Multiplex Canada Multiple Imputations

Fio Vialard

Imputation plan

Set-up

Packages

```
require(mice)
require(lattice)
require(tidyverse)
require(gridExtra)
require(pan)
require(naniar)
require(flextable)
#install.packages("miceadds")
require(miceadds)
require(readxl)
require(lme4)
set.seed(123)
theme_set(theme_bw())
```

Import dataset

```
multiplex<- read.csv("3_intermediate/missing_data/20230908_multiplex_na_final.csv")
head(multiplex, 3)
```

```
##       p_id DateRecuit            SexualActivity   CondomUse SexualPartners
## 1 NBU20001 2021-03-24                  1-Yes 2-Sometimes            0-0
## 2 NBU20002 2021-04-15 2-I do not wish to answer        <NA>           <NA>
## 3 NBU20003 2021-04-16                  1-Yes     0-Never            1-1
##           PastSTD TypePastSTD PastInjectDrugs ShareNeedles              Alcohol
## 1           1-Yes 3-Chlamydia           1-Yes        1-Yes                 0-No
## 2 2-I do not know        None           1-Yes        1-Yes                 0-No
## 3           1-Yes 3-Chlamydia           1-Yes        1-Yes 1-1-2 times per week
##                    PastHIVTest                   PastHCVTest
## 1   3-Yes; more than 1 year ago   3-Yes; more than 1 year ago
## 2 1-Yes; less than 6 months ago 1-Yes; less than 6 months ago
## 3 1-Yes; less than 6 months ago 1-Yes; less than 6 months ago
##                PastSyphilisTest PastSyphilis     Age   Gender
## 1   3-Yes; more than 1 year ago         0-No 3-35-44   1-Male
```

```
## 2 1-Yes; less than 6 months ago         0-No    <NA>     <NA>
## 3                          <NA>         0-No 3-35-44 2-Female
##                       Descent                            EducationStatus
## 1 4-European/ North American 2-High school to Pre-university education
## 2                         <NA>                                      <NA>
## 3 4-European/ North American      1-Did not complete primary schooling
##   WorkStatusStatus MonthlyIncomeStatus CurrentPartner     FearPOC     PainPOC
## 1   3-Not employed       1-<$2;000 CAD           0-No 0-1-no fear 0-1-no pain
## 2             <NA>                <NA>           <NA>        <NA>        <NA>
## 3   3-Not employed       1-<$2;000 CAD          1-Yes 0-1-no fear 0-1-no pain
##         DiscomfortPOC  SatisfactionPOC PainConventional FearConventional
## 1 0-1 - no discomfort      1-Satisfied      0-1-no pain      0-1-no fear
## 2               <NA>             <NA>             <NA>             <NA>
## 3 0-1 - no discomfort 0-Very satisfied      0-1-no pain      4-5-highest
##   DiscomfortConventional SatisfactionConventional   PreferenceTiming
## 1    0-1 - no discomfort      2-Neutral/ not sure 1-Less than 1 week
## 2                  <NA>                     <NA>              <NA>
## 3                   1-2      2-Neutral/ not sure            0-1 day
##   RetestPreference RecommendRapidTests PreferenceFollowup
## 1     0-Rapid test               1-Yes    4-No preference
## 2             <NA>                <NA>              <NA>
## 3     0-Rapid test               1-Yes       0-Phone call
##                                                                                  comments
## 1                                                                                    <NA>
## 2 Testing could have occured in a different site and positive tests reported in another province
## 3                                                                                    <NA>
##   Clinic.Location hcv.pos hiv.pos syphilis.pos baseline.hcv baseline.hiv
## 1           RECAP       1       0            0            0            1
## 2           RECAP       1       0            0            1            1
## 3           RECAP       0       0            0            0            1
##   baseline.syph new.hcv new.syphilis new.hiv
## 1             1       0            0       0
## 2             1       1            0       0
## 3             1       0            0       0
```

```
source("2_scripts/00_functions.R")
```

Factorize and remove baseline infection data

Since no NA and having the new infection and total infection data gives us enough information. Also remove the comments for now

```
multiplex <- multiplex %>%
  select(!c(comments, baseline.hcv:baseline.syph)) %>%
 mutate(across(SexualActivity:Clinic.Location, as.factor)) %>%
  rename(WorkStatus = WorkStatusStatus, Date.Recruited = DateRecuit)
summary(multiplex)
```

```
##      p_id           Date.Recruited                       SexualActivity
##  Length:401         Length:401         0-No                     : 88
##  Class :character   Class :character   1-Yes                    :289
##  Mode  :character   Mode  :character   2-I do not wish to answer: 18
```

```
##                                     NA's                    :  6
##
##
##
##       CondomUse        SexualPartners            PastSTD
##  0-Never    : 98   0-0         : 71    0-No           :204
##  1-Always   :105   1-1         : 94    1-Yes          :169
##  2-Sometimes:167   2-2 to 5    :159    2-I do not know: 12
##  NA's       : 31   3-6 to 10   : 38    NA's           : 16
##                    4-11 or more: 22
##                    NA's        : 17
##
##        TypePastSTD                   PastInjectDrugs ShareNeedles
##  None         :216   0-No                      :252    0-No : 55
##  3-Chlamydia  : 72   1-Yes                     :140    1-Yes: 85
##  4-Gonorrhoeae: 43   2-I do not wish to answer:  3    None :252
##  6-Other      : 25   Yes                       :  1    NA's :  9
##  0-HIV        : 10   NA's                      :  5
##  (Other)      : 13
##  NA's         : 22
##                 Alcohol                          PastHIVTest
##  0-No                :209   0-No                          : 57
##  1-1-2 times per week:122   1-Yes; less than 6 months ago:141
##  2-3-5 times per week: 65   2-Yes; 6 months to 1 year ago: 47
##  NA's                :  5   3-Yes; more than 1 year ago  :154
##                             NA's                          :  2
##
##
##                         PastHCVTest                         PastSyphilisTest
##  0-No                          : 44   0-No                          :  4
##  1-Yes; less than 6 months ago:140   1-Yes; less than 6 months ago:127
##  2-Yes; 6 months to 1 year ago: 48   2-Yes; 6 months to 1 year ago: 47
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :220
##  NA's                          :  2   NA's                          :  3
##
##
##          PastSyphilis             Age               Gender
##  0-No           :330   1-18-24       : 29   1-Male       :255
##  1-Yes          : 48   1-25-34       :127   2-Female     : 73
##  2-I do not know: 19   3-35-44       :103   3-Transgender:  7
##  NA's           :  4   4-45-54       : 54   4-Bisexual   :  7
##                        5-55 and above: 40   5-No response: 11
##                        NA's          : 48   NA's         : 48
##
##                                        Descent
##  1-North American Indigenous                 : 18
##  2-Asian/ South Asian                        : 10
##  3-Mediterranean                             : 13
##  4-European/ North American                  :254
##  4-Latin American/ Caribbean                 : 38
##  6-African (North; Central; West; East; South): 20
##  NA's                                        : 48
##                                  EducationStatus                  WorkStatus
##  1-Did not complete primary schooling     : 38    1-Employed (full time):127
```

```
##  2-High school to Pre-university education:143    2-Employed (part time): 61
##  3-Technical degree                         : 52    3-Not employed        :135
##  4-Post-graduate degree                     :120    4-Retired             : 11
##  NA's                                       : 48    5-Not willing to work : 19
##                                                     NA's                  : 48
##
##           MonthlyIncomeStatus CurrentPartner        FearPOC
##  1-<$2;000 CAD       :219      0-No :213      0-1-no fear:155
##  2-$2;001-$4;000 CAD :103      1-Yes:140      1-2        : 52
##  3-$4;001-$6;000 CAD : 22      NA's : 48      2-3        : 19
##  4-$6;001-$$8;000 CAD:  7                     3-4        :  9
##  5-$8;001 and above  :  2                     4-5-highest:  7
##  NA's                : 48                     NA's       :159
##
##        PainPOC                DiscomfortPOC            SatisfactionPOC
##  0-1-no pain:166   0-1 - no discomfort:181   0-Very satisfied   :207
##  1-2        : 59   1-2                : 46   1-Satisfied        : 30
##  2-3        : 15   2-3                : 11   2-Neutral/ not sure:  3
##  3-4        :  1   3-4                :  2   NA's               :161
##  NA's       :160   4-5-highest        :  1
##                    NA's               :160
##
##    PainConventional    FearConventional         DiscomfortConventional
##  0-1-no pain:159     0-1-no fear:150     0-1 - no discomfort:173
##  1-2        : 61     1-2        : 45     1-2                : 43
##  2-3        : 10     2-3        : 21     2-3                : 12
##  3-4        :  1     3-4        :  9     3-4                :  2
##  NA's       :170     4-5-highest:  5     4-5-highest        :  1
##                      NA's       :171     NA's               :170
##
##        SatisfactionConventional           PreferenceTiming
##  0-Very satisfied   :190         0-1 day            :137
##  1-Satisfied        : 33         1-Less than 1 week : 97
##  2-Neutral/ not sure:  7         2-Less than 2 weeks:  7
##  3-Dissatisfied     :  1         3-2 weeks or more  :  2
##  NA's               :170         NA's               :158
##
##
##        RetestPreference RecommendRapidTests
##  0-Rapid test   :144     0-No      :  7
##  1-Phlebotomy   : 66     1-Yes     :227
##  2-No preference: 32     2-Not Sure:  9
##  NA's           :159     NA's      :158
##
##
##
##                       PreferenceFollowup   Clinic.Location    hcv.pos
##  0-Phone call                   : 38      Outreach  : 51    Min.   :0.0000
##  1-Short-messaging service (SMS): 34      RECAP     : 87    1st Qu.:0.0000
##  2-App-based secure messaging   : 84      REZO      :237    Median :0.0000
##  3-Face-to-face                 : 54      Riverstone: 26    Mean   :0.1496
##  4-No preference                : 33                        3rd Qu.:0.0000
##  NA's                           :158                        Max.   :1.0000
##
```

```
##     hiv.pos        syphilis.pos        new.hcv         new.syphilis
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.000000
##  Mean   :0.0399   Mean   :0.01247   Mean   :0.03491   Mean   :0.004987
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000000
##
##     new.hiv
##  Min.   :0.000000
##  1st Qu.:0.000000
##  Median :0.000000
##  Mean   :0.004987
##  3rd Qu.:0.000000
##  Max.   :1.000000
##
```

Missing data pattern

```
total.na <- sum(is.na(multiplex))

indiv.na<- sum(!complete.cases(multiplex))

prop.indiv.na <- round(indiv.na/nrow(multiplex) *100, digits =2)

multiplex.no.pro <- multiplex %>%
  select(-c(FearPOC:PreferenceFollowup))

prop.indiv.na2 <- sum(!complete.cases(multiplex.no.pro))/nrow(multiplex.no.pro)*100
```

Total missing data: 2412

Total number of participants with at least one missing data: 199

Percentage of participants with missing data: 49.63 %

Percentage of participants with missing data when patient-reported outcomes are removed: 18.2 %

```
pattern<- md.pattern(multiplex, rotate.names = T)
```

```
na.patterns.total<- nrow(pattern)
```

Total numbers of patterns: 42 . There is 1 individuals with 26 (or 2/3) missing variables.

Which variables have similar patterns?

```
na.fear <-sum(pattern[, "FearConventional"] == 0)

na.timing <- sum(pattern[, "PreferenceTiming"] == 0)
```

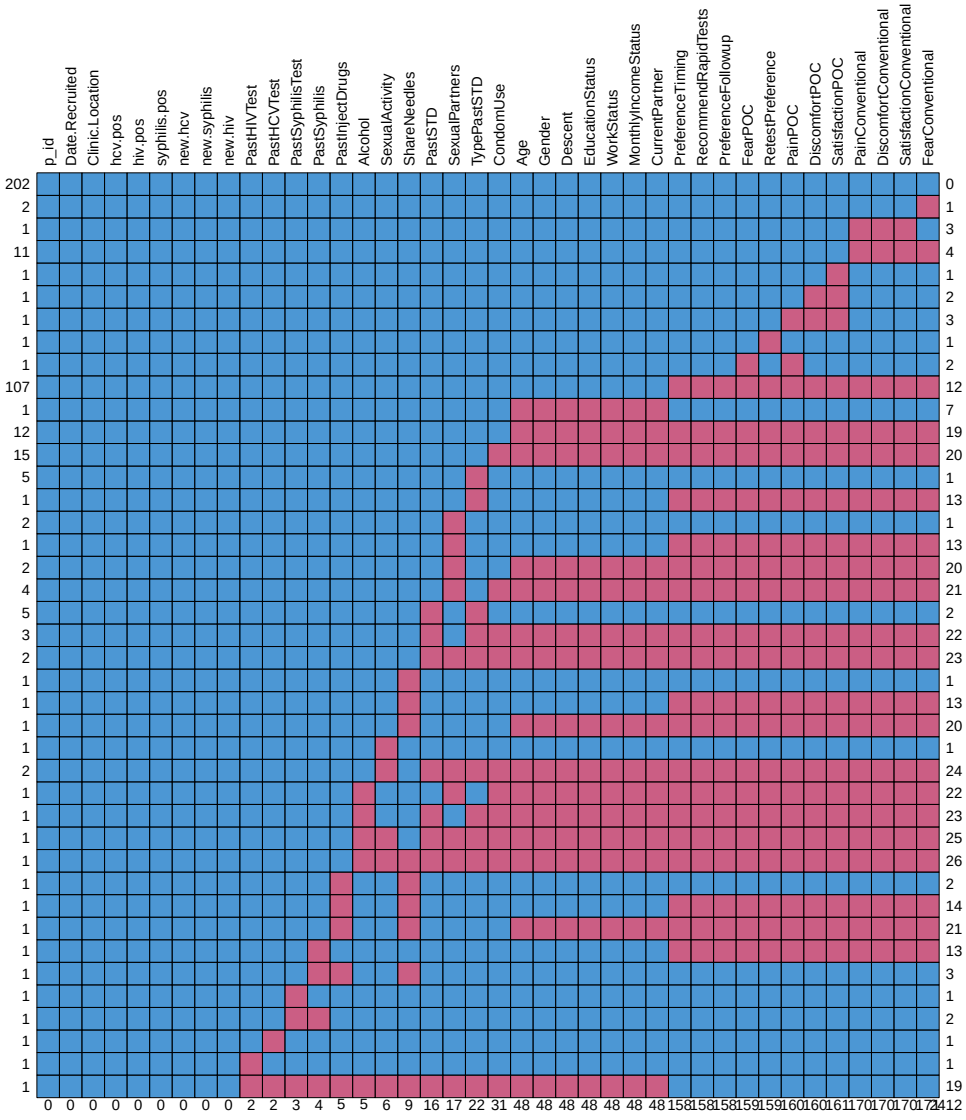


Figure 1: Pattern of missingness for variables in the Canada multiplex dataset. The numbers on the left indicate the number of rows that have the same pattern of missingness. The numbers on the right represent the number of missing data for each pattern. The numbers at the bottom represent the total number of missing values for each variable with the total number of missing data in the dataset at the bottom right

```
na.age <- sum(pattern[,"Age"]==0)

na.current.partners <- sum(pattern[, "CurrentPartner"]==0)

na.condom <- sum(pattern[,"CondomUse"]==0)

na.share.needles <- sum(pattern[,"ShareNeedles"]==0)
```

Patient-reported outcomes (e.g. fear of conventional strategy 21 and preference timing 19) have similar missing patterns.

Socio-demographic outcomes from age (15) to current partner (15) variable have the same pattern of missingness.

Risk factor outcomes from condom use (10) to sharing needles (9) variables have the same pattern of missingness.

Remove observations with 2/3 or more missing variables

```
multiplex1 <- multiplex %>%
  mutate(na_count = rowSums(is.na(multiplex))) %>%
  filter(na_count < 26) %>%
  select(!na_count)

# check
nrow(multiplex1) == nrow(multiplex) - 1
```

```
## [1] TRUE
```

Association between variables

Are there any associations between variables. Let's find out by using two representative examples for each group of similar missing pattern.

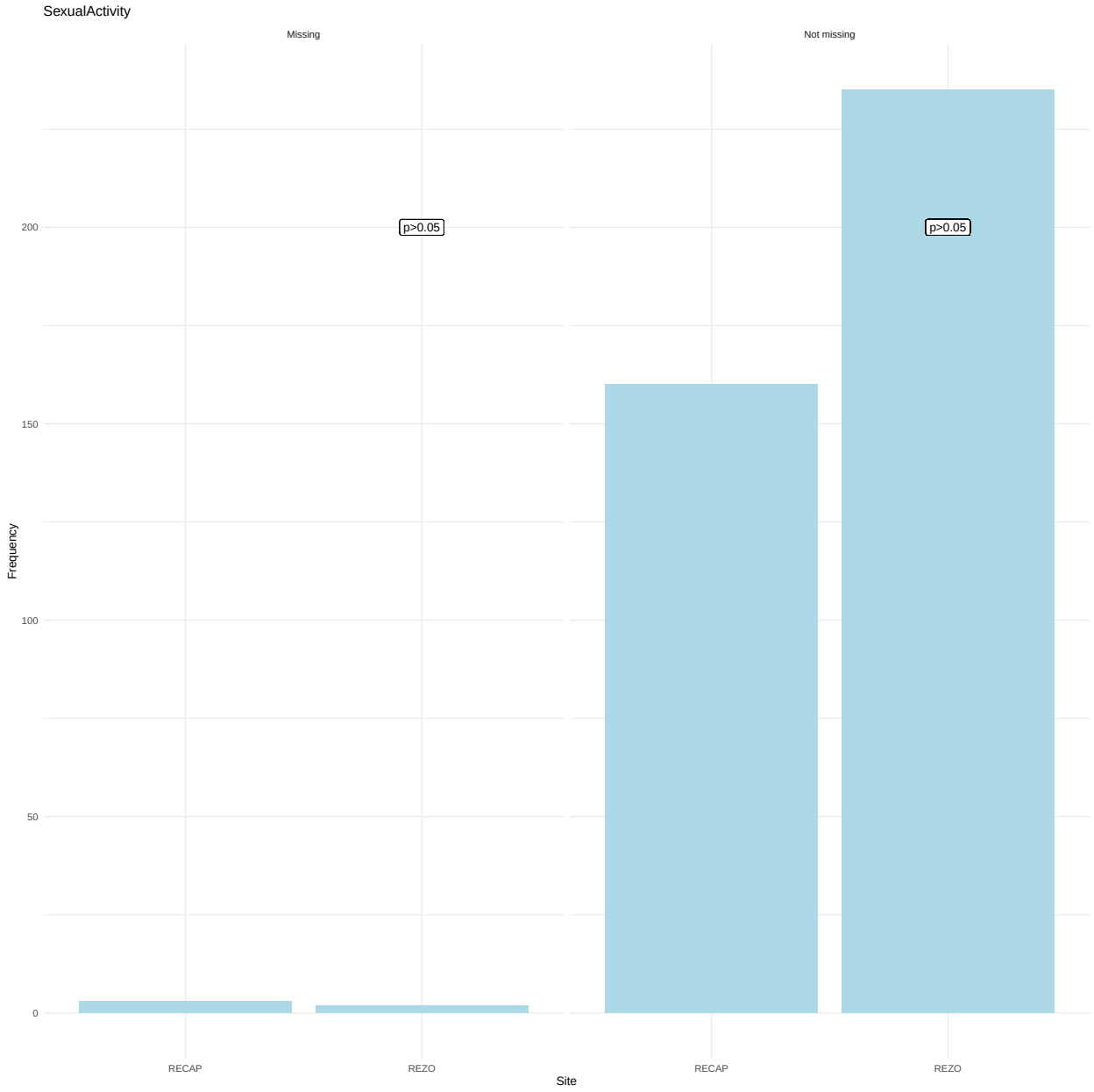
Create columns for clinic, stbbi status and missingness column to be populated later

```
multiplex4plots <- multiplex1 %>%
  mutate(site = ifelse(Clinic.Location == "REZO", "REZO", "RECAP"),
         stbbi.pos = as.factor(ifelse(hcv.pos == 1| hiv.pos ==1 | syphilis.pos == 1,
                                      "Infected","Non-infected")),
         missingness = NA)
```

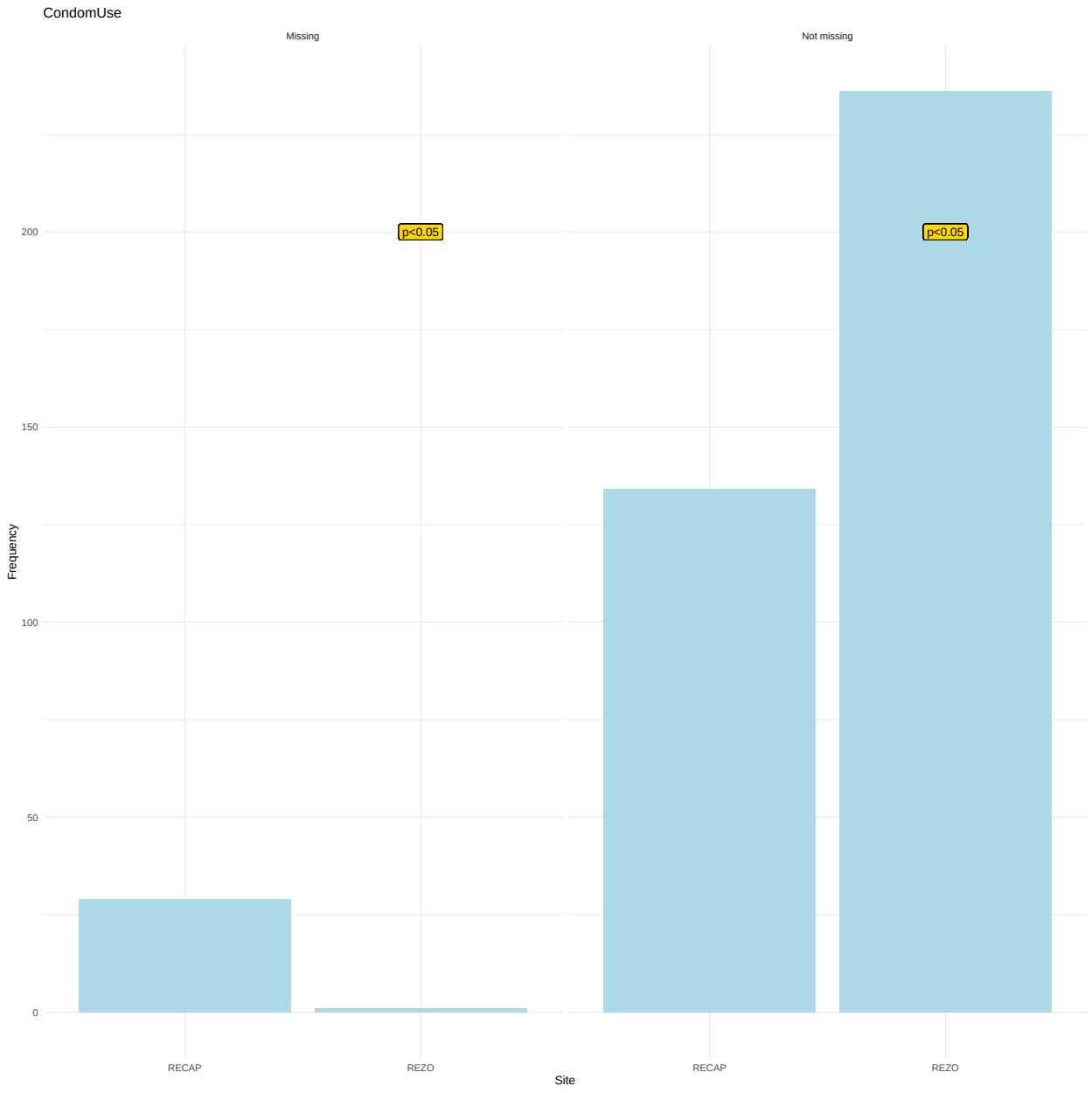
Does variable missingness depend on clinic?

```
variable <- colnames(multiplex4plots)[3:21]

plots.site <-  get.correlation.site(multiplex4plots, variable, title = variable)
plots.site
```

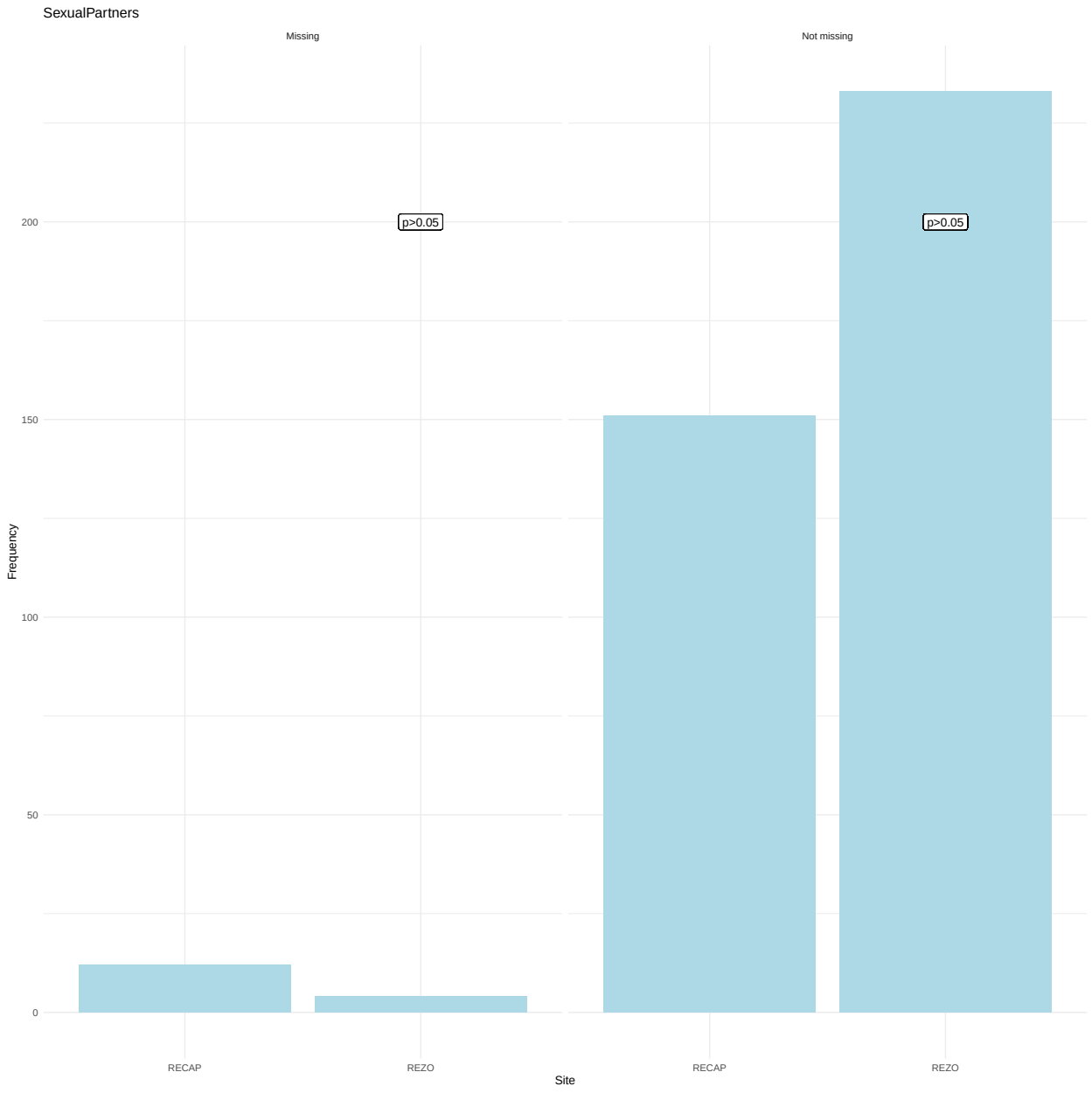
```
## [[1]]
```


SexualActivity
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

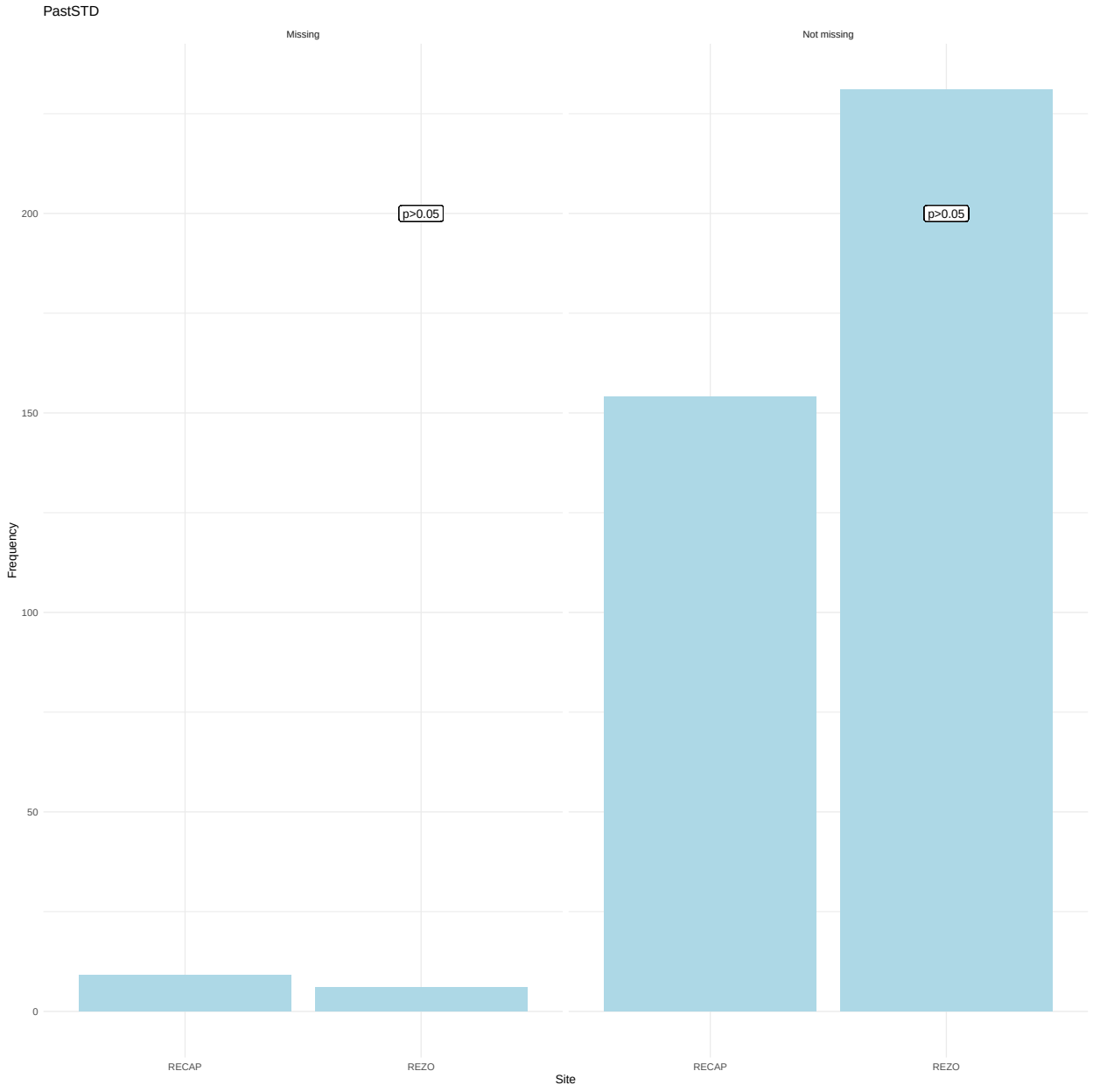
```
##
## [[2]]
```


CondomUse
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

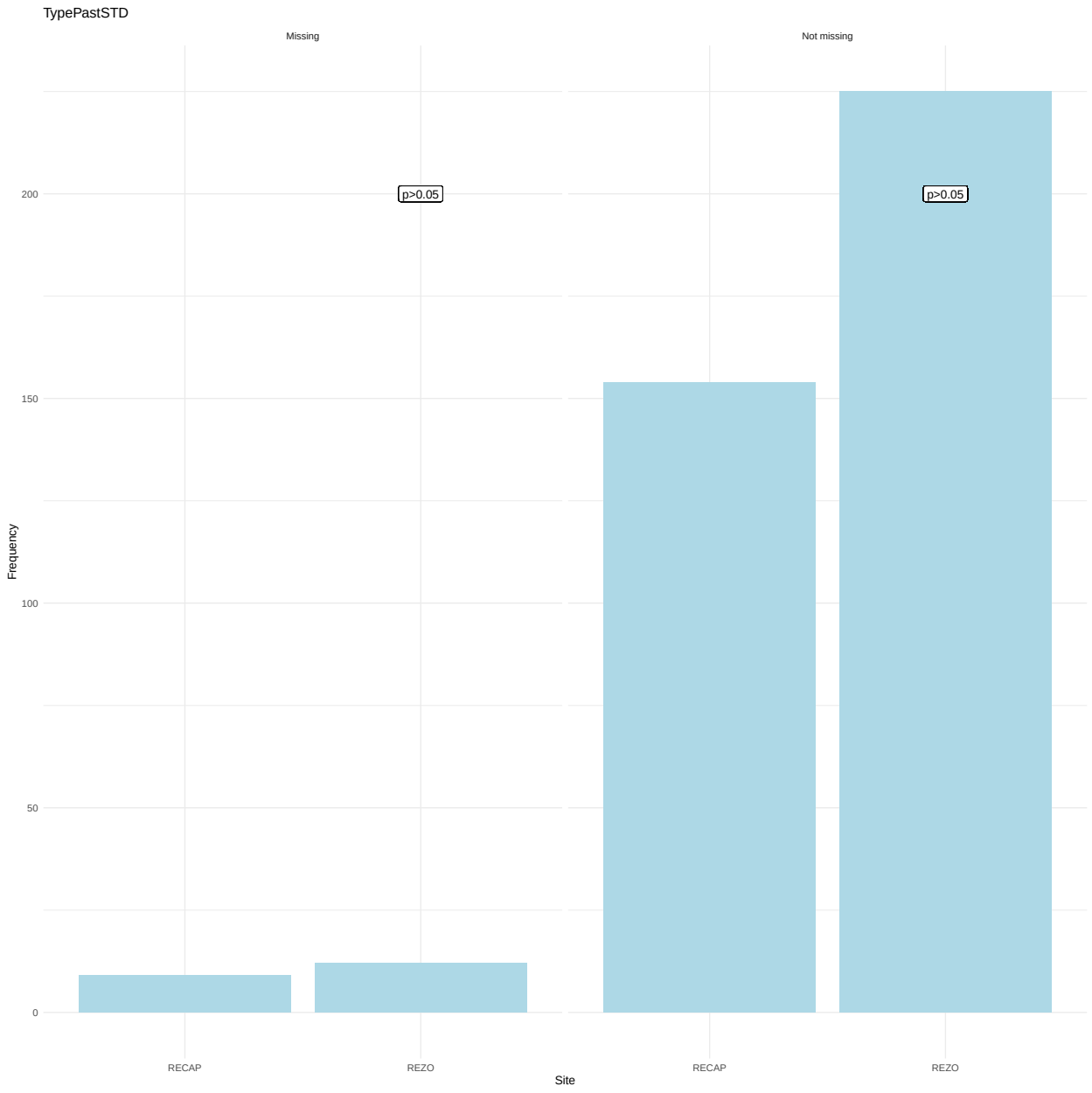
##
[[3]]


SexualPartners
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

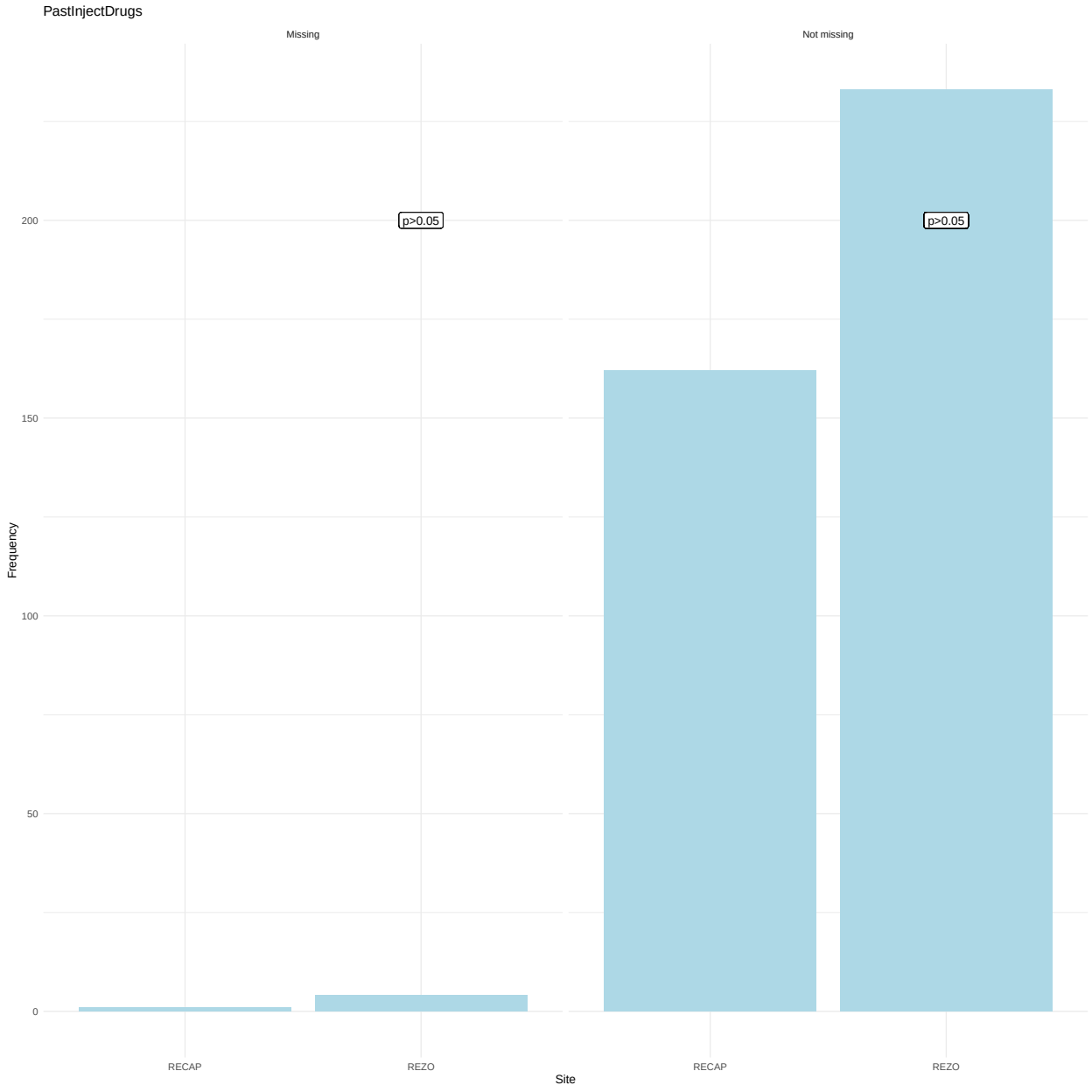
##
[[4]]


