Generating missing values with `ampute`

And why on earth you would do that

Rianne Schouten

1. PhD Candidate, Utrecht University
2. Developer Data & Analytics, Samen Veilig Midden-Nederland

September 1, 2018

Why on earth would you ampute?

For simulation purposes:

1. Generate complete data (or use real dataset)
2. Generate missing values in complete data
3. Apply missing data method
4. Perform analysis and compare with complete data

But also for:

- Planned missing data survey designs
- Investigating measurement errors
- Reproducing your missing data situation

Missing data in your dataset

```
head(inc_data)
```

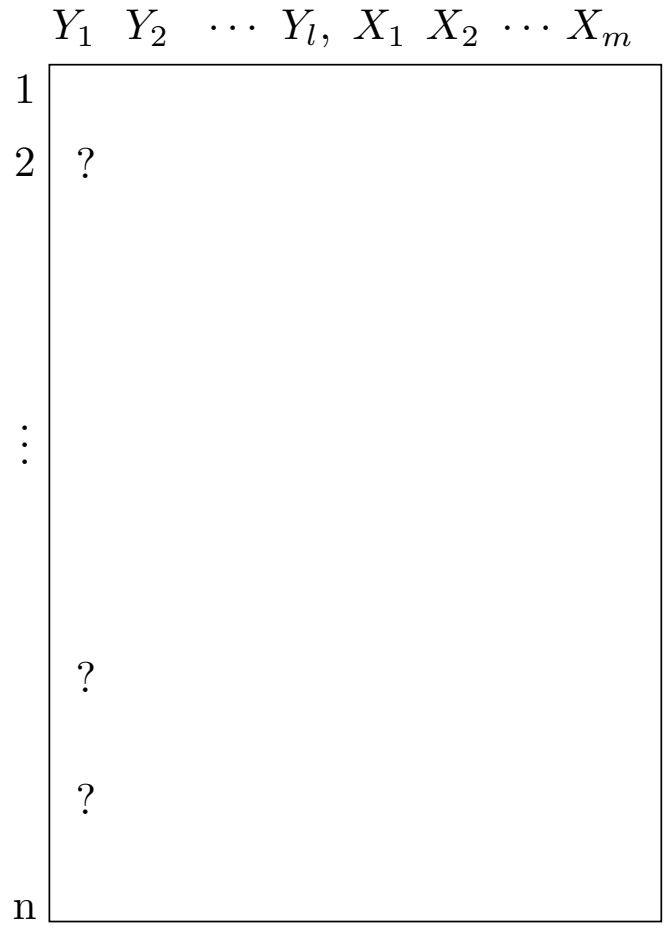
```
##         Income WorkingYears          Age
## 1 -0.08877292    1.5343721  1.33739681
## 2           NA           NA -0.41593616
## 3 -1.75818833           NA  0.06295286
## 4           NA           NA  1.73468904
## 5 -0.38850735           NA -1.22025110
## 6 -1.81223387    0.0950749  0.44283715
```

```
require(mice)
md.pattern(inc_data)
```

```
##     Age Income WorkingYears
## 206   1      1            1    0
## 412   1      1            0    1
## 382   1      0            0    2
##       0    382          794 1176
```

