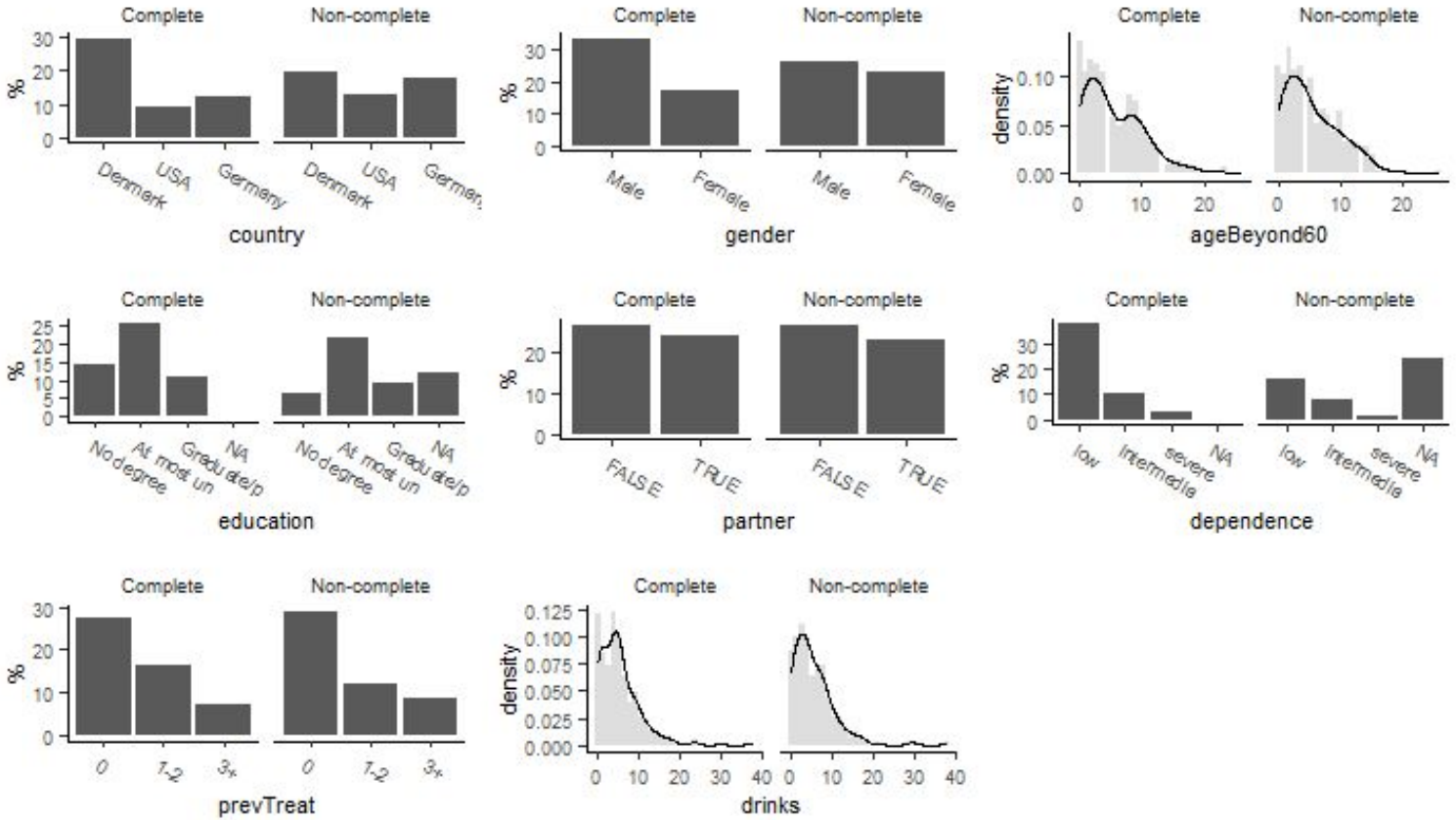
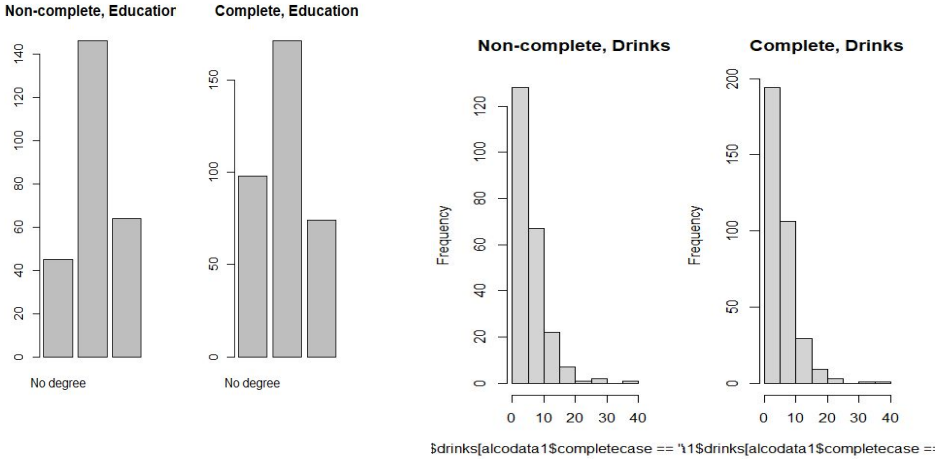
Alcodata1

