

# MAR 30% missing - MICE

```
# sample MCAR dataset from PUMS
source("../utils/sampleMAR30.R")
n = 10000
missing_col = c(1,3,7,9,10,11)
set.seed(2)

output_list <- sampleMAR30(n)
df <- output_list[['df']]
df_observed <- output_list[['df_observed']]

apply(is.na(df_observed), MARGIN = 2, mean)

##      VEH      MV      NP    RMSP      ENG    MARHT    SCHL RACNUM     AGEP      WKL    PINCP
## 0.3074 0.0000 0.2605 0.0000 0.0000 0.0000 0.3424 0.0000 0.3227 0.3078 0.3049
```

## MICE

Create 5 imputed dataset

```
library(mice)

##
## Attaching package: 'mice'
## The following objects are masked from 'package:base':
##       cbind, rbind
imputed_df <- mice(df_observed,m=5,print=F)
```

## Warning: Number of logged events: 50

Extract the 5 imputed dataset

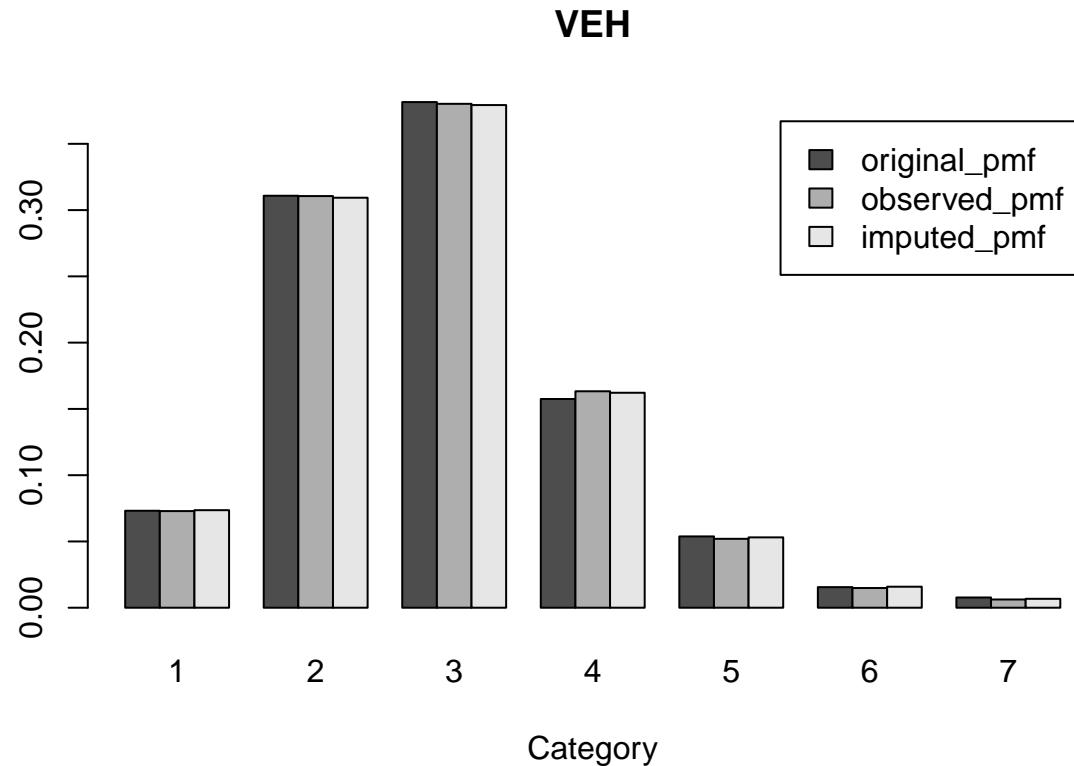
```
imputation_list = list()
for (i in 1:5) {
  imputation_list[[i]] = complete(imputed_df, i)
}
```