PastSTD
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

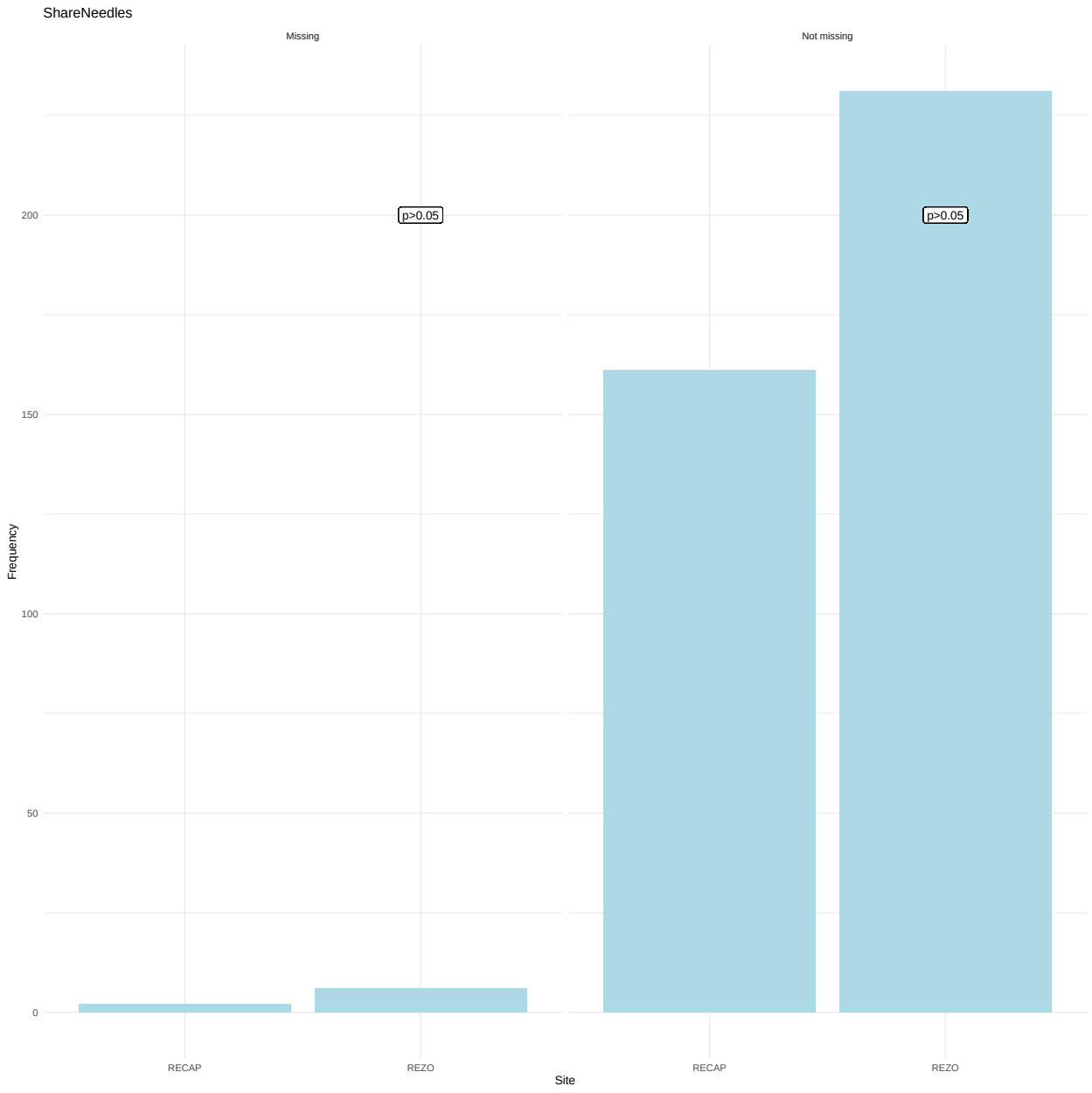
```
##
## [[5]]
```


TypePastSTD
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

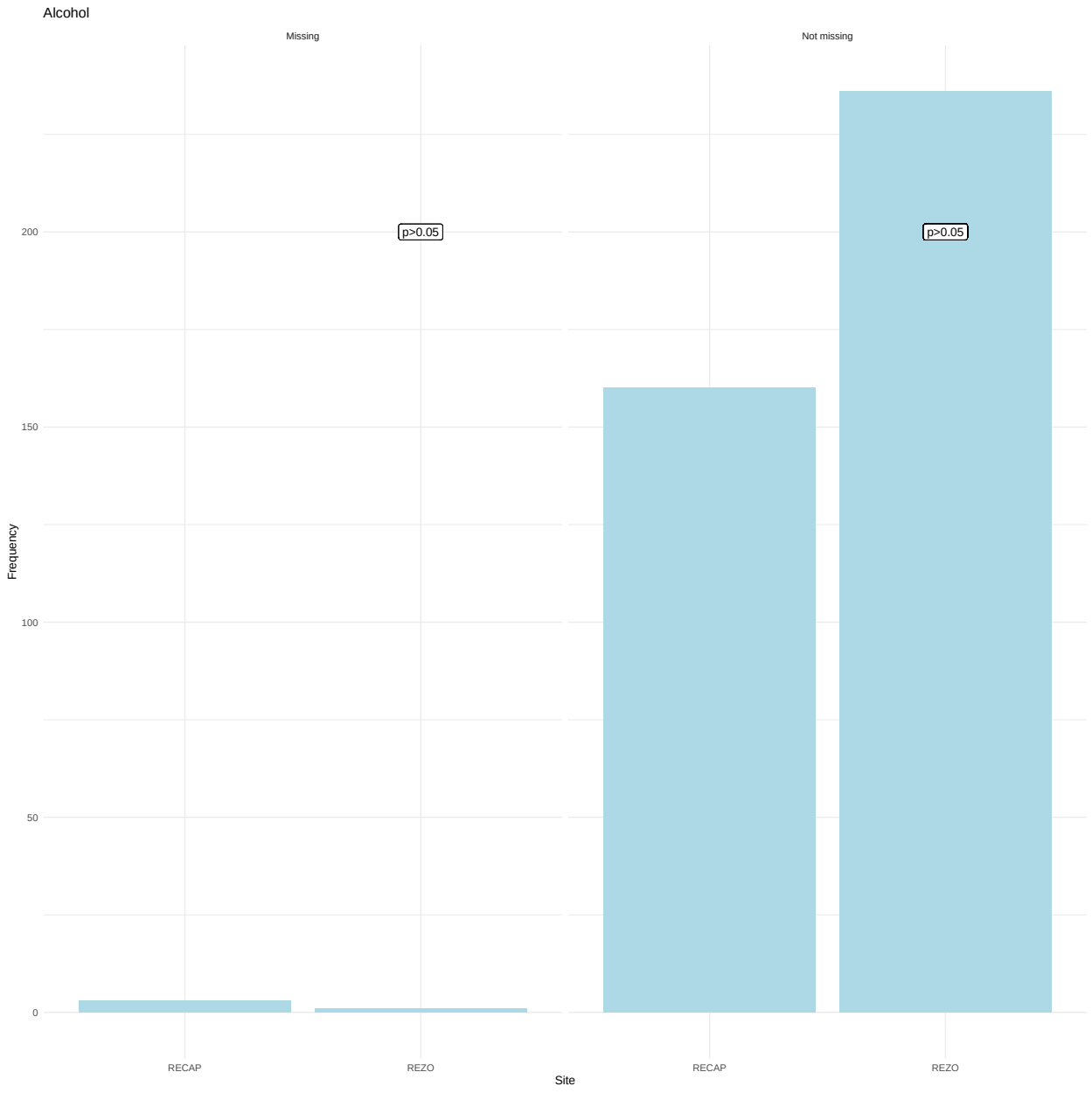
```
##
## [[6]]
```


PastInjectDrugs
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

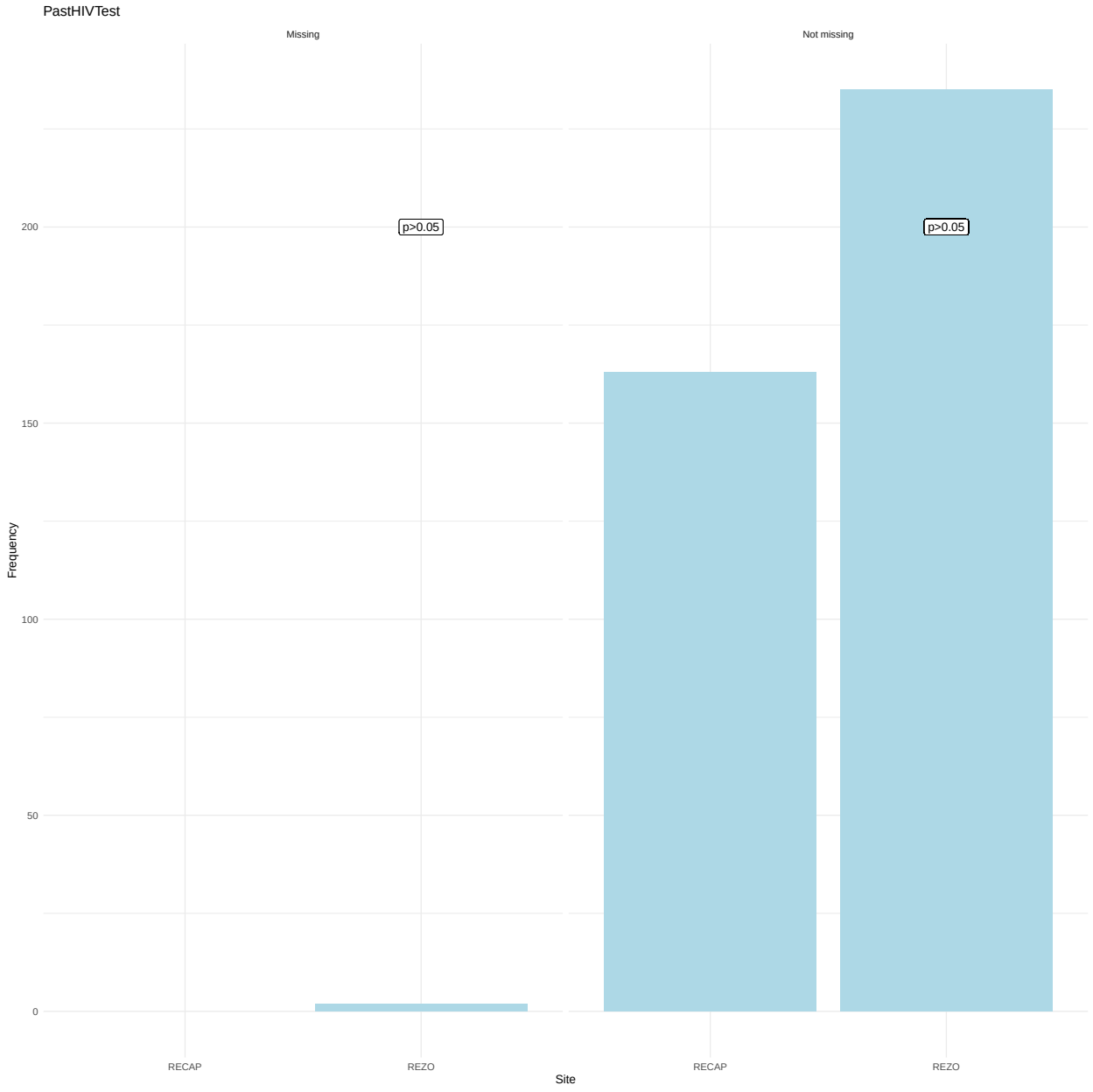
```
##
## [[7]]
```


ShareNeedles
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

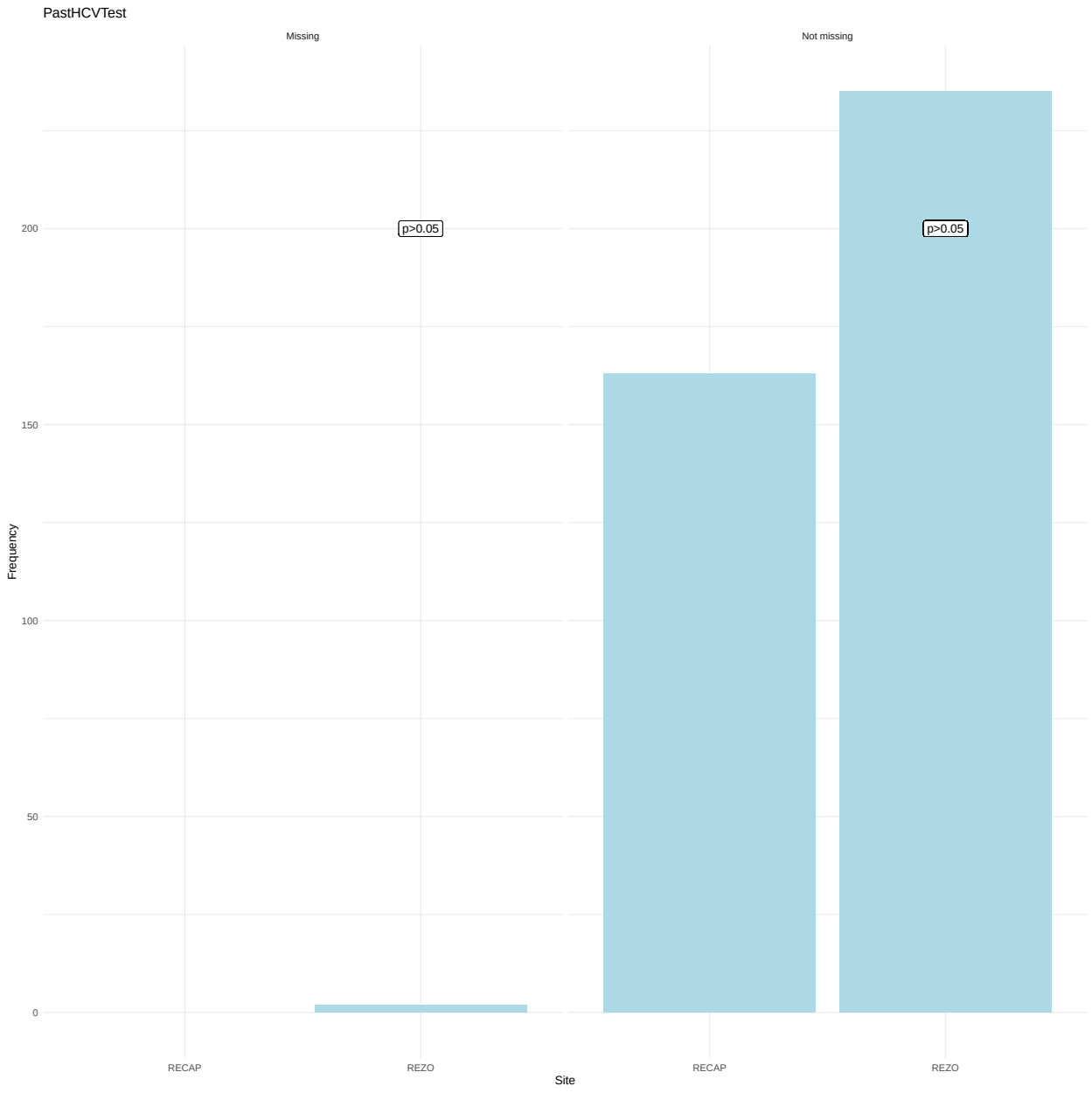
```
##
## [[8]]
```


Alcohol
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

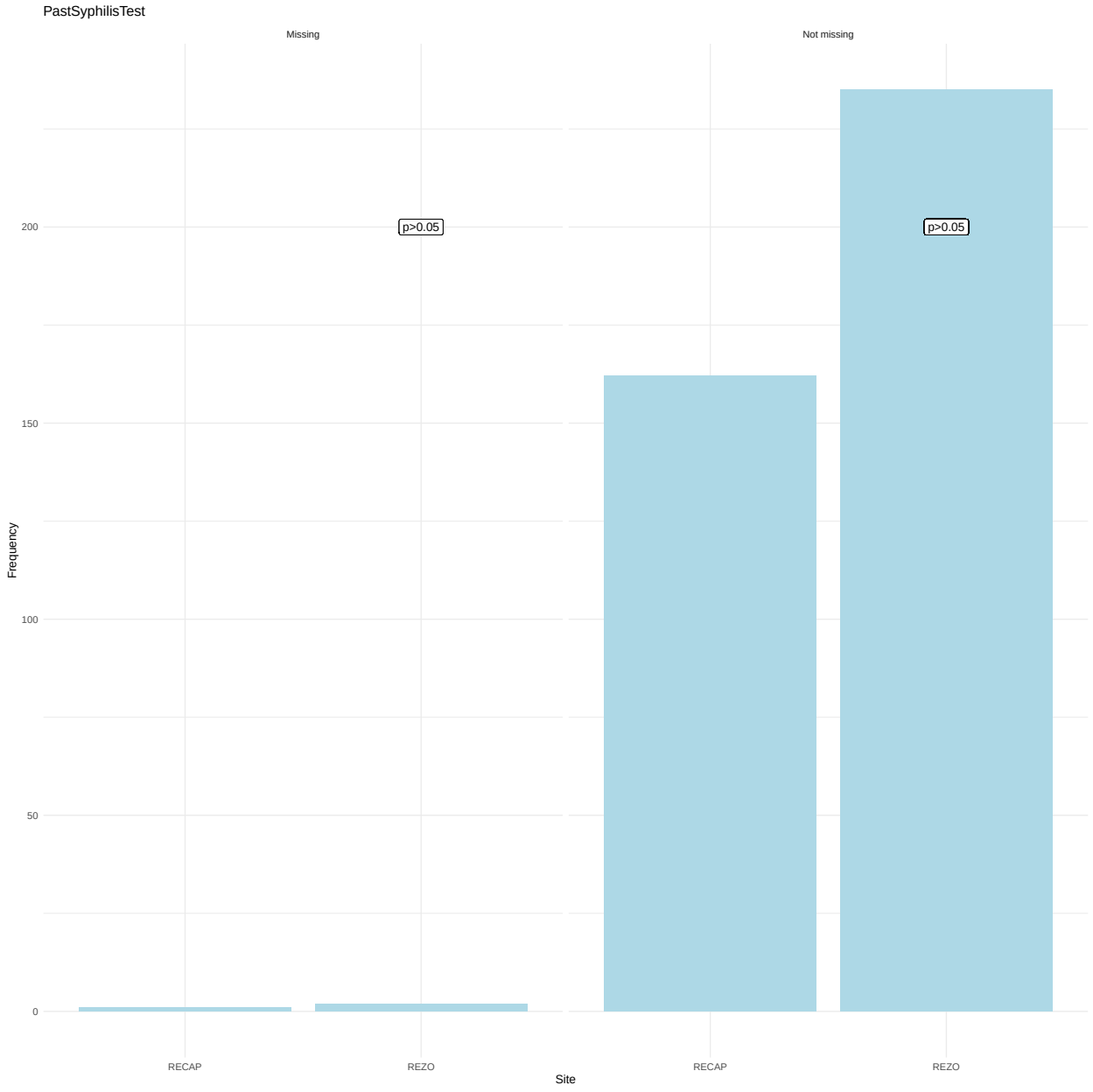
##
[[9]]


PastHIVTest
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

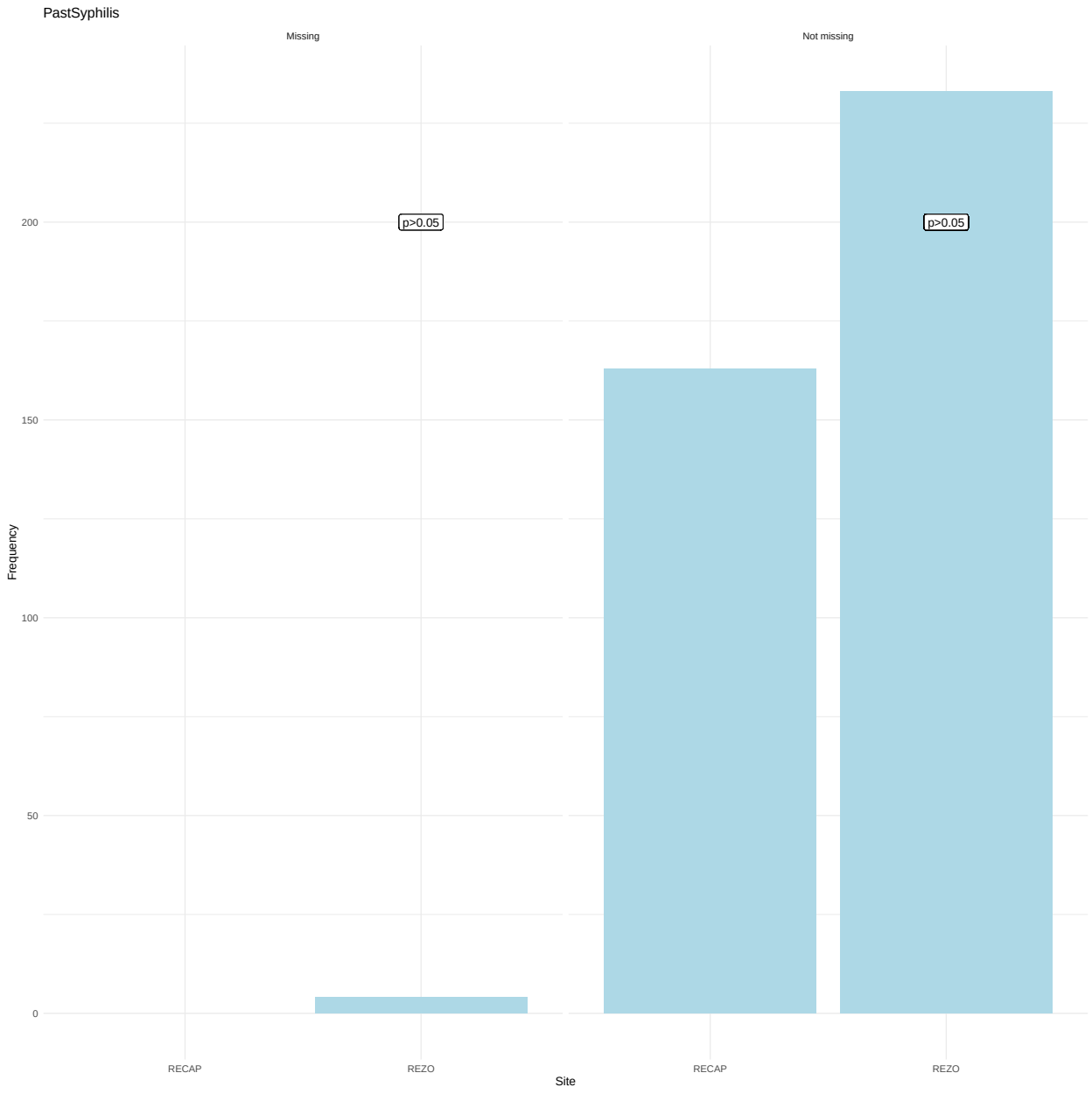
```
##
## [[10]]
```


PastHCVTest
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

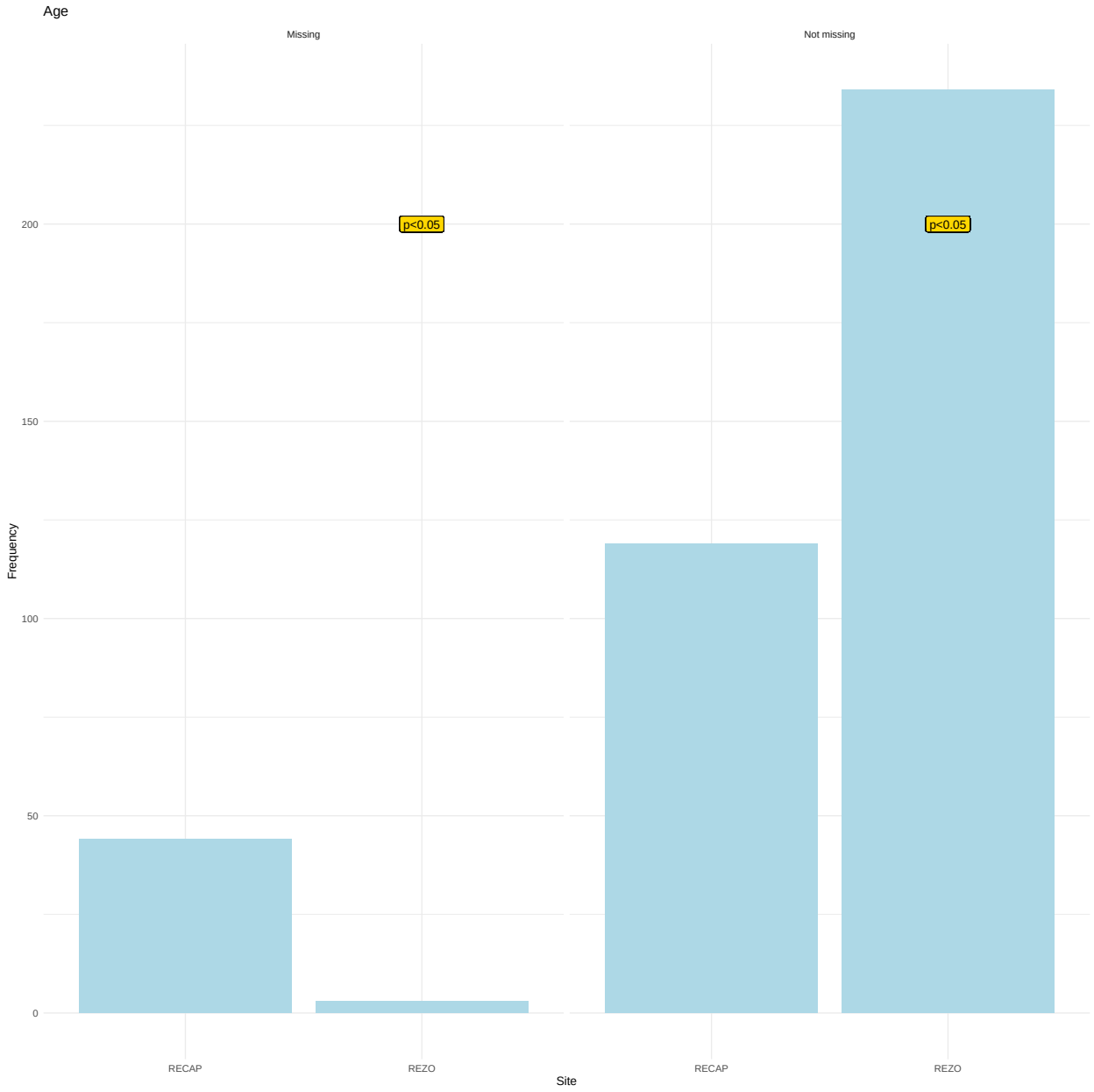
```
##
## [[11]]
```


PastSyphilisTest
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

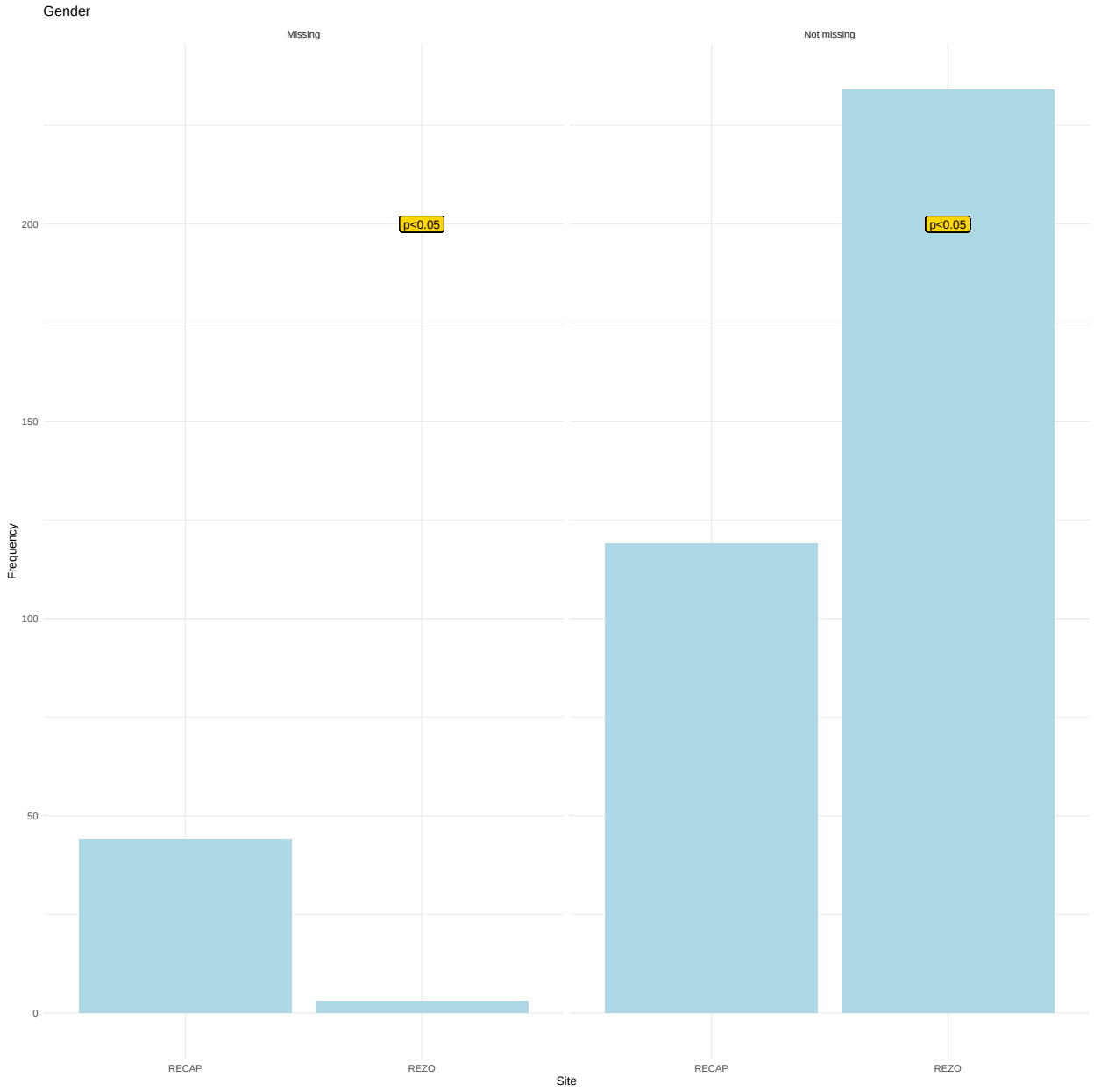
```
##
## [[12]]
```


PastSyphilis
Missing
Not missing
p>0.05
p>0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

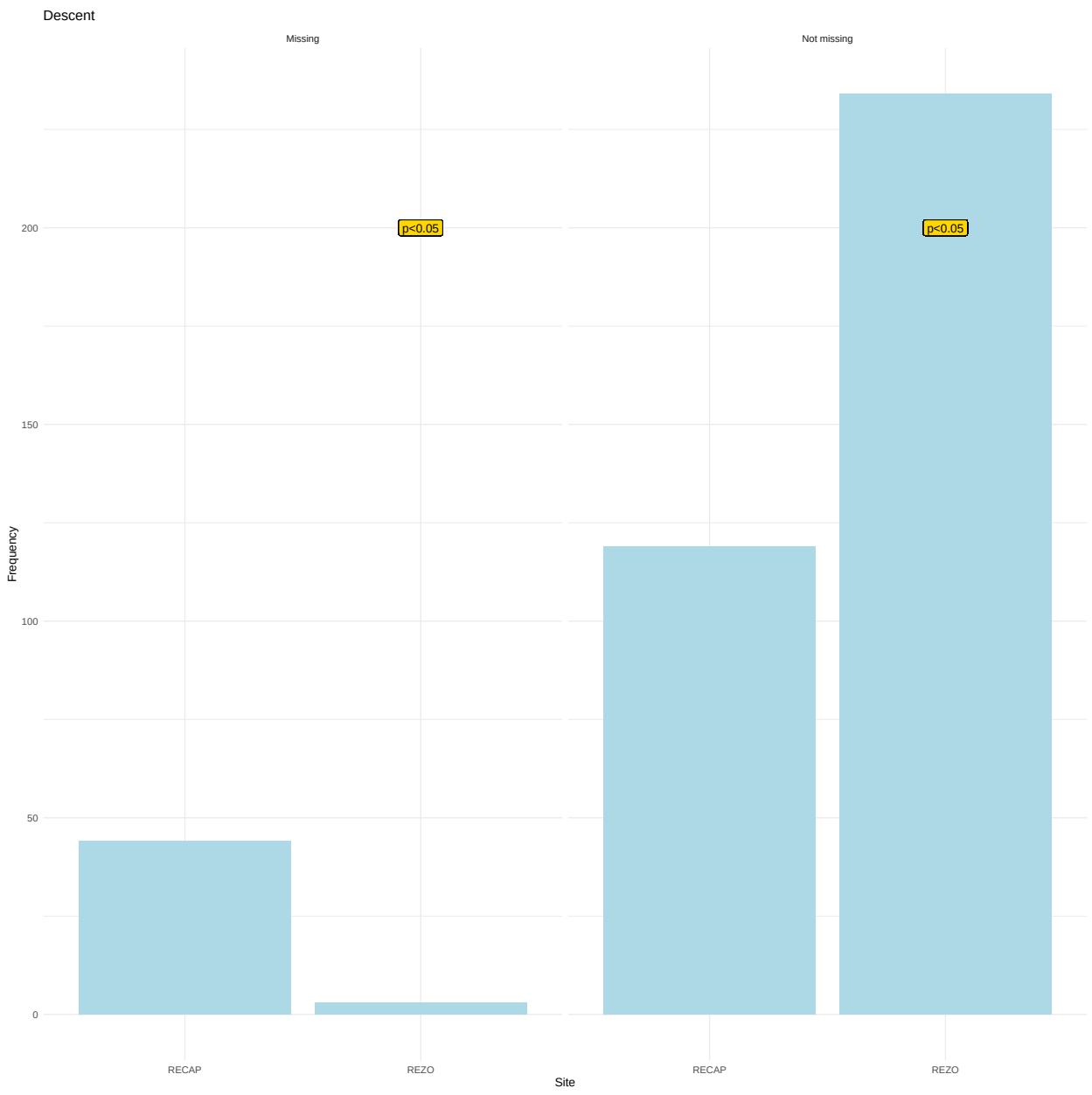
```
##
## [[13]]
```


Age
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

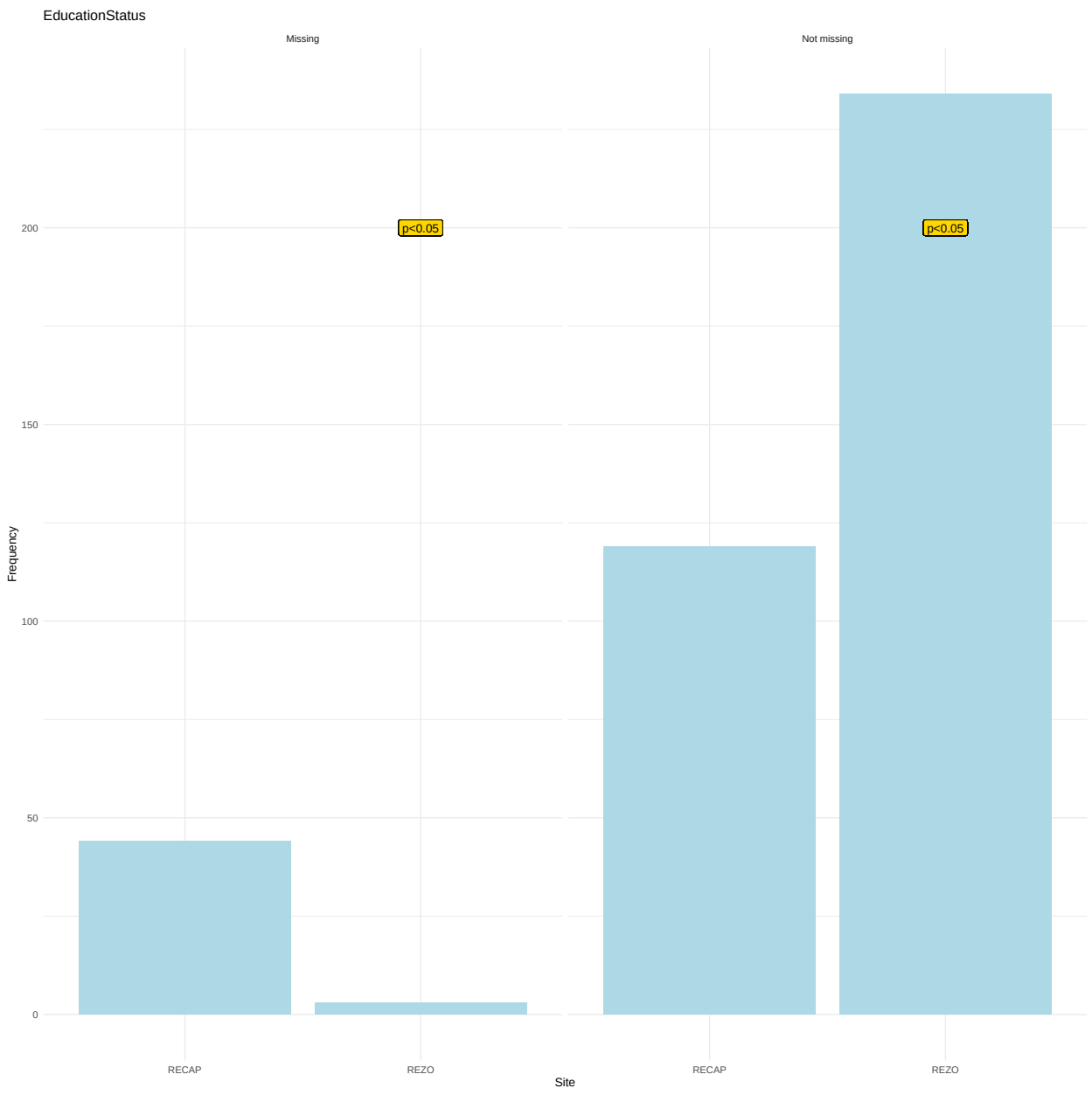
##
[[14]]