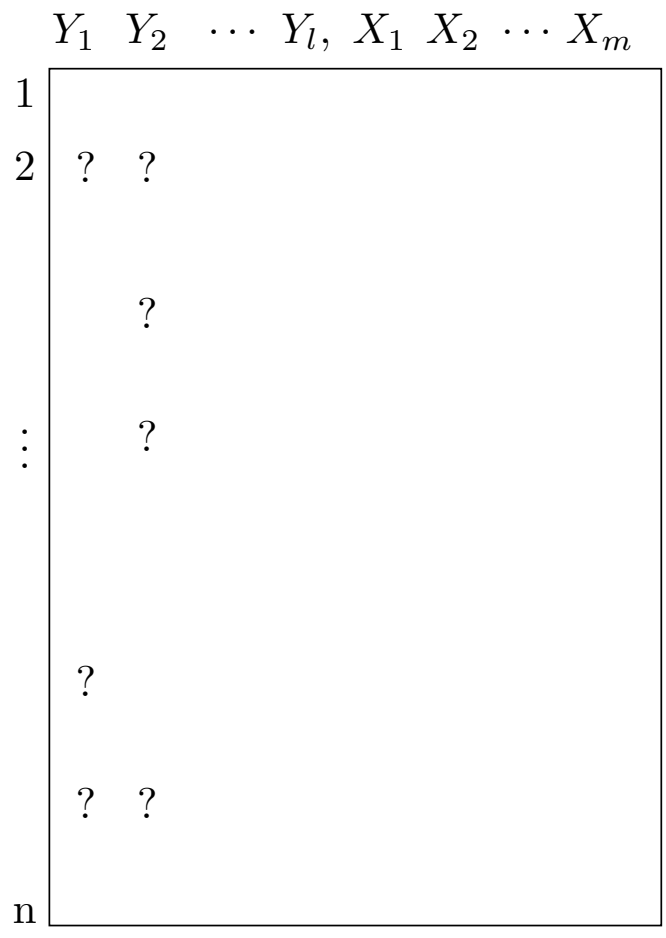
Generation of missing values

1. Y_1



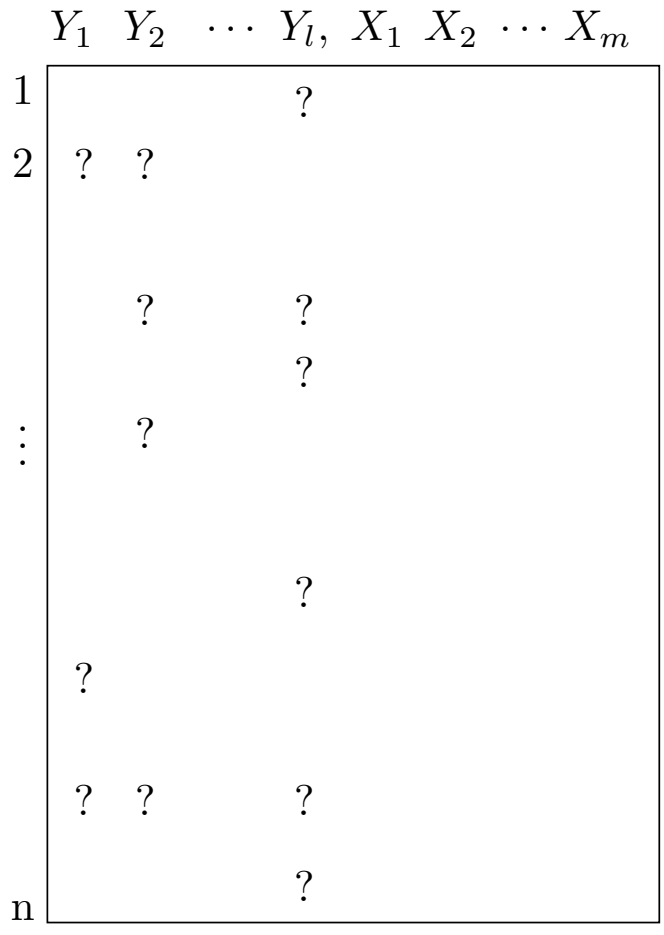
Generation of missing values

1. Y_1
2. Y_2

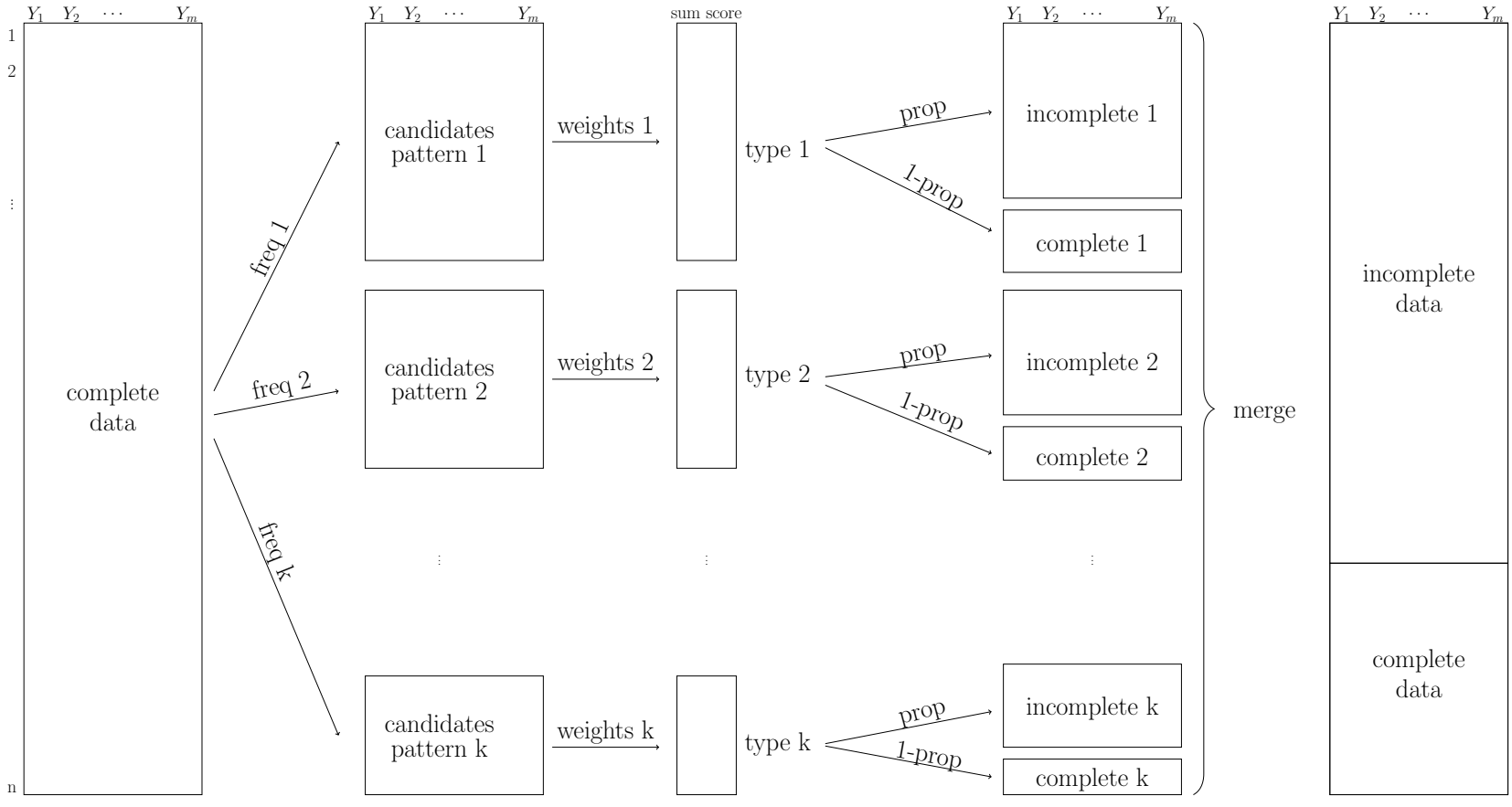


Generation of missing values

1. Y_1
2. Y_2
3. ...
4. Y_l



Multivariate amputation with `ampute`



Multivariate Amputation with `ampute`

Explanation of the method:

Rianne Margaretha Schouten, Peter Lugtig & Gerko Vink (2018) Generating missing values for simulation purposes: a multivariate amputation procedure, Journal of Statistical Computation and Simulation, 88:15, 2909-2930, DOI: 10.1080/00949655.2018.1491577

Multivariate amputation with ampute

```
amp <- ampute(data,
              patterns =  matrix(c(1, 0, 1,
                                   1, 0, 0,
                                   0, 0, 1),
                nrow = 3, byrow = TRUE),
              freq = c(0.6, 0.2, 0.2),
              prop = 0.5,
              mech = "MAR")
md.pattern(amp$amp)
```

```
##     Income Age WorkingYears
## 501      1   1             1   0
## 300      1   1             0   1
##  99      0   1             0   2
## 100      1   0             0   2
##         99 100           499 698
```

Multivariate amputation with ampute

```
require(mice)
?ampute
```

```
ampute(data, prop = 0.5, patterns = NULL, freq = NULL,
mech = "MAR", weights = NULL, cont = TRUE, type = NULL,
odds = NULL, bycases = TRUE, run = TRUE)
```

Explanation of all the arguments in vignette:

https://rianneschouten.github.io/mice_ampute/vignette/ampute.html

Missing data mechanisms

Missing Completely At Random (MCAR):

Missingness is not related to any variable

$\Pr(\text{Income} = \text{missing}) = 0.5$

Missing At Random (MAR):