Diagnostics

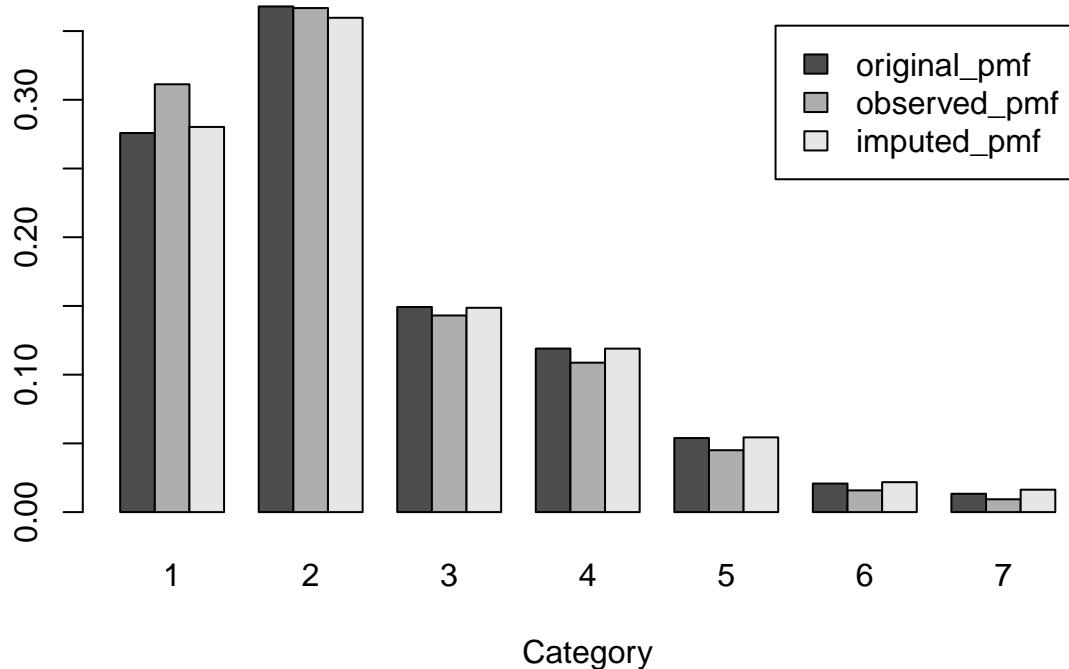
```
source("../utils/create_report.R")
create_report(imputation_list, max_nway=4, missing_col, df_observed)

## ##### Coverage #####
## Coverage 1 way: 92.59 percent
## Coverage 2 way: 93.07 percent
## Coverage 3 way: 96.34 percent
## Coverage 4 way: 98.18 percent
##
## ##### RMSE #####
## RMSE 1 way: 0.003019
## RMSE 2 way: 0.00166
## RMSE 3 way: 0.000626
```

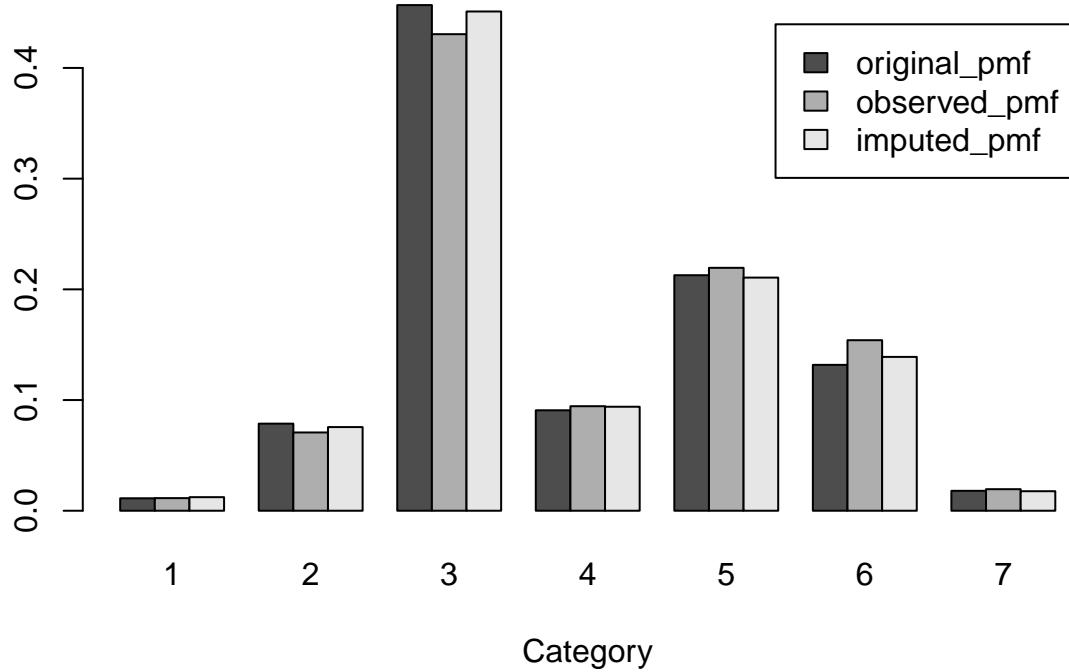
```
## RMSE 4 way: 0.000215
##
## ##### MAE #####
## MAE 1 way: 0.002201
## MAE 2 way: 0.00088
## MAE 3 way: 0.000238
## MAE 4 way: 6e-05
```