Gender
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

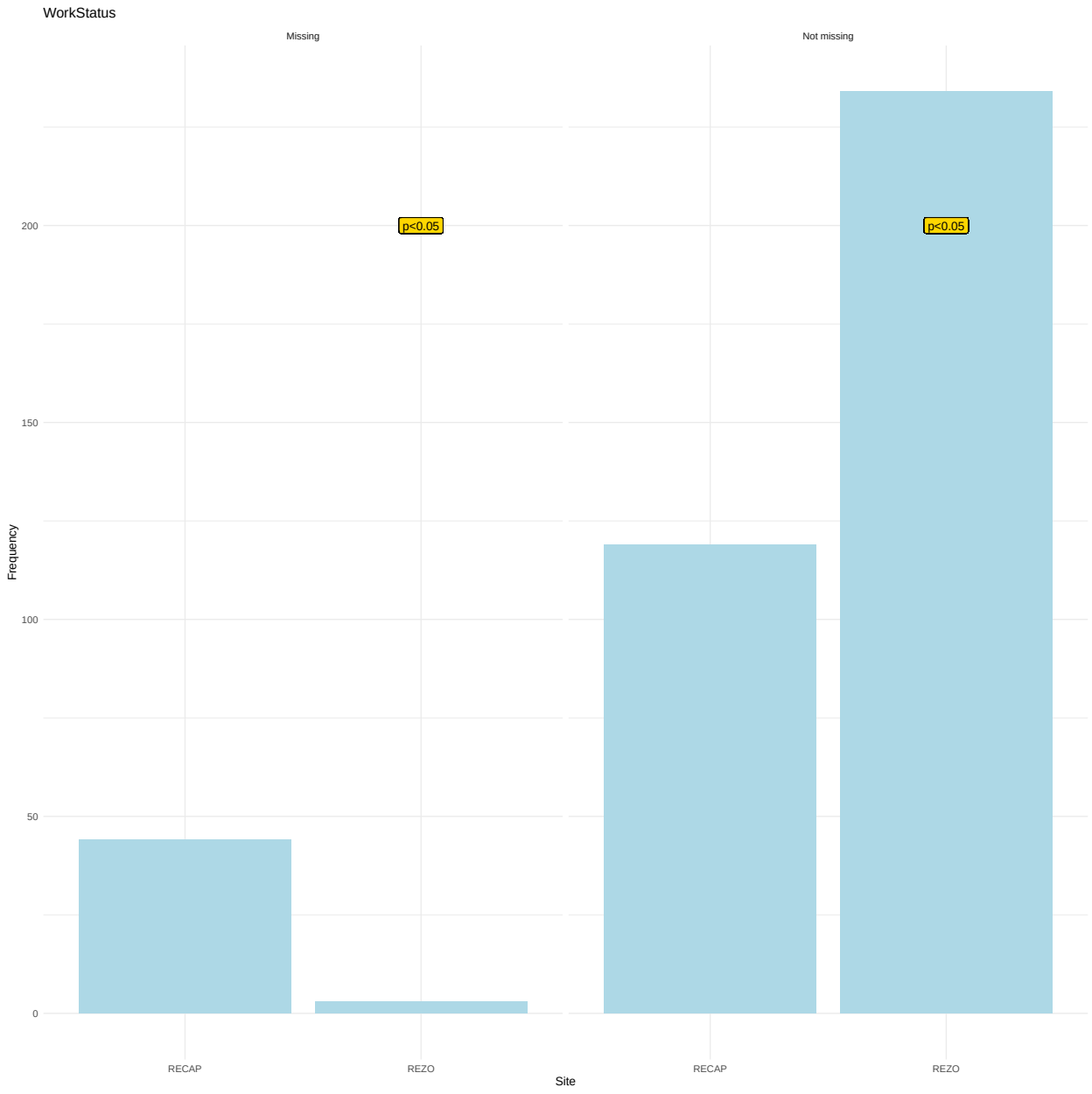
```
##
## [[15]]
```


Descent
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

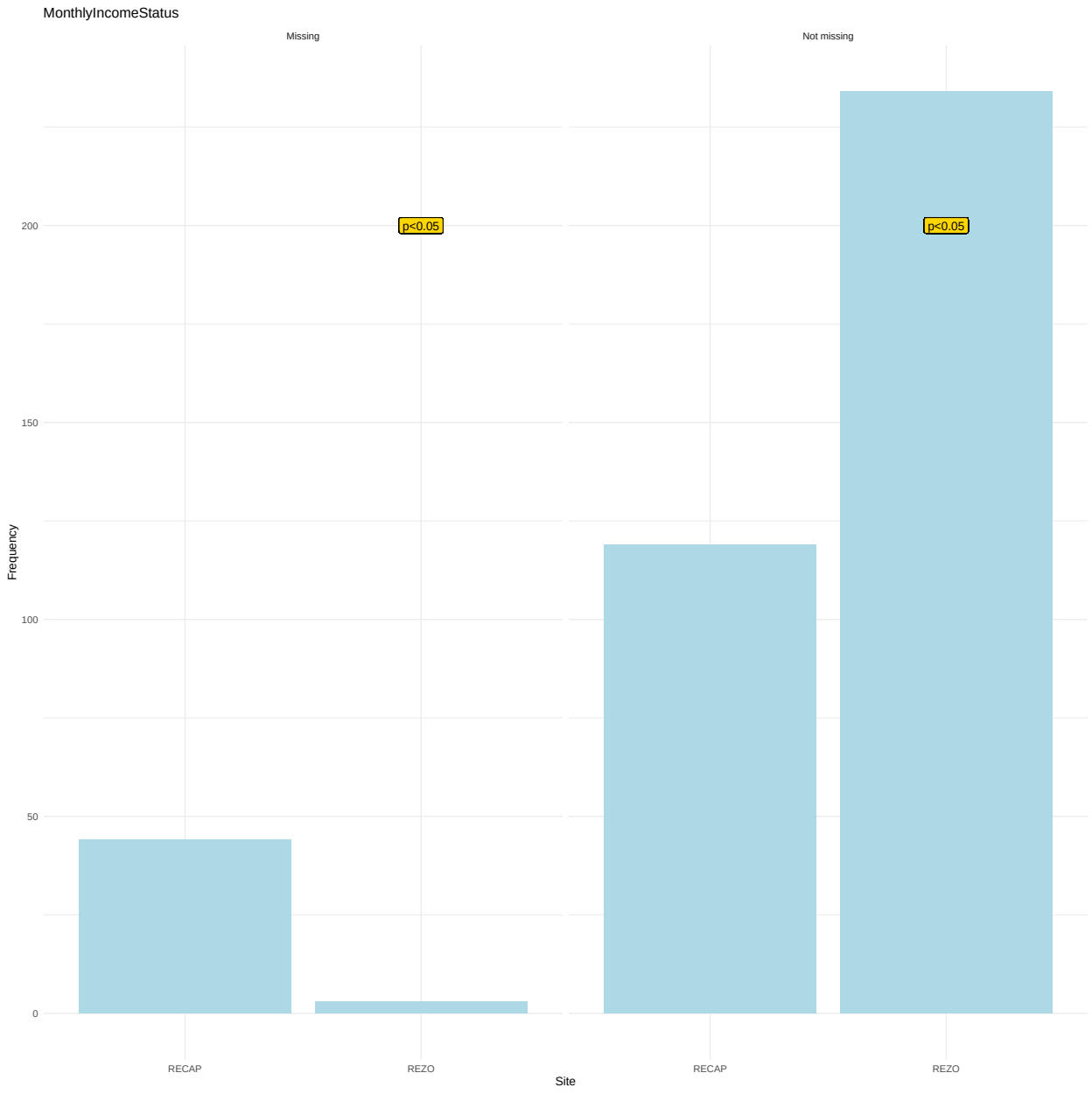
##
[[16]]


EducationStatus
Missing
Not missing
p<0.05
p<0.05
Frequency
200
150
100
50
0
RECAP
REZO
RECAP
REZO
Site

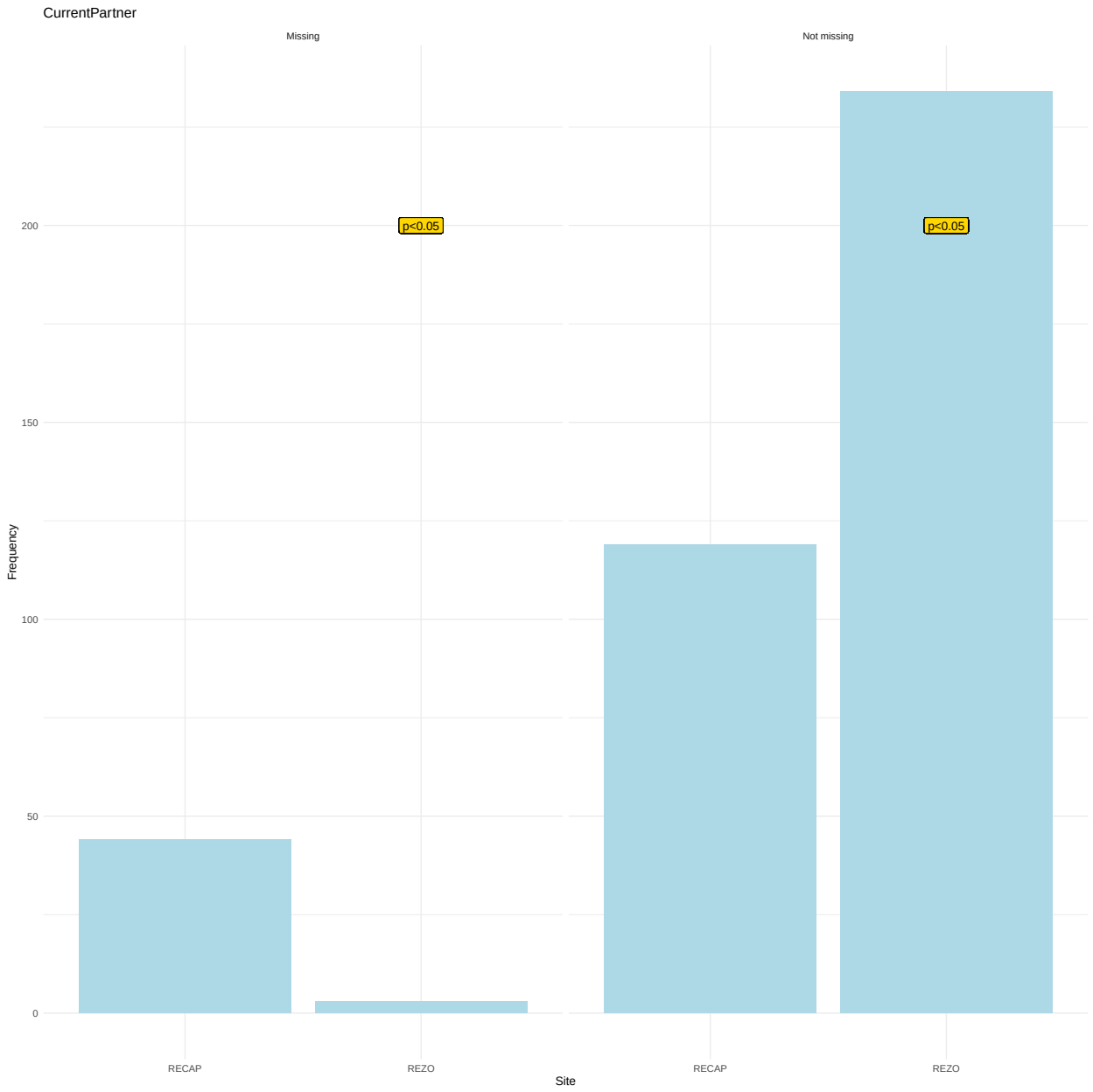
##
[[17]]


WorkStatus
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

##
[[18]]


MonthlyIncomeStatus
Missing
Not missing
p<0.05
p<0.05
200
150
100
50
0
Frequency
RECAP
REZO
RECAP
REZO
Site

##
[[19]]



Does variable missingness depend on infection status?

7. Fear of the conventional strategy

```
plot7 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$FearConventional,
                             title = "Fear of conventional strategy data")
```

8. Timing preference

```
plot8 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$PreferenceTiming,
                             title = "Preference timing data")
```

9. Age

```
plot9 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$Age,
                             title = "Age data")
```

10. Current partner

```
plot10 <- get.correlation.inf(dataset = multiplex4plots,
                              variable = multiplex4plots$CurrentPartner,
                              title = "Current partner data")
```

11. sharing needles

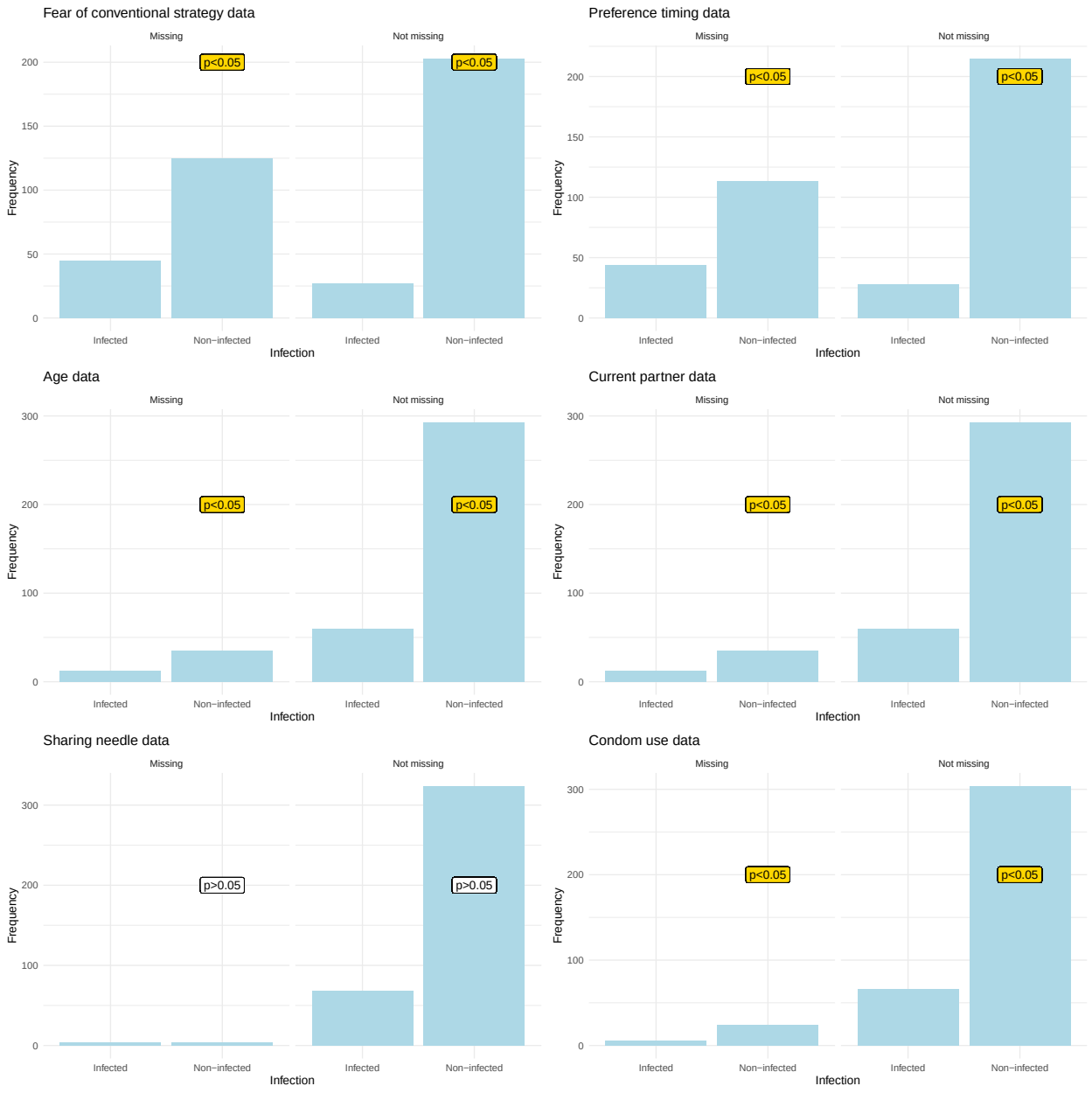
```
plot11 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$ShareNeedles,
                             title = "Sharing needle data")
```

12. Condom use

```
plot12 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$CondomUse,
                             title = "Condom use data")
```

Plots

```
all.plots <- grid.arrange( plot7,
                          plot8, plot9,  plot10,  plot11,  plot12, ncol=2)
```




```
ggsave(all.plots, filename = "3_intermediate/missing_data/missingness_correlation.png",
       device = "png", width = 14, height = 14)
```

There are more missing data at RECAP and in non-infected individuals. This pattern is MNAR. Starting from Sharing needles, the missing data does not depend on site or infection status (because number of missing data is small).

Multiple imputation methods

Preparation

Add site as a predictor

has to be integer for "2l.bin" method to work: 0 for REZO and 1 for RECAP

```
multiplex4imp <- multiplex1 %>%
  mutate(site = as.integer(ifelse(Clinic.Location == "REZO", 0, 1))) %>%
  select(-c(Clinic.Location))

summary(multiplex4imp$site)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4075  1.0000  1.0000
```

```
summary(multiplex4imp)
```

```
##      p_id           Date.Recruited                       SexualActivity
##  Length:400         Length:400         0-No                     : 88
##  Class :character   Class :character   1-Yes                    :289
##  Mode  :character   Mode  :character   2-I do not wish to answer: 18
##                                        NA's                     :  5
##
##
##
##        CondomUse        SexualPartners            PastSTD
##  0-Never    : 98   0-0         : 71    0-No           :204
##  1-Always   :105   1-1         : 94    1-Yes          :169
##  2-Sometimes:167   2-2 to 5    :159    2-I do not know: 12
##  NA's       : 30   3-6 to 10   : 38    NA's           : 15
##                    4-11 or more: 22
##                    NA's        : 16
##
##         TypePastSTD                   PastInjectDrugs ShareNeedles
##  None         :216   0-No                     :252    0-No : 55
##  3-Chlamydia  : 72   1-Yes                    :139    1-Yes: 85
##  4-Gonorrhoeae: 43   2-I do not wish to answer:  3    None :252
##  6-Other      : 25   Yes                      :  1    NA's :  8
##  0-HIV        : 10   NA's                     :  5
##  (Other)      : 13
##  NA's         : 21
##                 Alcohol                          PastHIVTest
##  0-No                :209   0-No                          : 57
##  1-1-2 times per week:122   1-Yes; less than 6 months ago:140
##  2-3-5 times per week: 65   2-Yes; 6 months to 1 year ago: 47
##  NA's                :  4   3-Yes; more than 1 year ago  :154
##                             NA's                          :  2
##
##
##                       PastHCVTest                       PastSyphilisTest
```

```
##  0-No                          : 44   0-No                          :  4
##  1-Yes; less than 6 months ago:139   1-Yes; less than 6 months ago:127
##  2-Yes; 6 months to 1 year ago: 48   2-Yes; 6 months to 1 year ago: 47
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :219
##  NA's                          :  2   NA's                          :  3
##
##
##          PastSyphilis             Age               Gender
##  0-No           :329   1-18-24       : 29   1-Male       :255
##  1-Yes          : 48   1-25-34       :127   2-Female     : 73
##  2-I do not know: 19   3-35-44       :103   3-Transgender:  7
##  NA's           :  4   4-45-54       : 54   4-Bisexual   :  7
##                        5-55 and above: 40   5-No response: 11
##                        NA's          : 47   NA's         : 47
##
##                                          Descent
##  1-North American Indigenous                  : 18
##  2-Asian/ South Asian                         : 10
##  3-Mediterranean                              : 13
##  4-European/ North American                   :254
##  4-Latin American/ Caribbean                  : 38
##  6-African (North; Central; West; East; South): 20
##  NA's                                         : 47
##                                    EducationStatus                   WorkStatus
##  1-Did not complete primary schooling     : 38    1-Employed (full time):127
##  2-High school to Pre-university education:143    2-Employed (part time): 61
##  3-Technical degree                       : 52    3-Not employed        :135
##  4-Post-graduate degree                   :120    4-Retired             : 11
##  NA's                                     : 47    5-Not willing to work : 19
##                                                   NA's                  : 47
##
##           MonthlyIncomeStatus CurrentPartner        FearPOC
##  1-<$2;000 CAD       :219      0-No :213      0-1-no fear:155
##  2-$2;001-$4;000 CAD :103      1-Yes:140      1-2        : 52
##  3-$4;001-$6;000 CAD : 22      NA's : 47      2-3        : 19
##  4-$6;001-$$8;000 CAD:  7                     3-4        :  9
##  5-$8;001 and above  :  2                     4-5-highest:  7
##  NA's                : 47                     NA's       :158
##
##        PainPOC               DiscomfortPOC            SatisfactionPOC
##  0-1-no pain:166   0-1 - no discomfort:181   0-Very satisfied   :207
##  1-2        : 59   1-2                : 46   1-Satisfied        : 30
##  2-3        : 15   2-3                : 11   2-Neutral/ not sure:  3
##  3-4        :  1   3-4                :  2   NA's               :160
##  NA's       :159   4-5-highest        :  1
##                    NA's               :159
##
##    PainConventional    FearConventional         DiscomfortConventional
##  0-1-no pain:159     0-1-no fear:150     0-1 - no discomfort:173
##  1-2        : 61     1-2        : 45     1-2                : 43
##  2-3        : 10     2-3        : 21     2-3                : 12
##  3-4        :  1     3-4        :  9     3-4                :  2
##  NA's       :169     4-5-highest:  5     4-5-highest        :  1
##                      NA's       :170     NA's               :169
```

```
##
##         SatisfactionConventional            PreferenceTiming
##  0-Very satisfied   :190         0-1 day            :137
##  1-Satisfied        : 33         1-Less than 1 week : 97
##  2-Neutral/ not sure:  7         2-Less than 2 weeks:  7
##  3-Dissatisfied     :  1         3-2 weeks or more  :  2
##  NA's               :169         NA's               :157
##
##
##         RetestPreference RecommendRapidTests
##  0-Rapid test   :144     0-No      :  7
##  1-Phlebotomy   : 66     1-Yes     :227
##  2-No preference: 32     2-Not Sure:  9
##  NA's           :158     NA's      :157
##
##
##
##                         PreferenceFollowup    hcv.pos          hiv.pos
##  0-Phone call                     : 38      Min.   :0.0000   Min.   :0.00
##  1-Short-messaging service (SMS): 34      1st Qu.:0.0000   1st Qu.:0.00
##  2-App-based secure messaging   : 84      Median :0.0000   Median :0.00
##  3-Face-to-face                 : 54      Mean   :0.1475   Mean   :0.04
##  4-No preference                : 33      3rd Qu.:0.0000   3rd Qu.:0.00
##  NA's                           :157      Max.   :1.0000   Max.   :1.00
##
##   syphilis.pos       new.hcv       new.syphilis      new.hiv
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.0000   Median :0.000   Median :0.000   Median :0.000
##  Mean   :0.0125   Mean   :0.035   Mean   :0.005   Mean   :0.005
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##
##       site
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.4075
##  3rd Qu.:1.0000
##  Max.   :1.0000
##
```

Initial imputation

Using default methods (5 iterations). No cluster

```
ini.imp <- mice(multiplex4imp, nnet.MaxNWts = 2000, print = F)
# adjust the weights to not get an error message
# do not print the results
```

Which variables were used as predictors?

Easier to visualize as a heatmap.

```
ini.pred <- ini.imp$predictorMatrix

prediction.matrix(ini.imp$predictorMatrix)
```

p_id and date were not predicted at all because all values were present in the dataset.

What methods were used for each?

```
ini.meth <- ini.imp$method
ini.meth
```

```
##                     p_id           Date.Recruited           SexualActivity
##                       ""                       ""                "polyreg"
##               CondomUse           SexualPartners                  PastSTD
##               "polyreg"                "polyreg"                "polyreg"
##             TypePastSTD          PastInjectDrugs              ShareNeedles
##               "polyreg"                "polyreg"                "polyreg"
##                 Alcohol              PastHIVTest              PastHCVTest
##               "polyreg"                "polyreg"                "polyreg"
##         PastSyphilisTest             PastSyphilis                      Age
##               "polyreg"                "polyreg"                "polyreg"
##                  Gender                  Descent         EducationStatus
##               "polyreg"                "polyreg"                "polyreg"
##              WorkStatus      MonthlyIncomeStatus           CurrentPartner
##               "polyreg"                "polyreg"                 "logreg"
##                 FearPOC                  PainPOC            DiscomfortPOC
##               "polyreg"                "polyreg"                "polyreg"
##          SatisfactionPOC         PainConventional        FearConventional
##               "polyreg"                "polyreg"                "polyreg"
##   DiscomfortConventional SatisfactionConventional        PreferenceTiming
##               "polyreg"                "polyreg"                "polyreg"
##         RetestPreference      RecommendRapidTests       PreferenceFollowup
##               "polyreg"                "polyreg"                "polyreg"
##                 hcv.pos                  hiv.pos             syphilis.pos
##                       ""                       ""                       ""
##                 new.hcv             new.syphilis                 new.hiv
##                       ""                       ""                       ""
##                    site
##                      ""
```

All used "polyreg" except for the variables without missing values and current partner used "logreg"

1-level imputation

Proposed imputation plan

```
param <- read_excel('1_data/parameters.xlsx')

param.1l <- param %>%
```

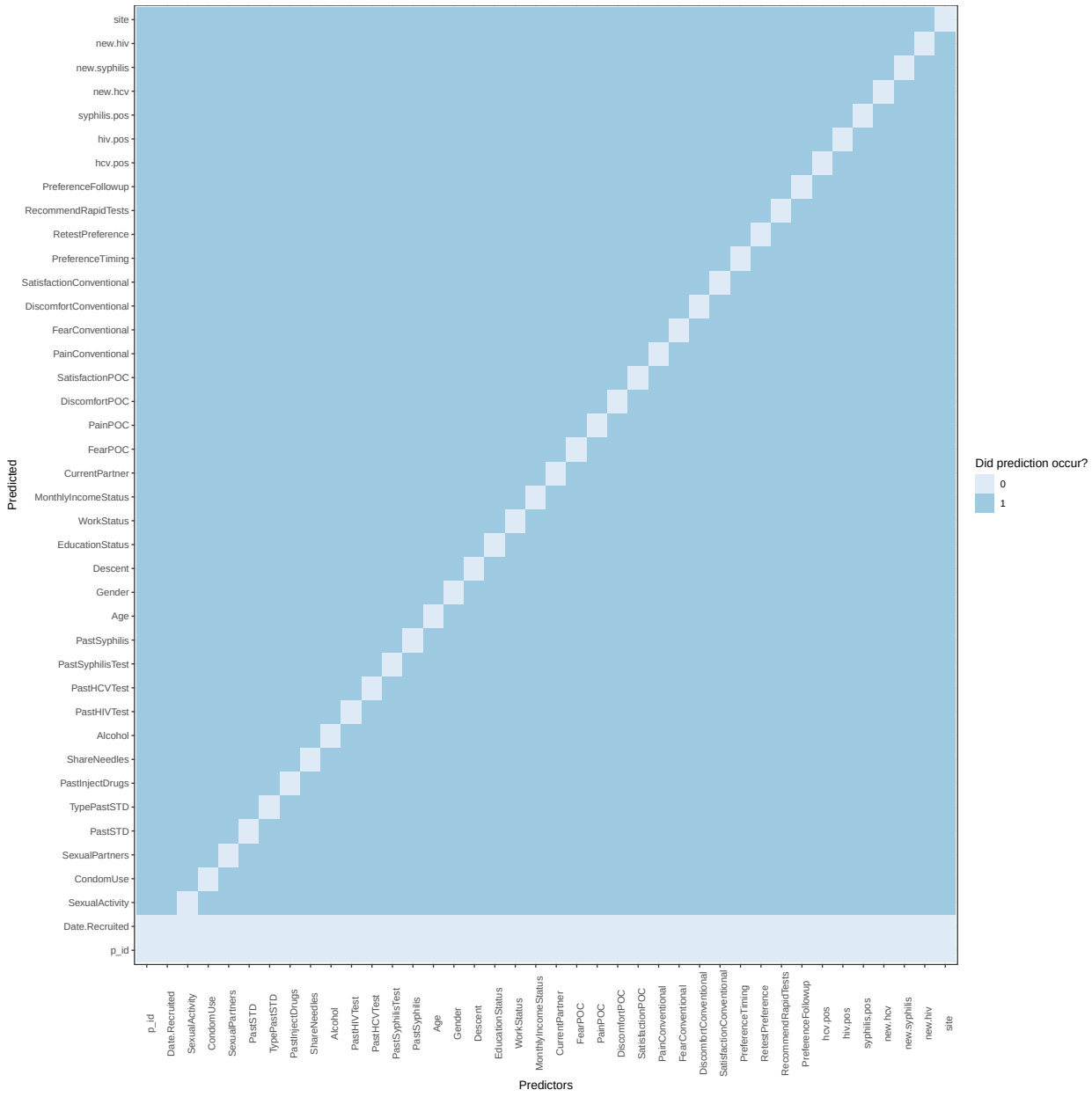


Figure 2: Heatmap of predictor matrix. The columns are the variables predicted and rows show which value were acting as predictors. A value of 0 means no predicion occurred

```
  select(-c(`Proposed.imputation.method (2-level)`))
# table
param.1l |>
  flextable(cwidth = 0.25) |> #turn into flextable object
  merge_v(j=1) |> #Merge cells in first column with same value
  theme_box() |> #Apply a theme for aesthetics
  autofit() |>#automatically set column widths to reasonable values
  fit_to_width(7.5) # size of the page
```

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method
Study ID, date and location	p_id	character	None	None
	Date.Recruited	character	None	None
	Clinic.Location	factor	None	None
Clinical	SexualActivity	factor	Binary	Logistic regression
	CondomUse	factor	Binary	Logistic regression
	SexualPartners	factor	Binary	Logistic regression
	PastSTD	factor	Binary	Logistic regression
	TypePastSTD	factor	None	Bayesian polytomous regression
	PastInjectDrugs	factor	Binary	Logistic regression
	ShareNeedles	factor	Binary	Logistic regression
	Alcohol	factor	Binary	Logistic regression
	PastHIVTest	factor	None	Bayesian polytomous regression
	PastHCVTest	factor	None	Bayesian polytomous regression
	PastSyphilisTest	factor	None	Bayesian polytomous regression
	PastSyphilis	factor	None	Bayesian polytomous regression
Socio-demographic	Age	factor	Binary	Logistic regression
	Gender	factor	None	Bayesian polytomous regression
	Descent	factor	Binary	Logistic regression
	EducationStatus	factor	Binary	Logistic regression
	WorkStatus	factor	Binary	Logistic regression
	MonthlyIncomeStatus	factor	Binary	Logistic regression
	CurrentPartner	binary	None	Logistic regression
	FearPOC	Likert scale	Not sure	Bayesian polytomous regression
	PainPOC	Likert scale	Not sure	Bayesian polytomous regression
	DiscomfortPOC	Likert scale	Not sure	Bayesian polytomous regression
	SatisfactionPOC	Likert scale	Not sure	Bayesian polytomous regression
	PainConventional	Likert scale	Not sure	Bayesian polytomous regression
	FearConventional	Likert scale	Not sure	Bayesian polytomous regression

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method
Patient-reported	DiscomfortConventional	Likert scale	Not sure	Bayesian polytomous regression
	SatisfactionConventional	Likert scale	Not sure	Bayesian polytomous regression
	PreferenceTiming	factor	None	Bayesian polytomous regression
	RetestPreference	factor	Binary	Logistic regression
	RecommendRapidTests	factor	Binary	Logistic regression
	PreferenceFollowup	factor	Binary	Logistic regression
Test results	hcv.pos	integer	None	None
	hiv.pos	integer	None	None
	syphilis.pos	integer	None	None
	new.hcv	integer	None	None
	new.syphilis	integer	None	None
	new.hiv	integer	None	None

Transformation

Transform the selected data to binary variables (see analysis plan)

```
multiplex4imp.1l <- multiplex4imp %>%
  mutate(
    SexualActivity = ifelse(
      SexualActivity %in% c("2-I do not wish to answer", "0-No"),
      "0-No or no wish to answer",
      as.character(SexualActivity)
    ),
    CondomUse = ifelse(
      CondomUse %in% c("1-Always", "2-Sometimes"),
      "1-Ever",
      as.character(CondomUse)
    ),
    SexualPartners = case_when(
      SexualPartners %in% c("0-0", "1-1", "2-2 to 5") ~ "0 to 5",
      SexualPartners %in% c("3-6 to 10", "4-11 or more") ~ "6 or more",
      TRUE ~ as.character(SexualPartners)
    ),
    PastSTD = ifelse(
     PastSTD  %in% c("2-I do not know", "0-No"),
      "0-No or do not know",
      as.character(PastSTD)
    ),
    PastInjectDrugs = ifelse(
      PastInjectDrugs %in% c("2-I do not wish to answer", "0-No"),
      "0-No or no wish to answer",
      as.character(PastInjectDrugs)
    ),
    ShareNeedles = ifelse(
      ShareNeedles %in% c("None", "0-No"),
```

```
    "0-No",
    as.character(ShareNeedles)
  ),
  Alcohol = ifelse(
    Alcohol %in% c("1-1-2 times per week", "2-3-5 times per week"),
    "1-5 times per week",
    as.character(Alcohol)
    ),
    # across(c(PastHIVTest, PastHCVTest, PastSyphilisTest),
    #        ~ifelse(c(.) %in% c("2-Yes; 6 months to 1 year ago",
  #"3-Yes; more than 1 year ago"),
    #                "more than 6 months ago",
    #                as.character(.)))
    # for now uncomment until we know how to deal with the variables
  PastSyphilis = ifelse(
   PastSyphilis %in% c("2-I do not know", "0-No"),
    "0-No or do not know",
    as.character(PastSyphilis)
  ),
  Age = case_when(
    Age %in% c("1-18-24", "1-25-34") ~ "18 to 34",
    Age %in% c("3-35-44", "4-45-54", "5-55 and above") ~ "35 and above",
    TRUE ~ as.character(Age)
  ),
  Gender = ifelse(
    Gender %in% "4-Bisexual",
    NA,
    as.character(Gender)
  ),
  Descent = case_when(
    Descent %in% c("4-European/ North American")
    ~ "Non-Mediterranean European/North American",
    !Descent %in% c("Non-Mediterranean European/North American", NA)
    ~ "Mediterranean European, African, Latin American, Caribbean and/or Asian",
    TRUE ~ as.character(Descent)
  ),
 EducationStatus= ifelse(
   !EducationStatus %in% c("4-Post-graduate degree", NA),
   "No post-graduate degree",
   as.character(EducationStatus)
 ),
 WorkStatus = case_when(
   WorkStatus %in% c("1-Employed (full time)", "2-Employed (part time)")
   ~ "Currently employed",
   !WorkStatus %in% c("Currently employed", NA) ~ "Not currently employed"
 ),
MonthlyIncomeStatus = ifelse(
  !MonthlyIncomeStatus %in% c("1-<$2;000 CAD", NA),
  "$2000 CAD and above",
  as.character(MonthlyIncomeStatus)
),
PreferenceTiming = ifelse(
  !PreferenceTiming %in% c("0-1 day", NA),
```

```
    "2 days or more",
    as.character(PreferenceTiming)
  ),
  RecommendRapidTests = ifelse(
    RecommendRapidTests %in% c("0-No", "2-Not Sure"),
    "No or not sure",
    as.character(RecommendRapidTests)
  ),
  PreferenceFollowup = case_when(
   PreferenceFollowup %in% c("3-Face-to-face", "4-No preference")
   ~ "No preference for digital follow-up",
   !PreferenceFollowup %in% c("No preference for digital follow-up", NA)
   ~"Digital follow-up",
   TRUE ~ as.character(PreferenceFollowup)
  )
      ) %>%
  mutate(across(SexualActivity:PreferenceFollowup, as.factor))

summary(multiplex4imp.1l)
```

```
##      p_id           Date.Recruited                      SexualActivity
##  Length:400         Length:400         0-No or no wish to answer:106
##  Class :character   Class :character   1-Yes                    :289
##  Mode  :character   Mode  :character   NA's                     :  5
##
##
##
##
##    CondomUse     SexualPartners               PastSTD           TypePastSTD
##  0-Never: 98   0 to 5   :324    0-No or do not know:216   None         :216
##  1-Ever :272   6 or more: 60    1-Yes              :169   3-Chlamydia  : 72
##  NA's   : 30   NA's     : 16    NA's               : 15   4-Gonorrhoeae: 43
##                                                           6-Other      : 25
##                                                           0-HIV        : 10
##                                                           (Other)      : 13
##                                                           NA's         : 21
##                   PastInjectDrugs ShareNeedles              Alcohol
##  0-No or no wish to answer:255    0-No :307    0-No              :209
##  1-Yes                    :139    1-Yes: 85    1-5 times per week:187
##  Yes                      :  1    NA's :  8    NA's              :  4
##  NA's                     :  5
##
##
##
##                         PastHIVTest                          PastHCVTest
##  0-No                          : 57   0-No                          : 44
##  1-Yes; less than 6 months ago:140   1-Yes; less than 6 months ago:139
##  2-Yes; 6 months to 1 year ago: 47   2-Yes; 6 months to 1 year ago: 48
##  3-Yes; more than 1 year ago  :154   3-Yes; more than 1 year ago  :167
##  NA's                          :  2   NA's                          :  2
##
##
##                       PastSyphilisTest             PastSyphilis
```

```
##  0-No                          :  4     0-No or do not know:348
##  1-Yes; less than 6 months ago:127     1-Yes               : 48
##  2-Yes; 6 months to 1 year ago: 47     NA's                :  4
##  3-Yes; more than 1 year ago  :219
##  NA's                          :  3
##
##
##           Age                Gender
##  18 to 34    :156   1-Male       :255
##  35 and above:197   2-Female     : 73
##  NA's        : 47   3-Transgender:  7
##                     5-No response: 11
##                     NA's         : 54
##
##
##                                                                    Descent
##  Mediterranean European, African, Latin American, Caribbean and/or Asian: 99
##  Non-Mediterranean European/North American                              :254
##  NA's                                                                   : 47
##
##
##
##
##                 EducationStatus                   WorkStatus
##  4-Post-graduate degree :120    Currently employed    :188
##  No post-graduate degree:233    Not currently employed:165
##  NA's                   : 47    NA's                  : 47
##
##
##
##
##          MonthlyIncomeStatus CurrentPartner        FearPOC
##  $2000 CAD and above:134      0-No :213      0-1-no fear:155
##  1-<$2;000 CAD      :219      1-Yes:140      1-2        : 52
##  NA's               : 47      NA's : 47      2-3        : 19
##                                              3-4        :  9
##                                              4-5-highest:  7
##                                              NA's       :158
##
##        PainPOC               DiscomfortPOC            SatisfactionPOC
##  0-1-no pain:166   0-1 - no discomfort:181   0-Very satisfied   :207
##  1-2        : 59   1-2                : 46   1-Satisfied        : 30
##  2-3        : 15   2-3                : 11   2-Neutral/ not sure:  3
##  3-4        :  1   3-4                :  2   NA's               :160
##  NA's       :159   4-5-highest        :  1
##                    NA's               :159
##
##    PainConventional    FearConventional         DiscomfortConventional
##  0-1-no pain:159     0-1-no fear:150     0-1 - no discomfort:173
##  1-2        : 61     1-2        : 45     1-2                : 43
##  2-3        : 10     2-3        : 21     2-3                : 12
##  3-4        :  1     3-4        :  9     3-4                :  2
##  NA's       :169     4-5-highest:  5     4-5-highest        :  1
##                      NA's       :170     NA's               :169
```

```
##
##         SatisfactionConventional       PreferenceTiming        RetestPreference
##  0-Very satisfied   :190         0-1 day       :137     0-Rapid test   :144
##  1-Satisfied        : 33         2 days or more:106     1-Phlebotomy   : 66
##  2-Neutral/ not sure:  7         NA's          :157     2-No preference: 32
##  3-Dissatisfied     :  1                                NA's           :158
##  NA's               :169
##
##
##      RecommendRapidTests                              PreferenceFollowup
##  1-Yes         :227      Digital follow-up                  :156
##  No or not sure: 16      No preference for digital follow-up: 87
##  NA's          :157      NA's                               :157
##
##
##
##
##     hcv.pos          hiv.pos      syphilis.pos       new.hcv
##  Min.   :0.0000   Min.   :0.00   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.00   Median :0.0000   Median :0.000
##  Mean   :0.1475   Mean   :0.04   Mean   :0.0125   Mean   :0.035
##  3rd Qu.:0.0000   3rd Qu.:0.00   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.00   Max.   :1.0000   Max.   :1.000
##
##   new.syphilis      new.hiv           site
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.000   Median :0.000   Median :0.0000
##  Mean   :0.005   Mean   :0.005   Mean   :0.4075
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000
##
```

Run imputation with logistic regression for transformed factors

```
# mice.1l <- mice(multiplex4imp.1l, nnet.MaxNWts = 2000, print = F, )
#
#
# summary(complete(mice.1l))
#
# # save results
# mice.1l.long<- complete(mice.1l, "long")
#
# mice.1l.broad <- complete(mice.1l, "broad")
#
# write.csv(mice.1l.long, "3_intermediate/imputed/imputed_data_long_1l.csv",
#           row.names = F)
#
# write.csv(mice.1l.broad, "3_intermediate/imputed/imputed_data_broad_1l.csv",
#           row.names = F)
```

Not working for now, try to fix later

Prediction matrix and methods

```
# prediction.matrix(mice.1l$predictorMatrix)
#
# meth.1l <- mice.1l$method
# meth.1l
```

The variables transformed into binary factors are now imputed using logistic regression

Diagnostic plots

```
# plot(mice.1l)
```

Variables with small amount of missing data (e.g. PastHIv/HCV/SyphilisTest) have unstable diagnostic plots.

