

MCAR 45% missing - RandomForest

```
# sample MCAR dataset from PUMS
source("../utils/sampleMCAR.R")
n = 10000
missing_col = c(1,3,7,9,10,11)
missing_prob = 0.45
set.seed(1)

output_list <- sampleMCAR(n, missing_prob)
df <- output_list[['df']]
df_observed <- output_list[['df_observed']]
```

missForest

```
df.imp <- missForest(df_observed, verbose = FALSE)
d1 <- df.imp$xim
df.imp <- missForest(df_observed, verbose = FALSE)
d2 <- df.imp$xim
df.imp <- missForest(df_observed, verbose = FALSE)
d3 <- df.imp$xim
df.imp <- missForest(df_observed, verbose = FALSE)
d4 <- df.imp$xim
df.imp <- missForest(df_observed, verbose = FALSE)
d5 <- df.imp$xim
imputed_sets = rbind(d1, d2, d3, d4, d5)
```

Diagnostics

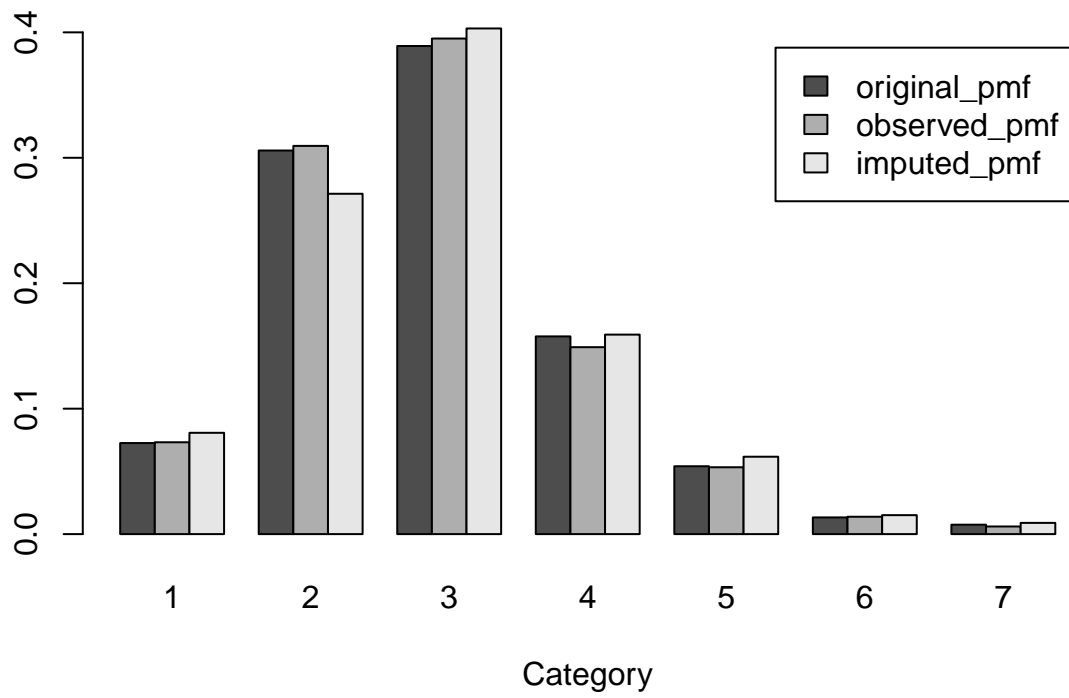
Assess bivariate joint distribution

Assess trivariate joint distribution

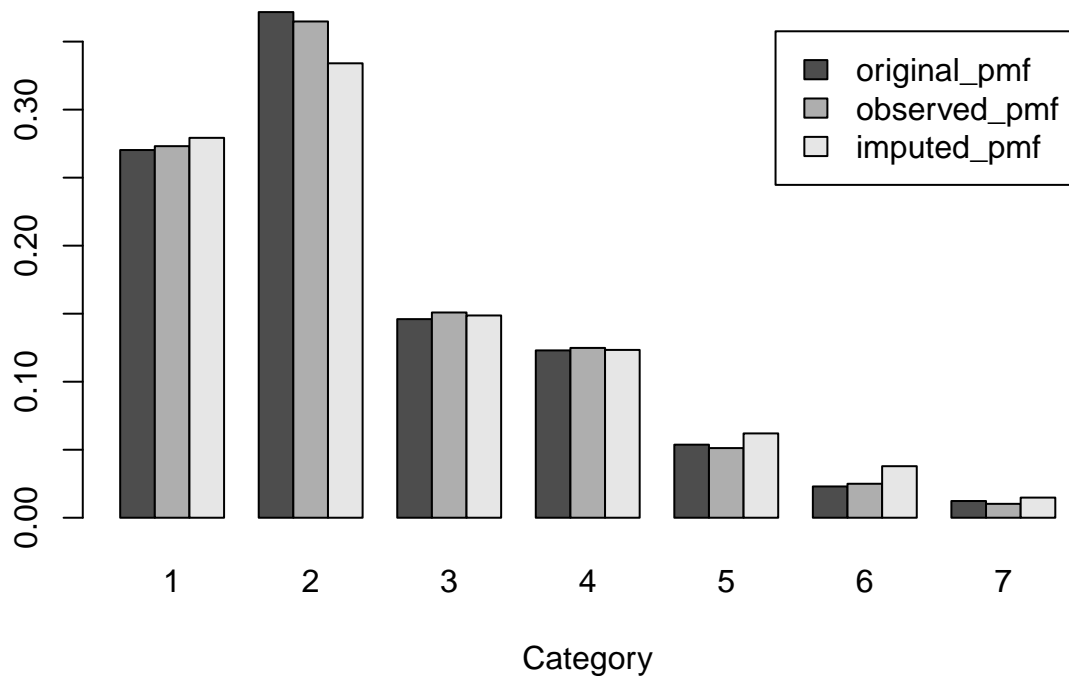
```
## [1] "rmse"
```

```
## [1] 0.2837011
```