Missingness is related to an observed variable

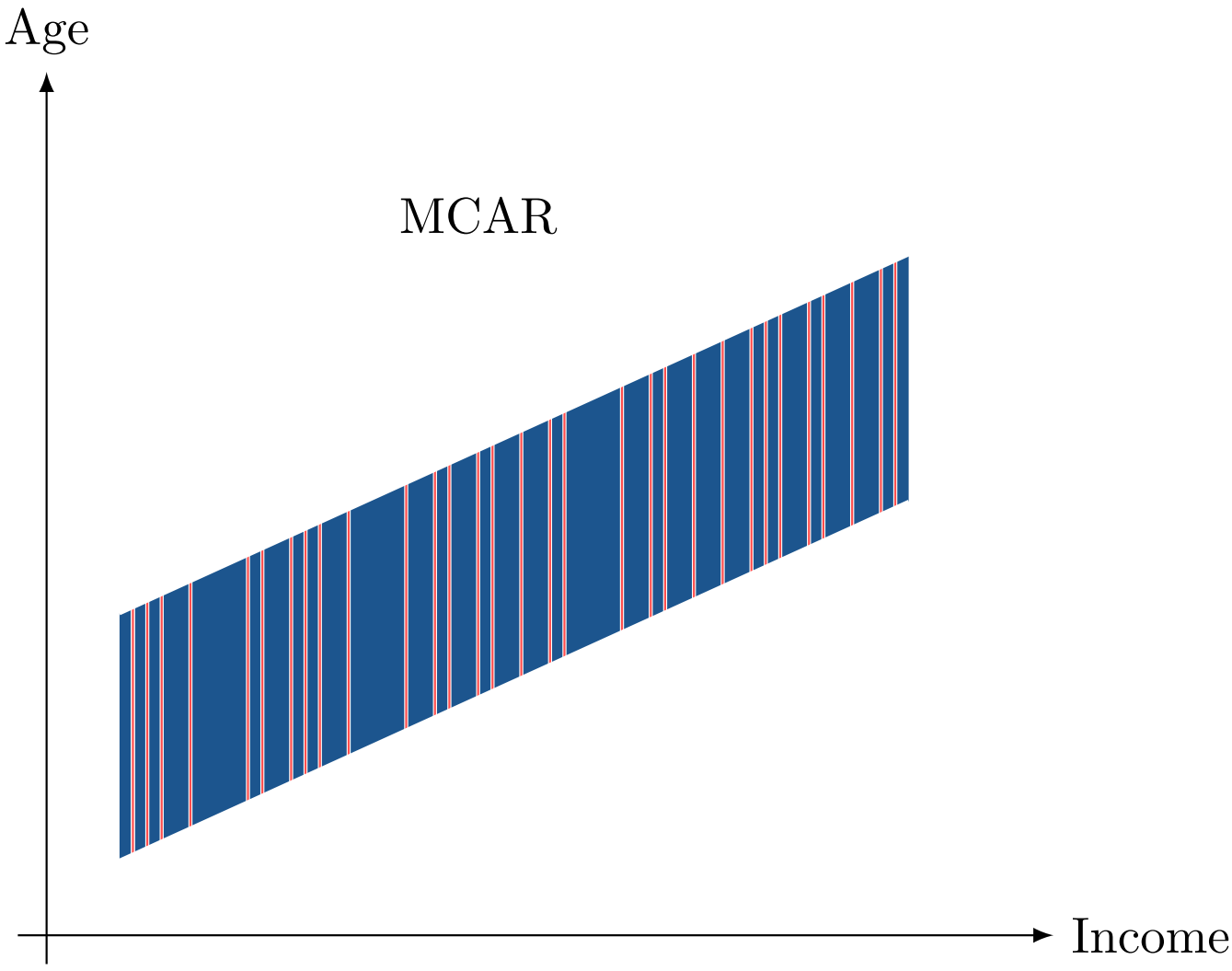
$\Pr(\text{Income} = \text{missing}) = \text{Age}$

Missing Not At Random (MNAR):