1-level predictive mean matching

Proposed imputation plan

```
param.pmm <- param.1l %>%
  mutate(Proposed.imputation.method = ifelse(
    Proposed.imputation.method %in% "None",
    "None",
    "Predictive mean matching"
  ))
# table
param.pmm |>
  flextable(cwidth = 0.25) |> #turn into flextable object
  merge_v(j=1) |> #Merge cells in first column with same value
  theme_box() |> #Apply a theme for aesthetics
  autofit() |>#automatically set column widths to reasonable values
  fit_to_width(7.5)
```

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method
Study ID, date and location	p_id	character	None	None
	Date.Recruited	character	None	None
	Clinic.Location	factor	None	None
	SexualActivity	factor	Binary	Predictive mean matching
	CondomUse	factor	Binary	Predictive mean matching
	SexualPartners	factor	Binary	Predictive mean matching
	PastSTD	factor	Binary	Predictive mean matching
	TypePastSTD	factor	None	Predictive mean matching

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method
Clinical	PastInjectDrugs	factor	Binary	Predictive mean matching
	ShareNeedles	factor	Binary	Predictive mean matching
	Alcohol	factor	Binary	Predictive mean matching
	PastHIVTest	factor	None	Predictive mean matching
	PastHCVTest	factor	None	Predictive mean matching
	PastSyphilisTest	factor	None	Predictive mean matching
	PastSyphilis	factor	None	Predictive mean matching
Socio-demographic	Age	factor	Binary	Predictive mean matching
	Gender	factor	None	Predictive mean matching
	Descent	factor	Binary	Predictive mean matching
	EducationStatus	factor	Binary	Predictive mean matching
	WorkStatus	factor	Binary	Predictive mean matching
	MonthlyIncomeStatus	factor	Binary	Predictive mean matching
	CurrentPartner	binary	None	Predictive mean matching
Patient-reported	FearPOC	Likert scale	Not sure	Predictive mean matching
	PainPOC	Likert scale	Not sure	Predictive mean matching
	DiscomfortPOC	Likert scale	Not sure	Predictive mean matching
	SatisfactionPOC	Likert scale	Not sure	Predictive mean matching
	PainConventional	Likert scale	Not sure	Predictive mean matching
	FearConventional	Likert scale	Not sure	Predictive mean matching
	DiscomfortConventional	Likert scale	Not sure	Predictive mean matching
	SatisfactionConventional	Likert scale	Not sure	Predictive mean matching
	PreferenceTiming	factor	None	Predictive mean matching
	RetestPreference	factor	Binary	Predictive mean matching
	RecommendRapidTests	factor	Binary	Predictive mean matching
	PreferenceFollowup	factor	Binary	Predictive mean matching
Test results	hcv.pos	integer	None	None
	hiv.pos	integer	None	None
	syphilis.pos	integer	None	None
	new.hcv	integer	None	None
	new.syphilis	integer	None	None
	new.hiv	integer	None	None

```
mice.pmm <- mice(multiplex4imp.1l, method = "pmm", nnet.MaxNWts = 2000, print = F)

summary(complete(mice.pmm))
```

```
##      p_id           Date.Recruited                       SexualActivity
##  Length:400         Length:400         0-No or no wish to answer:107
##  Class :character   Class :character   1-Yes                    :293
##  Mode  :character   Mode  :character
##
##
##
##
##    CondomUse     SexualPartners                PastSTD
##  0-Never:106   0 to 5   :336    0-No or do not know:223
##  1-Ever :294   6 or more: 64    1-Yes              :177
##
##
##
##
##
##            TypePastSTD                   PastInjectDrugs ShareNeedles
##  0-HIV           : 10   0-No or no wish to answer:259    0-No :313
##  2-Herpes simplex:  3   1-Yes                    :140    1-Yes: 87
##  3-Chlamydia     : 79   Yes                      :  1
##  4-Gonorrhoeae   : 47
##  5-Syphilis      : 10
##  6-Other         : 26
##  None            :225
##                Alcohol                           PastHIVTest
##  0-No              :212   0-No                          : 58
##  1-5 times per week:188   1-Yes; less than 6 months ago:140
##                           2-Yes; 6 months to 1 year ago: 47
##                           3-Yes; more than 1 year ago  :155
##
##
##
##                         PastHCVTest                        PastSyphilisTest
##  0-No                          : 45   0-No                           :  5
##  1-Yes; less than 6 months ago:139   1-Yes; less than 6 months ago:128
##  2-Yes; 6 months to 1 year ago: 49   2-Yes; 6 months to 1 year ago: 48
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :219
##
##
##
##               PastSyphilis           Age                Gender
##  0-No or do not know:351   18 to 34    :162   1-Male       :292
##  1-Yes              : 49   35 and above:238   2-Female     : 86
##                                               3-Transgender:  8
##                                               5-No response: 14
##
##
##
##                                                                    Descent
##  Mediterranean European, African, Latin American, Caribbean and/or Asian:115
##  Non-Mediterranean European/North American                               :285
##
##
##
```

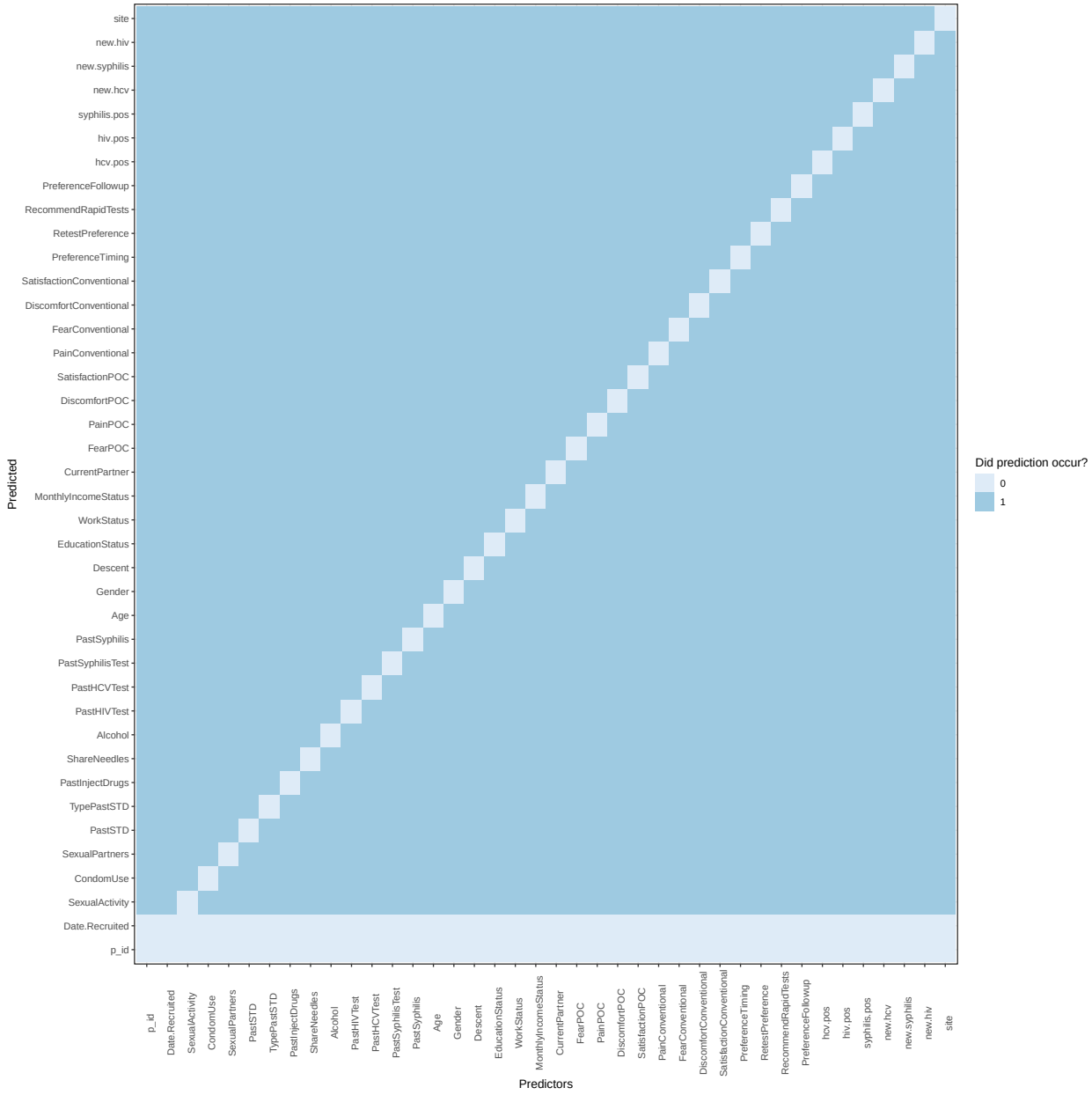
```
##
##
##                 EducationStatus                    WorkStatus
##  4-Post-graduate degree :140    Currently employed    :202
##  No post-graduate degree:260    Not currently employed:198
##
##
##
##
##
##           MonthlyIncomeStatus CurrentPartner        FearPOC
##  $2000 CAD and above:147      0-No :246      0-1-no fear:295
##  1-<$2;000 CAD      :253      1-Yes:154      1-2        : 55
##                                              2-3        : 23
##                                              3-4        : 16
##                                              4-5-highest: 11
##
##
##         PainPOC                DiscomfortPOC            SatisfactionPOC
##  0-1-no pain:178   0-1 - no discomfort:316   0-Very satisfied   :353
##  1-2        : 62   1-2                : 65   1-Satisfied        : 40
##  2-3        :137   2-3                : 15   2-Neutral/ not sure:  7
##  3-4        : 23   3-4                :  2
##                    4-5-highest        :  2
##
##
##    PainConventional    FearConventional         DiscomfortConventional
##  0-1-no pain:188     0-1-no fear:170     0-1 - no discomfort:199
##  1-2        : 96     1-2        : 49     1-2                : 60
##  2-3        : 93     2-3        : 25     2-3                : 62
##  3-4        : 23     3-4        : 37     3-4                : 54
##                      4-5-highest:119     4-5-highest        : 25
##
##
##         SatisfactionConventional       PreferenceTiming        RetestPreference
##  0-Very satisfied   :341         0-1 day       :256     0-Rapid test   :261
##  1-Satisfied        : 35         2 days or more:144     1-Phlebotomy   : 92
##  2-Neutral/ not sure: 22                               2-No preference: 47
##  3-Dissatisfied     :  2
##
##
##
##     RecommendRapidTests                            PreferenceFollowup
##  1-Yes         :351      Digital follow-up                  :184
##  No or not sure: 49      No preference for digital follow-up:216
##
##
##
##
##
##    hcv.pos          hiv.pos      syphilis.pos       new.hcv
##  Min.   :0.0000   Min.   :0.00   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000   Median :0.00   Median :0.0000   Median :0.000
```

```
##  Mean   :0.1475   Mean   :0.04   Mean   :0.0125   Mean   :0.035
##  3rd Qu.:0.0000   3rd Qu.:0.00   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.00   Max.   :1.0000   Max.   :1.000
##
##   new.syphilis      new.hiv           site
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.000   Median :0.000   Median :0.0000
##  Mean   :0.005   Mean   :0.005   Mean   :0.4075
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000
##
```

```
# save
mice.pmm.long<- complete(mice.pmm, "long")

mice.pmm.broad <- complete(mice.pmm, "broad")

write.csv(mice.pmm.long, "3_intermediate/imputed/imputed_data_long_pmm.csv",
          row.names = F)

write.csv(mice.pmm.broad, "3_intermediate/imputed/imputed_data_broad_pmm.csv",
           row.names = F)
```

Prediction matrix and methods

```
prediction.matrix(mice.pmm$predictorMatrix)
```



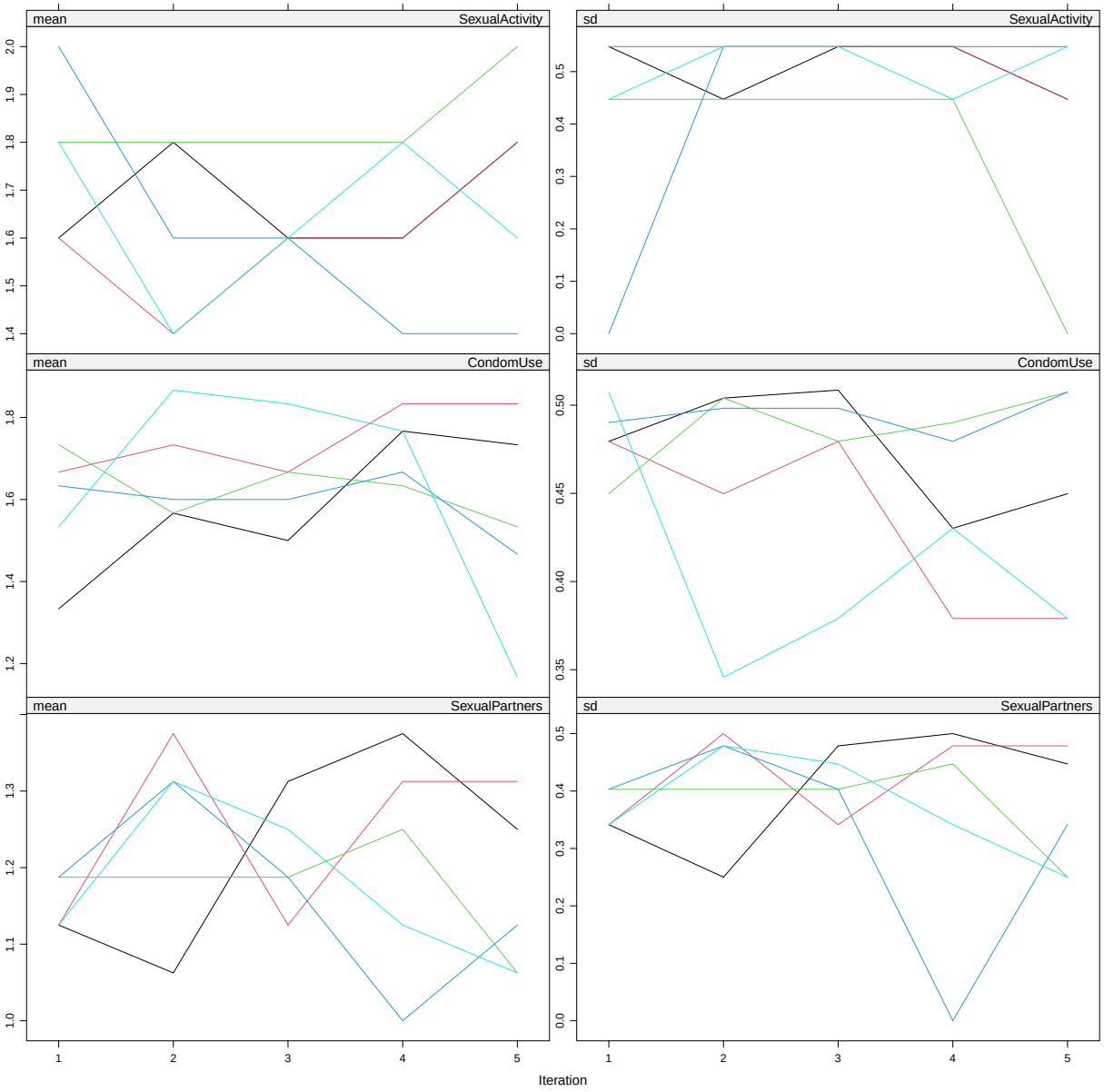

```
meth.pmm <- mice.pmm$method
meth.pmm
```

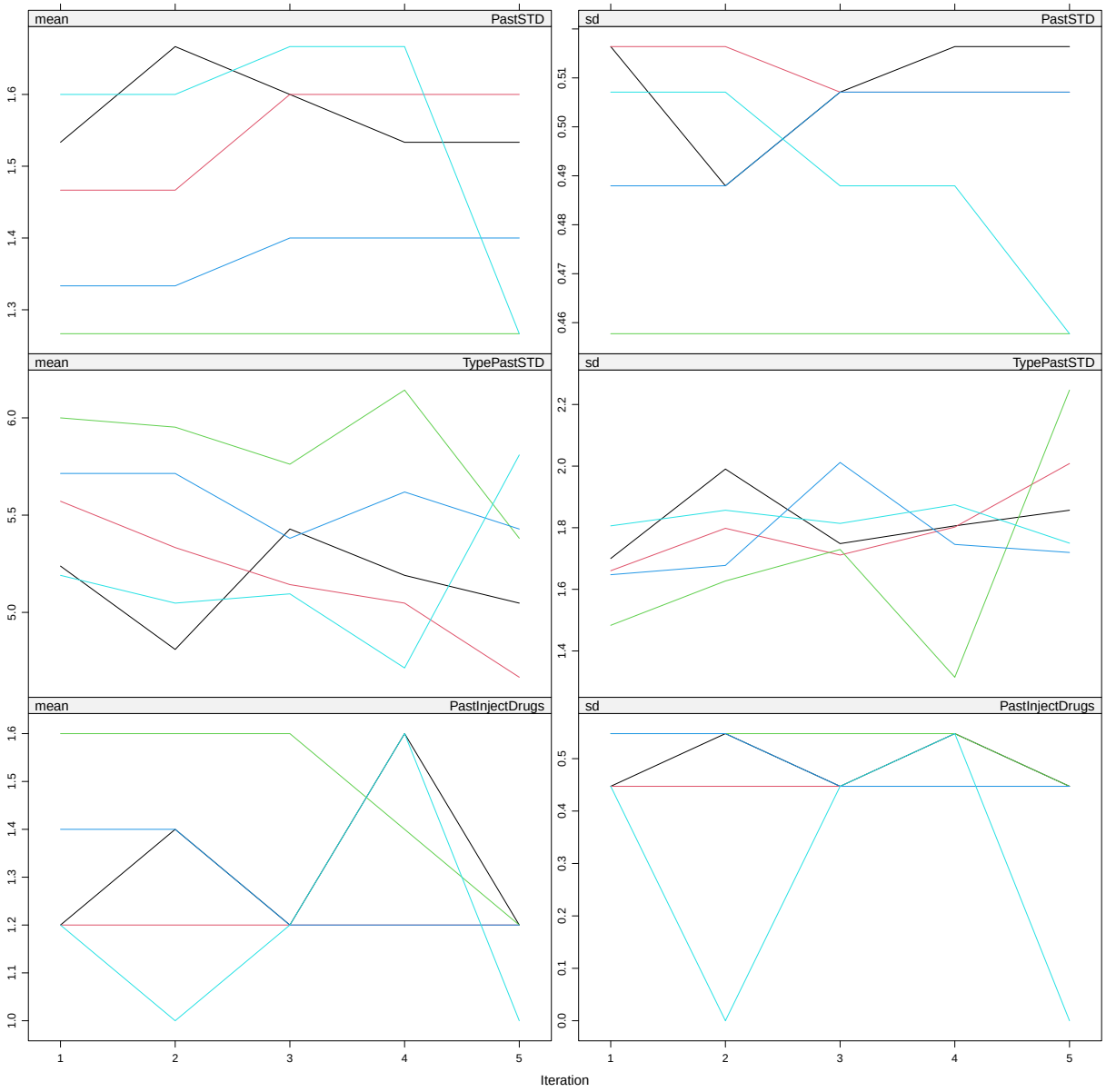
```
##                    p_id          Date.Recruited          SexualActivity
##                      ""                      ""                   "pmm"
##             CondomUse          SexualPartners                 PastSTD
##                   "pmm"                   "pmm"                   "pmm"
##           TypePastSTD          PastInjectDrugs            ShareNeedles
##                   "pmm"                   "pmm"                   "pmm"
##               Alcohol             PastHIVTest             PastHCVTest
##                   "pmm"                   "pmm"                   "pmm"
##     PastSyphilisTest             PastSyphilis                     Age
##                   "pmm"                   "pmm"                   "pmm"
```

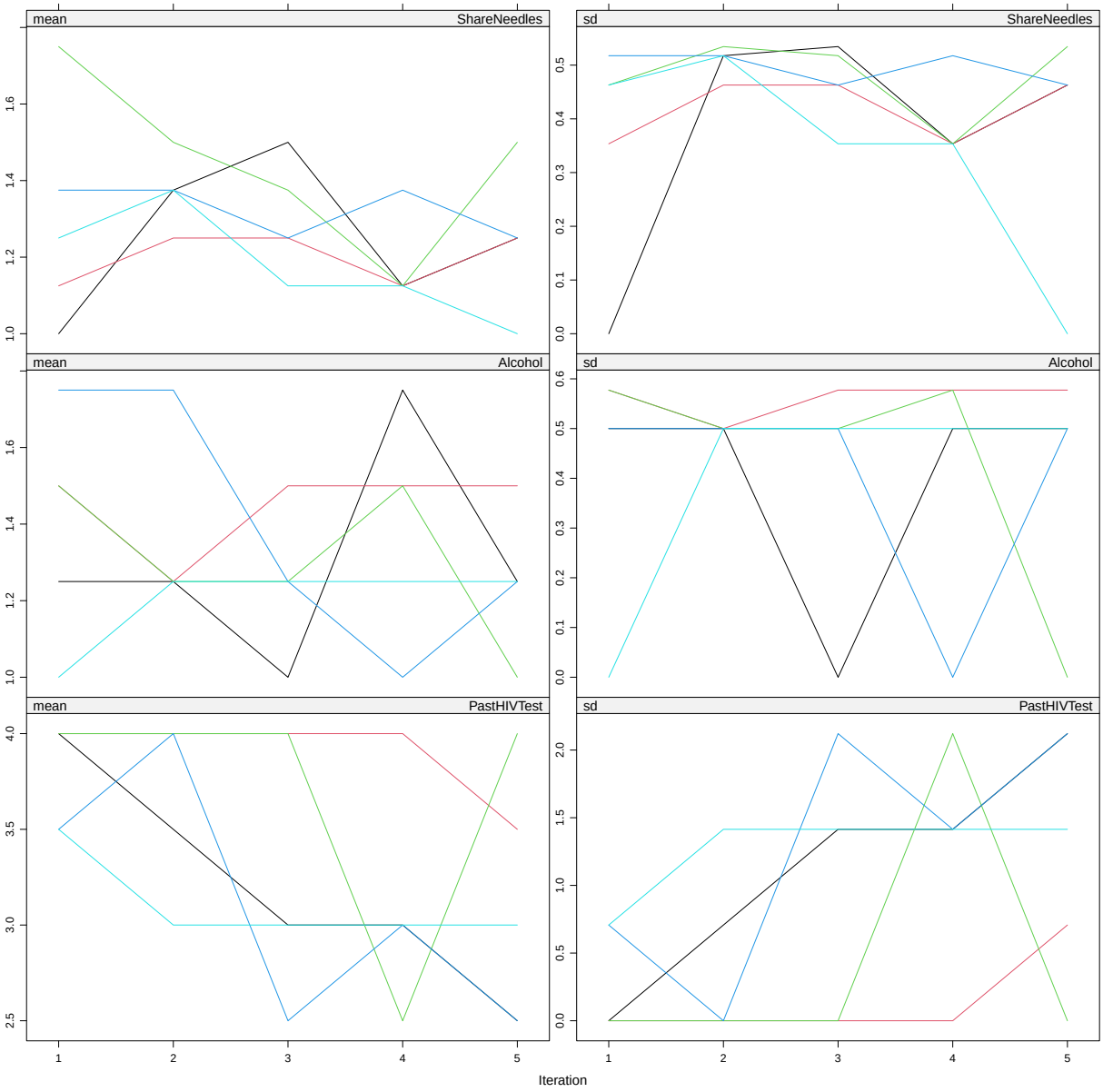
```
##                     Gender                    Descent           EducationStatus
##                      "pmm"                      "pmm"                      "pmm"
##                 WorkStatus       MonthlyIncomeStatus            CurrentPartner
##                      "pmm"                      "pmm"                      "pmm"
##                   FearPOC                   PainPOC             DiscomfortPOC
##                      "pmm"                      "pmm"                      "pmm"
##           SatisfactionPOC          PainConventional          FearConventional
##                      "pmm"                      "pmm"                      "pmm"
##    DiscomfortConventional SatisfactionConventional          PreferenceTiming
##                      "pmm"                      "pmm"                      "pmm"
##          RetestPreference       RecommendRapidTests       PreferenceFollowup
##                      "pmm"                      "pmm"                      "pmm"
##                   hcv.pos                   hiv.pos              syphilis.pos
##                         ""                         ""                         ""
##                   new.hcv              new.syphilis                   new.hiv
##                         ""                         ""                         ""
##                      site
##                         ""
```

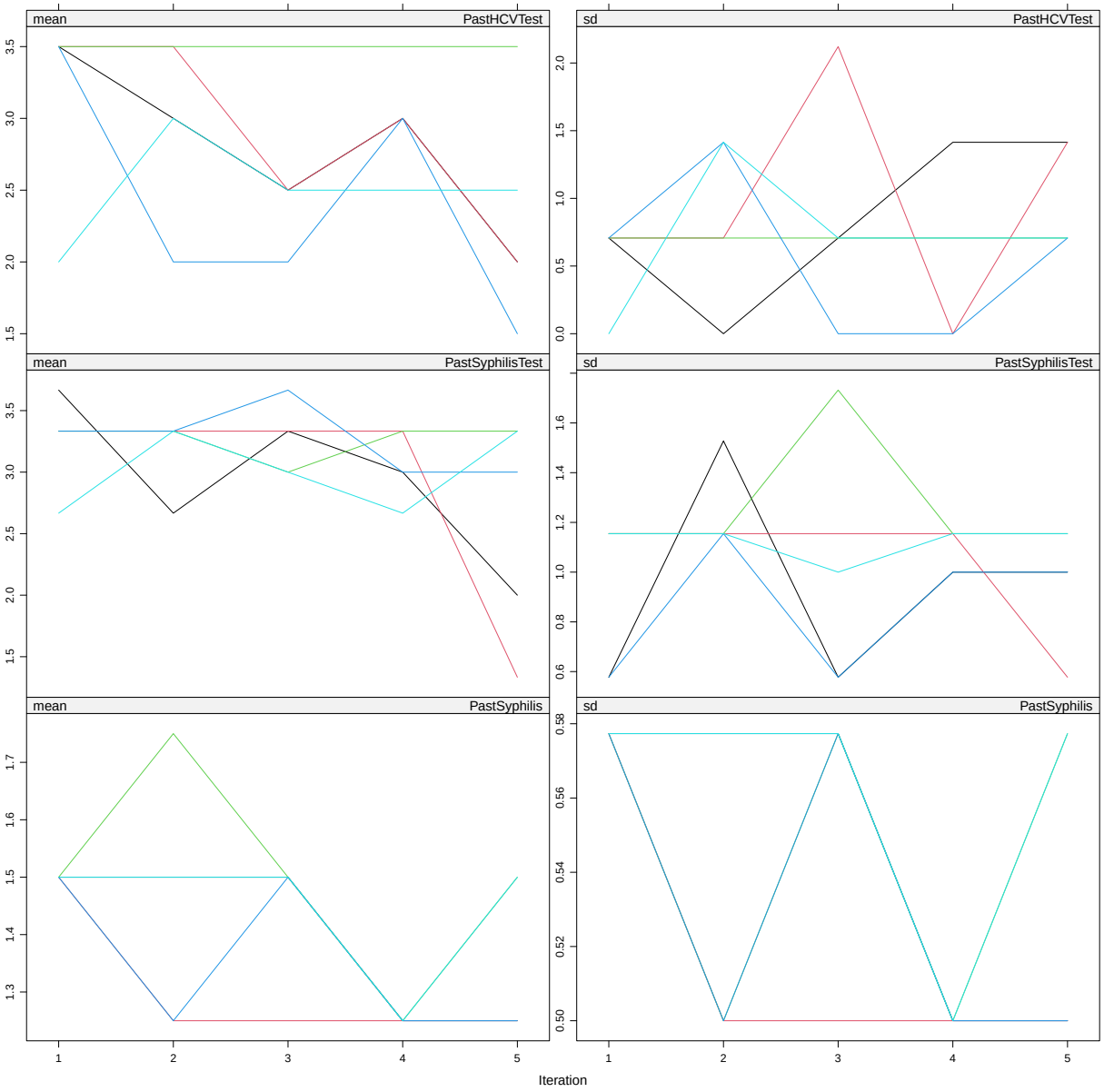
Diagnostic plots

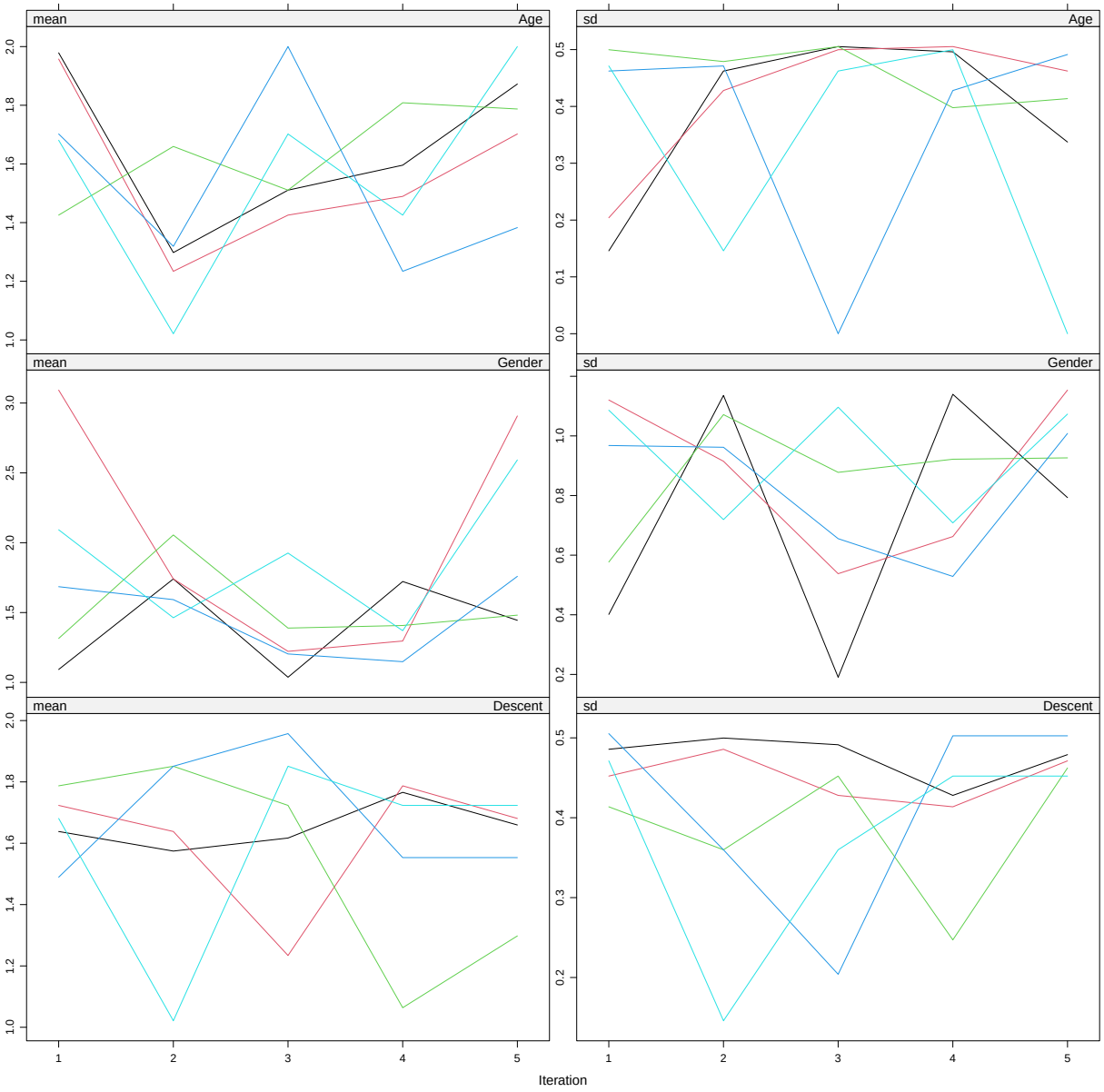
```
plot(mice.pmm)
```

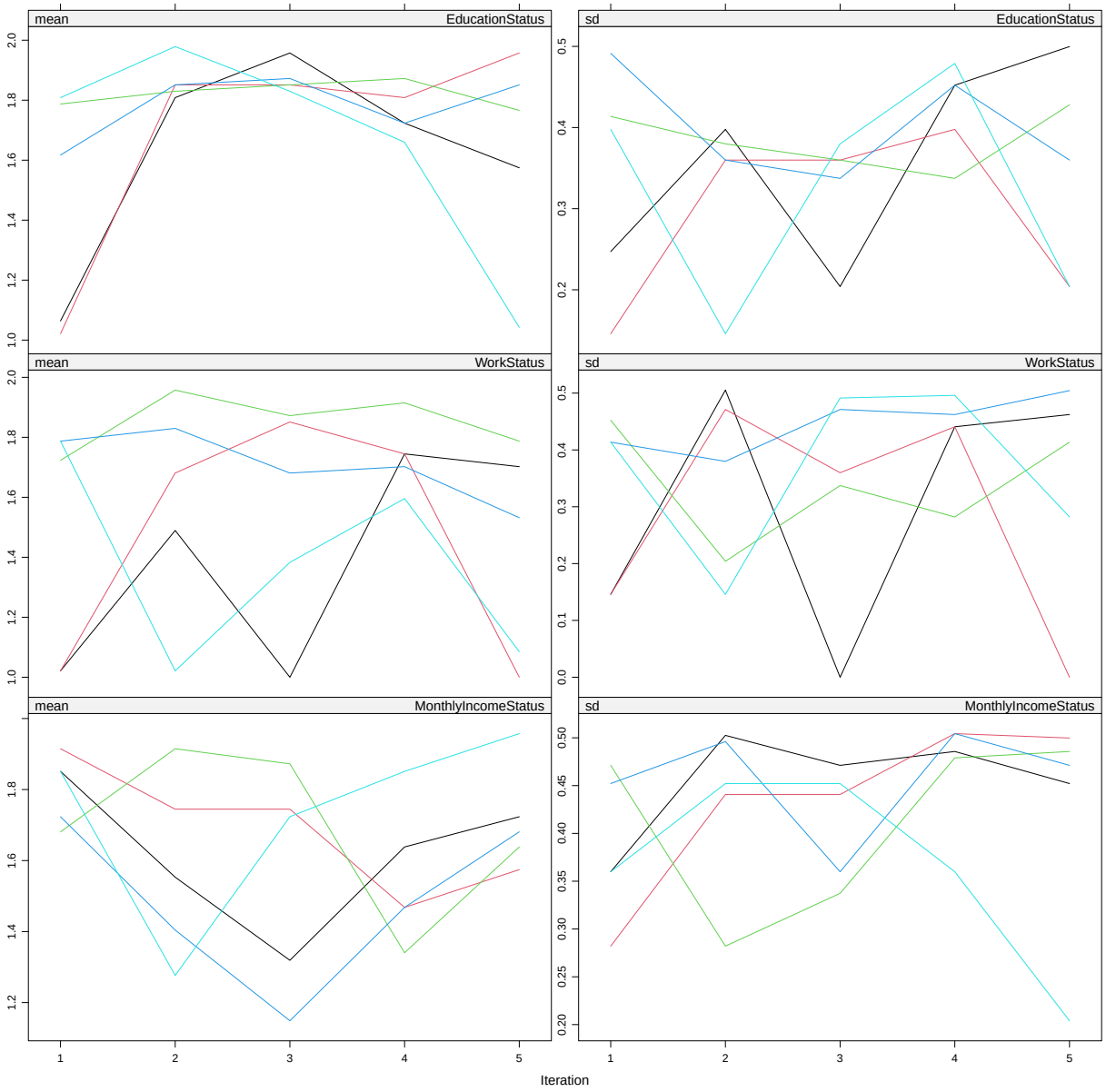

mean
SexualActivity
sd
SexualActivity
mean
CondomUse
sd
CondomUse
mean
SexualPartners
sd
SexualPartners
Iteration

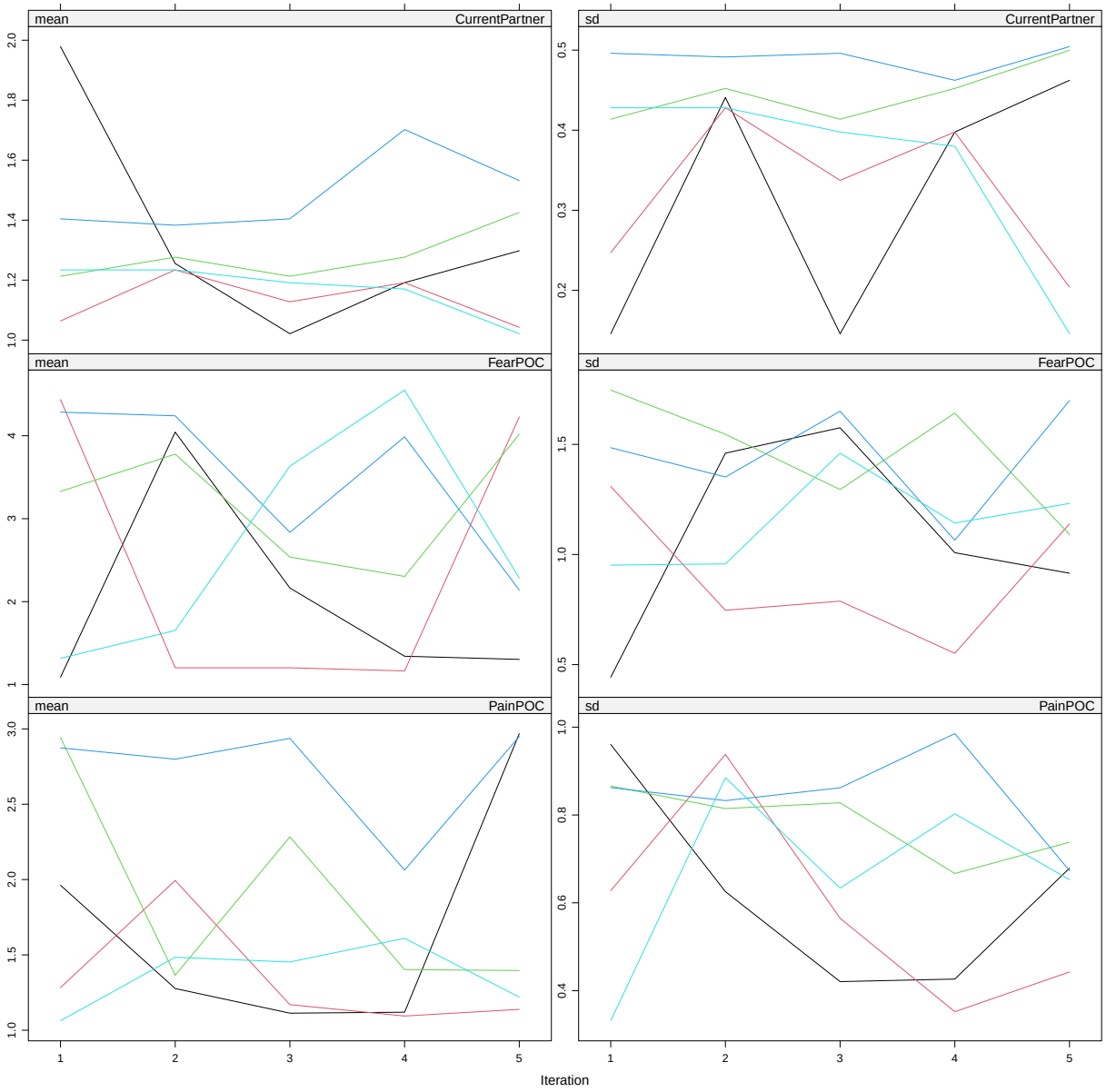

mean
PastSTD
sd
PastSTD
mean
TypePastSTD
sd
TypePastSTD
mean
PastInjectDrugs
sd
PastInjectDrugs
Iteration

