MIE237

Neil Montgomery

2016-04-05

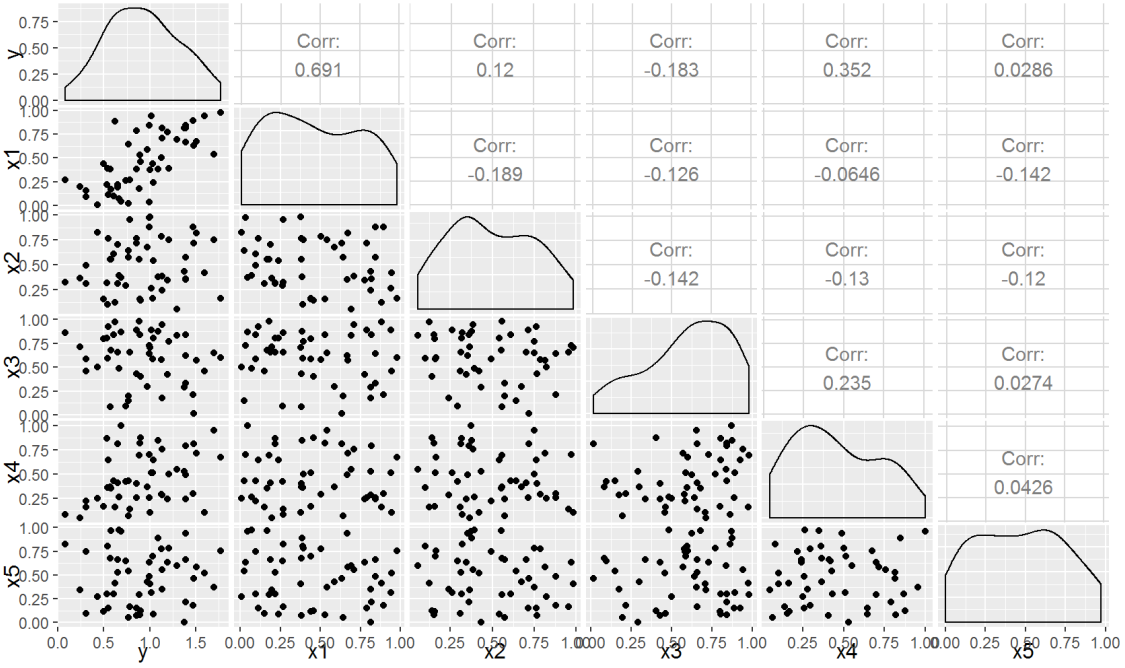
Example 1 - unintentially challenging!

```
set.seed(12)
n <- 50
x1 <- runif(n)
x2 <- runif(n)
x3 <- runif(n)
x4 <- runif(n)
x5 <- runif(n)
y <- x1 + 0.4*x2 + 0.6*x4 + rnorm(n, 0, 0.2)
f1 <- data.frame(y,x1,x2,x3,x4,x5)
```

But according to the forward strategy, all variables get included.

It turns out this happened by sheer chance (about 3 out of 1000 chance).

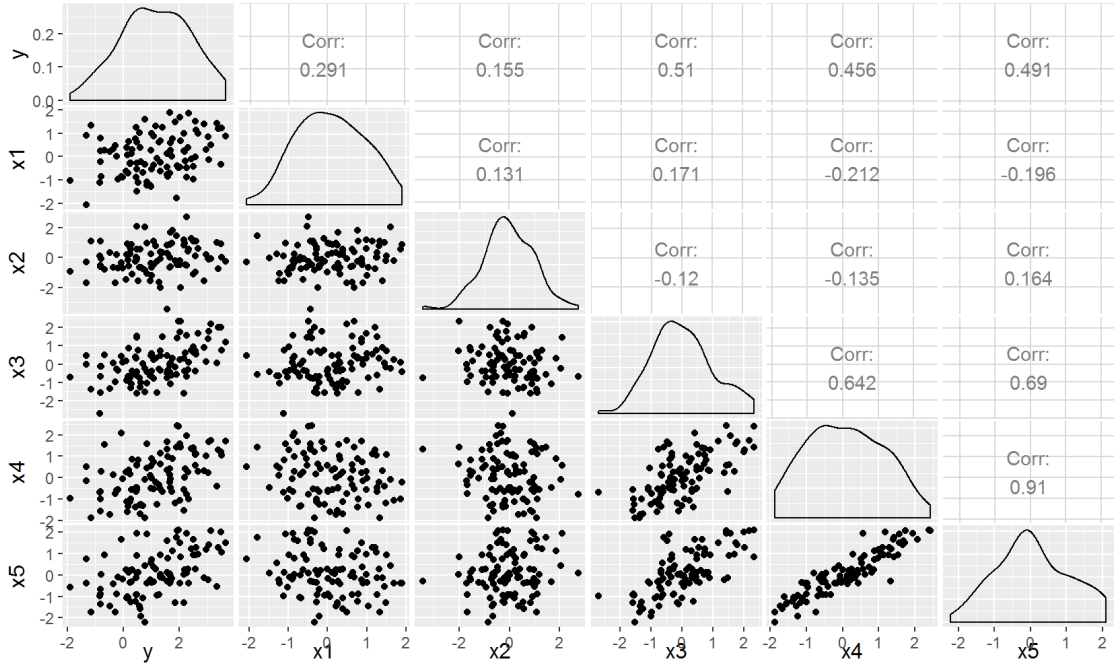
Otherwise example 1 was straightforward



Example 2

```
library(rio)
f2 <- import("problem.xlsx")
```

Ex 2 correlations



True model Ex 2

$$y = 1 + 0.5x_1 + 0.1x_2 + 0.3x_3 + 0.4x_4 + 0.5x_5 + \varepsilon$$

The x variables were created from a 5^d normal distribution withsome correlations put in.

The model fitting procedure doesn't end up with the ``truth'' in this case.