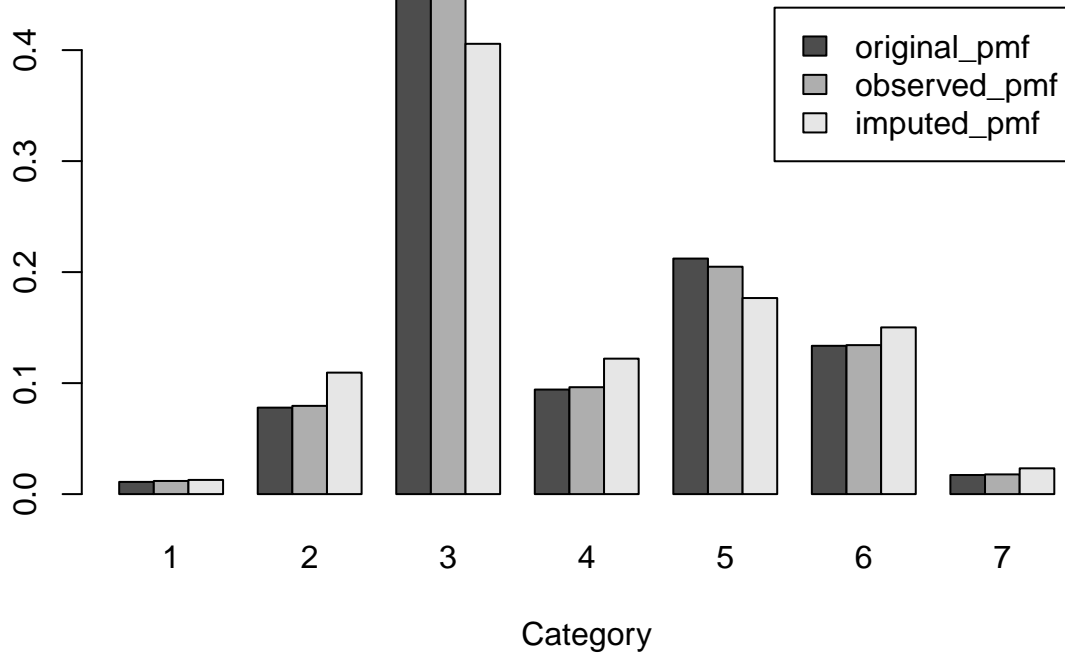
MICE: VEH



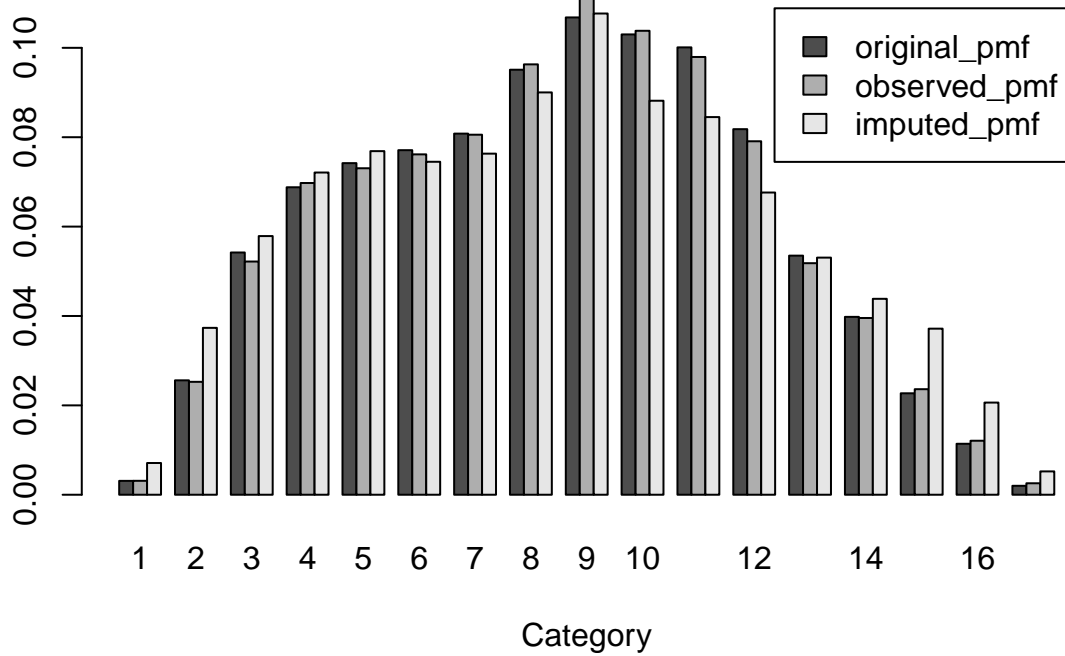
MICE: NP



