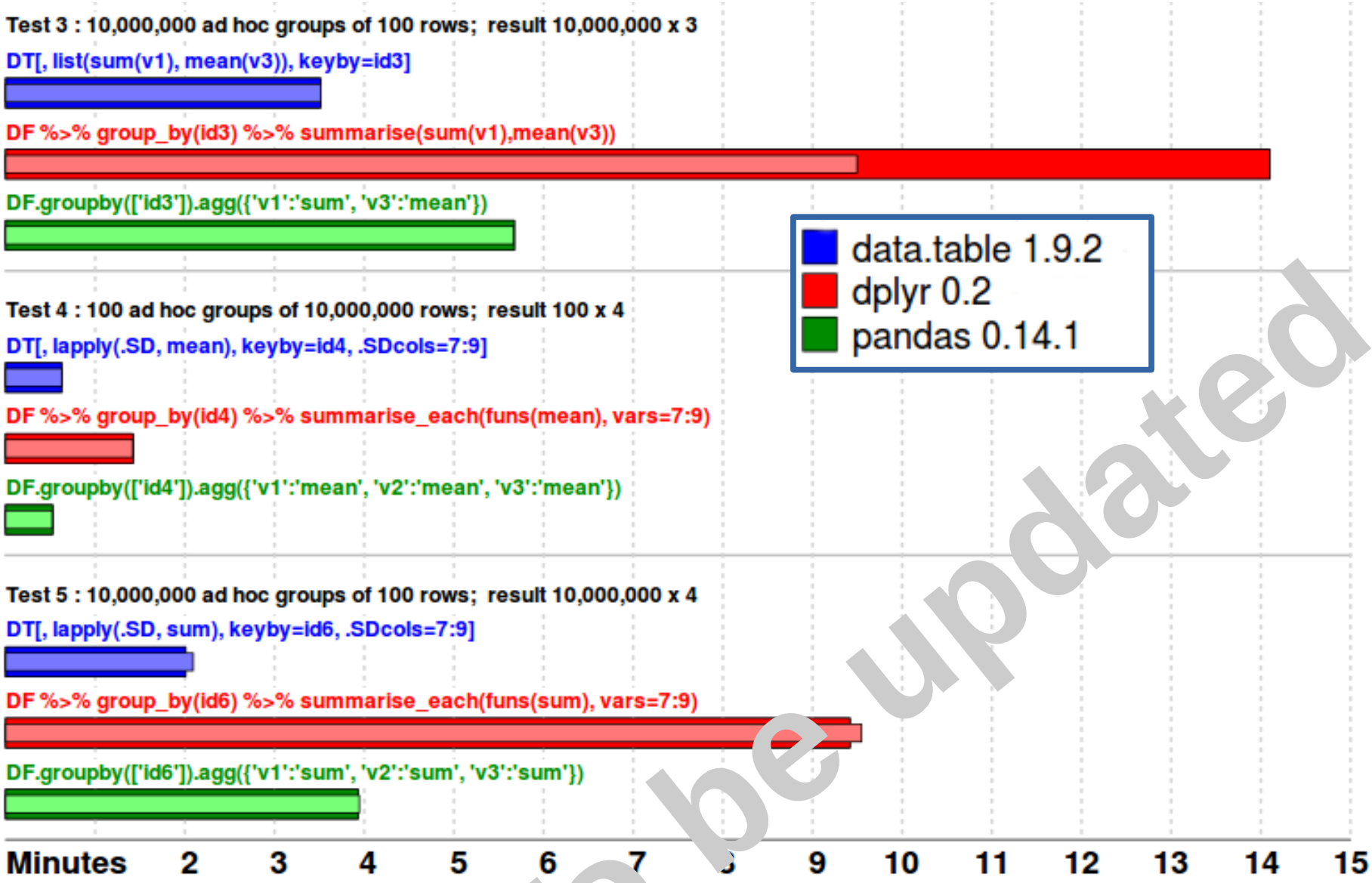
:= by reference



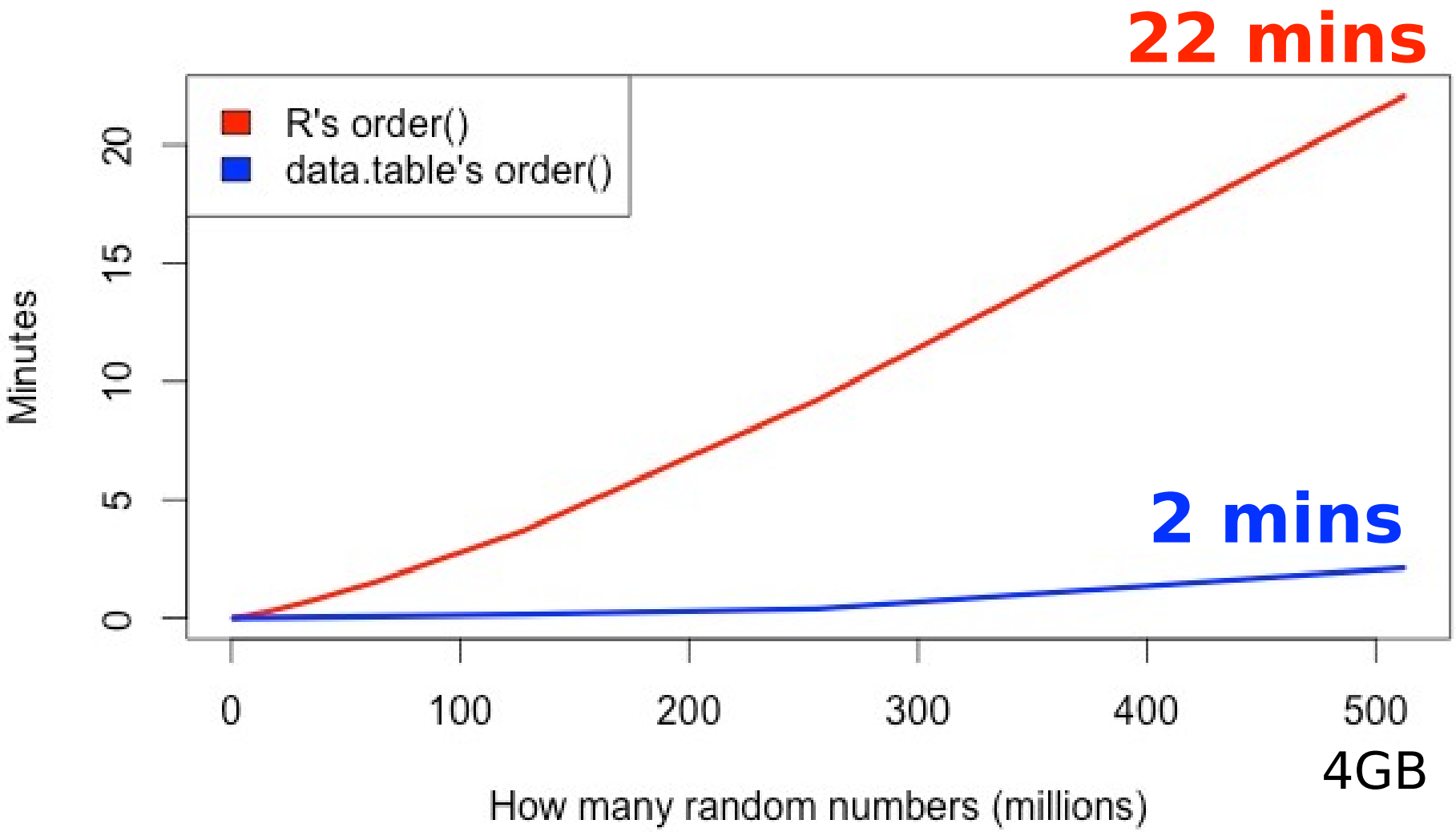
i.e. like a database.
We need this ability.



To be updated



To be updated



MacBook Pro 2.8GHz Intel Core i7 16GB
R 3.1.3 data.table 1.9.4

H2O.ai
Machine Intelligence

Non-speed reasons

Two S.O. questions

data.table wiki

?data.table examples

Thank you

https://github.com/Rdatatable/data.table/wiki

This presentation was recorded :
http://livestream.com/h2oai/events/4046257