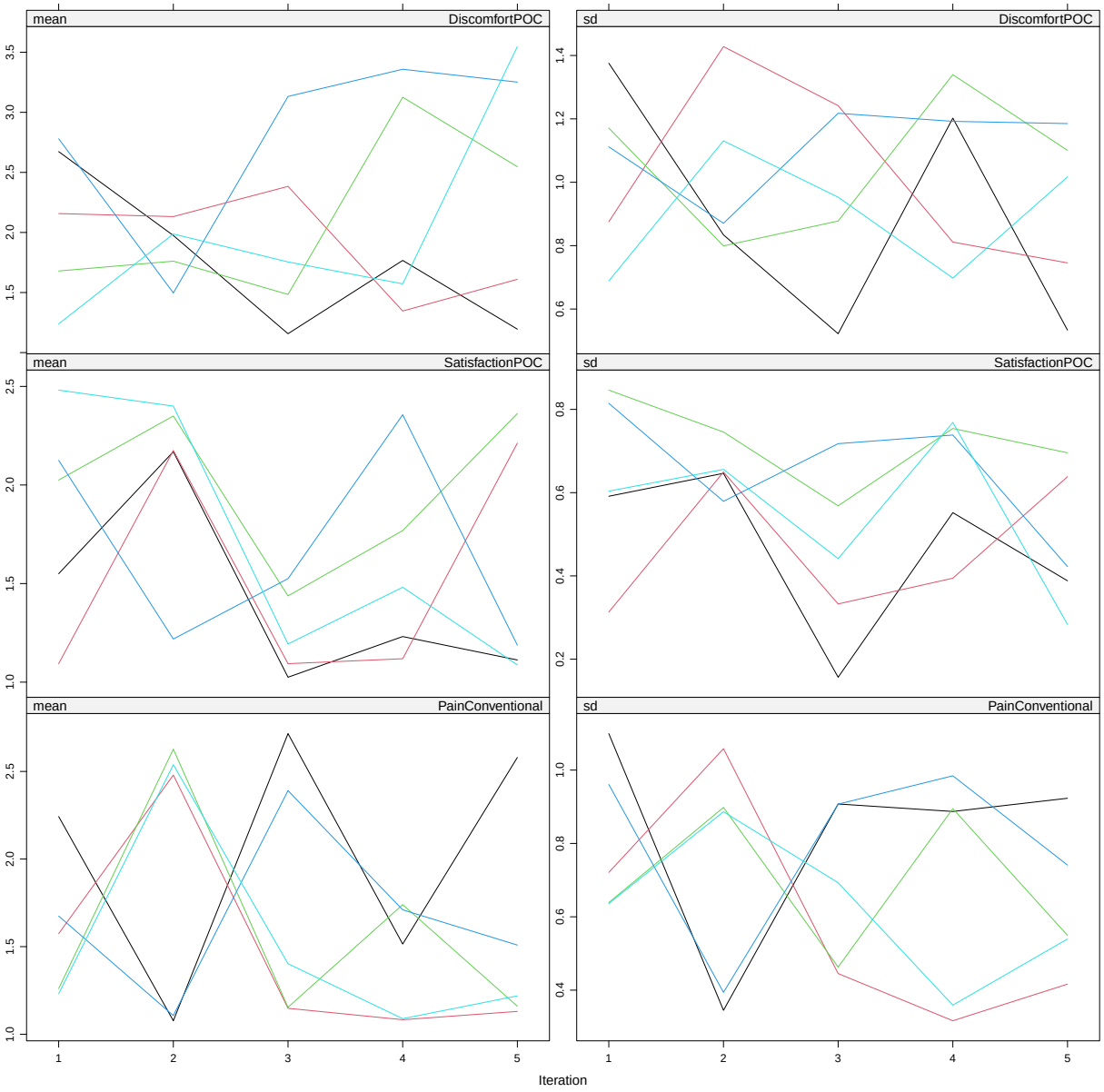

mean
ShareNeedles
sd
ShareNeedles
mean
Alcohol
sd
Alcohol
mean
PastHIVTest
sd
PastHIVTest
Iteration

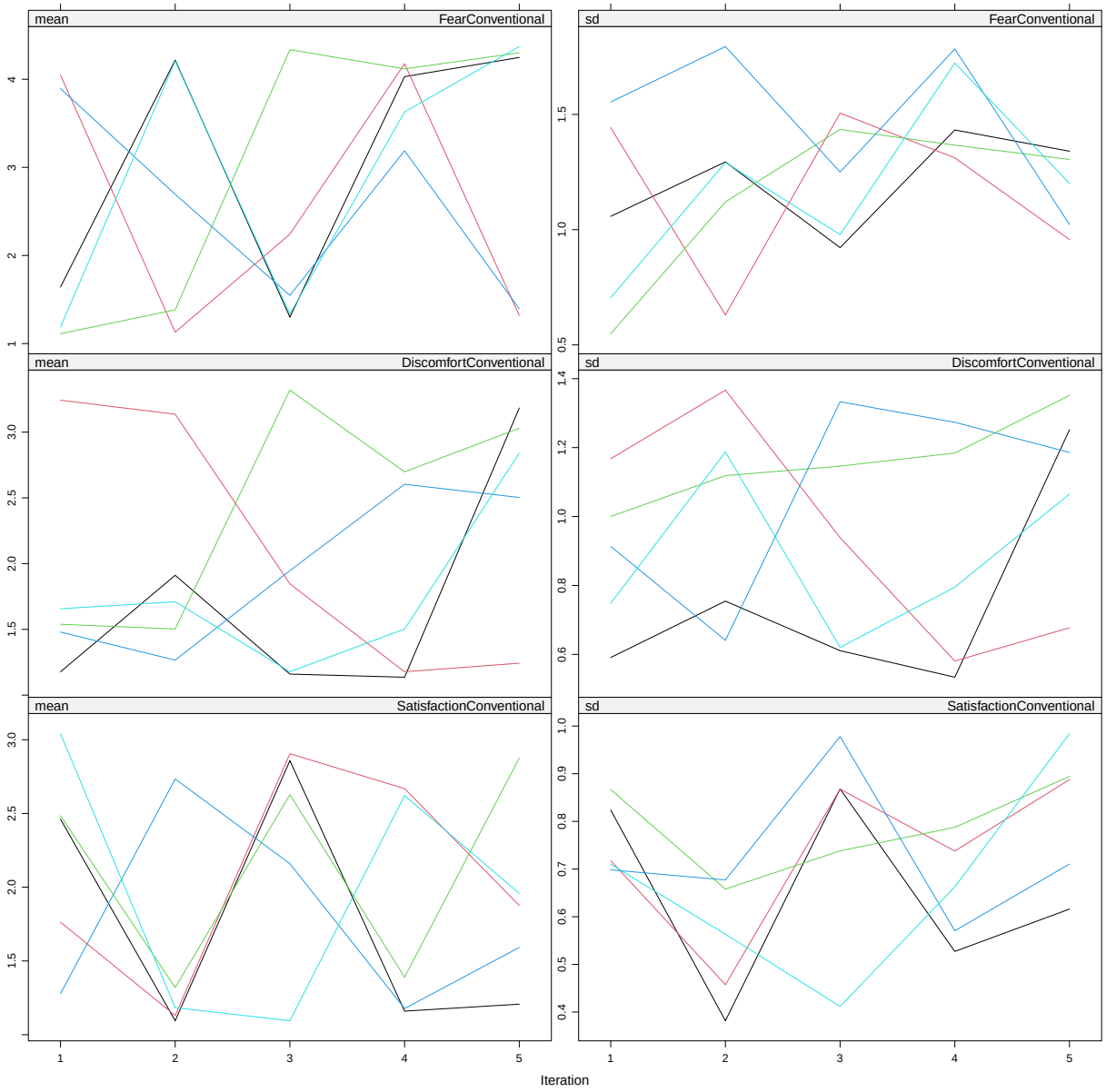

mean
PastHCVTest
sd
PastHCVTest
mean
PastSyphilisTest
sd
PastSyphilisTest
mean
PastSyphilis
sd
PastSyphilis
Iteration

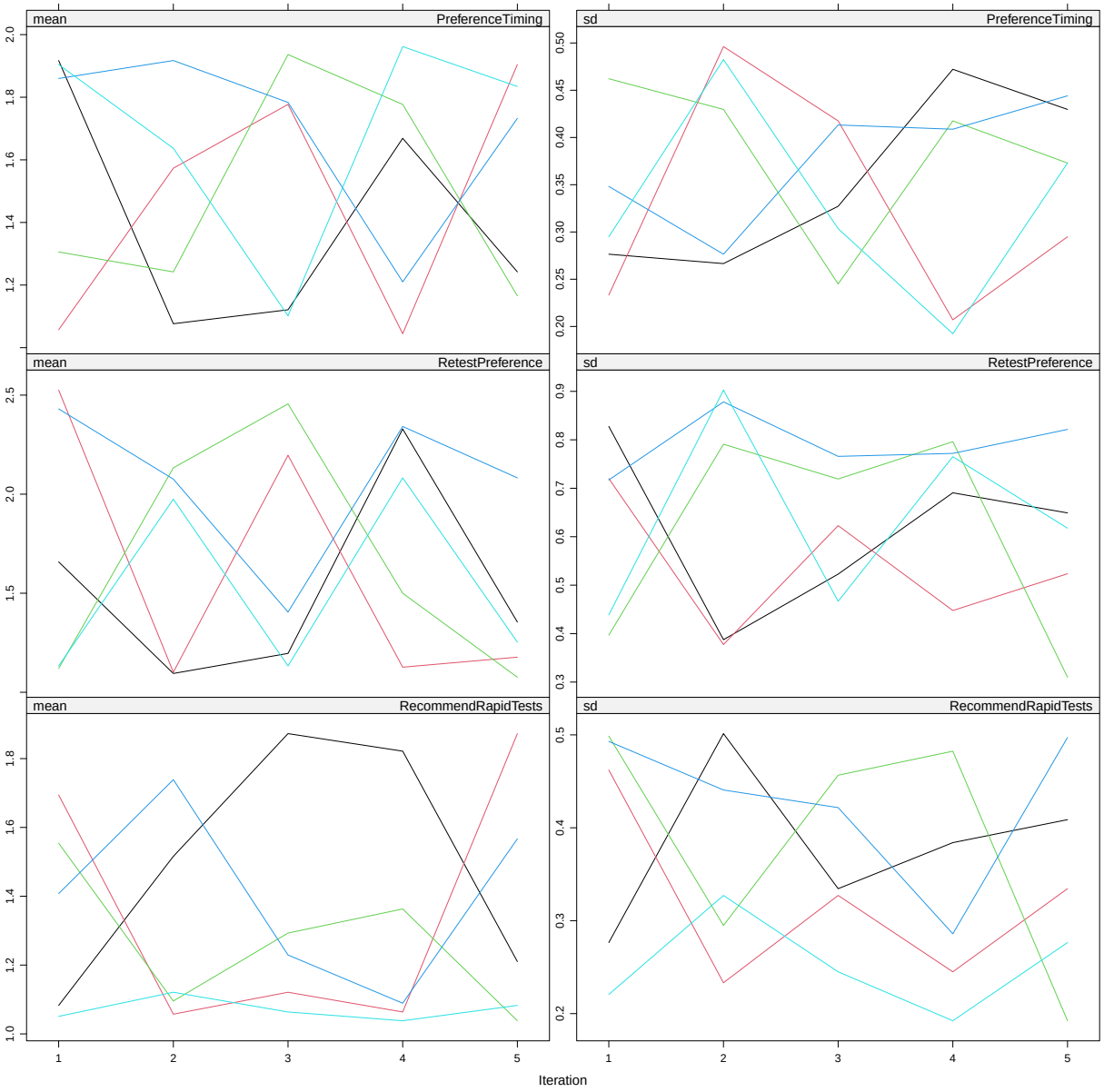

mean
Age
sd
Age
mean
Gender
sd
Gender
mean
Descent
sd
Descent
Iteration

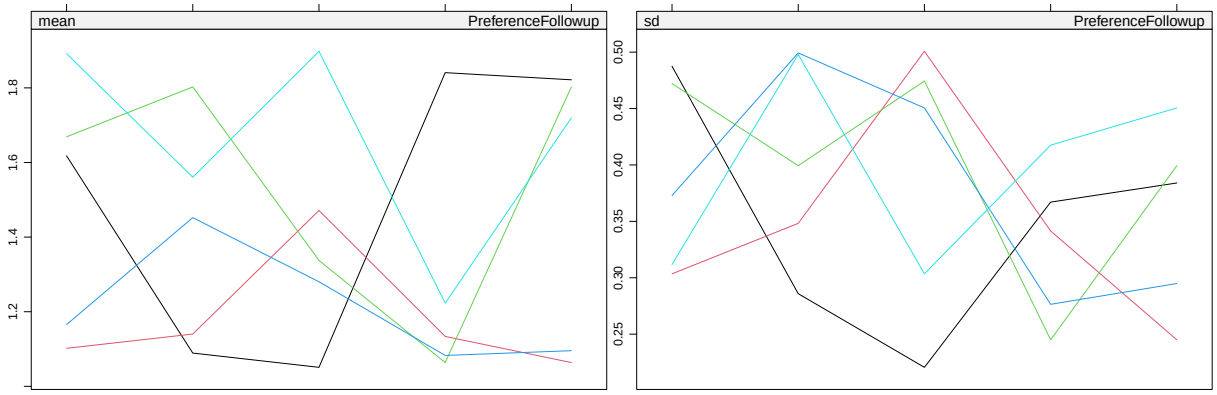

mean
EducationStatus
sd
EducationStatus
mean
WorkStatus
sd
WorkStatus
mean
MonthlyIncomeStatus
sd
MonthlyIncomeStatus
Iteration


mean
CurrentPartner
sd
CurrentPartner
mean
FearPOC
sd
FearPOC
mean
PainPOC
sd
PainPOC
Iteration


mean
DiscomfortPOC
sd
DiscomfortPOC
mean
SatisfactionPOC
sd
SatisfactionPOC
mean
PainConventional
sd
PainConventional
Iteration


mean
FearConventional
sd
FearConventional
mean
DiscomfortConventional
sd
DiscomfortConventional
mean
SatisfactionConventional
sd
SatisfactionConventional
Iteration


mean
PreferenceTiming
sd
PreferenceTiming
mean
RetestPreference
sd
RetestPreference
mean
RecommendRapidTests
sd
RecommendRapidTests
Iteration



2-level imputation

Proposed imputation method

```
param.2l <- param %>%
  select(-c(Proposed.imputation.method))
# table
param.2l |>
  flextable(cwidth = 0.25) |> #turn into flextable object
  merge_v(j=1) |> #Merge cells in first column with same value
  theme_box() |> #Apply a theme for aesthetics
```

```
autofit() |>#automatically set column widths to reasonable values
fit_to_width(7.5)
```

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method (2-level)
Study ID, date and location	p_id	character	None	None
	Date.Recruited	character	None	None
	Clinic.Location	factor	None	None (cluster)
Clinical	SexualActivity	factor	Binary	Binary (1-level)
	CondomUse	factor	Binary	Predictive mean matching
	SexualPartners	factor	Binary	Binary (1-level)
	PastSTD	factor	Binary	Binary (1-level)
	TypePastSTD	factor	None	Predictive mean matching
	PastInjectDrugs	factor	Binary	Binary (1-level)
	ShareNeedles	factor	Binary	Binary (1-level)
	Alcohol	factor	Binary	Not sure
	PastHIVTest	factor	None	Not sure
	PastHCVTest	factor	None	Not sure
	PastSyphilisTest	factor	None	Binary (1-level)
	PastSyphilis	factor	None	Binary (1-level)
Socio-demographic	Age	factor	Binary	Predictive mean matching
	Gender	factor	None	Predictive mean matching
	Descent	factor	Binary	Predictive mean matching
	EducationStatus	factor	Binary	Predictive mean matching
	WorkStatus	factor	Binary	Predictive mean matching
	MonthlyIncomeStatus	factor	Binary	Predictive mean matching
	CurrentPartner	binary	None	Predictive mean matching
	FearPOC	Likert scale	Not sure	Predictive mean matching
	PainPOC	Likert scale	Not sure	Predictive mean matching
	DiscomfortPOC	Likert scale	Not sure	Predictive mean matching
	SatisfactionPOC	Likert scale	Not sure	Predictive mean matching
	PainConventional	Likert scale	Not sure	Predictive mean matching
	FearConventional	Likert scale	Not sure	Predictive mean matching
	DiscomfortConventional	Likert scale	Not sure	Predictive mean matching
	SatisfactionConventional	Likert scale	Not sure	Predictive mean matching
	PreferenceTiming	factor	None	Predictive mean matching

Category	Parameters	Class	Proposed.transformation	Proposed.imputation.method (2-level)
	RetestPreference	factor	Binary	Predictive mean matching
	RecommendRapidTests	factor	Binary	Predictive mean matching
	PreferenceFollowup	factor	Binary	Predictive mean matching
Test results	hcv.pos	integer	None	None
	hiv.pos	integer	None	None
	syphilis.pos	integer	None	None
	new.hcv	integer	None	None
	new.syphilis	integer	None	None
	new.hiv	integer	None	None

As described in: Van Buuren, S. (2018) Flexible Imputation of Missing Data (2nd ed).

Transformation

For this method to work, the variables have to be transformed to numbers

```
multiplex4imp.2l <- multiplex4imp.1l %>%
  mutate(across(c(3:33), ~as.factor(case_when(
    . %in% c("0-No or no wish to answer", "0-Never", "0-No or do not know", "None", "0-No",
             "No post-graduate degree", "Not currently employed", "1-<$2;000 CAD",
             "0-Very satisfied", "2-No preference","No preference for digital follow-up",
             "No or not sure") ~0,
    . %in% c("1-Yes", "1-Ever", "0 to 5", "0-HIV", "Yes", "1-5 times per week",
             "1-Yes; less than 6 months ago", "18 to 34", "1-Male",
             "Mediterranean European, African, Latin American, Caribbean and/or Asian",
             "4-Post-graduate degree", "Currently employed", "$2000 CAD and above",
             "0-1-no fear","0-1-no pain", "0-1 - no discomfort", "1-Satisfied", "0-1 day",
             "1-Phlebotomy", "Digital follow-up") ~1,
    . %in% c("6 or more", "2-Herpes simplex", "2-Yes; 6 months to 1 year ago",
             "35 and above", "2-Female", "Non-Mediterranean European/North American", "1-2",
             "2-Neutral/ not sure", "2 days or more", "0-Rapid test") ~2,
    . %in% c("3-Chlamydia", "3-Yes; more than 1 year ago", "3-Transgender", "2-3",
             "3-Dissatisfied")~3,
    . %in% c("4-Gonorrhoeae", "5-No response", "3-4" )~4,
    . %in% c("5-Syphilis", "4-5-highest")~5,
    . %in% c("6-Other")~6,
    T~NA))))

summary(multiplex4imp.2l)
```

```
##      p_id           Date.Recruited     SexualActivity CondomUse  SexualPartners
##  Length:400         Length:400         0   :106       0   : 98   1   :324
##  Class :character   Class :character   1   :289       1   :272   2   : 60
##  Mode  :character   Mode  :character   NA's:  5       NA's: 30   NA's: 16
##
```

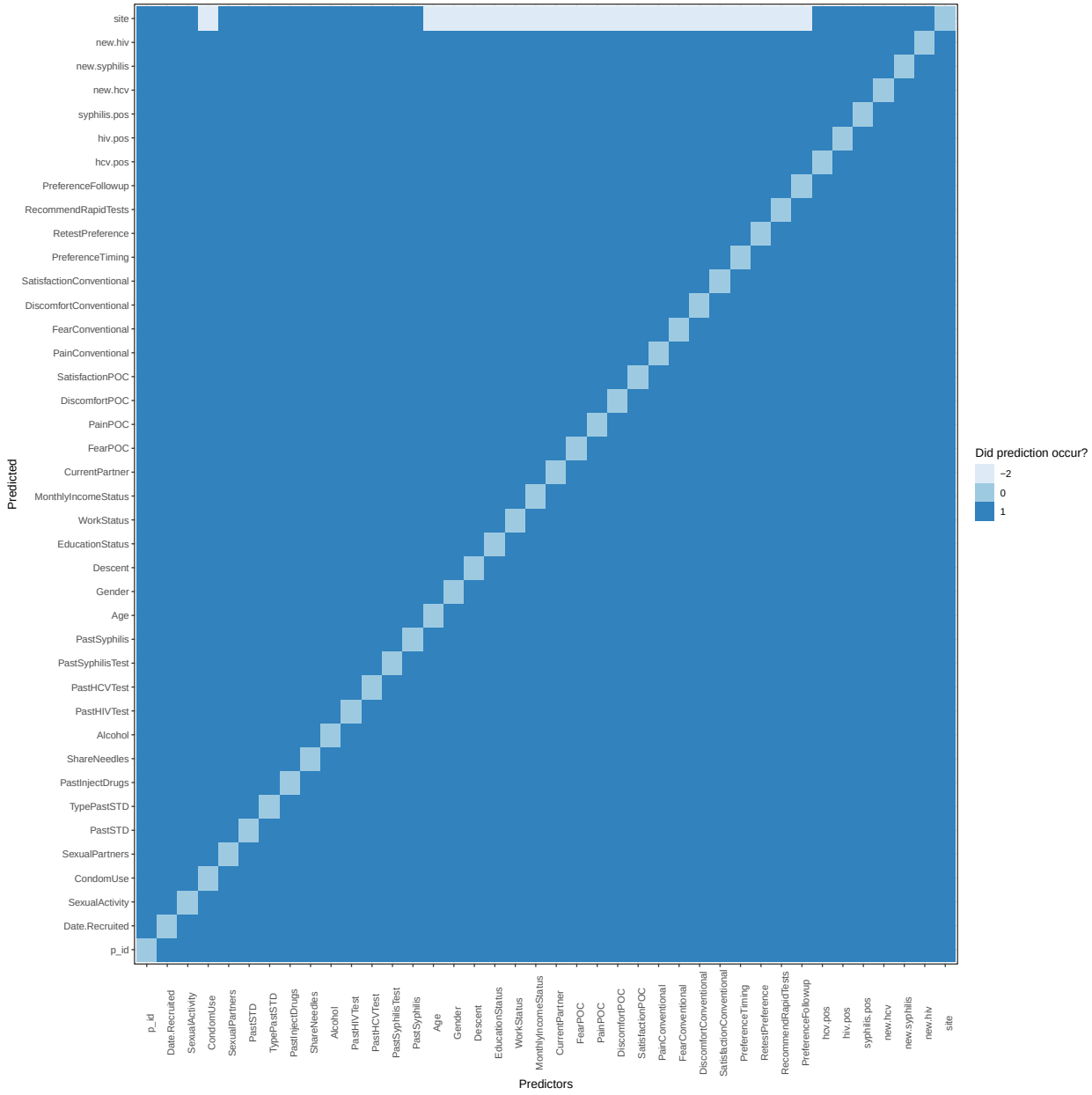
```
##
##
##
##  PastSTD     TypePastSTD  PastInjectDrugs ShareNeedles Alcohol    PastHIVTest
##  0   :216   0      :216   0   :255        0   :307     0   :209   0   : 57
##  1   :169   3      : 72   1   :140        1   : 85     1   :187   1   :140
##  NA's: 15   4      : 43   NA's:  5        NA's:  8     NA's:  4   2   : 47
##             6      : 25                                           3   :154
##             1      : 10                                           NA's:  2
##             (Other): 13
##             NA's   : 21
##  PastHCVTest PastSyphilisTest PastSyphilis   Age       Gender    Descent
##  0   : 44    0   :  4         0   :348     1   :156   1   :255   1   : 99
##  1   :139    1   :127         1   : 48     2   :197   2   : 73   2   :254
##  2   : 48    2   : 47         NA's:  4     NA's: 47   3   :  7   NA's: 47
##  3   :167    3   :219                                 4   : 11
##  NA's:  2    NA's:  3                                 NA's: 54
##
##
##  EducationStatus WorkStatus MonthlyIncomeStatus CurrentPartner FearPOC
##  0   :233        0   :165   0   :219            0   :213       1   :155
##  1   :120        1   :188   1   :134            1   :140       2   : 52
##  NA's: 47        NA's: 47   NA's: 47            NA's: 47       3   : 19
##                                                                4   :  9
##                                                                5   :  7
##                                                                NA's:158
##
##  PainPOC    DiscomfortPOC SatisfactionPOC PainConventional FearConventional
##  1   :166   1   :181      0   :207        1   :159         1   :150
##  2   : 59   2   : 46      1   : 30        2   : 61         2   : 45
##  3   : 15   3   : 11      2   :  3        3   : 10         3   : 21
##  4   :  1   4   :  2      NA's:160        4   :  1         4   :  9
##  NA's:159   5   :  1                      NA's:169         5   :  5
##             NA's:159                                       NA's:170
##
##  DiscomfortConventional SatisfactionConventional PreferenceTiming
##  1   :173               0   :190                 1   :137
##  2   : 43               1   : 33                 2   :106
##  3   : 12               2   :  7                 NA's:157
##  4   :  2               3   :  1
##  5   :  1               NA's:169
##  NA's:169
##
##  RetestPreference RecommendRapidTests PreferenceFollowup    hcv.pos
##  0   : 32         0   : 16            0   : 87            Min.   :0.0000
##  1   : 66         1   :227            1   :156            1st Qu.:0.0000
##  2   :144         NA's:157            NA's:157            Median :0.0000
##  NA's:158                                                 Mean   :0.1475
##                                                           3rd Qu.:0.0000
##                                                           Max.   :1.0000
##
##     hiv.pos      syphilis.pos       new.hcv       new.syphilis
##  Min.   :0.00   Min.   :0.0000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.00   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000
```

```
##  Median :0.00   Median :0.0000   Median :0.000   Median :0.000
##  Mean   :0.04   Mean   :0.0125   Mean   :0.035   Mean   :0.005
##  3rd Qu.:0.00   3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.00   Max.   :1.0000   Max.   :1.000   Max.   :1.000
##
##     new.hiv           site
##  Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.000   Median :0.0000
##  Mean   :0.005   Mean   :0.4075
##  3rd Qu.:0.000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.0000
##
```

Change prediction matrix

to include 2-level pmm for selected variables (see analysis plan)

```
# Create the predictor matrix
pred.2l <- make.predictorMatrix(multiplex4imp.2l)

# set site as cluster variables
pred.2l[c(4, 15:33), "site"] <- -2

#check
prediction.matrix(pred.2l)
```



Adapt methods

For now we will keep variables not predicted by cluster variables as the initial method

```
# run quick imputation
ini.quick <- mice(multiplex4imp.2l, maxit = 0)

# copy initial methods
meth.2l<- ini.quick$method

#replace selected variables to 2l_pmm
meth.2l[c(4, 15:33)] <- "2l.pmm"
```

```
# check
meth.2l
```

```
##                     p_id           Date.Recruited           SexualActivity
##                       ""                       ""                 "logreg"
##               CondomUse           SexualPartners                  PastSTD
##                "2l.pmm"                 "logreg"                 "logreg"
##              TypePastSTD          PastInjectDrugs              ShareNeedles
##                "polyreg"                 "logreg"                 "logreg"
##                  Alcohol             PastHIVTest             PastHCVTest
##                "logreg"                "polyreg"                "polyreg"
##         PastSyphilisTest             PastSyphilis                      Age
##                "polyreg"                 "logreg"                "2l.pmm"
##                   Gender                 Descent          EducationStatus
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##               WorkStatus      MonthlyIncomeStatus           CurrentPartner
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##                  FearPOC                 PainPOC           DiscomfortPOC
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##          SatisfactionPOC         PainConventional         FearConventional
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##   DiscomfortConventional SatisfactionConventional         PreferenceTiming
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##         RetestPreference      RecommendRapidTests       PreferenceFollowup
##                "2l.pmm"                "2l.pmm"                 "2l.pmm"
##                  hcv.pos                 hiv.pos            syphilis.pos
##                       ""                       ""                      ""
##                  new.hcv             new.syphilis                 new.hiv
##                       ""                       ""                      ""
##                     site
##                       ""
```

Re-run imputation

```
# mice.2l <- mice(multiplex4imp.2l, nnet.MaxNWts = 2000, pred= pred.2l, meth = meth.2l, print =F)
```

also not working

View imputed data

```
# summary(complete(mice.2l))
#
# summary(multiplex4imp.2l)
```

Extract imputed data and save as csv

```
# mice.2l.long<- complete(mice.2l, "long")
#
# mice.2l.broad <- complete(mice.2l, "broad")
#
# write.csv(mice.2l.long, "3_intermediate/imputed/imputed_data_long_2l.csv",
#           row.names = F)
#
# write.csv(mice.2l.broad, "3_intermediate/imputed/imputed_data_broad_2l.csv",
#            row.names = F)
```

Diagnostic plot

With 5 iteration

```
# plot(mice.2l)
```

with 20 (using the same object and running 15 more iterations)

```
# mice.2l.20 <- mice.mids(mice.2l, nnet.MaxNWts = 2000, print =F, maxit=15)
#
# plot(mice.2l.20)
```

Sensitivity analysis

Density plot of selected variables 1l imputation

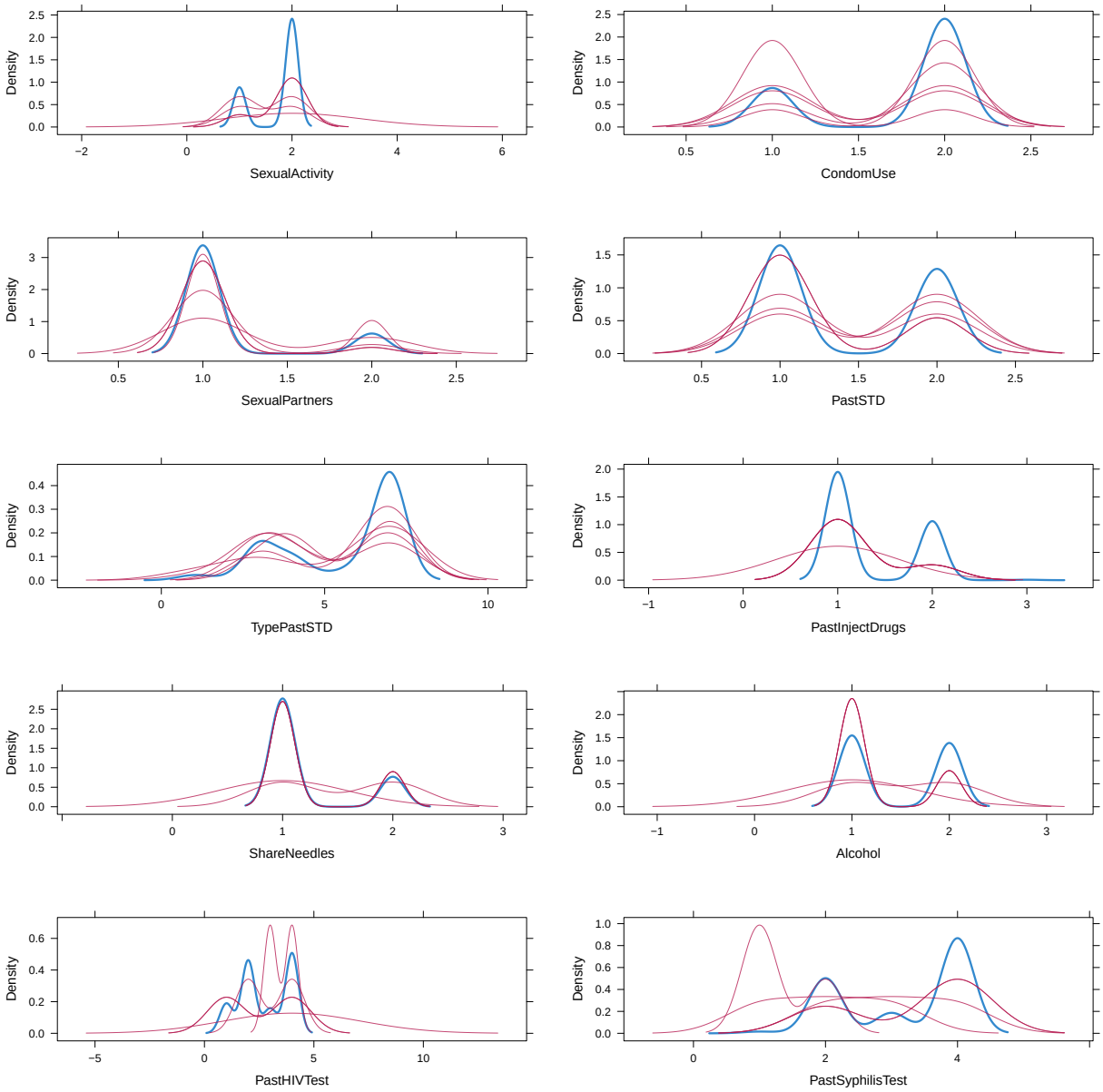
```
# dplot1 <- densityplot(mice.1l, ~SexualActivity)
# dplot2 <- densityplot(mice.1l, ~CondomUse)
# dplot3 <- densityplot(mice.1l, ~SexualPartners)
# dplot4 <- densityplot(mice.1l, ~PastSTD)
# dplot5 <- densityplot(mice.1l, ~TypePastSTD)
# dplot6 <- densityplot(mice.1l, ~PastInjectDrugs)
# dplot7 <- densityplot(mice.1l, ~ShareNeedles)
# dplot8 <- densityplot(mice.1l, ~Alcohol)
# dplot9 <- densityplot(mice.1l, ~PastHIVTest)
# dplot10 <- densityplot(mice.1l, ~PastSyphilisTest)
#
# dplots.clin <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
#                         dplot7, dplot8, dplot9, dplot10, ncol =2 )
# dplots.clin
#
# ggsave (dplots.clin, filename = "3_intermediate/imputed/Densityplots_observed_vs_1l-imputed_data_clin
#           device = "png", width = 14, height = 14)
#
# dplot1 <- densityplot(mice.1l, ~Age)
# dplot2 <- densityplot(mice.1l, ~Gender)
# dplot3 <- densityplot(mice.1l, ~Descent)
# dplot4 <- densityplot(mice.1l, ~EducationStatus)
# dplot5 <- densityplot(mice.1l, ~WorkStatus)
```

```
# dplot6 <- densityplot(mice.1l, ~MonthlyIncomeStatus)
# dplot7 <- densityplot(mice.1l, ~CurrentPartner)
#
# dplots.ses <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
#                         dplot7, ncol =2 )
#
# ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_1l-imputed_data_ses.p
#          device = "png", width = 14, height = 14)
#
# dplot1 <- densityplot(mice.1l, ~DiscomfortPOC)
# dplot2 <- densityplot(mice.1l, ~FearConventional)
# dplot3 <- densityplot(mice.1l, ~RetestPreference)
# dplot4 <- densityplot(mice.1l, ~SatisfactionConventional)
# dplot5 <- densityplot(mice.1l, ~PreferenceTiming)
# dplot6 <- densityplot(mice.1l, ~PainPOC)
#
# dplot.pro <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6)
#
# dplot.pro
#
# ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_1l-imputed_data_pro.p
#          device = "png", width = 14, height = 14)
```

For some variables, the imputed data follows the shape of the observed data.