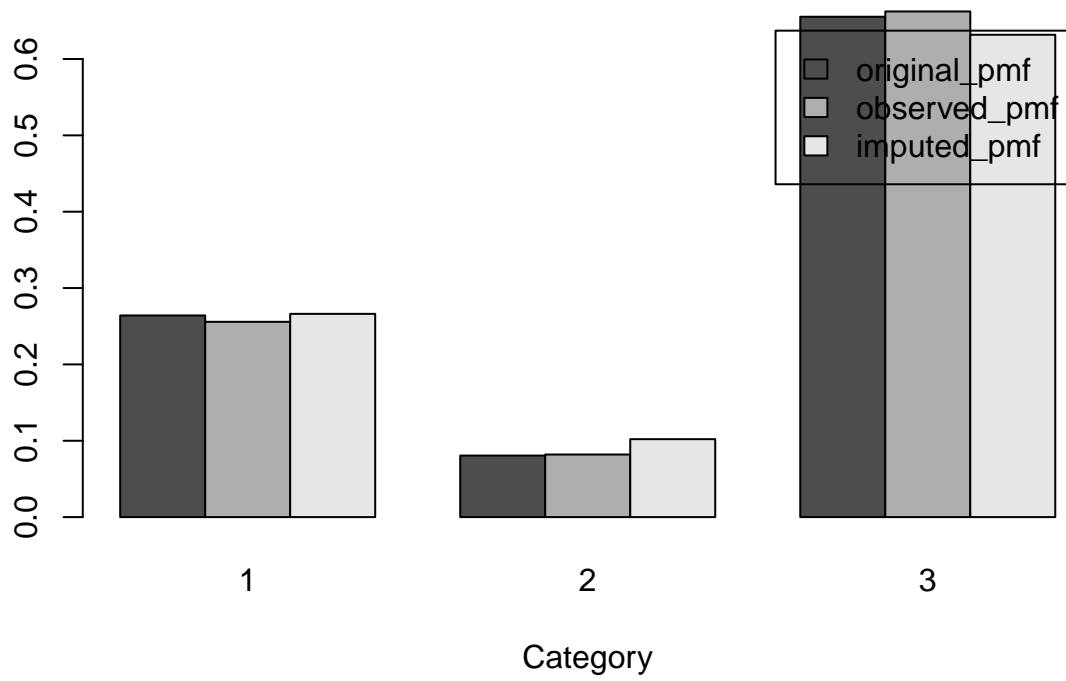
MICE: SCHL



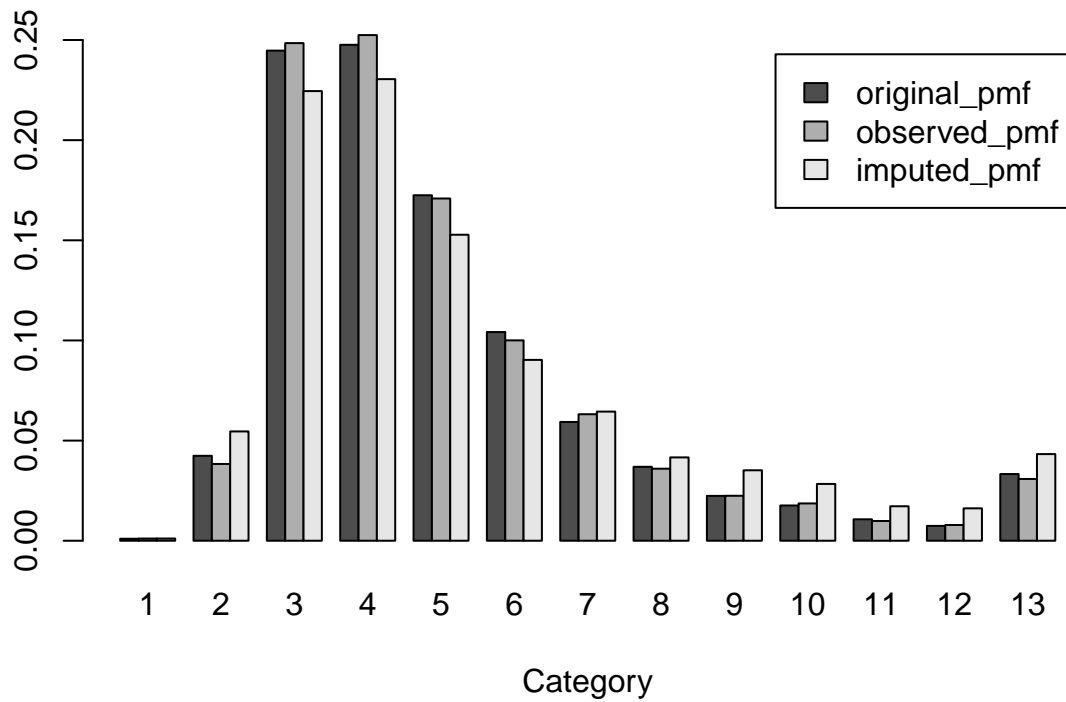
MICE: AGEP