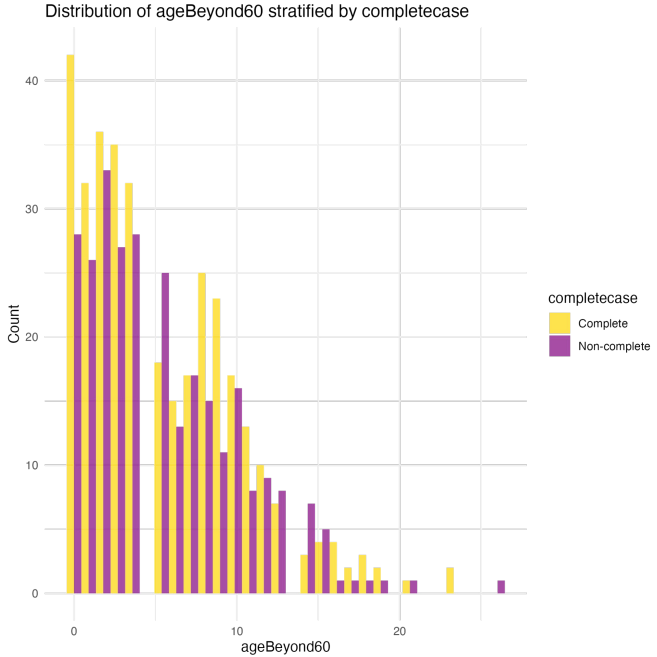
Exploratory analysis

Distributions of variables, stratified by complete case status

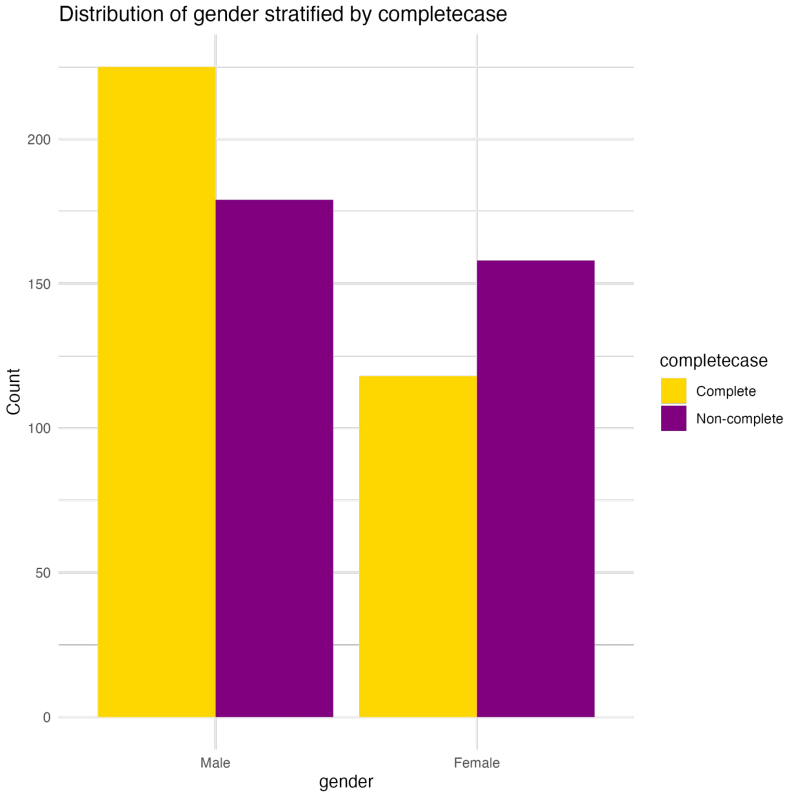


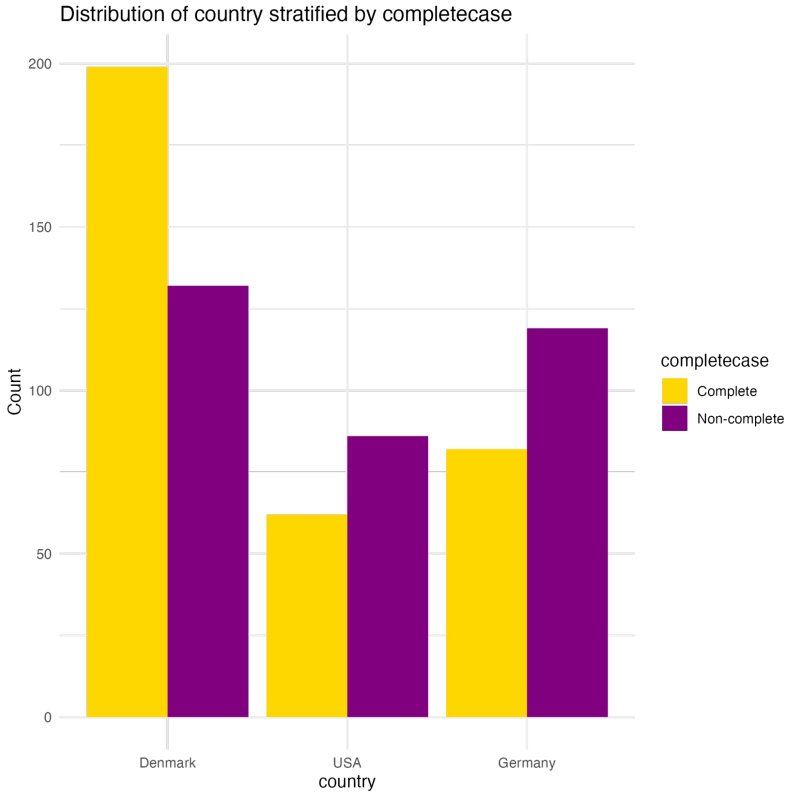
Distributions of variables, stratified by complete case status