Density plot of selected variables 1l imputation - pmm

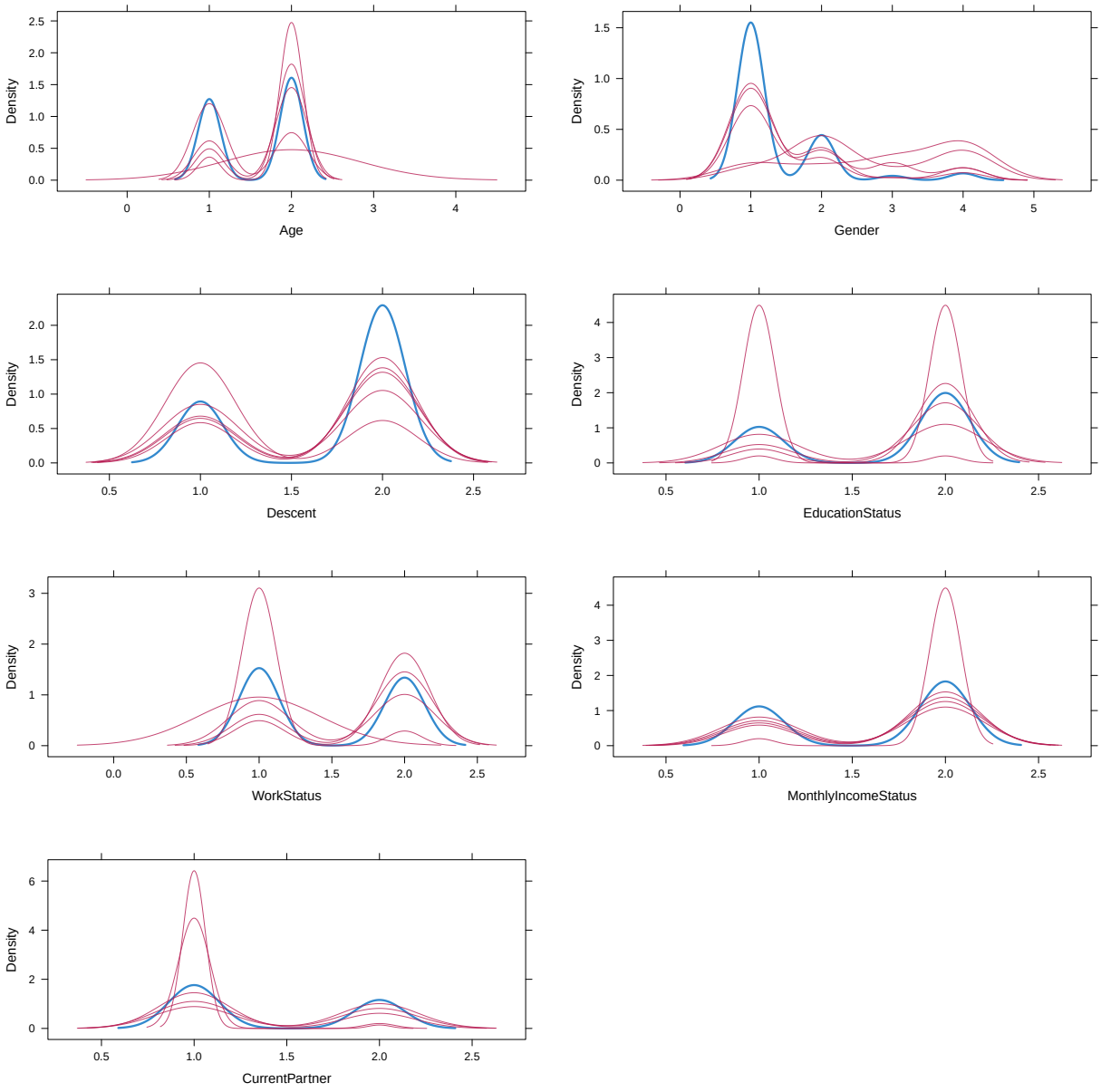
```
dplot1 <- densityplot(mice.pmm, ~SexualActivity)
dplot2 <- densityplot(mice.pmm, ~CondomUse)
dplot3 <- densityplot(mice.pmm, ~SexualPartners)
dplot4 <- densityplot(mice.pmm, ~PastSTD)
dplot5 <- densityplot(mice.pmm, ~TypePastSTD)
dplot6 <- densityplot(mice.pmm, ~PastInjectDrugs)
dplot7 <- densityplot(mice.pmm, ~ShareNeedles)
dplot8 <- densityplot(mice.pmm, ~Alcohol)
dplot9 <- densityplot(mice.pmm, ~PastHIVTest)
dplot10 <- densityplot(mice.pmm, ~PastSyphilisTest)

dplots.clin <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
                        dplot7, dplot8, dplot9, dplot10, ncol =2 )
```



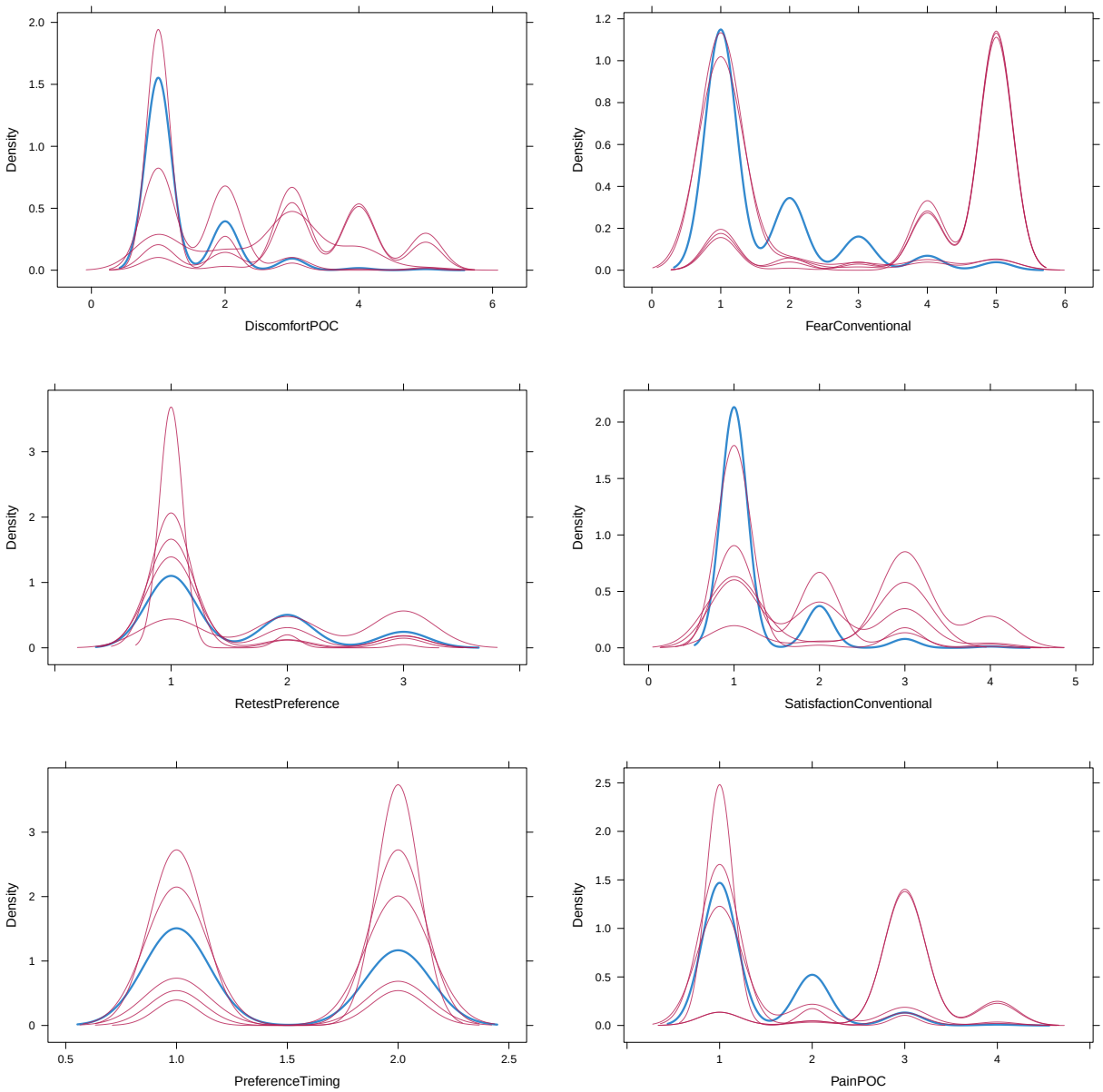

```
dplots.clin
```

```
## TableGrob (5 x 2) "arrange": 10 grobs
##     z     cells    name                           grob
## 1   1 (1-1,1-1) arrange lattice[GRID.lattice.32727]
## 2   2 (1-1,2-2) arrange lattice[GRID.lattice.32728]
## 3   3 (2-2,1-1) arrange lattice[GRID.lattice.32729]
## 4   4 (2-2,2-2) arrange lattice[GRID.lattice.32730]
## 5   5 (3-3,1-1) arrange lattice[GRID.lattice.32731]
## 6   6 (3-3,2-2) arrange lattice[GRID.lattice.32732]
## 7   7 (4-4,1-1) arrange lattice[GRID.lattice.32733]
## 8   8 (4-4,2-2) arrange lattice[GRID.lattice.32734]
## 9   9 (5-5,1-1) arrange lattice[GRID.lattice.32735]
## 10 10 (5-5,2-2) arrange lattice[GRID.lattice.32736]
```

```
ggsave (dplots.clin, filename = "3_intermediate/imputed/Densityplots_observed_vs_pmm-imputed_data_clini
         device = "png", width = 14, height = 14)

dplot1 <- densityplot(mice.pmm, ~Age)
dplot2 <- densityplot(mice.pmm, ~Gender)
dplot3 <- densityplot(mice.pmm, ~Descent)
dplot4 <- densityplot(mice.pmm, ~EducationStatus)
dplot5 <- densityplot(mice.pmm, ~WorkStatus)
dplot6 <- densityplot(mice.pmm, ~MonthlyIncomeStatus)
dplot7 <- densityplot(mice.pmm, ~CurrentPartner)

dplots.ses <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
                           dplot7, ncol =2 )
```



```r
ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_pmm-imputed_data_ses.png
         device = "png", width = 14, height = 14)

dplot1 <- densityplot(mice.pmm, ~DiscomfortPOC)
dplot2 <- densityplot(mice.pmm, ~FearConventional)
dplot3 <- densityplot(mice.pmm, ~RetestPreference)
dplot4 <- densityplot(mice.pmm, ~SatisfactionConventional)
dplot5 <- densityplot(mice.pmm, ~PreferenceTiming)
dplot6 <- densityplot(mice.pmm, ~PainPOC)

dplot.pro <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6)
```



```
dplot.pro
```

```
## TableGrob (3 x 2) "arrange": 6 grobs
##   z     cells    name                           grob
## 1 1 (1-1,1-1) arrange lattice[GRID.lattice.32880]
## 2 2 (1-1,2-2) arrange lattice[GRID.lattice.32881]
## 3 3 (2-2,1-1) arrange lattice[GRID.lattice.32882]
## 4 4 (2-2,2-2) arrange lattice[GRID.lattice.32883]
## 5 5 (3-3,1-1) arrange lattice[GRID.lattice.32884]
## 6 6 (3-3,2-2) arrange lattice[GRID.lattice.32885]
```

```
ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_pmm-imputed_data_pro.png
         device = "png", width = 14, height = 14)
```

pmm predicts the missing data to be closer to the predicted data

Density plot of selected variables pmm2-level imputation

```
# dplot1 <- densityplot(mice.2l, ~SexualActivity)
# dplot2 <- densityplot(mice.2l, ~CondomUse)
# dplot3 <- densityplot(mice.2l, ~SexualPartners)
# dplot4 <- densityplot(mice.2l, ~PastSTD)
# dplot5 <- densityplot(mice.2l, ~TypePastSTD)
# dplot6 <- densityplot(mice.2l, ~PastInjectDrugs)
# dplot7 <- densityplot(mice.2l, ~ShareNeedles)
# dplot8 <- densityplot(mice.2l, ~Alcohol)
# dplot9 <- densityplot(mice.2l, ~PastHIVTest)
# dplot10 <- densityplot(mice.2l, ~PastSyphilisTest)
#
# dplots.clin <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
#                         dplot7, dplot8, dplot9, dplot10, ncol =2 )
# dplots.clin
#
# ggsave (dplots.clin, filename = "3_intermediate/imputed/Densityplots_observed_vs_2l-imputed_data_clin
#          device = "png", width = 14, height = 14)
#
# dplot1 <- densityplot(mice.2l, ~Age)
# dplot2 <- densityplot(mice.2l, ~Gender)
# dplot3 <- densityplot(mice.2l, ~Descent)
# dplot4 <- densityplot(mice.2l, ~EducationStatus)
# dplot5 <- densityplot(mice.2l, ~WorkStatus)
# dplot6 <- densityplot(mice.2l, ~MonthlyIncomeStatus)
# dplot7 <- densityplot(mice.2l, ~CurrentPartner)
#
# dplots.ses <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
#                         dplot7, ncol =2 )
#
# ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_2l-imputed_data_ses.p
#          device = "png", width = 14, height = 14)
#
# dplot1 <- densityplot(mice.2l, ~DiscomfortPOC)
```

```
# dplot2 <- densityplot(mice.2l, ~FearConventional)
# dplot3 <- densityplot(mice.2l, ~RetestPreference)
# dplot4 <- densityplot(mice.2l, ~SatisfactionConventional)
# dplot5 <- densityplot(mice.2l, ~PreferenceTiming)
# dplot6 <- densityplot(mice.2l, ~PainPOC)
#
# dplot.pro <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6)
#
# dplot.pro
#
# ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_2l-imputed_data_pro.p
#          device = "png", width = 14, height = 14)
```

Prediction is even better.

Random intercept mixed model regression

Original data

We will conduct a random intercept model multiple log regression with HCV as the dependent variable, site as the cluster variable and selected variables as independent variables :

$y_i = \beta_0 + \beta_1 X_{ij} + u_j$ with $u_j \sim N(0, \sigma_j^2)$ - random intercept at site level

1. SexualActivity
2. PastSTD
3. PastInjectDrugs
4. Age
5. Gender

```
fit.original <- glm(hcv.pos~
                      SexualActivity+PastSTD+PastInjectDrugs+Age+Gender+ (1:site),
                    data = multiplex4imp.1l)

summary(fit.original)
```

```
##
## Call:
## glm(formula = hcv.pos ~ SexualActivity + PastSTD + PastInjectDrugs +
##     Age + Gender + (1:site), data = multiplex4imp.1l)
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.01041    0.04936   0.211   0.8330
## SexualActivity1-Yes   -0.02648    0.04226  -0.627   0.5313
## PastSTD1-Yes           0.07418    0.03335   2.224   0.0268 *
## PastInjectDrugs1-Yes   0.32877    0.03738   8.797   <2e-16 ***
## Age35 and above        0.01508    0.03540   0.426   0.6704
## Gender2-Female        -0.02807    0.04144  -0.677   0.4987
## Gender3-Transgender   -0.01572    0.11752  -0.134   0.8936
## Gender5-No response   -0.02404    0.09857  -0.244   0.8075
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.0906839)
##
##     Null deviance: 39.009  on 337  degrees of freedom
## Residual deviance: 29.926  on 330  degrees of freedom
##   (62 observations deleted due to missingness)
## AIC: 157.78
##
## Number of Fisher Scoring iterations: 2
```

Some values have huge standard errors. This probably corresponds to variables with a lot of missing data or levels with little amounts of data (e.g. transgender). But the binary trasnformation got rid of most of them.

Imputed dataset

```
# fit.imp <- with(mice.1l,
#             glm(hcv.pos~
#                       SexualActivity+PastSTD+PastInjectDrugs+Age+Gender+(1:site)))
#
# # pool the data
# pool.fit.imp <- pool(fit.imp)
#
# summary(pool.fit.imp, conf.int = T)
```

Pmm results

```
fit.imp <- with(mice.pmm,
            glm(hcv.pos~
                      SexualActivity+PastSTD+PastInjectDrugs+Age+Gender+ (1:site)))

# pool the data
pool.fit.imp <- pool(fit.imp)

summary(pool.fit.imp, conf.int = T)
```

```
##                   term      estimate  std.error   statistic        df
## 1          (Intercept)  0.0005374442 0.04675124  0.01149583 340.54073
## 2  SexualActivity1-Yes -0.0232496295 0.03860428 -0.60225524 378.80856
## 3         PastSTD1-Yes  0.0760217985 0.03225098  2.35719357 373.55506
## 4 PastInjectDrugs1-Yes  0.3549868672 0.03601224  9.85739484 267.94694
## 5   PastInjectDrugsYes -0.0126736603 0.31503117 -0.04022986 384.36542
## 6      Age35 and above  0.0200981986 0.03462061  0.58052692 296.14977
## 7       Gender2-Female -0.0289831384 0.04525567 -0.64043107  55.86183
## 8  Gender3-Transgender -0.0150710106 0.12581765 -0.11978455  21.69956
## 9  Gender5-No response  0.0354086158 0.10461658  0.33846083  15.15180
##        p.value       2.5 %     97.5 %
## 1 9.908346e-01 -0.09142012 0.09249501
## 2 5.473646e-01 -0.09915515 0.05265589
```

```
## 3 1.892887e-02  0.01260558 0.13943802
## 4 9.271167e-20  0.28408392 0.42588982
## 5 9.679308e-01 -0.63207378 0.60672646
## 6 5.620008e-01 -0.04803540 0.08823180
## 7 5.245108e-01 -0.11964609 0.06167981
## 8 9.057559e-01 -0.27621046 0.24606843
## 9 7.396605e-01 -0.18738193 0.25819916
```

2l-results

```
# fit.imp <- with(mice.2l,
#              glm(hcv.pos~
#                        SexualActivity+PastSTD+PastInjectDrugs+Age+Gender+ (1:site)))
#
# # pool the data
# pool.fit.imp <- pool(fit.imp)
#
# summary(pool.fit.imp, conf.int = T)
```

Results are all very similar. Which method is best?

Updates

Added code for dataset that excludes PROMs

Keep 1-level no PROMs for projpred for now

```
multiplex_no_prom <- read.csv("3_intermediate/missing_data/multiplex_na_final_no_prom.csv")
head(multiplex_no_prom, 3)
```

```
##       p_id DateRecuit            SexualActivity   CondomUse SexualPartners
## 1 NBU20001 2021-03-24                  1-Yes 2-Sometimes            0-0
## 2 NBU20002 2021-04-15 2-I do not wish to answer        <NA>           <NA>
## 3 NBU20003 2021-04-16                  1-Yes     0-Never            1-1
##           PastSTD TypePastSTD PastInjectDrugs ShareNeedles              Alcohol
## 1           1-Yes 3-Chlamydia           1-Yes        1-Yes                 0-No
## 2 2-I do not know        None           1-Yes        1-Yes                 0-No
## 3           1-Yes 3-Chlamydia           1-Yes        1-Yes 1-1-2 times per week
##                    PastHIVTest                   PastHCVTest
## 1   3-Yes; more than 1 year ago   3-Yes; more than 1 year ago
## 2 1-Yes; less than 6 months ago 1-Yes; less than 6 months ago
## 3 1-Yes; less than 6 months ago 1-Yes; less than 6 months ago
##                PastSyphilisTest PastSyphilis     Age   Gender
## 1   3-Yes; more than 1 year ago         0-No 3-35-44   1-Male
## 2 1-Yes; less than 6 months ago         0-No    <NA>     <NA>
## 3                         <NA>         0-No 3-35-44 2-Female
##                      Descent                            EducationStatus
## 1 4-European/ North American 2-High school to Pre-university education
## 2                       <NA>                                      <NA>
## 3 4-European/ North American      1-Did not complete primary schooling
```

```
##   WorkStatusStatus MonthlyIncomeStatus CurrentPartner Clinic.Location hcv.pos
## 1   3-Not employed       1-<$2;000 CAD           0-No           RECAP       1
## 2             <NA>                <NA>           <NA>           RECAP       1
## 3   3-Not employed       1-<$2;000 CAD          1-Yes           RECAP       0
##   hiv.pos syphilis.pos baseline.hcv baseline.hiv baseline.syph new.hcv
## 1       0            0            0            1             1       0
## 2       0            0            1            1             1       1
## 3       0            0            0            1             1       0
##   new.syphilis new.hiv
## 1            0       0
## 2            0       0
## 3            0       0
```

```
multiplex_no_prom <- multiplex_no_prom %>%
  select(!c(baseline.hcv:baseline.syph)) %>%
 mutate(across(SexualActivity:Clinic.Location, as.factor)) %>%
  rename(WorkStatus = WorkStatusStatus, Date.Recruited = DateRecuit)
summary(multiplex_no_prom)
```

```
##      p_id           Date.Recruited                       SexualActivity
##  Length:401         Length:401         0-No                     : 88
##  Class :character   Class :character   1-Yes                    :289
##  Mode  :character   Mode  :character   2-I do not wish to answer: 18
##                                        NA's                     :  6
##
##
##
##        CondomUse        SexualPartners            PastSTD
##  0-Never    : 98   0-0         : 71    0-No           :204
##  1-Always   :105   1-1         : 94    1-Yes          :169
##  2-Sometimes:167   2-2 to 5    :159    2-I do not know: 12
##  NA's       : 31   3-6 to 10   : 38    NA's           : 16
##                    4-11 or more: 22
##                    NA's        : 17
##
##         TypePastSTD                   PastInjectDrugs ShareNeedles
##  None         :216   0-No                     :252    0-No : 55
##  3-Chlamydia  : 72   1-Yes                    :140    1-Yes: 85
##  4-Gonorrhoeae: 43   2-I do not wish to answer:  3    None :252
##  6-Other      : 25   Yes                      :  1    NA's :  9
##  0-HIV        : 10   NA's                     :  5
##  (Other)      : 13
##  NA's         : 22
##                  Alcohol                           PastHIVTest
##  0-No                :209   0-No                          : 57
##  1-1-2 times per week:122   1-Yes; less than 6 months ago:141
##  2-3-5 times per week: 65   2-Yes; 6 months to 1 year ago: 47
##  NA's                :  5   3-Yes; more than 1 year ago  :154
##                             NA's                          :  2
##
##
##                         PastHCVTest                         PastSyphilisTest
##  0-No                          : 44   0-No                          :  4
##  1-Yes; less than 6 months ago:140   1-Yes; less than 6 months ago:127
```

```
##  2-Yes; 6 months to 1 year ago: 48   2-Yes; 6 months to 1 year ago: 47
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :220
##  NA's                         :  2   NA's                         :  3
##
##
##          PastSyphilis             Age                Gender
##  0-No           :330   1-18-24       : 29   1-Male       :255
##  1-Yes          : 48   1-25-34       :127   2-Female     : 73
##  2-I do not know: 19   3-35-44       :103   3-Transgender:  7
##  NA's           :  4   4-45-54       : 54   4-Bisexual   :  7
##                        5-55 and above: 40   5-No response: 11
##                        NA's          : 48   NA's         : 48
##
##                                          Descent
##  1-North American Indigenous                  : 18
##  2-Asian/ South Asian                         : 10
##  3-Mediterranean                              : 13
##  4-European/ North American                   :254
##  4-Latin American/ Caribbean                  : 38
##  6-African (North; Central; West; East; South): 20
##  NA's                                         : 48
##                                   EducationStatus                  WorkStatus
##  1-Did not complete primary schooling     : 38    1-Employed (full time):127
##  2-High school to Pre-university education:143    2-Employed (part time): 61
##  3-Technical degree                       : 52    3-Not employed        :135
##  4-Post-graduate degree                   :120    4-Retired             : 11
##  NA's                                     : 48    5-Not willing to work : 19
##                                                   NA's                  : 48
##
##           MonthlyIncomeStatus CurrentPartner   Clinic.Location
##  1-<$2;000 CAD       :219      0-No :213      Outreach  : 51
##  2-$2;001-$4;000 CAD :103      1-Yes:140      RECAP     : 87
##  3-$4;001-$6;000 CAD : 22      NA's : 48      REZO      :237
##  4-$6;001-$$8;000 CAD:  7                     Riverstone: 26
##  5-$8;001 and above  :  2
##  NA's                : 48
##
##     hcv.pos          hiv.pos        syphilis.pos        new.hcv
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.1496   Mean   :0.0399   Mean   :0.01247   Mean   :0.03491
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##
##   new.syphilis         new.hiv
##  Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.000000   Median :0.000000
##  Mean   :0.004987   Mean   :0.004987
##  3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.000000
##
```

```
pattern<- md.pattern(multiplex_no_prom, rotate.names = T)
```

```
na.patterns.total<- nrow(pattern)
```

There are a total of 28 patterns

```
large.na<- as.integer(2/3*ncol(multiplex_no_prom))

multiplex_no_prom <- multiplex_no_prom %>%
  mutate(na_count = rowSums(is.na(multiplex_no_prom))) %>%
  filter(na_count < large.na) %>%
  select(!na_count)
```

Create columns for clinic, stbbi status and missingness column to be populated later

```
multiplex4plots <- multiplex_no_prom %>%
  mutate(site = ifelse(Clinic.Location == "REZO", "REZO", "RECAP"),
         stbbi.pos = as.factor(ifelse(hcv.pos == 1| hiv.pos ==1 | syphilis.pos == 1,
                                      "Infected","Non-infected")),
         missingness = NA)
```

Does variable missingness depend on clinic?

same as above but without the PROMs.

Use get correlation function to get all plots at once

```
variable <- colnames(multiplex4plots)[3:21]

plots <-  get.correlation.site(multiplex4plots, variable, title = variable)
```

Does variable missingness depend on infection status?

9. Age

```
plot9 <- get.correlation.inf(dataset = multiplex4plots,
                             variable = multiplex4plots$Age,
                             title = "Age data")
```

10. Current partner

```
plot10 <- get.correlation.inf(dataset = multiplex4plots,
                              variable = multiplex4plots$CurrentPartner,
                              title = "Current partner data")
```

11. sharing needles

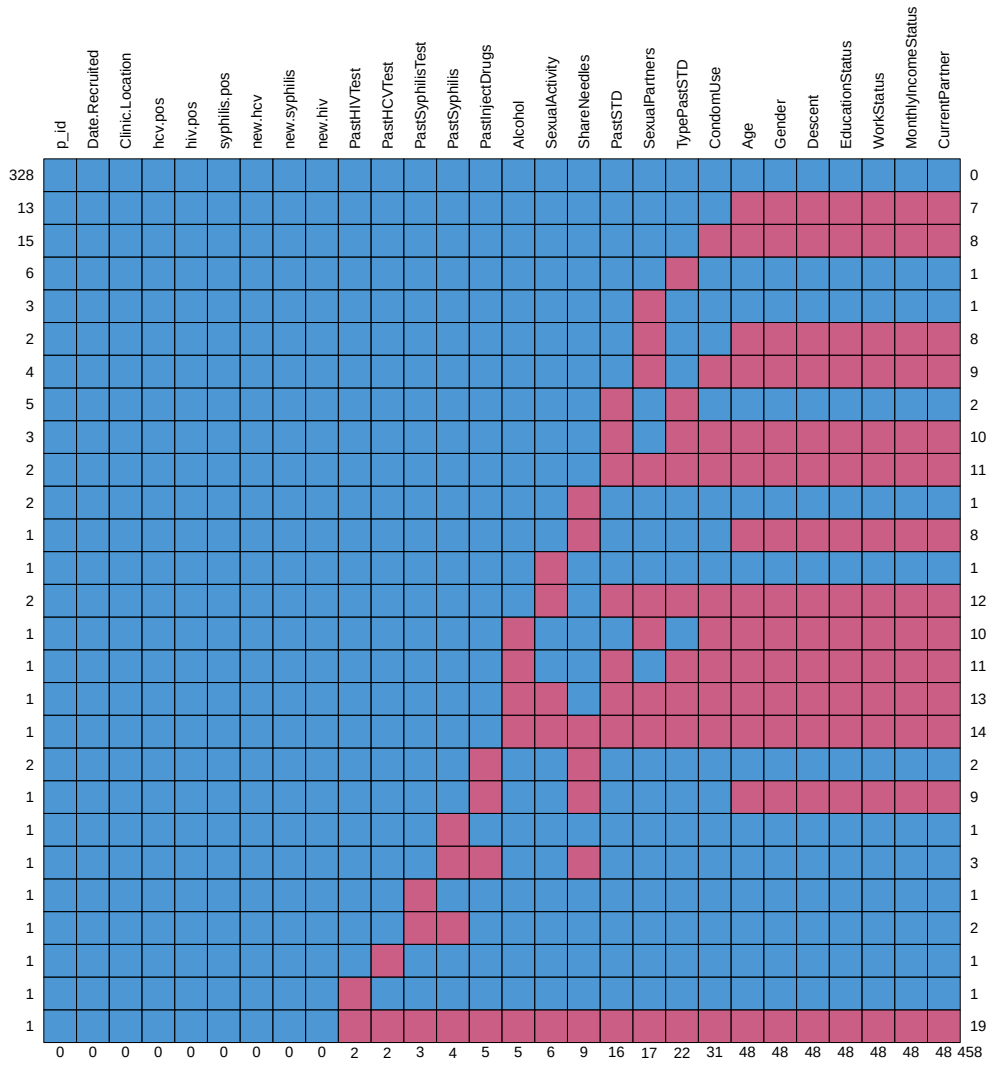


Figure 3: Pattern of missingness for variables in the Canada multiplex dataset (without PROM). The numbers on the left indicate the number of rows that have the same pattern of missingness. The numbers on the right represent the number of missing data for each pattern. The numbers at the bottom represent the total number of missing values for each variable with the total number of missing data in the dataset at the bottom right

```
plot11 <- get.correlation.inf(dataset = multiplex4plots,
                              variable = multiplex4plots$ShareNeedles,
                              title = "Sharing needle data")
```

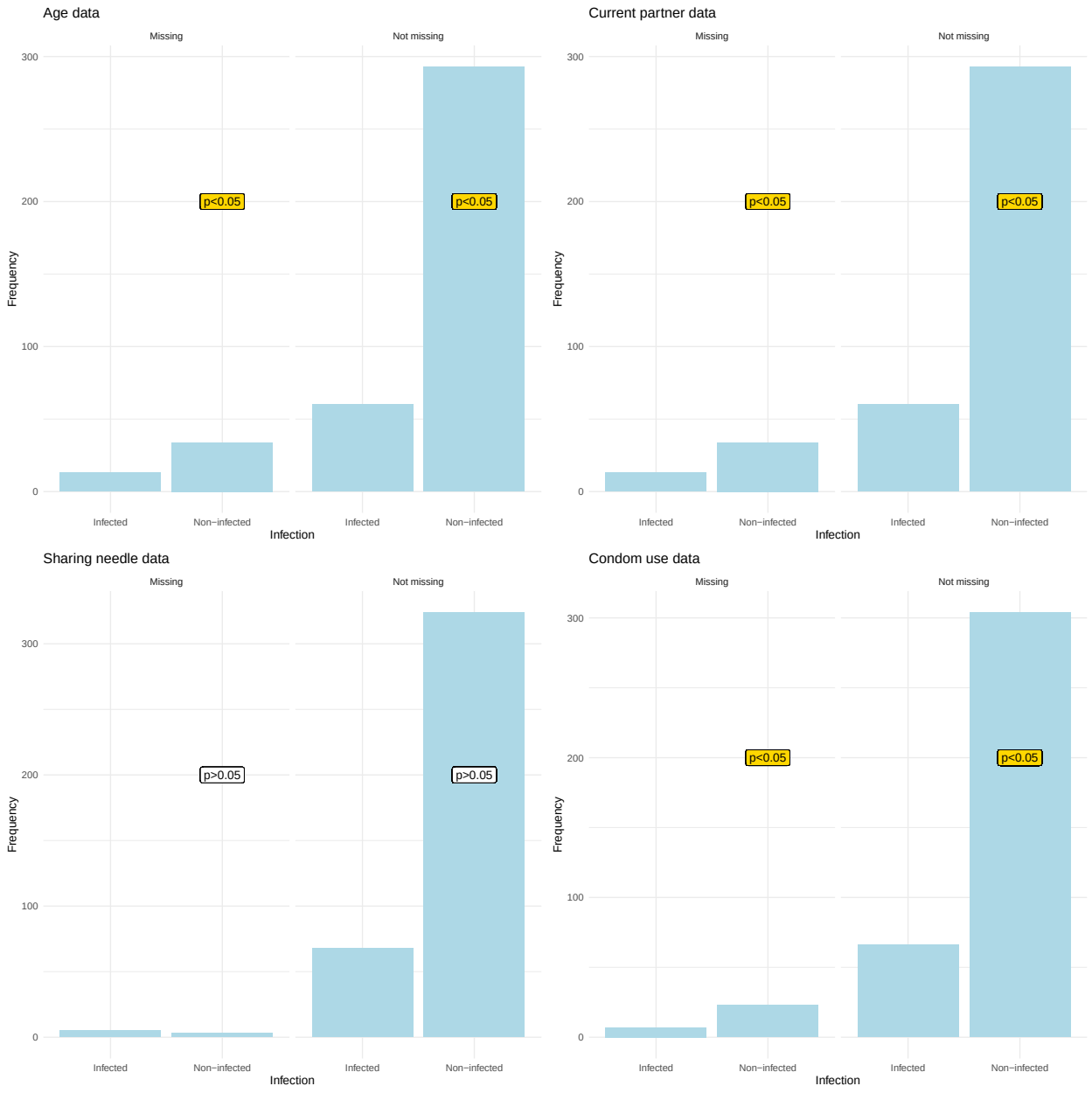
12. Condom use

```
plot12 <- get.correlation.inf(dataset = multiplex4plots,
                              variable = multiplex4plots$CondomUse,
                              title = "Condom use data")
```

Plots

```
plots.per.site <- grid.arrange(plots[[1]], plots[[2]], plots[[3]],
                               plots[[4]], plots[[5]], plots[[6]],
                               plots[[7]], plots[[8]], plots[[9]],
                               plots[[10]], plots[[11]], plots[[12]],
                               plots[[13]], plots[[14]], plots[[15]],
                               plots[[16]], plots[[17]], plots[[18]],
                               plots[[19]],
                               ncol=3)
```

```
plots.per.status <- grid.arrange(plot9, plot10, plot11, plot12, ncol= 2)
```




```
ggsave(plots.per.status, filename = "3_intermediate/missing_data/missingness_no_prom.png", device = png

ggsave(plots.per.site, filename = "4_outputs/missingness_site.png",
       device = "png", width = 14, height = 14)
```

There are more missing data at RECAP and in non-infected individuals. This pattern is MNAR. Starting from Sharing needles, the missing data does not depend on site or infection status (because number of missing data is small).

Imputation

```
multiplex4imp2 <- multiplex_no_prom %>%
  mutate(site = as.integer(ifelse(Clinic.Location == "REZO", 0, 1))) %>%
  select(-c(Clinic.Location))

summary(multiplex4imp2$site)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    0.41    1.00    1.00
```

```
summary(multiplex4imp2)
```

```
##      p_id           Date.Recruited                      SexualActivity
##  Length:400         Length:400         0-No                     : 88
##  Class :character   Class :character   1-Yes                    :289
##  Mode  :character   Mode  :character   2-I do not wish to answer: 18
##                                        NA's                     :  5
##
##
##
##        CondomUse        SexualPartners            PastSTD
##  0-Never    : 98   0-0         : 71    0-No           :204
##  1-Always   :105   1-1         : 94    1-Yes          :169
##  2-Sometimes:167   2-2 to 5    :159    2-I do not know: 12
##  NA's       : 30   3-6 to 10   : 38    NA's           : 15
##                    4-11 or more: 22
##                    NA's        : 16
##
##         TypePastSTD                   PastInjectDrugs ShareNeedles
##  None         :216   0-No                     :252    0-No : 55
##  3-Chlamydia  : 72   1-Yes                    :140    1-Yes: 85
##  4-Gonorrhoeae: 43   2-I do not wish to answer:  3    None :252
##  6-Other      : 25   Yes                      :  1    NA's :  8
##  0-HIV        : 10   NA's                     :  4
##  (Other)      : 13
##  NA's         : 21
##                  Alcohol                          PastHIVTest
##  0-No                :209   0-No                         : 57
##  1-1-2 times per week:122   1-Yes; less than 6 months ago:141
##  2-3-5 times per week: 65   2-Yes; 6 months to 1 year ago: 47
##  NA's                :  4   3-Yes; more than 1 year ago  :154
##                             NA's                         :  1
##
##
##                         PastHCVTest                       PastSyphilisTest
##  0-No                          : 44   0-No                          :  4
##  1-Yes; less than 6 months ago:140   1-Yes; less than 6 months ago:127
##  2-Yes; 6 months to 1 year ago: 48   2-Yes; 6 months to 1 year ago: 47
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :220
##  NA's                          :  1   NA's                          :  2
##
```

```
##
##           PastSyphilis             Age                Gender
##  0-No            :330   1-18-24       : 29   1-Male       :255
##  1-Yes           : 48   1-25-34       :127   2-Female     : 73
##  2-I do not know: 19    3-35-44       :103   3-Transgender:  7
##  NA's            :  3   4-45-54       : 54   4-Bisexual   :  7
##                         5-55 and above: 40   5-No response: 11
##                         NA's          : 47   NA's         : 47
##
##                                          Descent
##  1-North American Indigenous                  : 18
##  2-Asian/ South Asian                         : 10
##  3-Mediterranean                              : 13
##  4-European/ North American                   :254
##  4-Latin American/ Caribbean                  : 38
##  6-African (North; Central; West; East; South): 20
##  NA's                                         : 47
##                                    EducationStatus                   WorkStatus
##  1-Did not complete primary schooling     : 38    1-Employed (full time):127
##  2-High school to Pre-university education:143    2-Employed (part time): 61
##  3-Technical degree                       : 52    3-Not employed        :135
##  4-Post-graduate degree                   :120    4-Retired             : 11
##  NA's                                     : 47    5-Not willing to work : 19
##                                                   NA's                  : 47
##
##            MonthlyIncomeStatus CurrentPartner    hcv.pos        hiv.pos
##  1-<$2;000 CAD       :219      0-No :213      Min.   :0.00   Min.   :0.00
##  2-$2;001-$4;000 CAD :103      1-Yes:140      1st Qu.:0.00   1st Qu.:0.00
##  3-$4;001-$6;000 CAD : 22      NA's : 47      Median :0.00   Median :0.00
##  4-$6;001-$$8;000 CAD:  7                     Mean   :0.15   Mean   :0.04
##  5-$8;001 and above  :  2                     3rd Qu.:0.00   3rd Qu.:0.00
##  NA's                : 47                     Max.   :1.00   Max.   :1.00
##
##   syphilis.pos       new.hcv       new.syphilis      new.hiv
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.0000   Median :0.000   Median :0.000   Median :0.000
##  Mean   :0.0125   Mean   :0.035   Mean   :0.005   Mean   :0.005
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##
##       site
##  Min.   :0.00
##  1st Qu.:0.00
##  Median :0.00
##  Mean   :0.41
##  3rd Qu.:1.00
##  Max.   :1.00
##
```

sing default methods (5 iterations). No cluster

```
ini.imp <- mice(multiplex4imp2, nnet.MaxNWts = 2000, print = F)
# adjust the weights to not get an error message
# do not print the results
```

Which variables were used as predictors?

```
ini.pred <- ini.imp$predictorMatrix

prediction.matrix(ini.imp$predictorMatrix)
```

What methods were used for each?

```
ini.meth <- ini.imp$method
ini.meth
```

```
##                p_id      Date.Recruited      SexualActivity           CondomUse
##                  ""                  ""           "polyreg"           "polyreg"
##      SexualPartners             PastSTD         TypePastSTD     PastInjectDrugs
##           "polyreg"           "polyreg"           "polyreg"           "polyreg"
##        ShareNeedles             Alcohol         PastHIVTest         PastHCVTest
##           "polyreg"           "polyreg"           "polyreg"           "polyreg"
##    PastSyphilisTest        PastSyphilis                 Age              Gender
##           "polyreg"           "polyreg"           "polyreg"           "polyreg"
##             Descent     EducationStatus          WorkStatus MonthlyIncomeStatus
##           "polyreg"           "polyreg"           "polyreg"           "polyreg"
##      CurrentPartner             hcv.pos             hiv.pos        syphilis.pos
##            "logreg"                  ""                  ""                  ""
##             new.hcv        new.syphilis             new.hiv                site
##                  ""                  ""                  ""                  ""
```

Transformation

Transform the selected data to binary variables (see analysis plan)

```
multiplex4imp.1l2 <- multiplex4imp2 %>%
  mutate(
    SexualActivity = ifelse(
      SexualActivity %in% c("2-I do not wish to answer", "0-No"),
      "0-No or no wish to answer",
      as.character(SexualActivity)
    ),
    CondomUse = ifelse(
      CondomUse %in% c("1-Always", "2-Sometimes"),
      "1-Ever",
      as.character(CondomUse)
    ),
    SexualPartners = case_when(
```