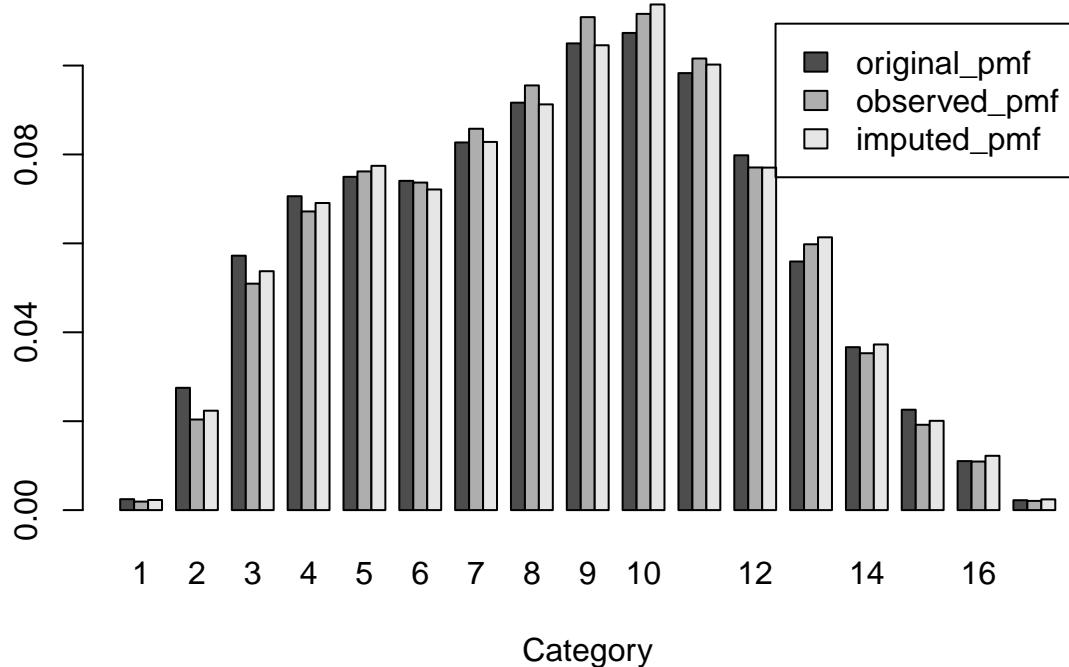
## NP



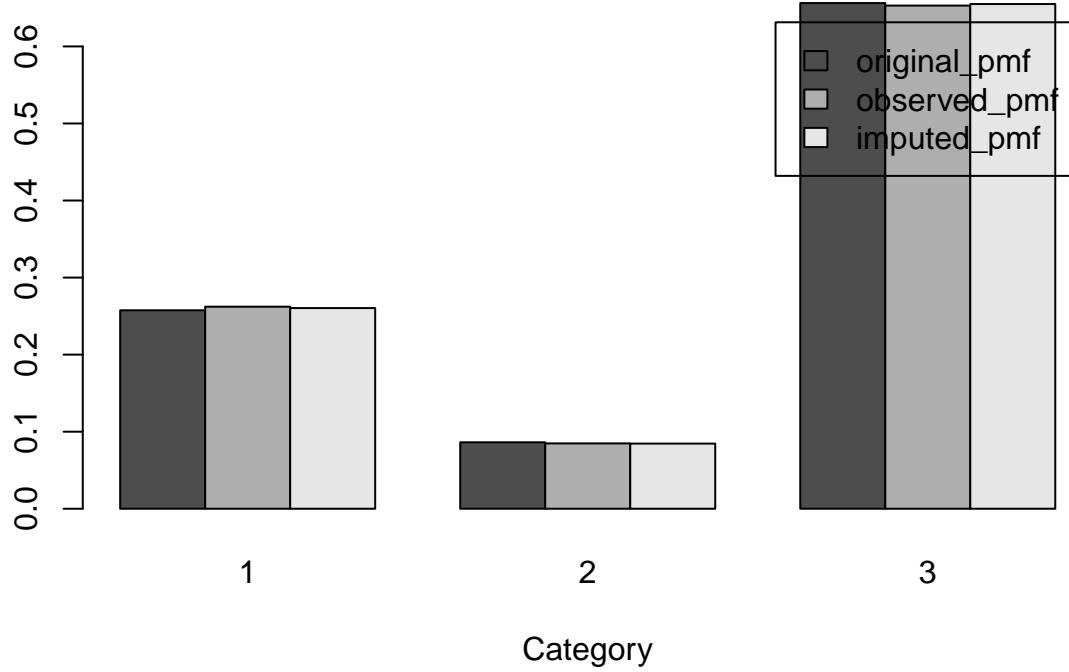