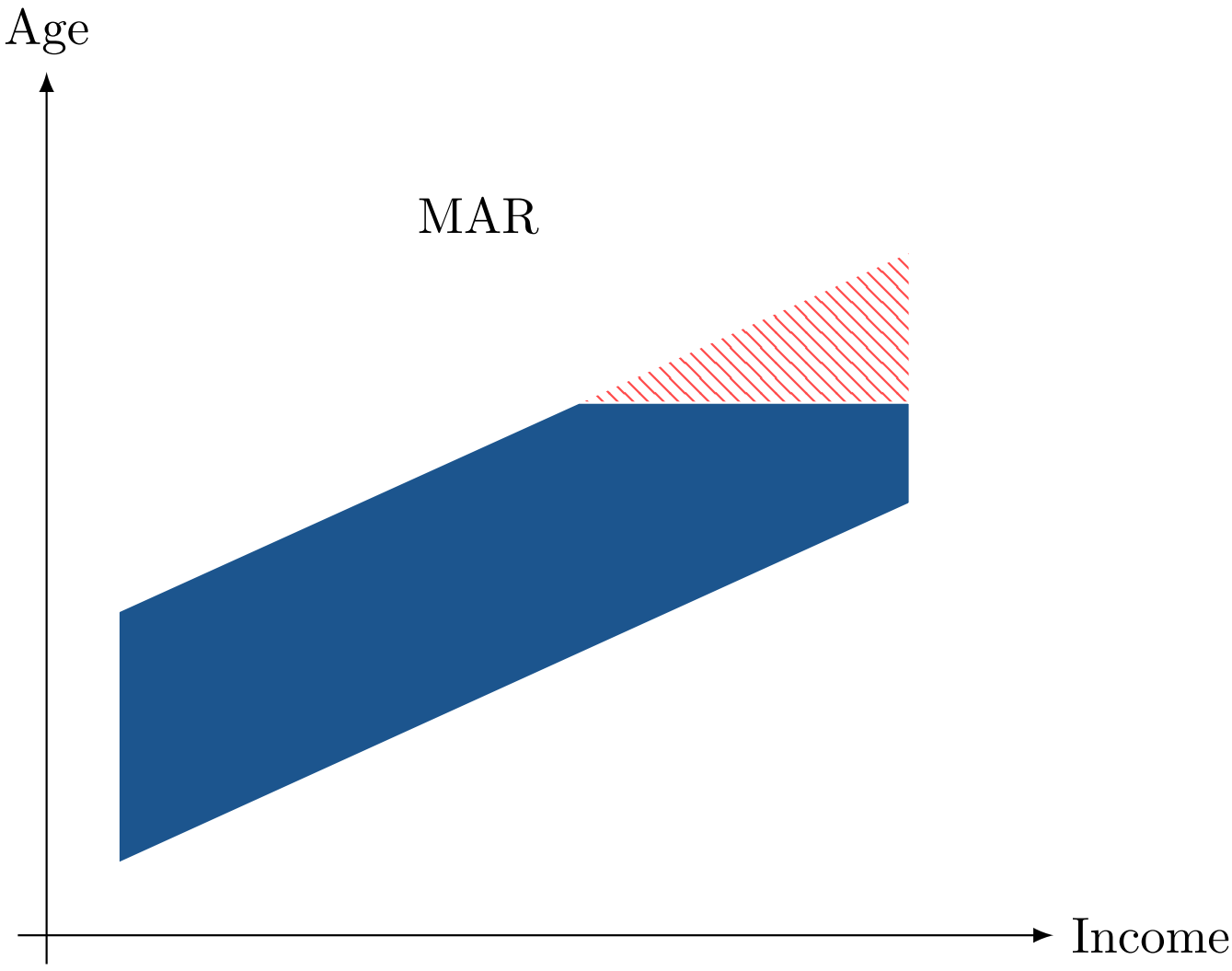
Missingness is related to the missingness itself or to an unobserved variable

$\Pr(\text{Income} = \text{missing}) = \text{Income}$

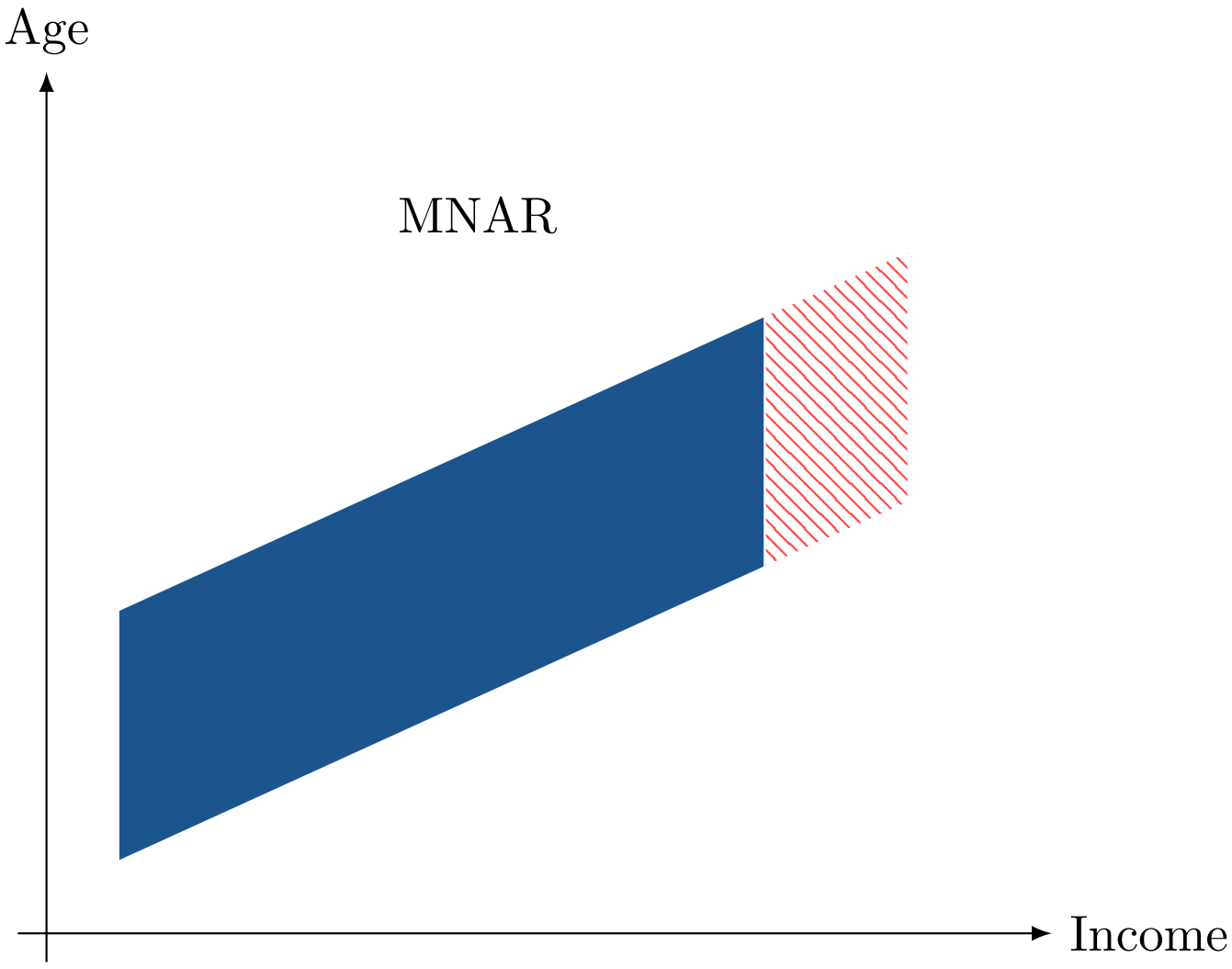
Missing data mechanisms: Missing Completely At Random