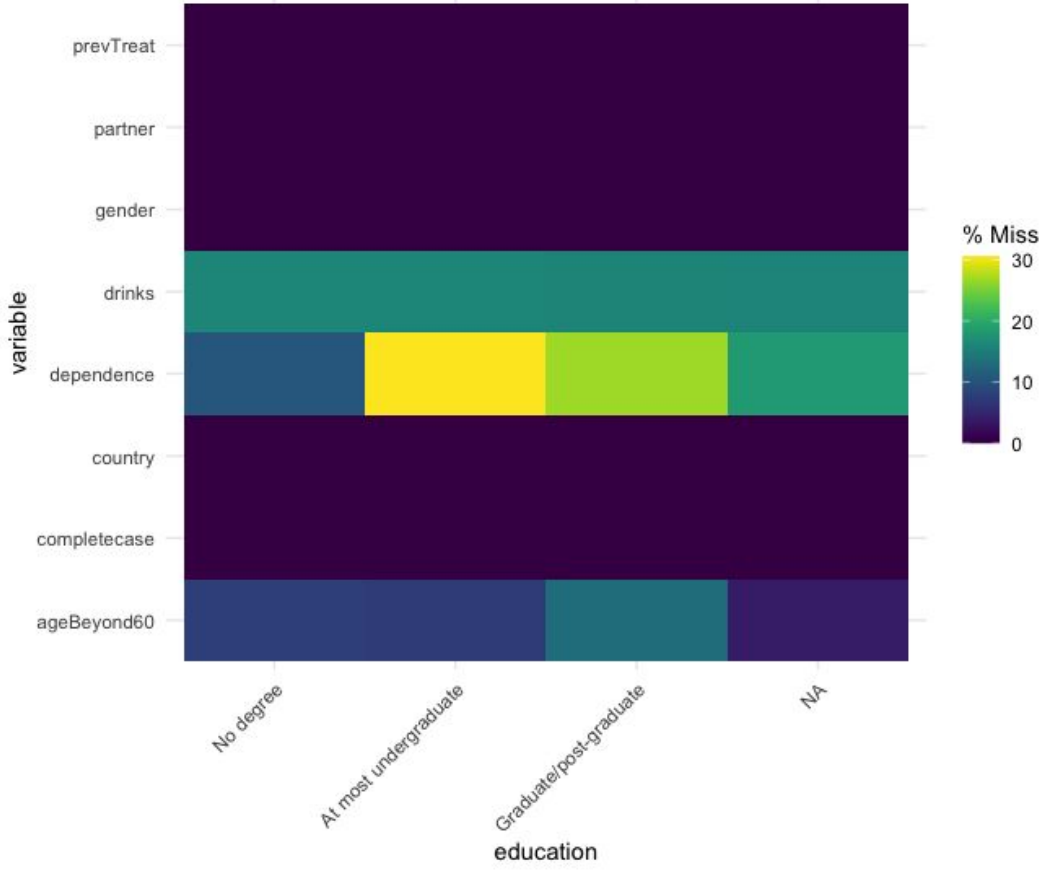


Distributions of variables, stratified by complete case status

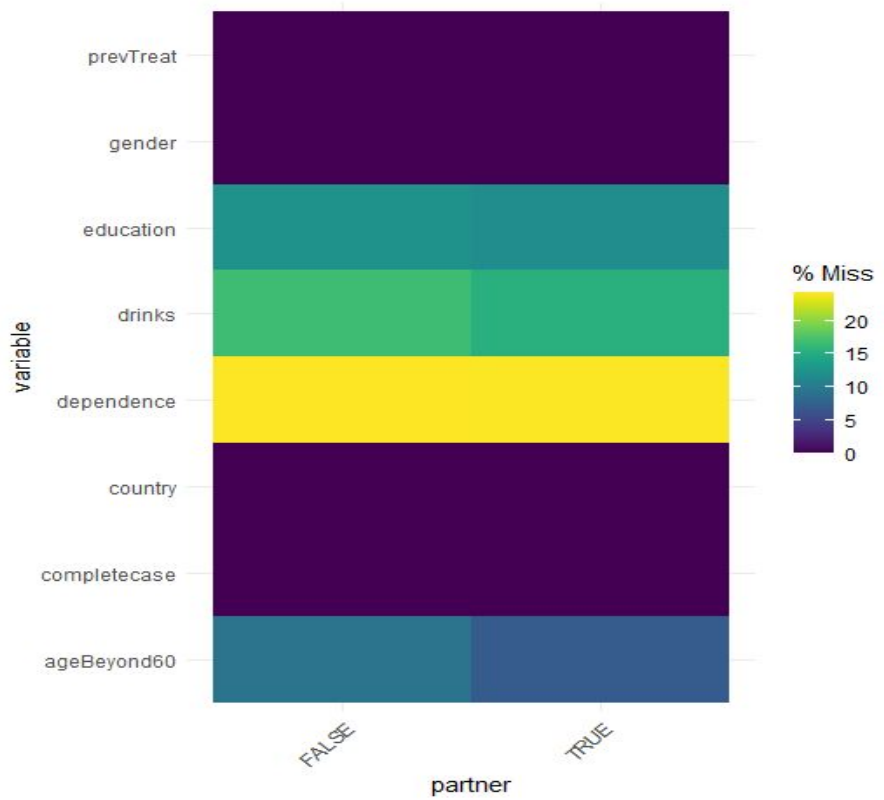




