Imputation-KNNBased

SP

15/11/2019

```
ds<-read.csv("fitness.csv")
ds
```

```
##    Oxygen RunTime RunPulse
## 1  44.609   11.37      178
## 2  45.313   10.07      185
## 3  54.297    8.65      156
## 4  59.571      NA     <NA>
## 5  49.874    9.22     NA
## 6  44.811   11.63      176
## 7  45.681   11.95      176
## 8  49.091   10.85     <NA>
## 9  39.442   13.08      174
## 10 60.055    8.63      170
## 11 50.541      NA     NA
## 12 37.388   14.03      186
## 13 44.754   11.12      176
## 14 47.273      NA     <NA>
## 15 51.855   10.33      166
## 16 49.156    8.95      180
## 17 40.836   10.95      168
## 18 46.672   10.00     <NA>
## 19 46.774   10.25      NA
## 20 50.388   10.08      168
## 21 39.407   12.63      174
## 22 46.080   11.17      156
## 23 45.441    9.63      164
## 24 54.625    8.92      146
## 25 45.118   11.08     NA
## 26 39.203   12.88      168
## 27 45.790   10.47      186
## 28 50.545    9.93      148
## 29 48.673    9.40      186
## 30 47.920   11.50      170
## 31 47.467   10.50      170
```

RunPulse was read as the factor, so we will convert it into int

```
ds$RunPulse = as.double(ds$RunPulse)
ds
```

```
##    Oxygen RunTime RunPulse
## 1  44.609   11.37       10
## 2  45.313   10.07       12
## 3  54.297    8.65        3
## 4  59.571      NA       NA
```

```
## 5  49.874    9.22       15
## 6  44.811   11.63        9
## 7  45.681   11.95        9
## 8  49.091   10.85       NA
## 9  39.442   13.08        8
## 10 60.055    8.63        7
## 11 50.541      NA       15
## 12 37.388   14.03       13
## 13 44.754   11.12        9
## 14 47.273      NA       NA
## 15 51.855   10.33        5
## 16 49.156    8.95       11
## 17 40.836   10.95        6
## 18 46.672   10.00       NA
## 19 46.774   10.25       14
## 20 50.388   10.08        6
## 21 39.407   12.63        8
## 22 46.080   11.17        3
## 23 45.441    9.63        4
## 24 54.625    8.92        1
## 25 45.118   11.08       15
## 26 39.203   12.88        6
## 27 45.790   10.47       13
## 28 50.545    9.93        2
## 29 48.673    9.40       13
## 30 47.920   11.50        7
## 31 47.467   10.50        7
```

Missing values

Oxygen - has complete data

RunTime - has three observations missing

RunPulse - has three observations (4, 11, 14) missing together with RunTime and five on its own (5, 8, 18,19, 25)

```
library(yaImpute)
```

```
## Warning: package 'yaImpute' was built under R version 3.6.1
```

```
set.seed(3)
refs=sample(rownames(ds),c(1,2,3,6,7,9,10,12,13,15,16,17,20:24))
```

```
refs
```

```
## [1] "5"
```

```
x <- as.matrix(ds[, 1])
x
```

```
##          [,1]
##  [1,] 44.609
##  [2,] 45.313
##  [3,] 54.297
##  [4,] 59.571
##  [5,] 49.874
##  [6,] 44.811
##  [7,] 45.681
##  [8,] 49.091
##  [9,] 39.442
## [10,] 60.055
## [11,] 50.541
## [12,] 37.388
## [13,] 44.754
## [14,] 47.273
## [15,] 51.855
## [16,] 49.156
## [17,] 40.836
## [18,] 46.672
## [19,] 46.774
## [20,] 50.388
## [21,] 39.407
## [22,] 46.080
## [23,] 45.441
## [24,] 54.625
## [25,] 45.118
## [26,] 39.203
## [27,] 45.790
## [28,] 50.545
## [29,] 48.673
## [30,] 47.920
## [31,] 47.467
```

```
y <- ds[, 2:3]
y
```

```
##    RunTime RunPulse
## 1    11.37       10
## 2    10.07       12
## 3     8.65        3
## 4       NA       NA
## 5     9.22       15
## 6    11.63        9
## 7    11.95        9
## 8    10.85       NA
## 9    13.08        8
## 10    8.63        7
## 11      NA       15
## 12   14.03       13
## 13   11.12        9
## 14      NA       NA
## 15   10.33        5
## 16    8.95       11
## 17   10.95        6
```

```
## 18   10.00       NA
## 19   10.25       14
## 20   10.08        6
## 21   12.63        8
## 22   11.17        3
## 23    9.63        4
## 24    8.92        1
## 25   11.08       15
## 26   12.88        6
## 27   10.47       13
## 28    9.93        2
## 29    9.40       13
## 30   11.50        7
## 31   10.50        7
```