## SCHL



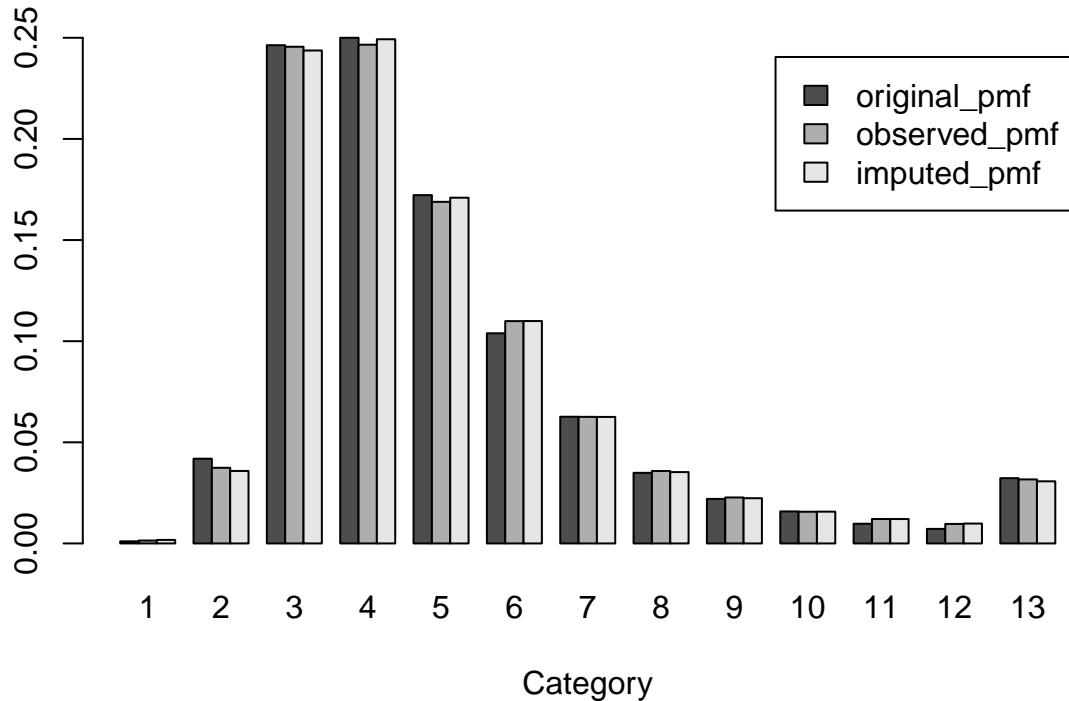
## AGEP



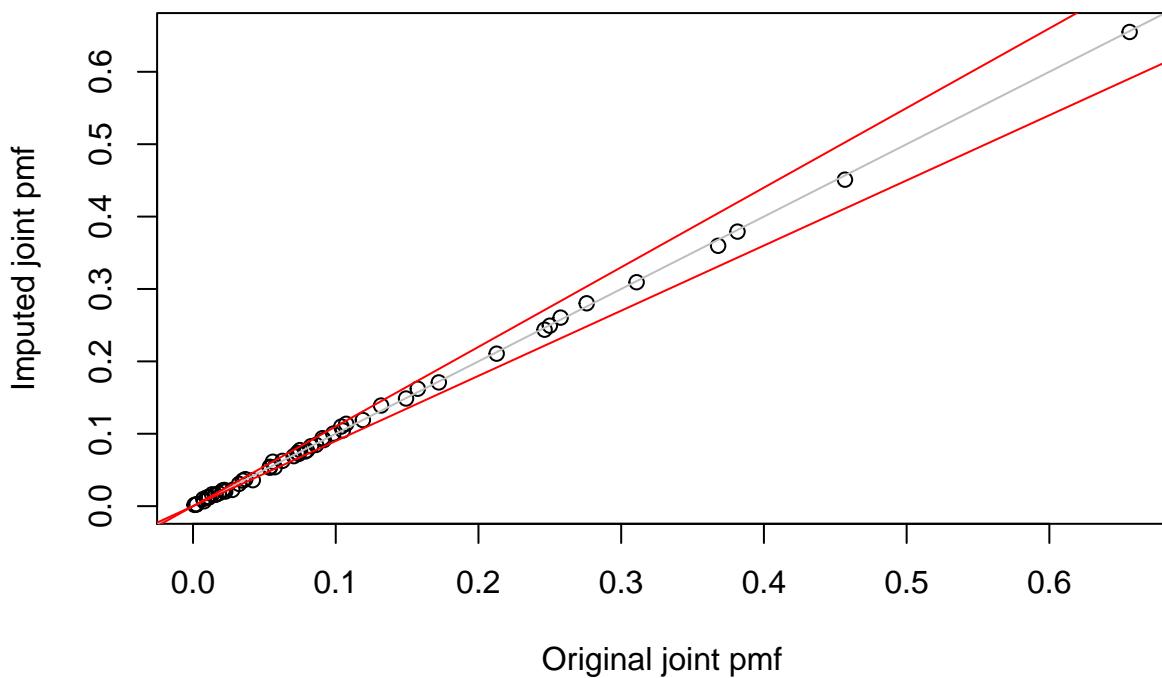
## WKL