Missing data mechanisms: Missing At Random



Missing data mechanisms: Missing Not At Random



Multivariate amputation with ampute

```
require(mice)
?ampute
```

```
ampute(data, prop = 0.5, patterns = NULL, freq = NULL,
mech = "MAR", weights = NULL, cont = TRUE, type = NULL,
odds = NULL, bycases = TRUE, run = TRUE)
```

Explanation of all the arguments in vignette:

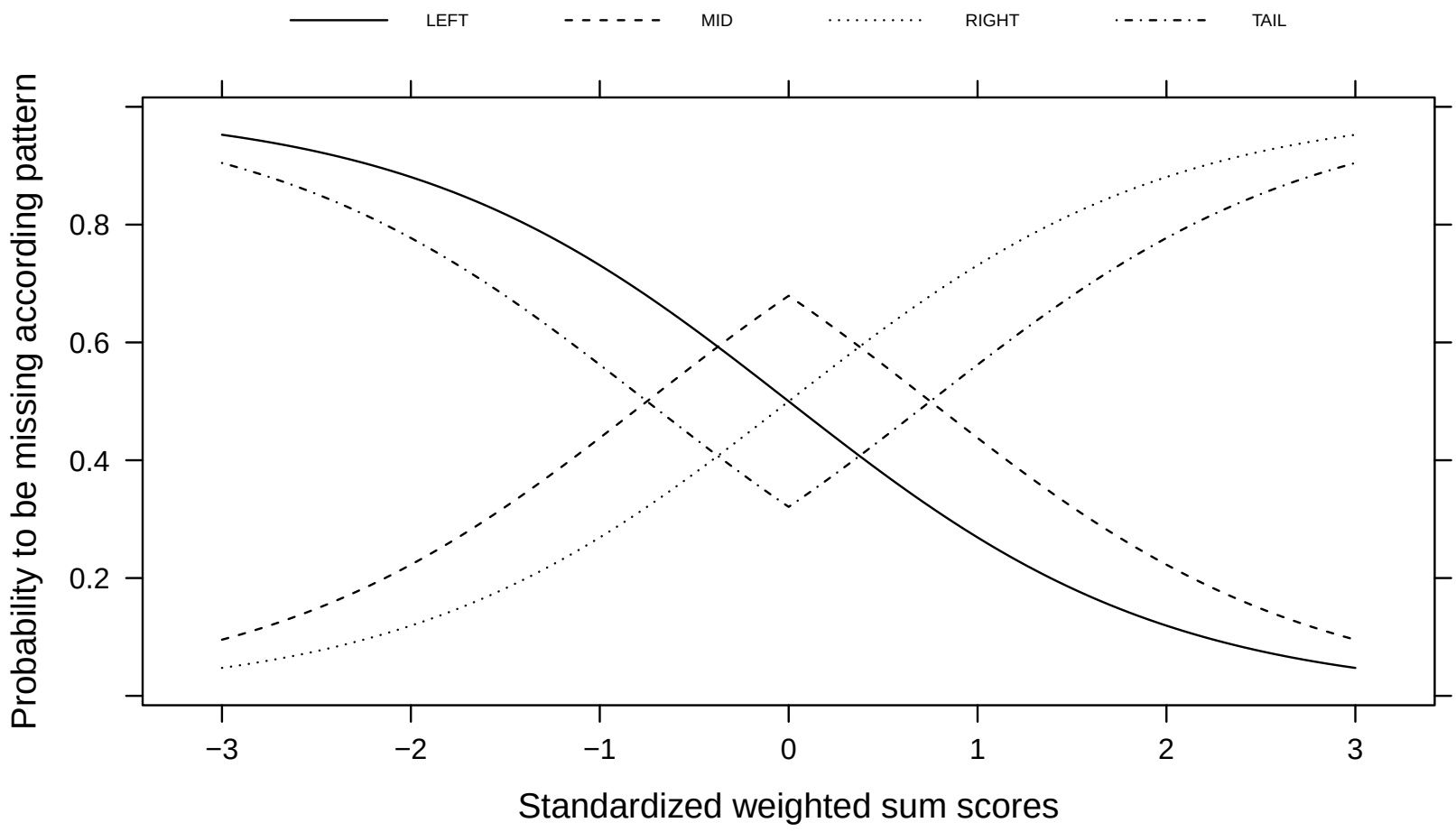
https://rianneschouten.github.io/mice_ampute/vignette/ampute.html