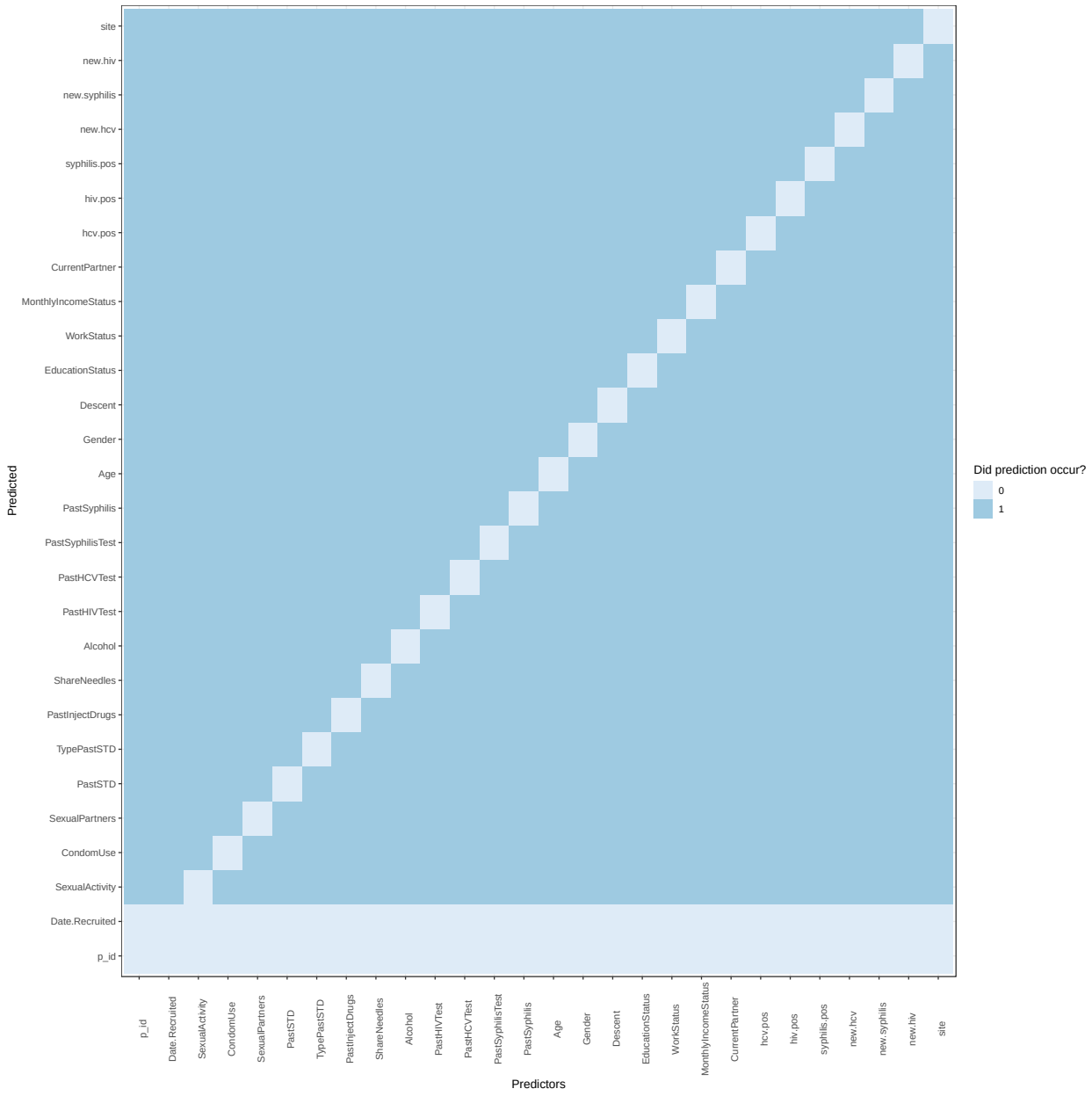


Figure 4: Heatmap of predictor matrix. The columns are the variables predicted and rows show which value were acting as predictors. A value of 0 means no predicion occurred

```
    SexualPartners %in% c("0-0", "1-1", "2-2 to 5") ~ "0 to 5",
    SexualPartners %in% c("3-6 to 10", "4-11 or more") ~ "6 or more",
    TRUE ~ as.character(SexualPartners)
  ),
  PastSTD = ifelse(
   PastSTD  %in% c("2-I do not know", "0-No"),
    "0-No or do not know",
    as.character(PastSTD)
  ),
  PastInjectDrugs = case_when(
    PastInjectDrugs %in% c("2-I do not wish to answer", "0-No")~
    "0-No or no wish to answer",
    PastInjectDrugs %in% "Yes"~ "1-Yes",
    T~as.character(PastInjectDrugs)
  ),
  ShareNeedles = ifelse(
    ShareNeedles %in% c("None", "0-No"),
    "0-No",
    as.character(ShareNeedles)
  ),
  Alcohol = ifelse(
    Alcohol %in% c("1-1-2 times per week", "2-3-5 times per week"),
    "1-5 times per week",
    as.character(Alcohol)
    ),
    # across(c(PastHIVTest, PastHCVTest, PastSyphilisTest),
    #        ~ifelse(c(.) %in% c("2-Yes; 6 months to 1 year ago",
  #"3-Yes; more than 1 year ago"),
    #                "more than 6 months ago",
    #                as.character(.)))
    # for now uncomment until we know how to deal with the variables
  PastSyphilis = ifelse(
   PastSyphilis %in% c("2-I do not know", "0-No"),
    "0-No or do not know",
    as.character(PastSyphilis)
  ),
  Age = case_when(
    Age %in% c("1-18-24", "1-25-34") ~ "18 to 34",
    Age %in% c("3-35-44", "4-45-54", "5-55 and above") ~ "35 and above",
    TRUE ~ as.character(Age)
  ),
  Gender = ifelse(
    Gender %in% "4-Bisexual",
    NA,
    as.character(Gender)
  ),
  Descent = case_when(
    Descent %in% c("4-European/ North American")
    ~ "Non-Mediterranean European/North American",
    !Descent %in% c("Non-Mediterranean European/North American", NA)
    ~ "Mediterranean European, African, Latin American, Caribbean and/or Asian",
    TRUE ~ as.character(Descent)
  ),
```

```
   EducationStatus= ifelse(
     !EducationStatus %in% c("4-Post-graduate degree", NA),
     "No post-graduate degree",
     as.character(EducationStatus)
   ),
   WorkStatus = case_when(
     WorkStatus %in% c("1-Employed (full time)", "2-Employed (part time)")
     ~ "Currently employed",
     !WorkStatus %in% c("Currently employed", NA) ~ "Not currently employed"
   ),
  MonthlyIncomeStatus = ifelse(
    !MonthlyIncomeStatus %in% c("1-<$2;000 CAD", NA),
    "$2000 CAD and above",
    as.character(MonthlyIncomeStatus)
  # ),
  # PreferenceTiming = ifelse(
  #   !PreferenceTiming %in% c("0-1 day", NA),
  #   "2 days or more",
  #   as.character(PreferenceTiming)
  # ),
  # RecommendRapidTests = ifelse(
  #   RecommendRapidTests %in% c("0-No", "2-Not Sure"),
  #   "No or not sure",
  #   as.character(RecommendRapidTests)
  # ),
  # PreferenceFollowup = case_when(
  #  PreferenceFollowup %in% c("3-Face-to-face", "4-No preference")
  #  ~ "No preference for digital follow-up",
  #  !PreferenceFollowup %in% c("No preference for digital follow-up", NA)
  #  ~"Digital follow-up",
  #  TRUE ~ as.character(PreferenceFollowup)
  )
      ) %>%
  mutate(across(SexualActivity:CurrentPartner, as.factor))

summary(multiplex4imp.1l2)
```

```
##      p_id           Date.Recruited                      SexualActivity
##  Length:400         Length:400         0-No or no wish to answer:106
##  Class :character   Class :character   1-Yes                     :289
##  Mode  :character   Mode  :character   NA's                      :  5
##
##
##
##
##    CondomUse     SexualPartners                PastSTD           TypePastSTD
##  0-Never: 98   0 to 5   :324    0-No or do not know:216   None         :216
##  1-Ever :272   6 or more: 60    1-Yes              :169   3-Chlamydia  : 72
##  NA's   : 30   NA's     : 16    NA's               : 15   4-Gonorrhoeae: 43
##                                                          6-Other      : 25
##                                                          0-HIV        : 10
##                                                          (Other)      : 13
##                                                          NA's         : 21
```

```
##                   PastInjectDrugs ShareNeedles               Alcohol
##  0-No or no wish to answer:255    0-No :307    0-No              :209
##  1-Yes                    :141    1-Yes: 85    1-5 times per week:187
##  NA's                     :  4    NA's :  8    NA's              :  4
##
##
##
##
##                         PastHIVTest                          PastHCVTest
##  0-No                           : 57   0-No                           : 44
##  1-Yes; less than 6 months ago:141   1-Yes; less than 6 months ago:140
##  2-Yes; 6 months to 1 year ago: 47   2-Yes; 6 months to 1 year ago: 48
##  3-Yes; more than 1 year ago  :154   3-Yes; more than 1 year ago  :167
##  NA's                           :  1   NA's                           :  1
##
##
##                       PastSyphilisTest              PastSyphilis
##  0-No                           :  4     0-No or do not know:349
##  1-Yes; less than 6 months ago:127     1-Yes              : 48
##  2-Yes; 6 months to 1 year ago: 47     NA's               :  3
##  3-Yes; more than 1 year ago  :220
##  NA's                           :  2
##
##
##            Age                Gender
##  18 to 34    :156   1-Male       :255
##  35 and above:197   2-Female     : 73
##  NA's        : 47   3-Transgender:  7
##                     5-No response: 11
##                     NA's         : 54
##
##
##                                                                    Descent
##  Mediterranean European, African, Latin American, Caribbean and/or Asian: 99
##  Non-Mediterranean European/North American                              :254
##  NA's                                                                   : 47
##
##
##
##
##                 EducationStatus                   WorkStatus
##  4-Post-graduate degree :120    Currently employed    :188
##  No post-graduate degree:233    Not currently employed:165
##  NA's                   : 47    NA's                  : 47
##
##
##
##
##           MonthlyIncomeStatus CurrentPartner    hcv.pos        hiv.pos
##  $2000 CAD and above:134      0-No :213      Min.   :0.00   Min.   :0.00
##  1-<$2;000 CAD      :219      1-Yes:140      1st Qu.:0.00   1st Qu.:0.00
##  NA's               : 47      NA's : 47      Median :0.00   Median :0.00
##                                              Mean   :0.15   Mean   :0.04
##                                              3rd Qu.:0.00   3rd Qu.:0.00
```

```
##                                           Max.   :1.00   Max.   :1.00
##
##   syphilis.pos       new.hcv       new.syphilis      new.hiv
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.0000   Median :0.000   Median :0.000   Median :0.000
##  Mean   :0.0125   Mean   :0.035   Mean   :0.005   Mean   :0.005
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##
##       site
##  Min.   :0.00
##  1st Qu.:0.00
##  Median :0.00
##  Mean   :0.41
##  3rd Qu.:1.00
##  Max.   :1.00
##
```

Run imputation with logistic regression for transformed factors

```
mice.1l <- mice(multiplex4imp.1l2, nnet.MaxNWts = 2000, print = F)

summary(complete(mice.1l))
```

```
##      p_id           Date.Recruited                       SexualActivity
##  Length:400         Length:400         0-No or no wish to answer:107
##  Class :character   Class :character   1-Yes                    :293
##  Mode  :character   Mode  :character
##
##
##
##
##    CondomUse     SexualPartners                PastSTD
##  0-Never:106   0 to 5   :335    0-No or do not know:222
##  1-Ever :294   6 or more: 65    1-Yes              :178
##
##
##
##
##
##            TypePastSTD                   PastInjectDrugs ShareNeedles
##  0-HIV           : 11   0-No or no wish to answer:258    0-No :312
##  2-Herpes simplex:  4   1-Yes                    :142    1-Yes: 88
##  3-Chlamydia     : 78
##  4-Gonorrhoeae   : 47
##  5-Syphilis      : 12
##  6-Other         : 25
##  None            :223
##                Alcohol                           PastHIVTest
##  0-No              :213   0-No                          : 57
##  1-5 times per week:187   1-Yes; less than 6 months ago:141
```

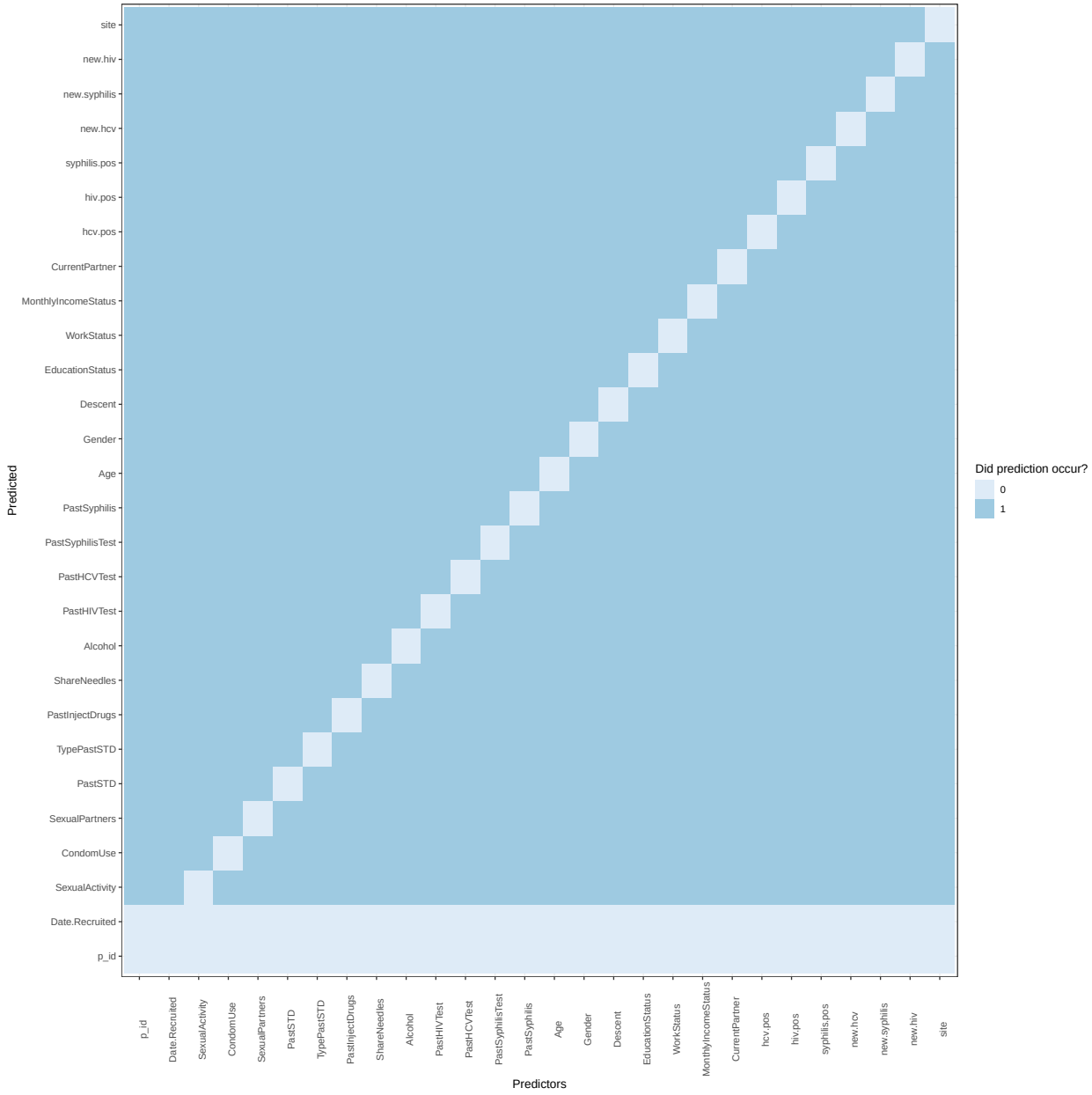
```
##                           2-Yes; 6 months to 1 year ago: 47
##                           3-Yes; more than 1 year ago  :155
##
##
##
##                         PastHCVTest                         PastSyphilisTest
##  0-No                           : 44   0-No                           :  4
##  1-Yes; less than 6 months ago:140   1-Yes; less than 6 months ago:128
##  2-Yes; 6 months to 1 year ago: 49   2-Yes; 6 months to 1 year ago: 47
##  3-Yes; more than 1 year ago  :167   3-Yes; more than 1 year ago  :221
##
##
##
##               PastSyphilis           Age                 Gender
##  0-No or do not know:351   18 to 34    :177   1-Male       :284
##  1-Yes               : 49   35 and above:223   2-Female     : 89
##                                               3-Transgender: 10
##                                               5-No response: 17
##
##
##
##                                                                    Descent
##  Mediterranean European, African, Latin American, Caribbean and/or Asian:106
##  Non-Mediterranean European/North American                               :294
##
##
##
##
##
##                 EducationStatus                   WorkStatus
##  4-Post-graduate degree :127    Currently employed    :195
##  No post-graduate degree:273    Not currently employed:205
##
##
##
##
##
##           MonthlyIncomeStatus CurrentPartner    hcv.pos        hiv.pos
##  $2000 CAD and above:141      0-No :247      Min.   :0.00   Min.   :0.00
##  1-<$2;000 CAD      :259      1-Yes:153      1st Qu.:0.00   1st Qu.:0.00
##                                              Median :0.00   Median :0.00
##                                              Mean   :0.15   Mean   :0.04
##                                              3rd Qu.:0.00   3rd Qu.:0.00
##                                              Max.   :1.00   Max.   :1.00
##
##   syphilis.pos       new.hcv       new.syphilis      new.hiv
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.0000   Median :0.000   Median :0.000   Median :0.000
##  Mean   :0.0125   Mean   :0.035   Mean   :0.005   Mean   :0.005
##  3rd Qu.:0.0000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##
##       site
```

```
##  Min.   :0.00
##  1st Qu.:0.00
##  Median :0.00
##  Mean   :0.41
##  3rd Qu.:1.00
##  Max.   :1.00
##
```

```
# save results
mice.1l.long<- complete(mice.1l, "long", include=TRUE)

mice.1l.broad <- complete(mice.1l, "broad", include = T)

write.csv(mice.1l.long, "3_intermediate/imputed/imputed_data_long_no_prom_1l.csv", row.names = F)

write.csv(mice.1l.broad, "3_intermediate/imputed/imputed_data_broad_no_prom_1l.csv")
```

Prediction matrix and methods

```
prediction.matrix(mice.1l$predictorMatrix)
```




```
meth.1l <- mice.1l$method
meth.1l
```

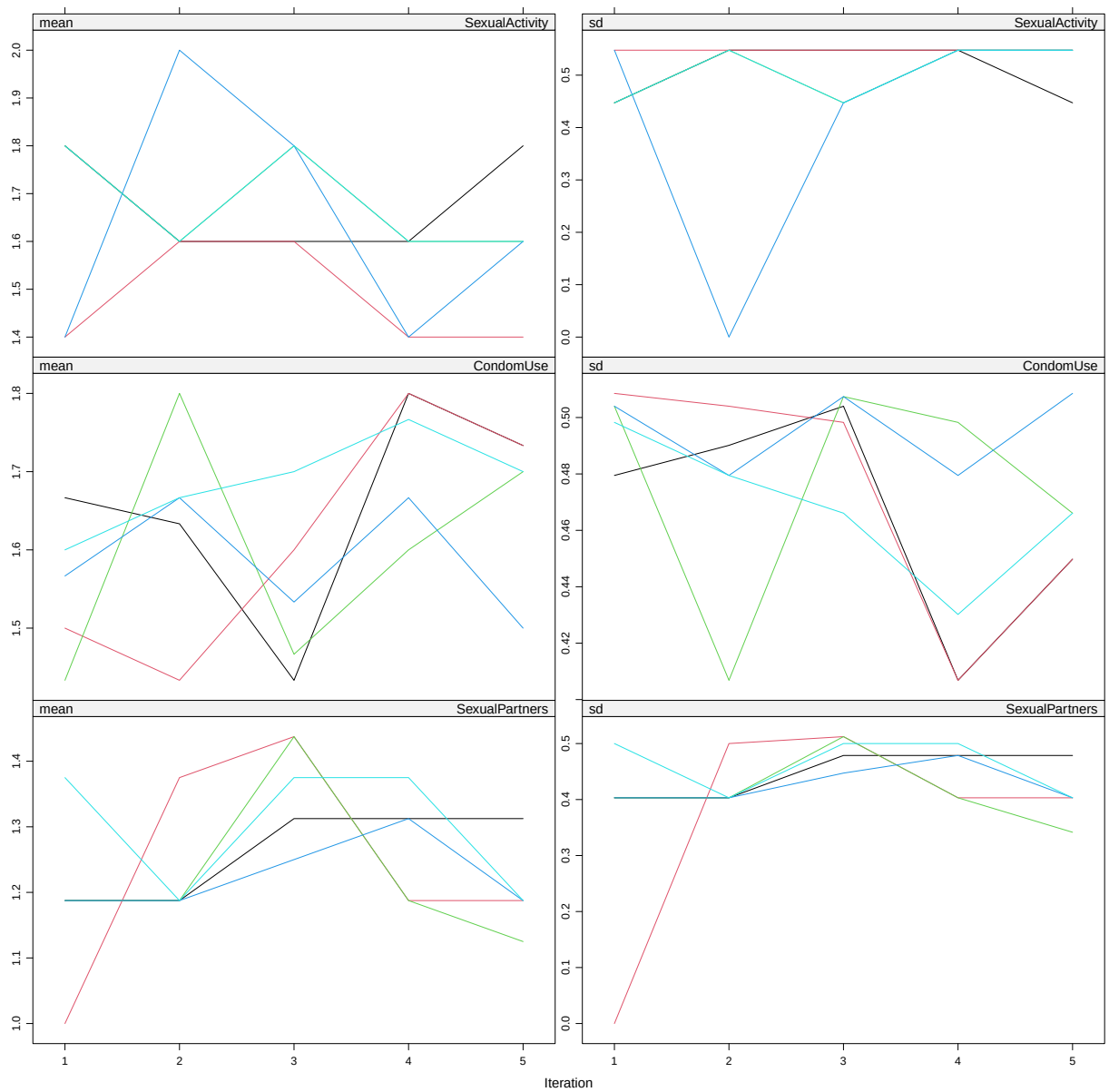
```
##                 p_id      Date.Recruited      SexualActivity          CondomUse
##                   ""                  ""            "logreg"           "logreg"
##      SexualPartners             PastSTD         TypePastSTD     PastInjectDrugs
##            "logreg"           "logreg"           "polyreg"           "logreg"
##        ShareNeedles             Alcohol        PastHIVTest        PastHCVTest
##            "logreg"           "logreg"           "polyreg"           "polyreg"
##    PastSyphilisTest        PastSyphilis                Age             Gender
##           "polyreg"           "logreg"           "logreg"           "polyreg"
##            Descent     EducationStatus        WorkStatus MonthlyIncomeStatus
##            "logreg"           "logreg"           "logreg"           "logreg"
##      CurrentPartner             hcv.pos            hiv.pos        syphilis.pos
```

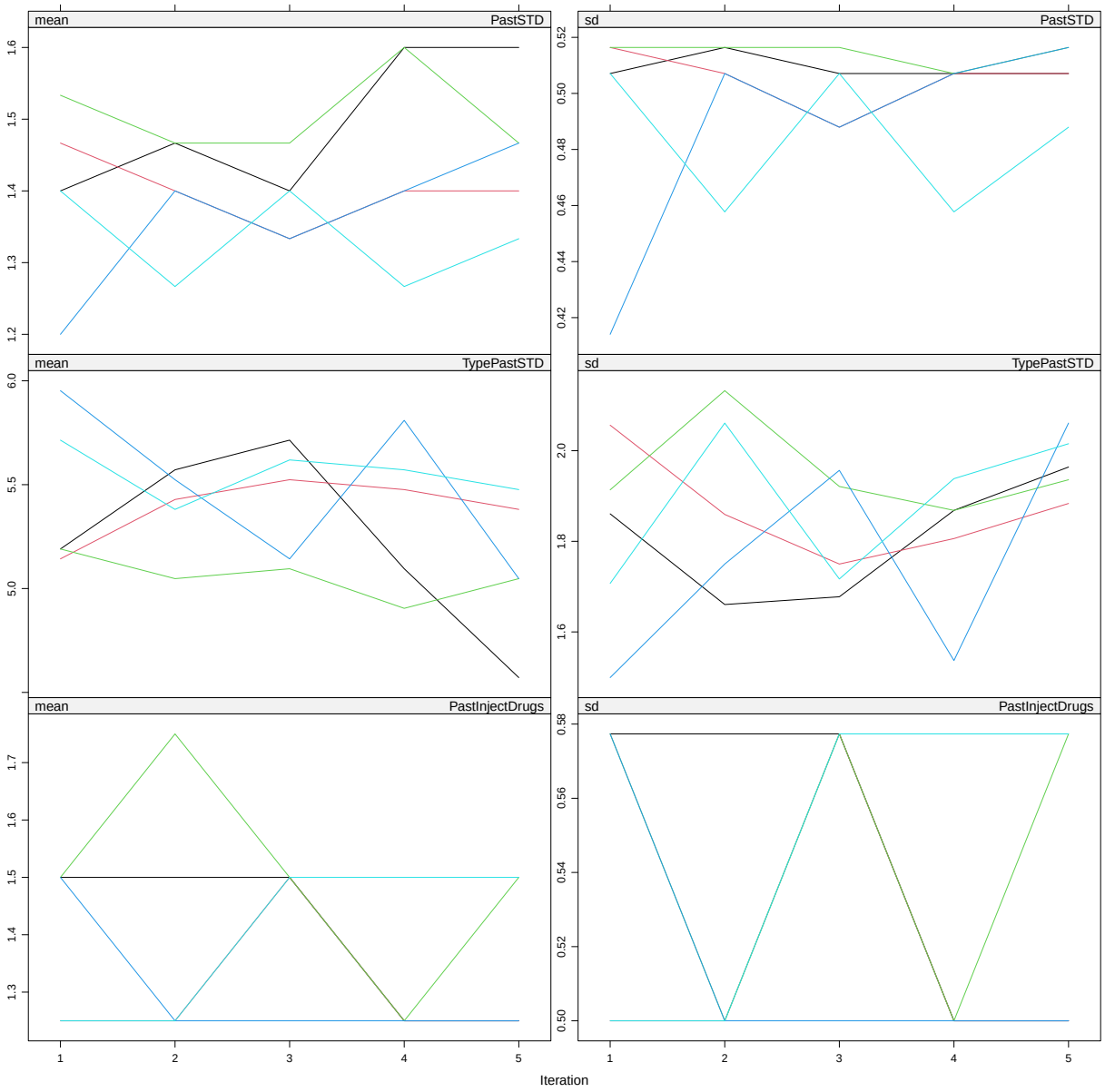
```
##          "logreg"                ""                ""                ""
##           new.hcv      new.syphilis           new.hiv              site
##                ""                ""                ""                ""
```

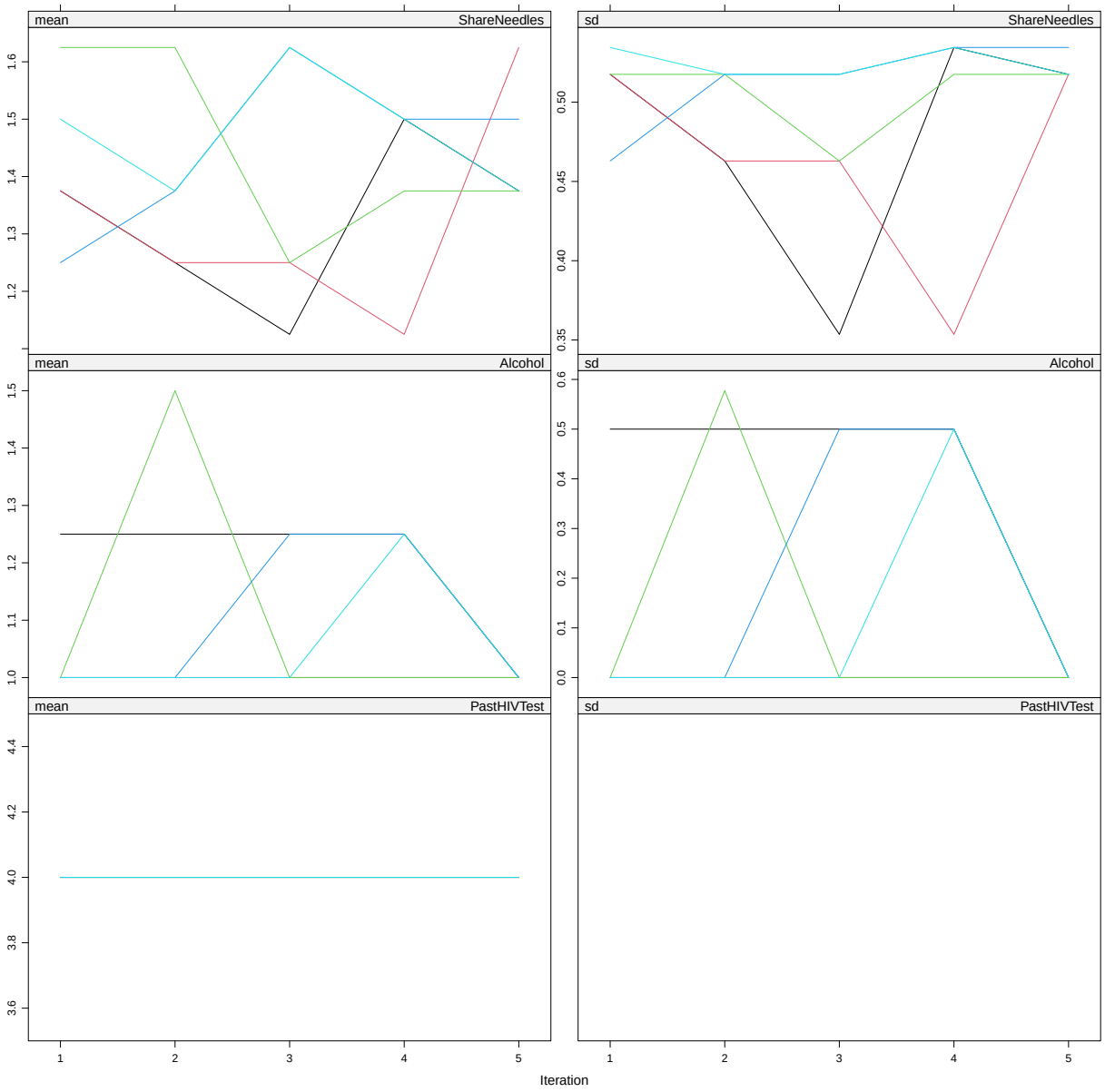
The variables transformed into binary factors are now imputed using logistic regression

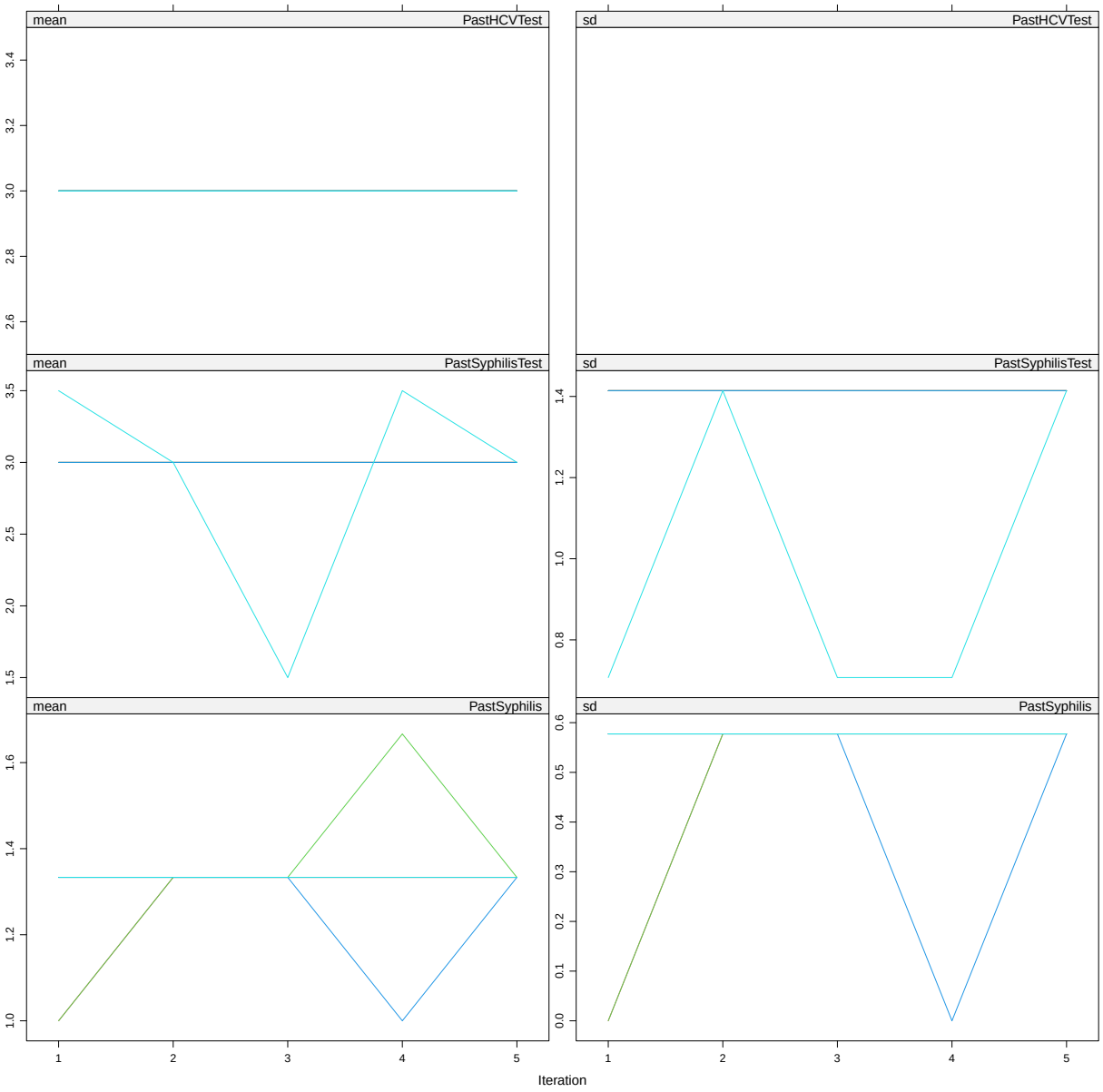
Diagnostic plots

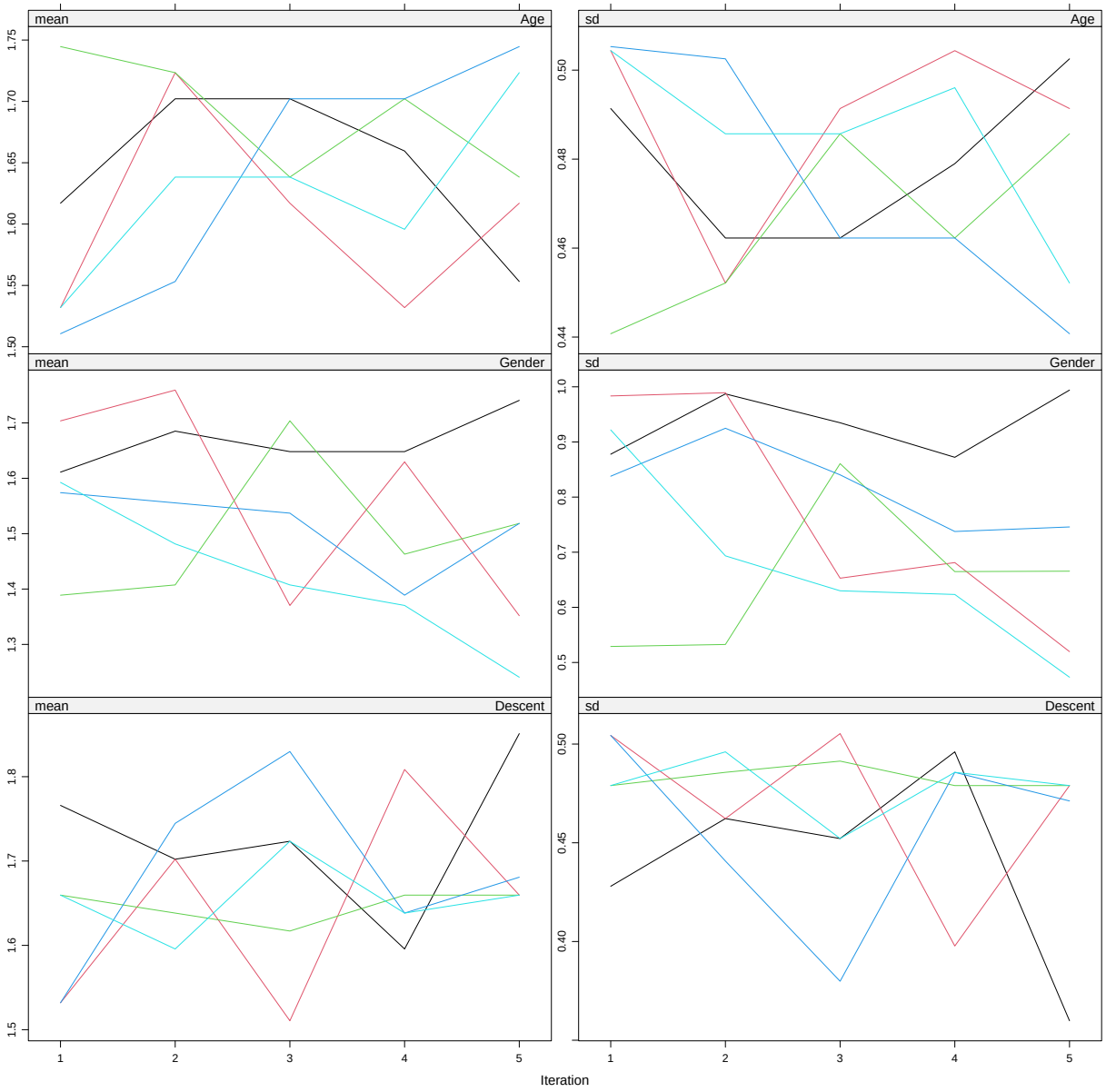
```
plot(mice.1l)
```

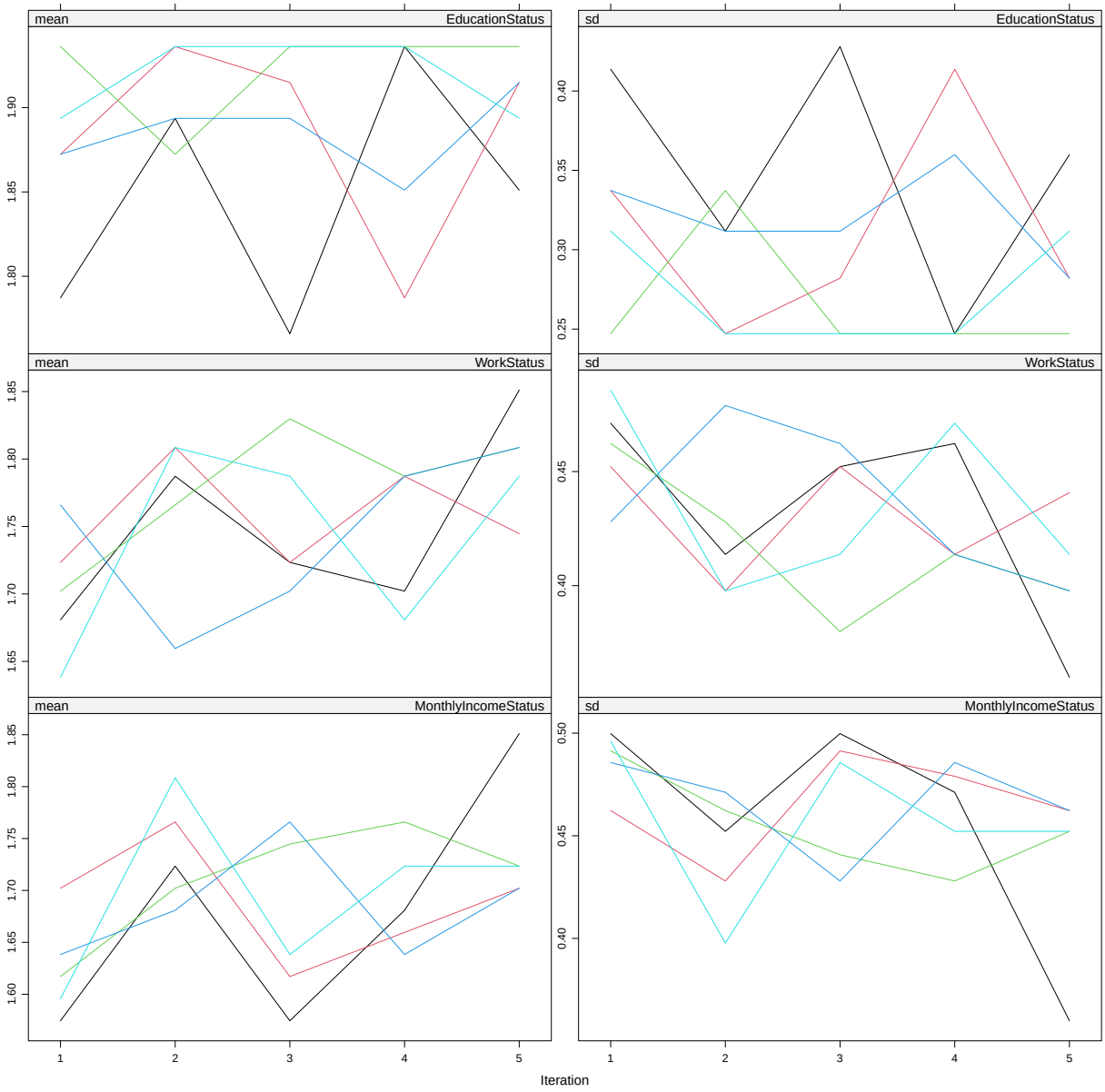


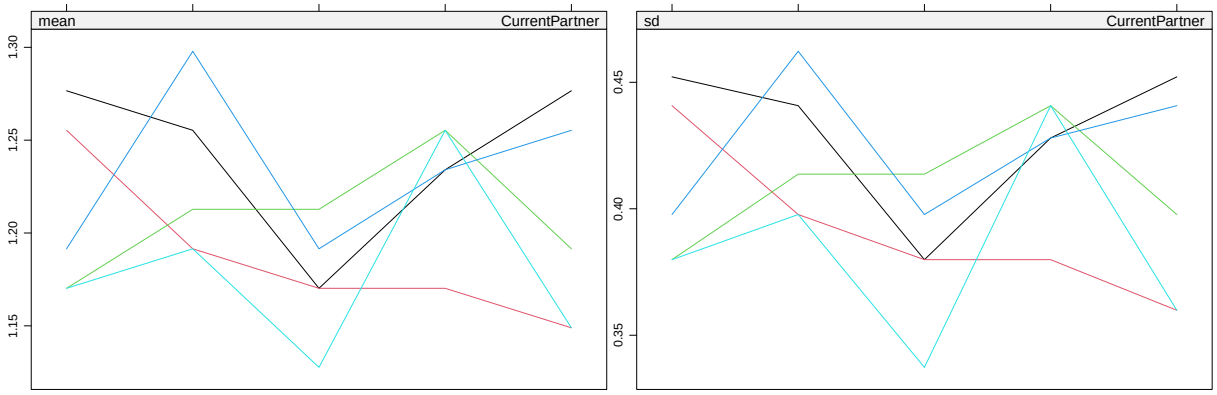

mean
PastSTD
sd
PastSTD
mean
TypePastSTD
sd
TypePastSTD
mean
PastInjectDrugs
sd
PastInjectDrugs
Iteration


mean
ShareNeedles
sd
ShareNeedles
mean
Alcohol
sd
Alcohol
mean
PastHIVTest
sd
PastHIVTest
Iteration


mean
PastHCVTest
sd
PastHCVTest
mean
PastSyphilisTest
sd
PastSyphilisTest
mean
PastSyphilis
sd
PastSyphilis
Iteration