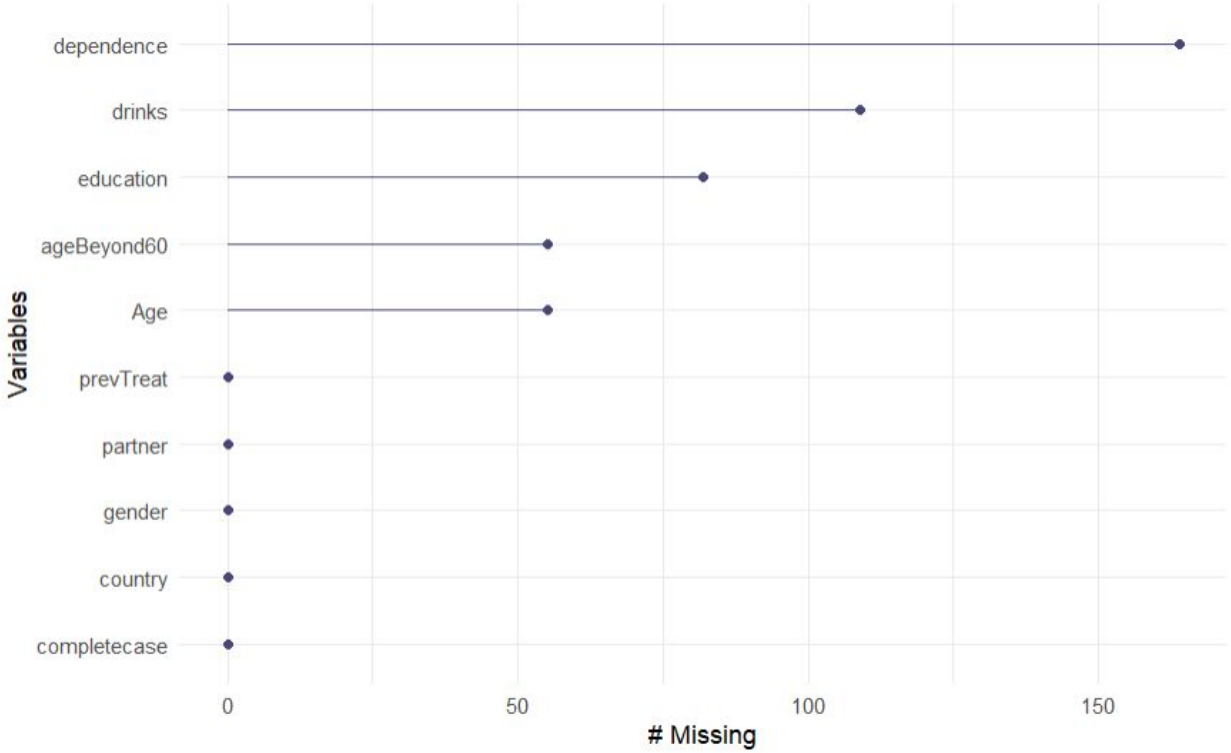
Patterns of missing information

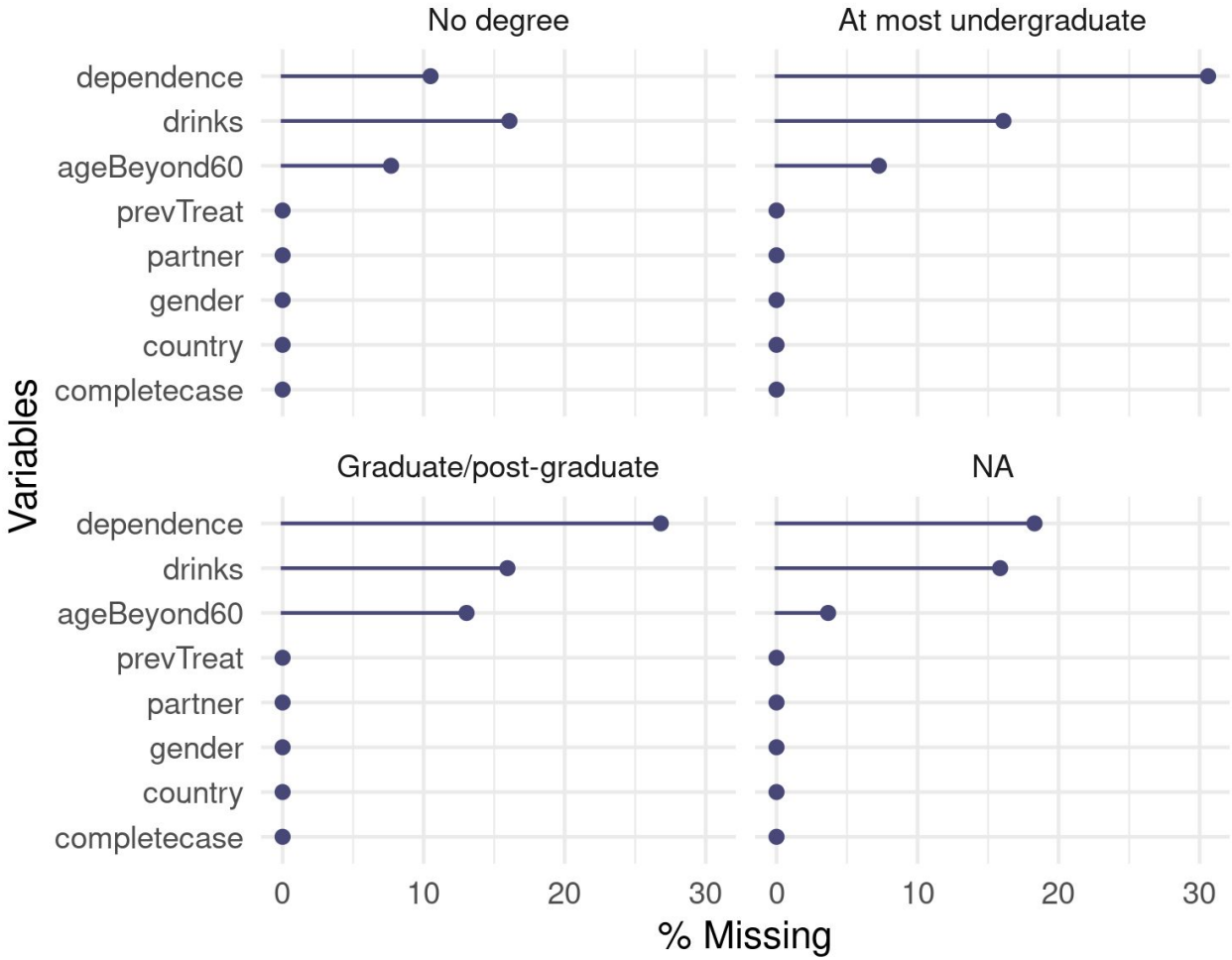


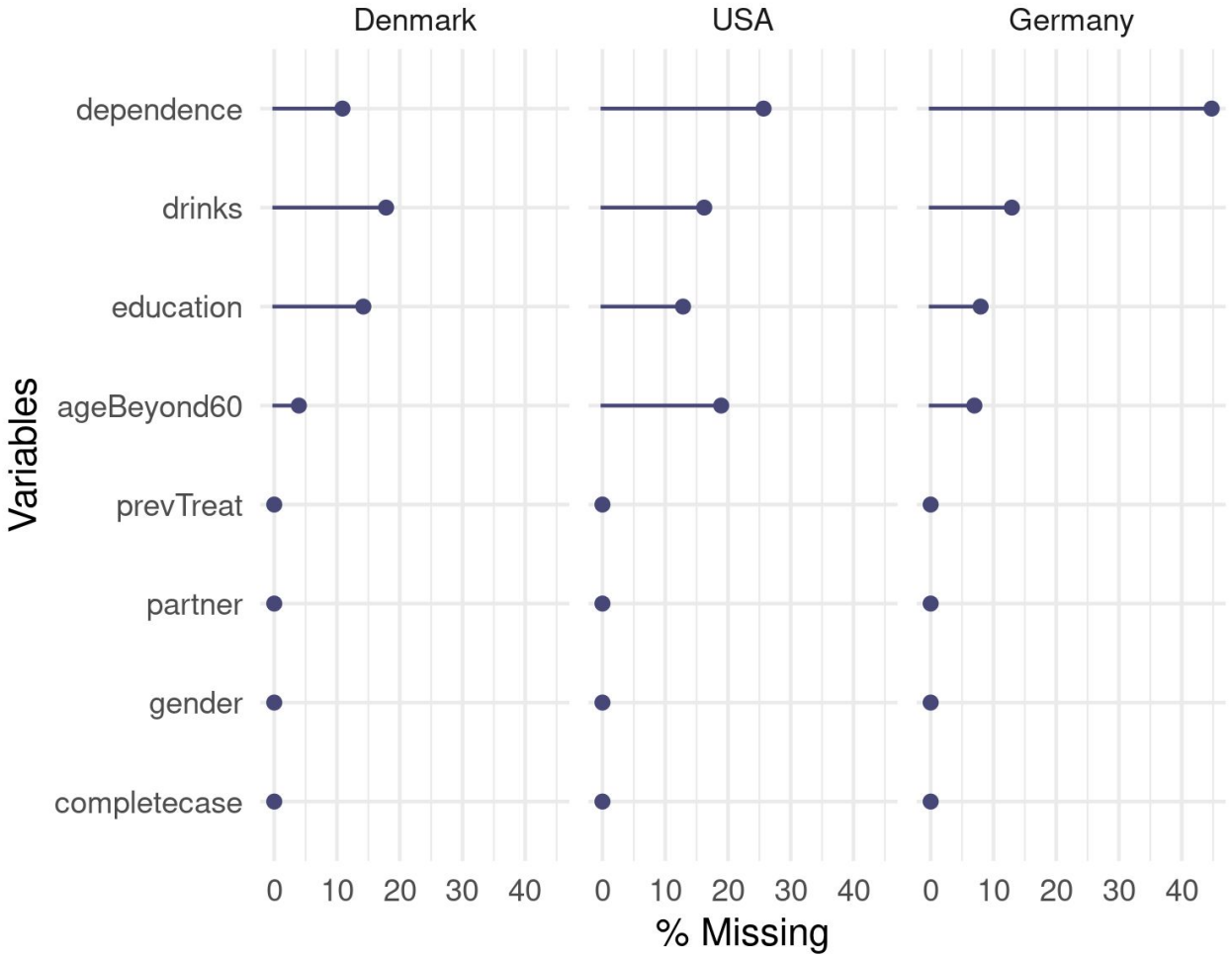