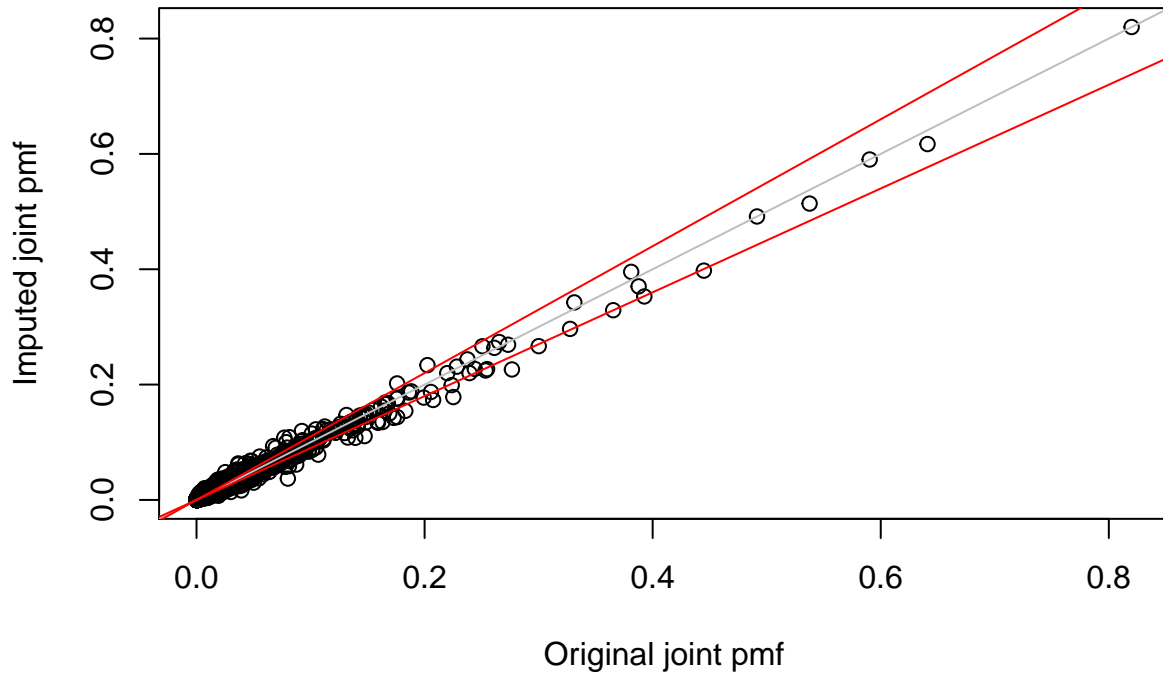
MICE: WKL



MICE: PINCP



Bivariate pmf



Trivariate pmf