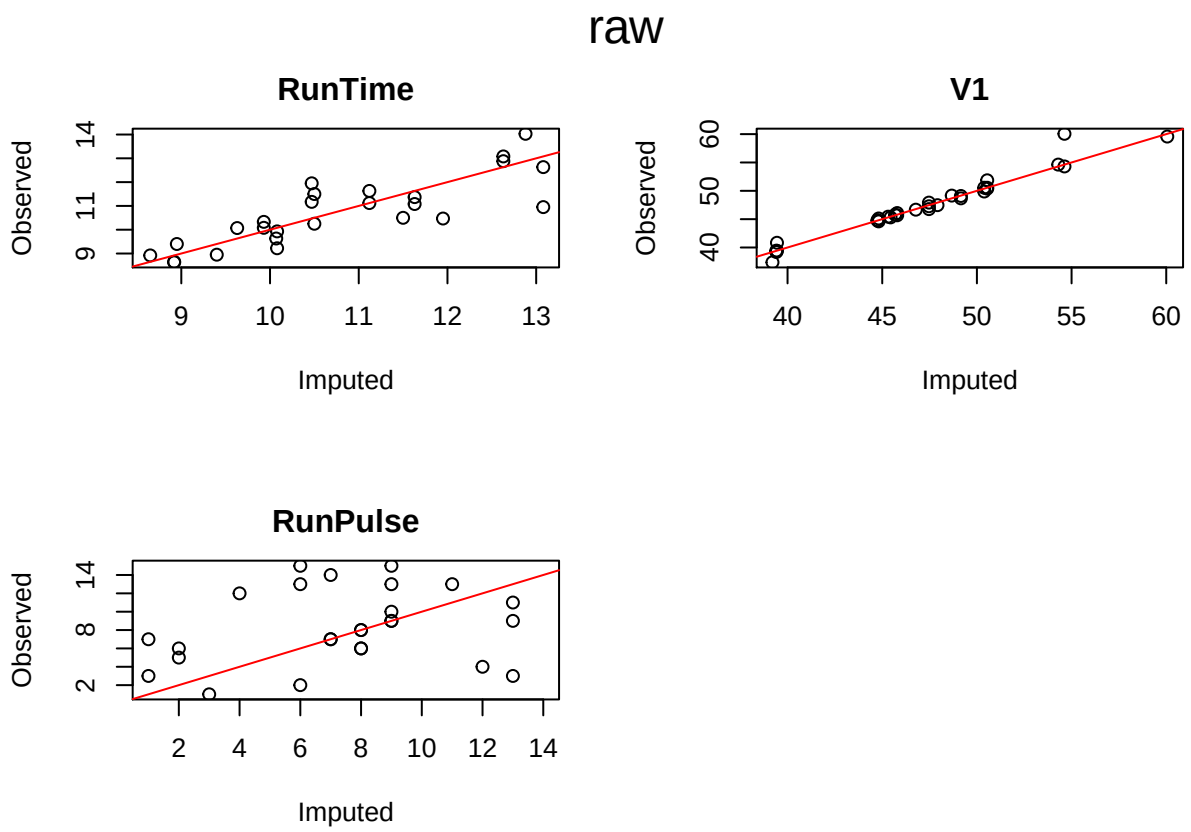
```
raw <- yai(x = x, y = y, method = "randomForest")
```

```
## Loading required namespace: randomForest

## Warning in yai(x = x, y = y, method = "randomForest"): 5 y observation(s)
## removed
```

```
#raw
```

```
plot(raw)
```



```r
tail(impute(raw))
```

```
##    RunTime RunPulse RunTime.o RunPulse.o     V1   V1.o
## 31   11.50        7      10.5          7 47.920 47.467
## 4     8.63        7        NA         NA 60.055 59.571
## 8     8.95       11        NA         NA 49.156 49.091
## 11    9.93        2        NA         NA 50.545 50.541
## 14   10.50        7        NA         NA 47.467 47.273
## 18   10.25       14        NA         NA 46.774 46.672
```