Patterns of missing information



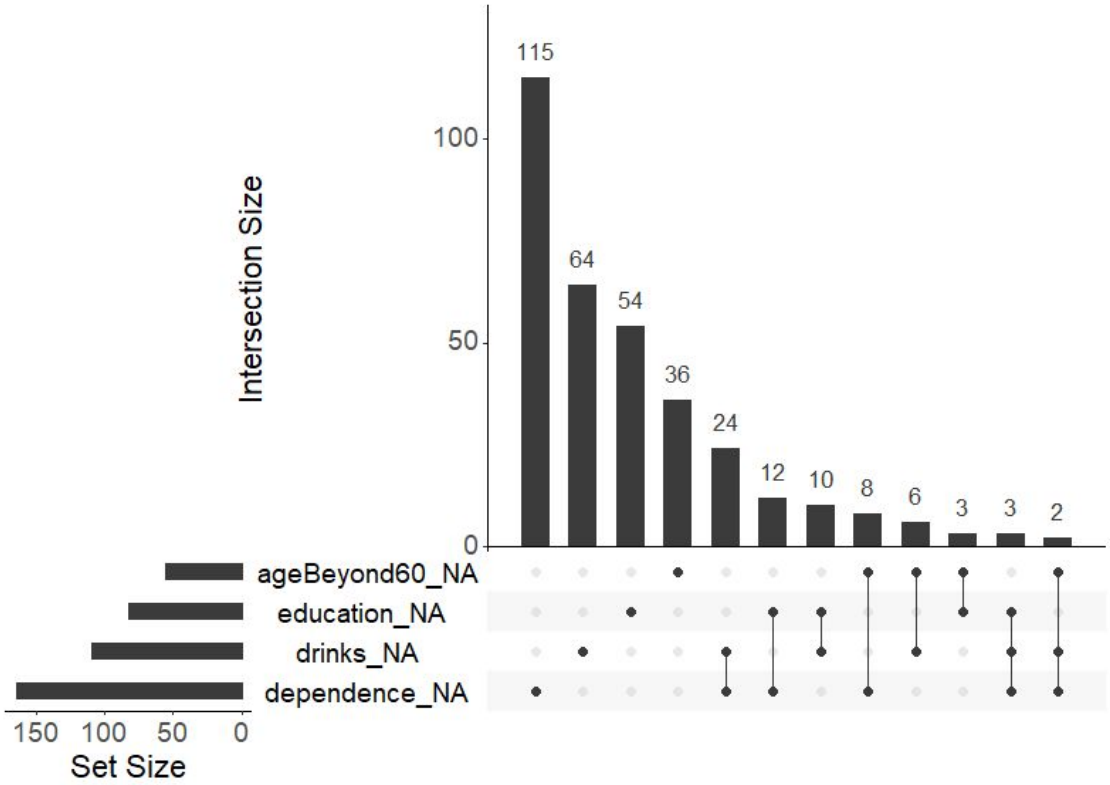
Patterns of missing information



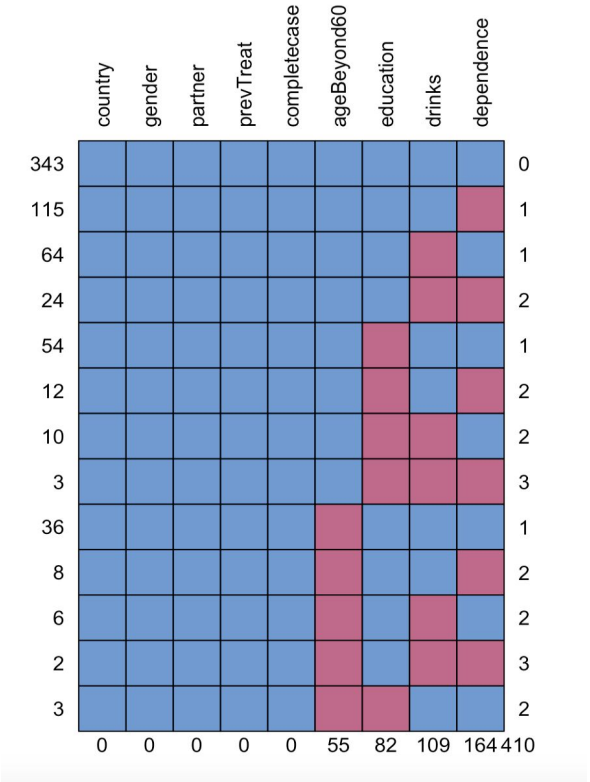

No degree
At most undergraduate
Graduate/post-graduate
NA
dependence
drinks
ageBeyond60
prevTreat
partner
gender
country
completecase
Variables
0
10
20
30
% Missing