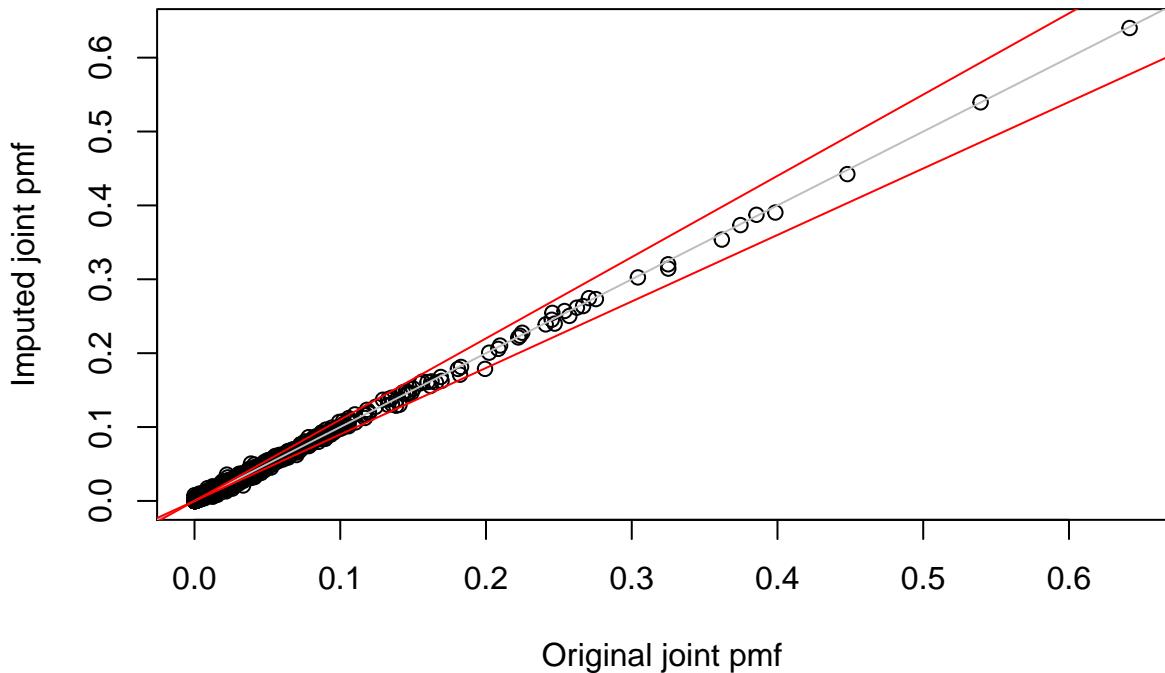
## PINCP



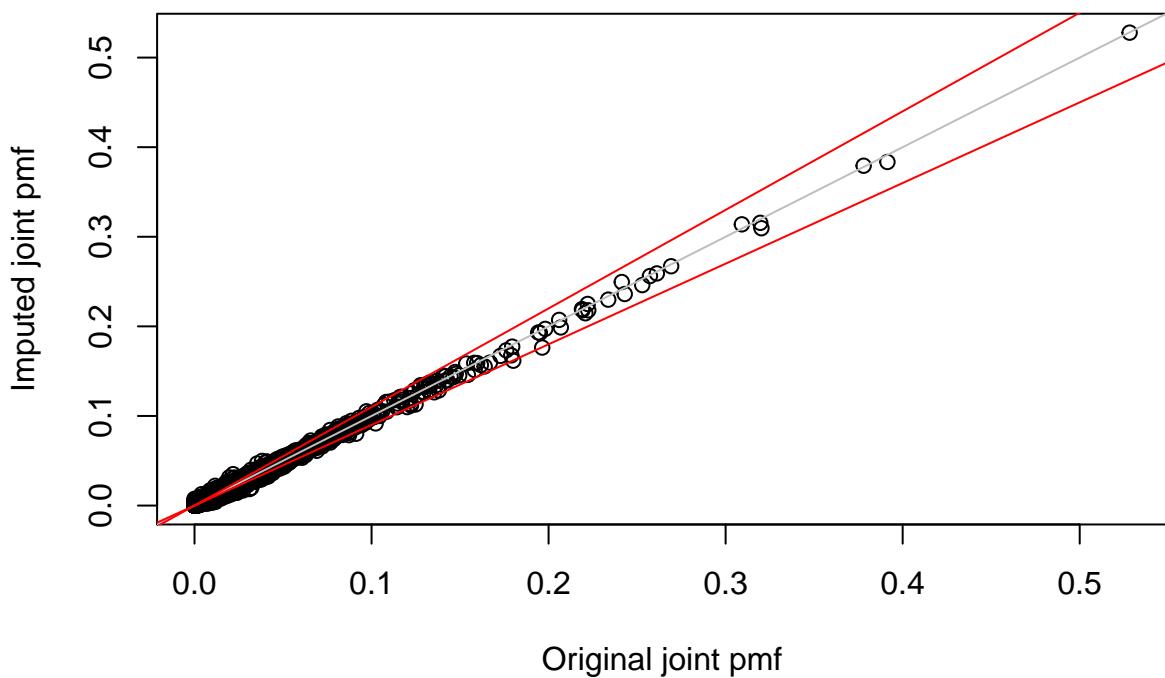
## Assess imputed pmf: 1 way



**Assess imputed pmf: 2 way**



**Assess imputed pmf: 3 way**



### Assess imputed pmf: 4 way

