ffbase, statistical functions for large datasets

Jan Wijffels and Edwin de Jonge

UseR2013 July 10 2013

Large data

Data files tend to become bigger and bigger.

ff

- Speedy access to large files
- Up to 10^9

ff

`ff` is nice, but:

- you have to rewrite most code into ff code
- example. . . .
- mainly efficient storage of numeric types
- no statistical functions

Enter: ffbase

`ffbase` tries to add # What is difficult?

- value vs reference

```
x <- ff(0, length=10^7)
print(x[1])
f <- function(y){
   y[1] <- 1
}
f(x)
print(x[1])
```

Idiosyncratic R

- `with`, `within`
- `transform`
- `subset`

filtering

- `within`

Under the hood

chunked processing