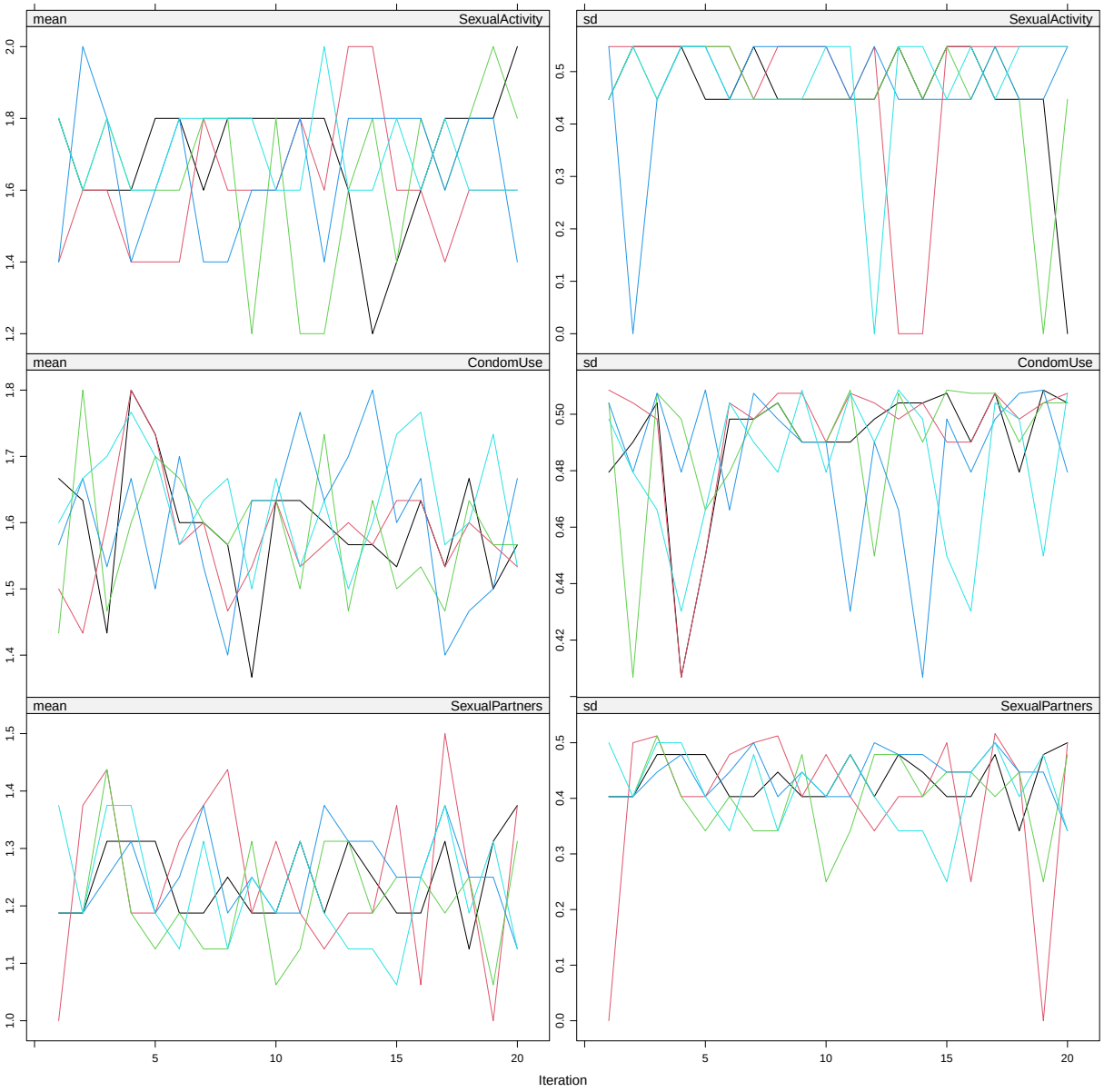

mean
Age
sd
Age
mean
Gender
sd
Gender
mean
Descent
sd
Descent
Iteration

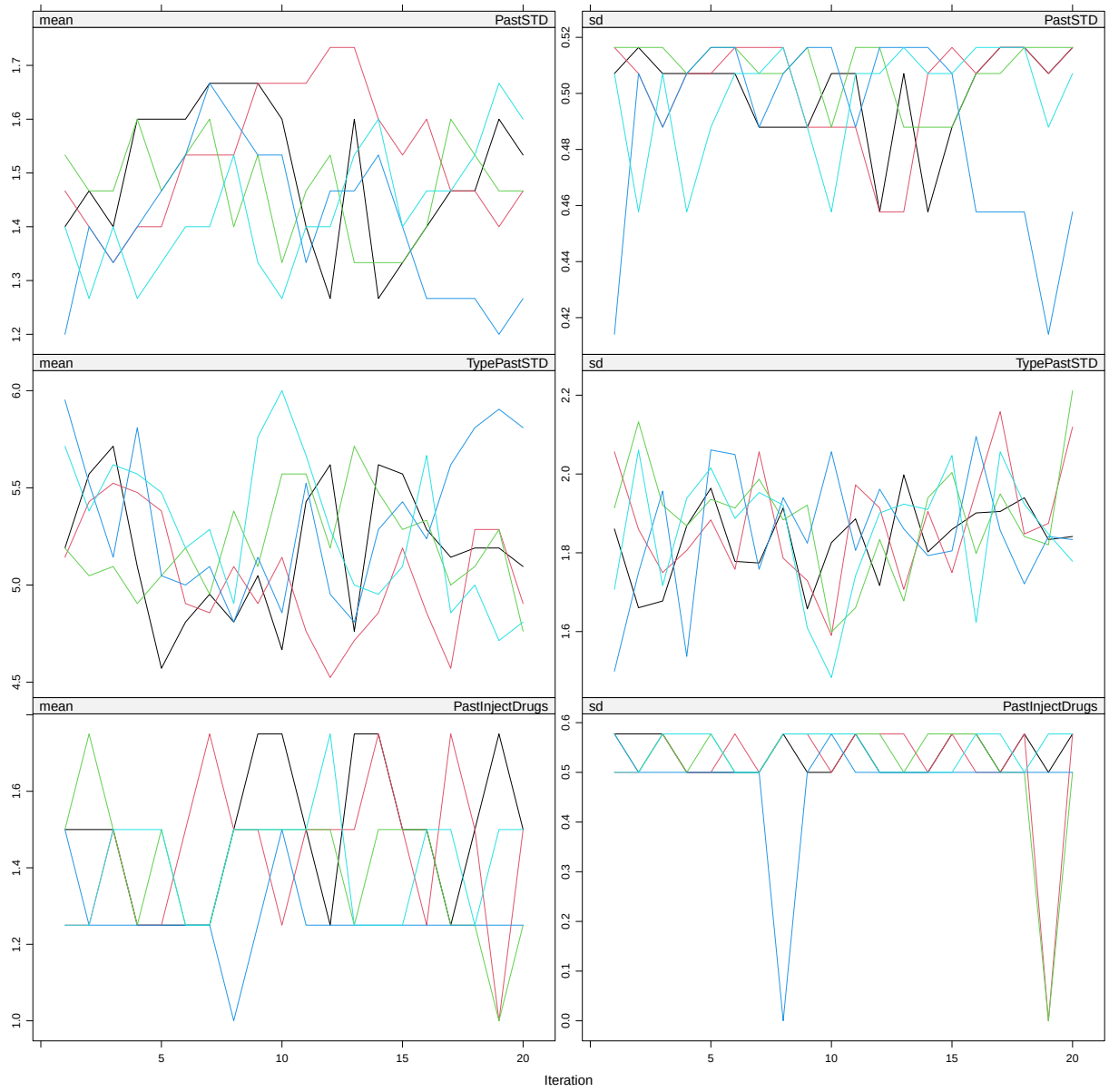

mean
EducationStatus
sd
EducationStatus
mean
WorkStatus
sd
WorkStatus
mean
MonthlyIncomeStatus
sd
MonthlyIncomeStatus
Iteration

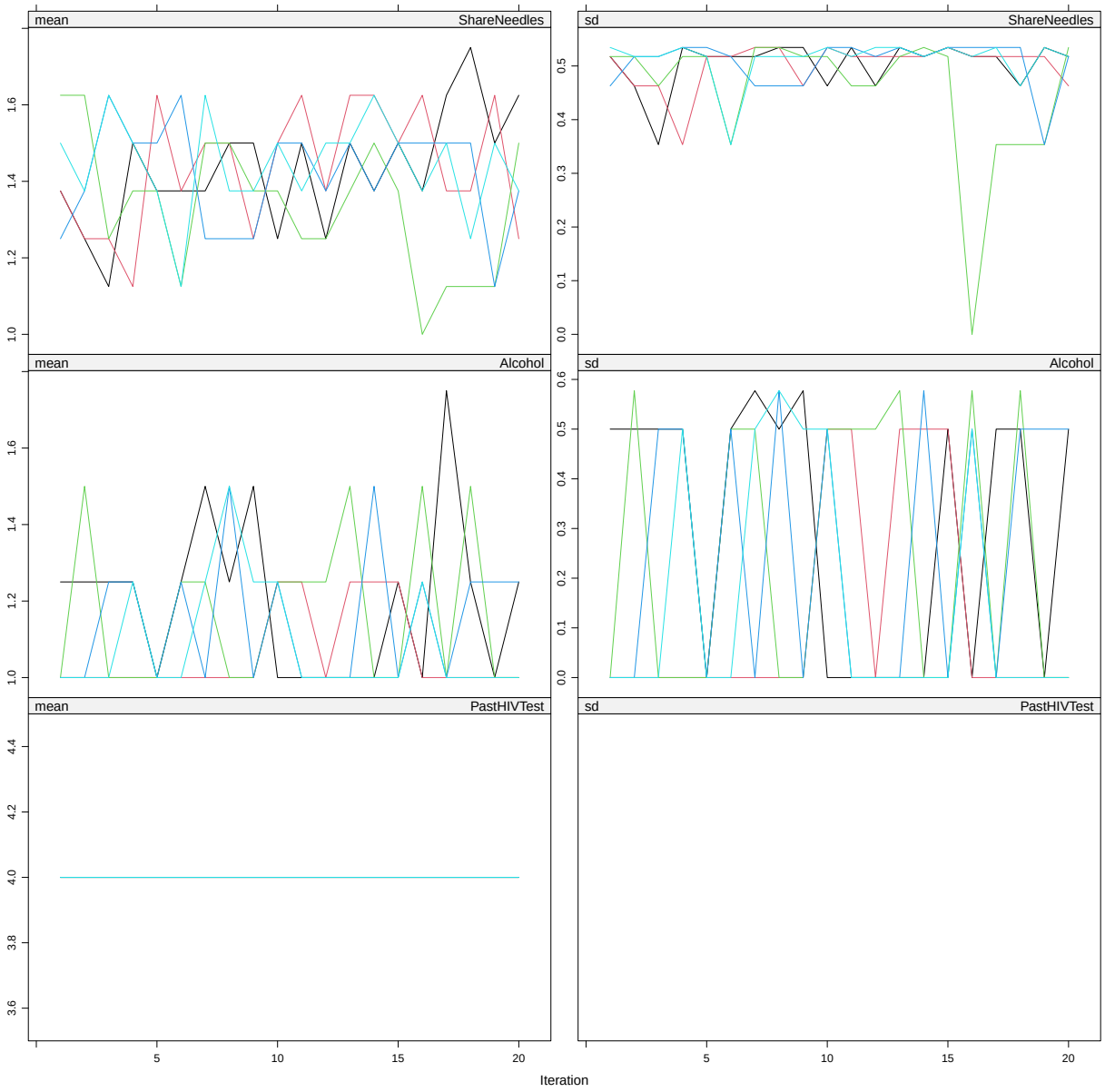


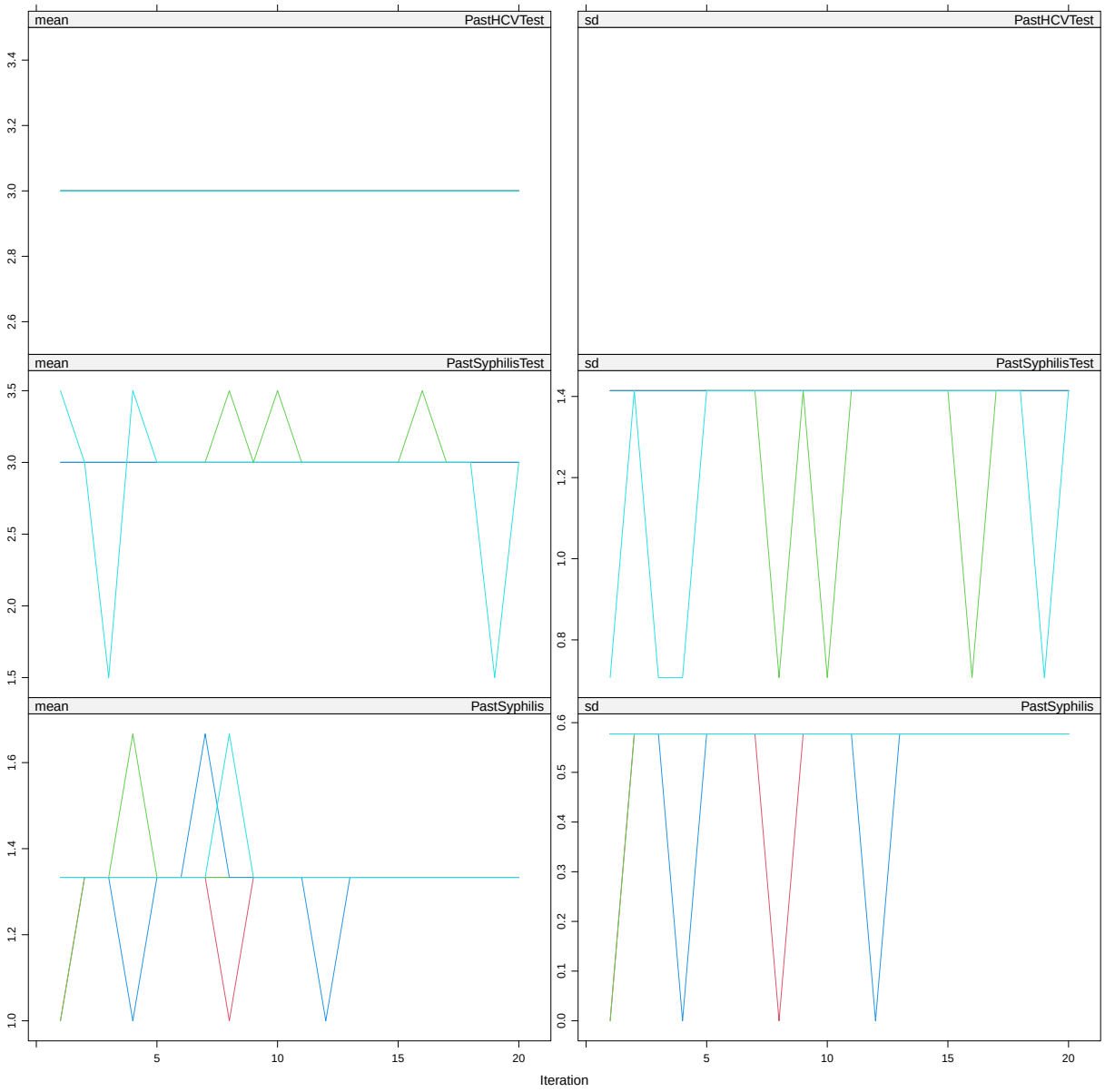
add more

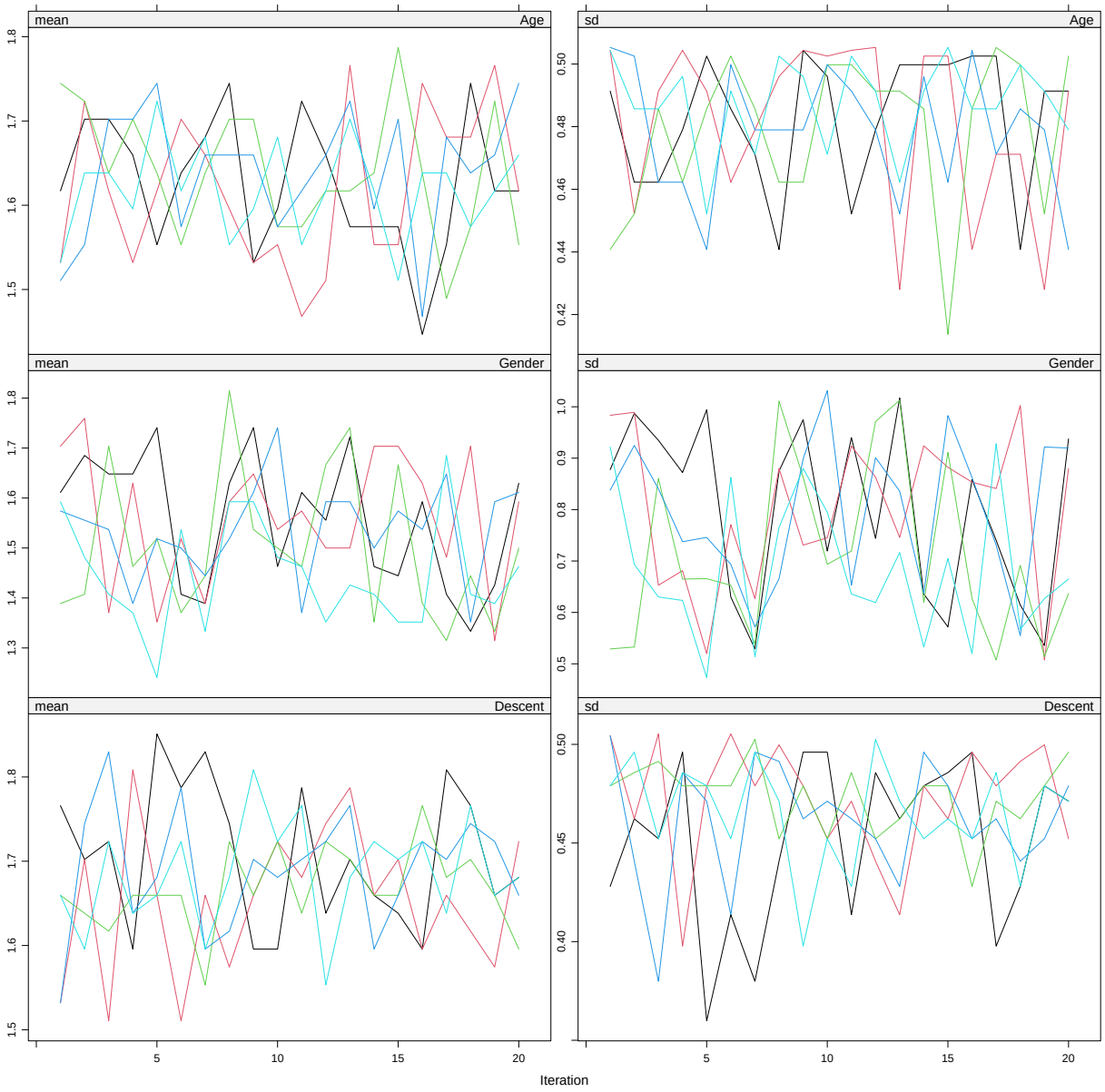
```
mice.1l.20 <- mice.mids(mice.1l, nnet.MaxNWts = 2000, print =F, maxit=15)

plot(mice.1l.20)
```

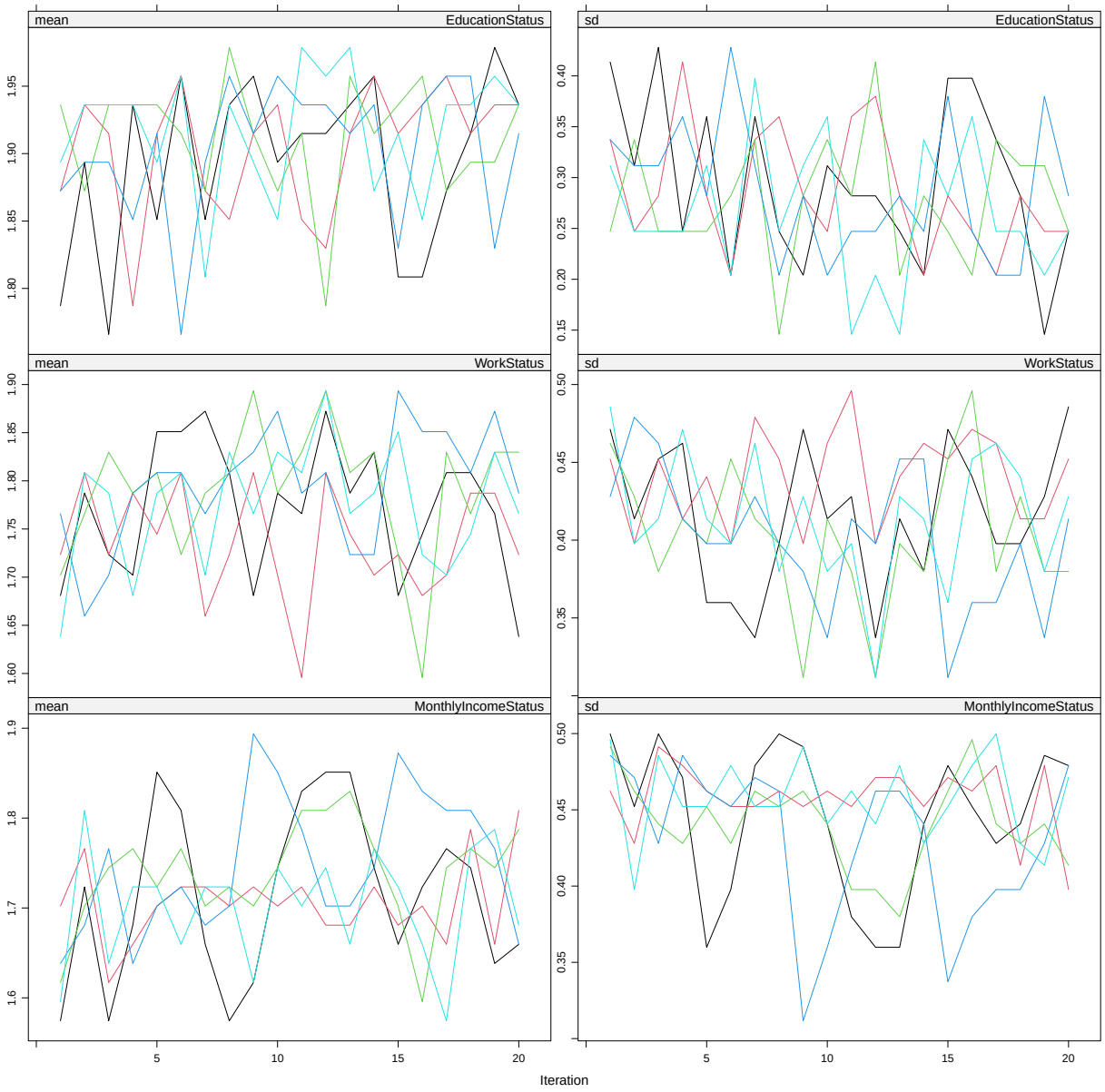

mean
SexualActivity
sd
SexualActivity
mean
CondomUse
sd
CondomUse
mean
SexualPartners
sd
SexualPartners
Iteration

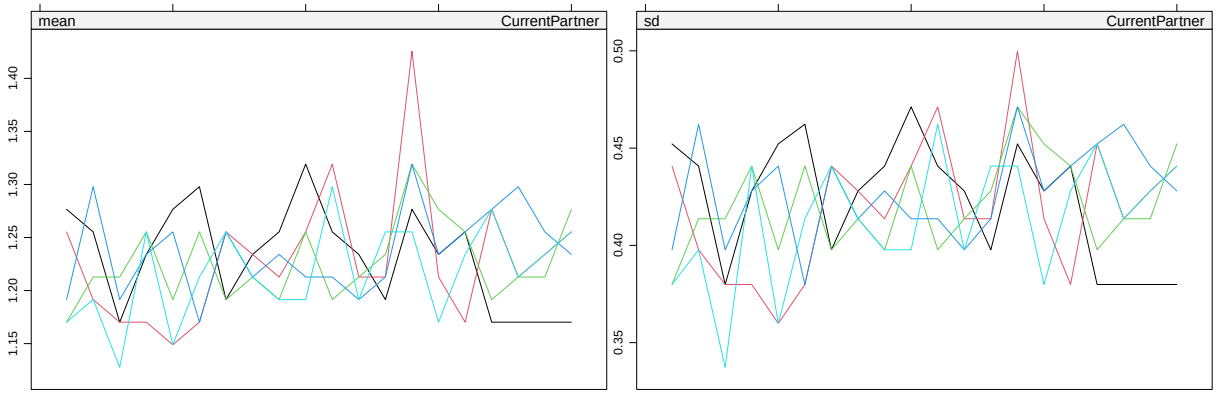

mean
PastSTD
sd
PastSTD
mean
TypePastSTD
sd
TypePastSTD
mean
PastInjectDrugs
sd
PastInjectDrugs
Iteration


mean
ShareNeedles
sd
ShareNeedles
mean
Alcohol
sd
Alcohol
mean
PastHIVTest
sd
PastHIVTest
Iteration


mean
PastHCVTest
sd
PastHCVTest
mean
PastSyphilisTest
sd
PastSyphilisTest
mean
PastSyphilis
sd
PastSyphilis
Iteration


mean
Age
sd
Age
mean
Gender
sd
Gender
mean
Descent
sd
Descent
Iteration


mean
EducationStatus
sd
EducationStatus
mean
WorkStatus
sd
WorkStatus
mean
MonthlyIncomeStatus
sd
MonthlyIncomeStatus
Iteration



Sensitivity analysis

Density plot of selected variables 1l imputation

```
dplot1 <- densityplot(mice.1l, ~SexualActivity)

dplot2 <- densityplot(mice.1l, ~CondomUse)
dplot3 <- densityplot(mice.1l, ~SexualPartners)
dplot4 <- densityplot(mice.1l, ~PastSTD)
dplot5 <- densityplot(mice.1l, ~TypePastSTD)
dplot6 <- densityplot(mice.1l, ~PastInjectDrugs)
dplot7 <- densityplot(mice.1l, ~ShareNeedles)
```

```
dplot8 <- densityplot(mice.1l, ~Alcohol)
#dplot9 <- densityplot(mice.1l, ~PastHIVTest)
dplot10 <- densityplot(mice.1l, ~PastSyphilisTest)

dplots.clin <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
                        dplot7, dplot8, dplot10, ncol =2 )
```

```
dplots.clin
```

```
## TableGrob (5 x 2) "arrange": 9 grobs
##   z     cells    name                           grob
## 1 1 (1-1,1-1) arrange lattice[GRID.lattice.63106]
## 2 2 (1-1,2-2) arrange lattice[GRID.lattice.63107]
## 3 3 (2-2,1-1) arrange lattice[GRID.lattice.63108]
## 4 4 (2-2,2-2) arrange lattice[GRID.lattice.63109]
## 5 5 (3-3,1-1) arrange lattice[GRID.lattice.63110]
## 6 6 (3-3,2-2) arrange lattice[GRID.lattice.63111]
## 7 7 (4-4,1-1) arrange lattice[GRID.lattice.63112]
## 8 8 (4-4,2-2) arrange lattice[GRID.lattice.63113]
## 9 9 (5-5,1-1) arrange lattice[GRID.lattice.63114]
```

```
ggsave (dplots.clin, filename = "3_intermediate/imputed/Densityplots_observed_vs_1l-imputed_data_clinica
         device = "png", width = 14, height = 14)

dplot11 <- densityplot(mice.1l, ~Age)
dplot12 <- densityplot(mice.1l, ~Gender)
dplot13 <- densityplot(mice.1l, ~Descent)
dplot14 <- densityplot(mice.1l, ~EducationStatus)
dplot15 <- densityplot(mice.1l, ~WorkStatus)
dplot16 <- densityplot(mice.1l, ~MonthlyIncomeStatus)
dplot17 <- densityplot(mice.1l, ~CurrentPartner)

dplots.ses <- grid.arrange(dplot11, dplot12, dplot13, dplot14, dplot15, dplot16,
                        dplot17, ncol =2)
```

```
ggsave (dplots.ses, filename = "3_intermediate/imputed/Densityplots_observed_vs_1l-imputed_data_ses_no_
         device = "png", width = 14, height = 14)

all.dplots <- grid.arrange(dplot1, dplot2, dplot3, dplot4, dplot5, dplot6,
                        dplot7, dplot8, dplot10, dplot11, dplot12, dplot13, dplot14, dplot15, dplot16,
                        dplot17, ncol = 2)
```

```
ggsave(all.dplots, filename = "4_outputs/density_1l_imp.png",
       device = "png", width = 14, height = 20)
```

References

https://www.rdocumentation.org/packages/mice/versions/3.16.0

See vignettes for examples

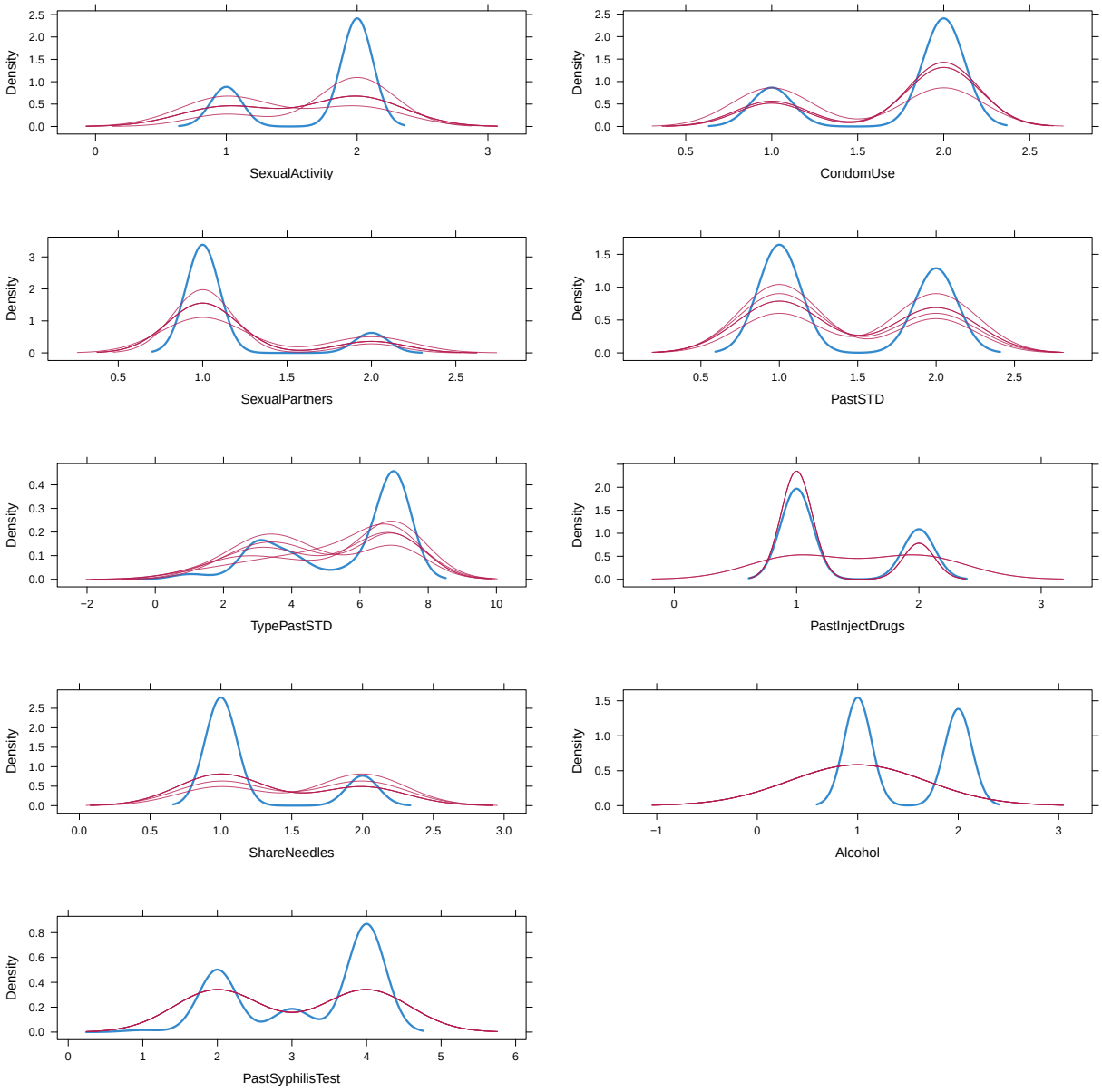


Figure 5: density plot of the values selected observed variables (blue) and their corresponding imputed values (red).

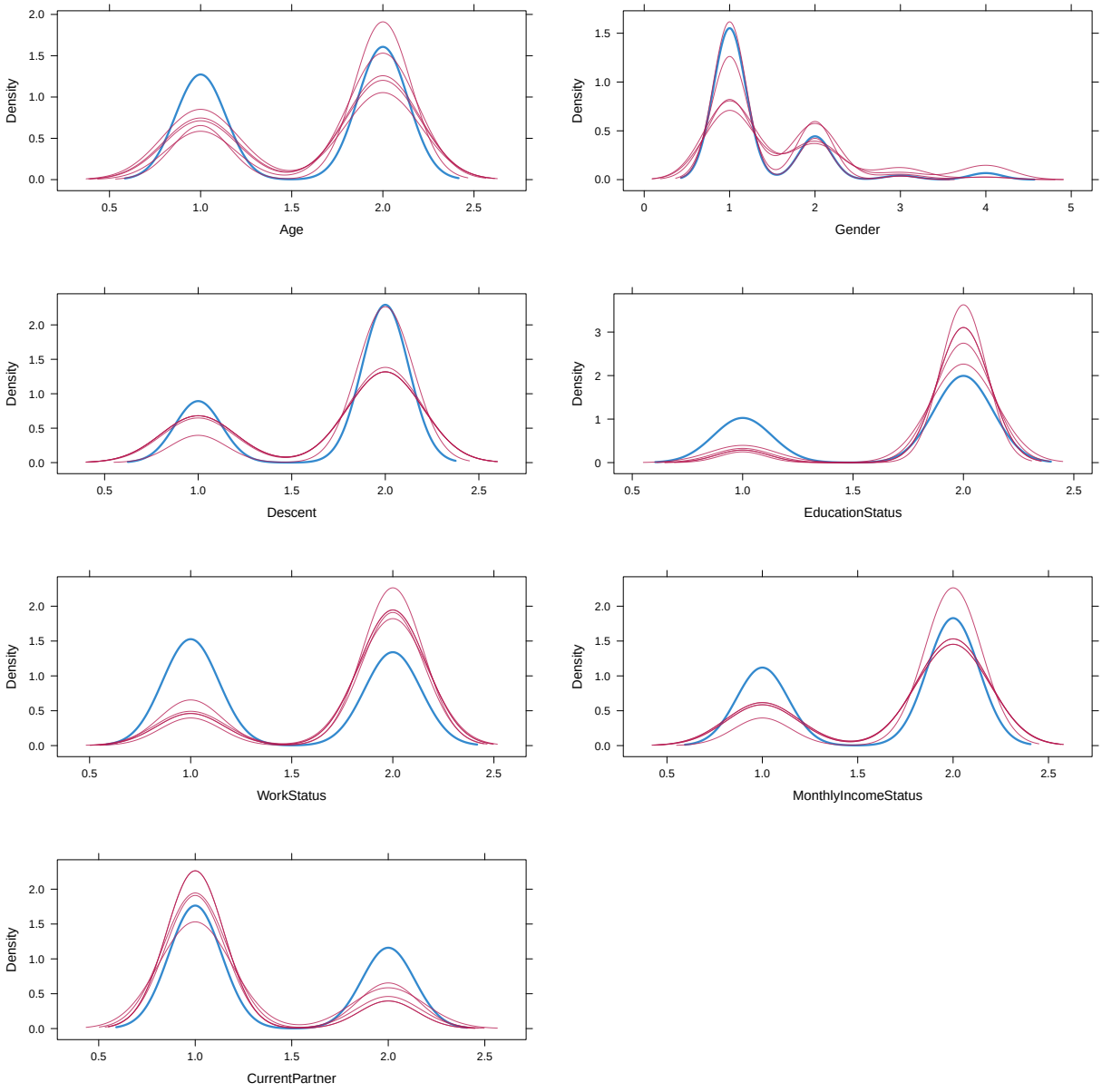


Figure 6: density plot of the values selected observed variables (blue) and their corresponding imputed values (red).

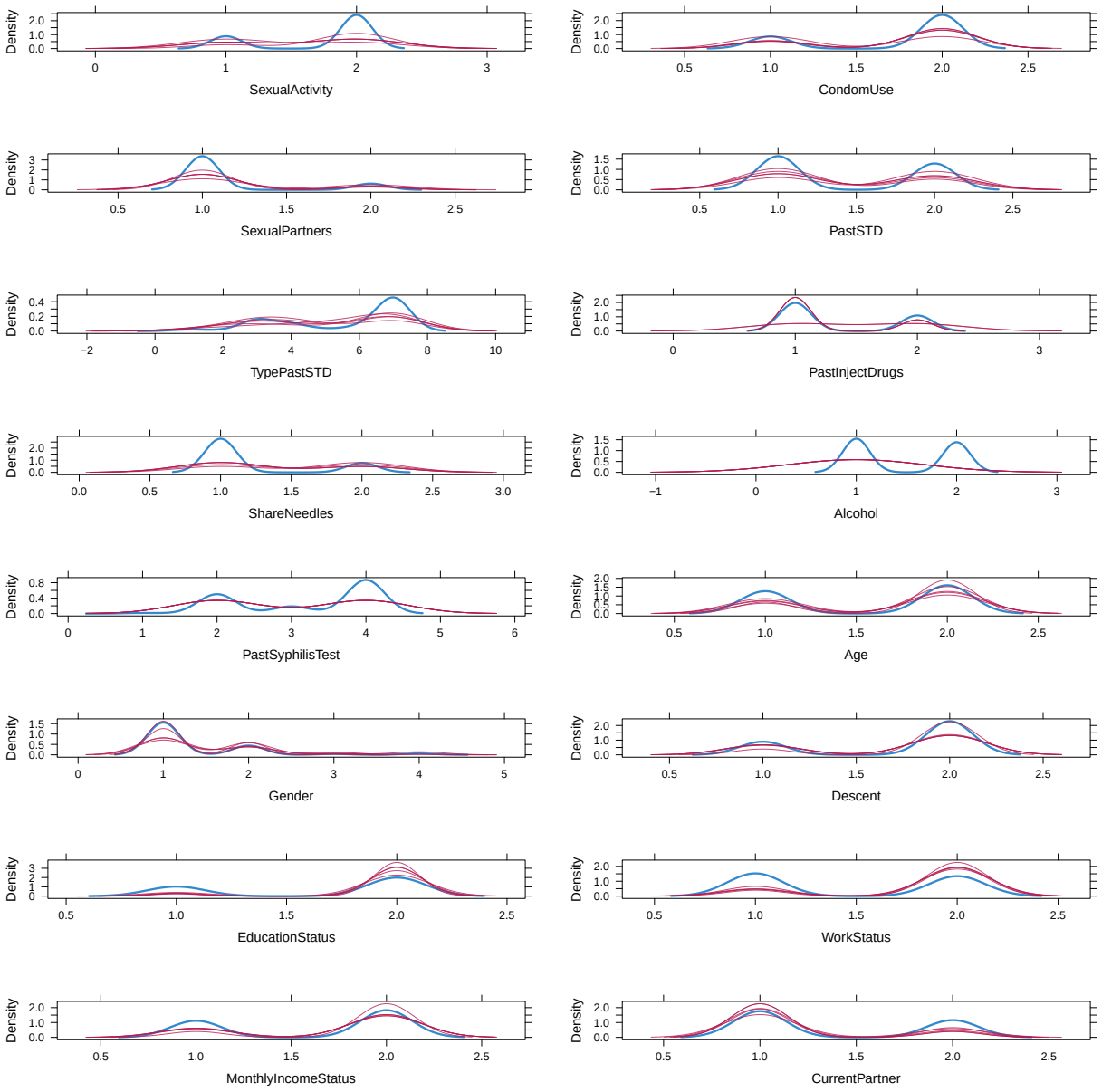


Figure 7: density plot of the values selected observed variables (blue) and their corresponding imputed values (red).