Contact information

Rianne Schouten: r.m.schouten@uu.nl

Follow my work: rianneschouten.github.io

Missingness types



Multivariate Amputation: Weighted sum scores

- Missing values in multiple variables

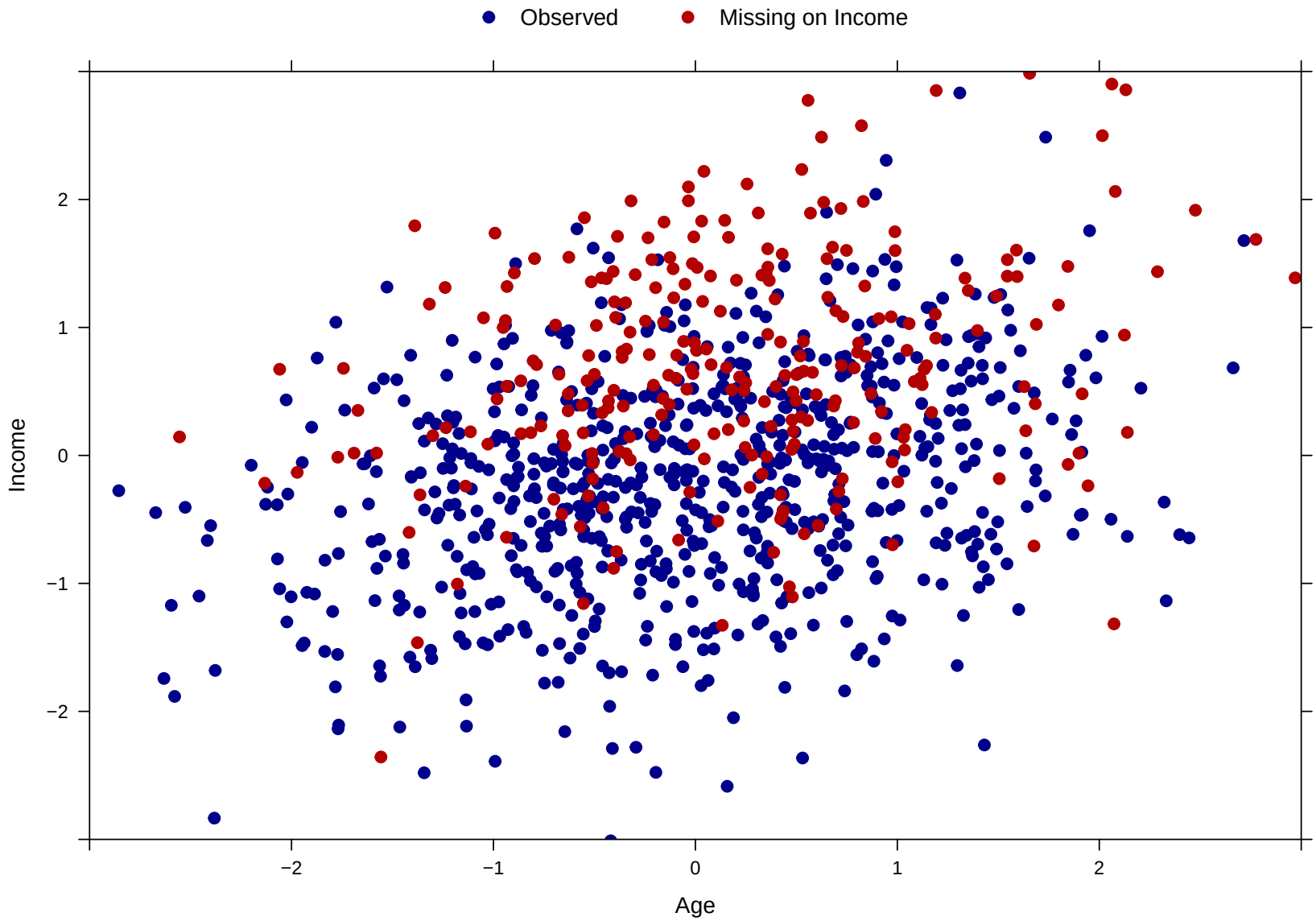
$$\begin{array}{c} \\ P_1 \\ P_2 \end{array} \begin{array}{c} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{array}$$