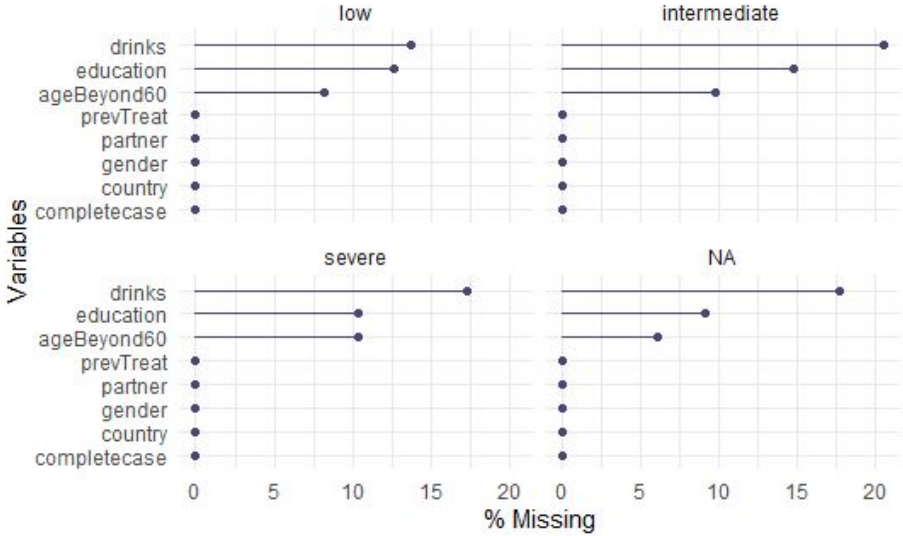

Denmark
USA
Germany
dependence
drinks
education
ageBeyond60
prevTreat
partner
gender
completecase
Variables
0
10
20
30
40
% Missing

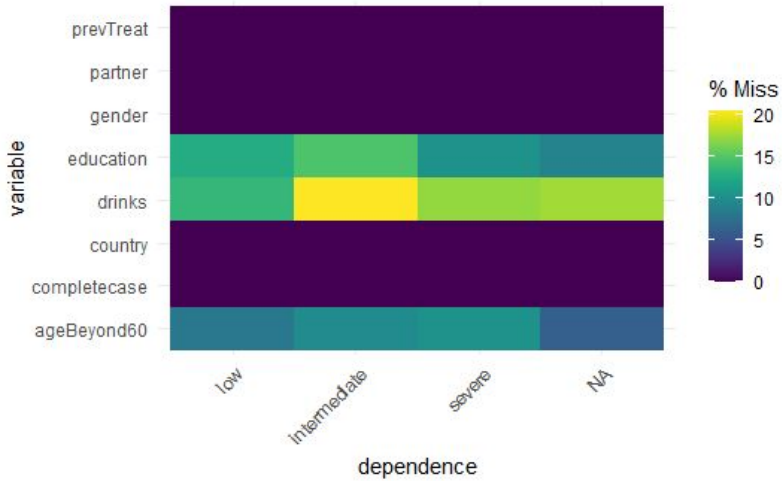
Patterns of missing information

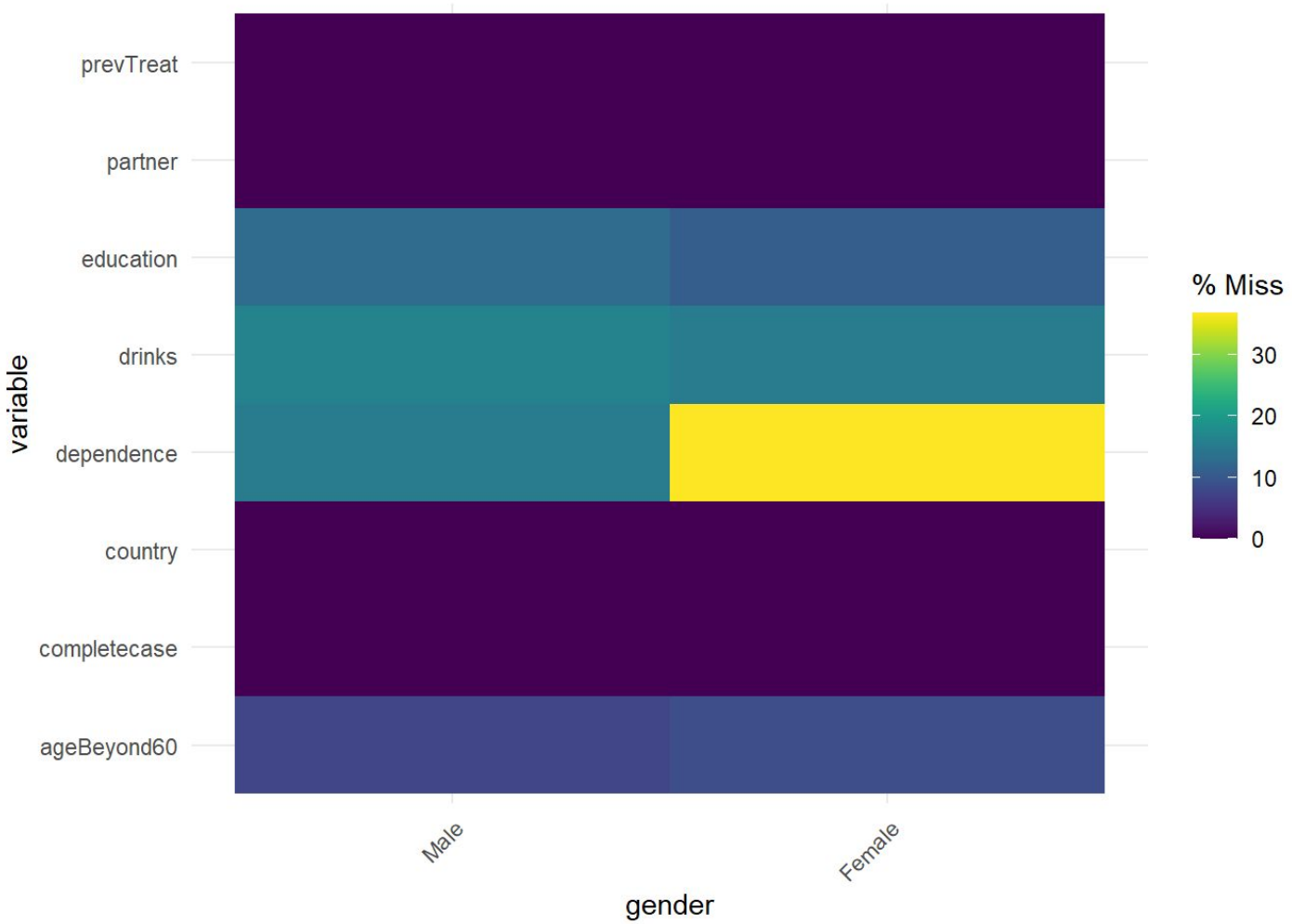


Patterns of missing information




low
intermediate
severe
NA
drinks
education
ageBeyond60
prevTreat
partner
gender
country
completecase
Variables
0
5
10
15
20
% Missing