- Based on multiple variables

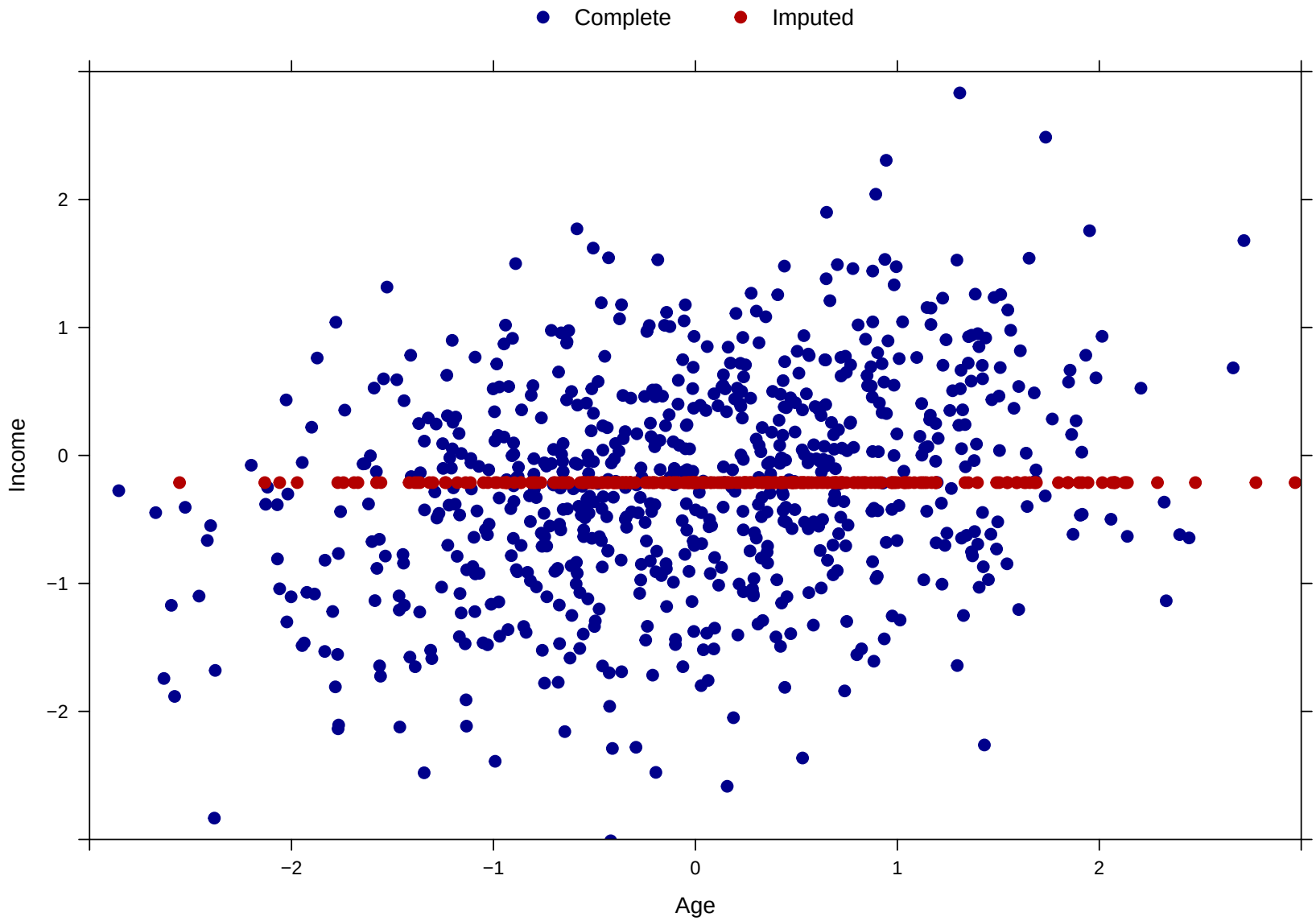
  $wss_i = w_{1,1} \cdot y_{1i} + w_{1,2} \cdot y_{2i} + w_{1,3} \cdot x_i$ if case i is in pattern 1

$$\begin{array}{c} \\ W_1 \\ W_2 \end{array} \begin{array}{c} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 5 \end{bmatrix} \end{array}$$

Missing data methods



Mean imputation



Regression imputation