prevTreat
partner
gender
education
drinks
country
completecase
ageBeyond60
variable
low
intermediate
severe
NA
dependence
% Miss
20
15
10
5
0


prevTreat
partner
education
drinks
dependence
country
completecase
ageBeyond60
variable
Male
Female
gender
% Miss
30
20
10
0

Overall non-response analysis

```
Call:
glm(formula = completecase == "Non-complete" ~ country + gender +
    partner + prevTreat, family = "binomial", data = alcodata1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7015  -1.1052  -0.8669   1.1250   1.5237

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.72058    0.19747  -3.649 0.000263 ***
countryUSA      0.80893    0.20991   3.854 0.000116 ***
countryGermany  0.82387    0.20081   4.103 4.08e-05 ***
genderFemale    0.49174    0.16453   2.989 0.002800 **
partnerTRUE     0.03628    0.16445   0.221 0.825389
prevTreat1-2   -0.06457    0.19246  -0.336 0.737238
prevTreat3+     0.54823    0.24582   2.230 0.025733 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 942.63  on 679  degrees of freedom
Residual deviance: 903.09  on 673  degrees of freedom
AIC: 917.09

Number of Fisher Scoring iterations: 4
```

Overall non-response analysis

```
Single term deletions

Model:
completecase == "Non-complete" ~ country + education + partner +
    prevTreat
           Df Deviance    AIC     LRT  Pr(>Chi)
<none>          770.19 786.19
country     2   801.49 813.49 31.3001 1.597e-07 ***
education   2   772.21 784.21  2.0146   0.36520
partner     1   770.30 784.30  0.1037   0.74743
prevTreat   2   777.16 789.16  6.9653   0.03073 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Overall non-response analysis

```
Single term deletions

Model:
completecase == "Non-complete" ~ country + gender + partner +
    prevTreat
          Df Deviance    AIC     LRT  Pr(>Chi)
<none>         903.09 917.09
country    2   926.45 936.45 23.3602 8.461e-06 ***
gender     1   912.09 924.09  8.9999   0.00270 **
partner    1   903.14 915.14  0.0487   0.82536
prevTreat  2   909.56 919.56  6.4714   0.03933 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable-specific non-response analysis: drinks

```
Single term deletions

Model:
is.na(drinks) ~ country + gender + ageBeyond60 + partner + education +
    dependence + prevTreat
             Df Deviance    AIC     LRT Pr(>Chi)
<none>            336.45 360.45
country       2   336.47 356.47  0.0211 0.989482
gender        1   336.45 358.45  0.0043 0.947890
ageBeyond60   1   336.61 358.61  0.1653 0.684302
partner       1   337.38 359.38  0.9295 0.334988
education     2   336.52 356.52  0.0777 0.961911
dependence    2   340.91 360.91  4.4608 0.107488
prevTreat     2   347.15 367.15 10.7039 0.004739 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable-specific non-response analysis: age

```
Single term deletions

Model:
is.na(ageBeyond60) ~ country + gender + drinks + partner + education +
    dependence + prevTreat
           Df Deviance    AIC     LRT Pr(>Chi)
<none>          212.79 236.79
country     2   226.36 246.36 13.5682 0.001132 **
gender      1   213.30 235.30  0.5135 0.473618
drinks      1   214.20 236.20  1.4051 0.235865
partner     1   212.81 234.81  0.0240 0.876826
education   2   214.82 234.82  2.0301 0.362386
dependence  2   215.57 235.57  2.7775 0.249392
prevTreat   2   214.44 234.44  1.6441 0.439534
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable-specific non-response analysis: dependence

```
Single term deletions

Model:
is.na(dependence) ~ country + gender + drinks + partner + prevTreat
          Df Deviance    AIC    LRT  Pr(>Chi)
<none>         508.65 524.65
country    2   583.55 595.55 74.904 < 2.2e-16 ***
gender     1   534.73 548.73 26.082 3.271e-07 ***
drinks     1   508.75 522.75  0.109    0.7417
partner    1   508.65 522.65  0.001    0.9792
prevTreat  2   510.99 522.99  2.346    0.3094
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable-specific non-response analysis: education

```
Single term deletions

Model:
is.na(education) ~ country + gender + partner + prevTreat
          Df Deviance    AIC    LRT Pr(>Chi)
<none>         492.94 506.94
country    2   499.01 509.01 6.0688   0.0481 *
gender     1   493.32 505.32 0.3764   0.5396
partner    1   492.95 504.95 0.0081   0.9284
prevTreat  2   495.28 505.28 2.3320   0.3116
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Other things

Cat for happy smiles

Conclusions so far

- Best guesses for type of missing data mechanism (MCAR/MAR/MNAR):
  - Age: MCAR
  - Dependence: MAR –
  - Education: MCAR ()
  - Drinks: MCAR
-