

MAR 30% missing - Generative Adversarial Imputation Nets (GAIN)

```
# 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
```

Generative Adversarial Imputation Nets (GAIN)

reference: <https://arxiv.org/abs/1806.02920>

```
# Load imputed dataset
d1 = read.csv('../GAIN/imputed_dataset/MAR_30percent_1.csv', header = FALSE, sep = ',')
d2 = read.csv('../GAIN/imputed_dataset/MAR_30percent_2.csv', header = FALSE, sep = ',')
d3 = read.csv('../GAIN/imputed_dataset/MAR_30percent_3.csv', header = FALSE, sep = ',')
d4 = read.csv('../GAIN/imputed_dataset/MAR_30percent_4.csv', header = FALSE, sep = ',')
d5 = read.csv('../GAIN/imputed_dataset/MAR_30percent_5.csv', header = FALSE, sep = ',')

# Format imputed dataset into list
imputation_df = list(d1, d2, d3, d4, d5)
imputation_list = list()
levels = c(7,7,7,19,5,4,7,2,17,3,13)
for (i in 1:5) {
  # need to plus 1 here because the class index of DP function starts at 0
  d = imputation_df[[i]]
  colnames(d) = colnames(df_observed)
  # format columns of d
  for (col_index in 1:ncol(df_observed)) {
    d[,col_index] = factor(d[,col_index], levels = 1:levels[col_index], ordered = TRUE)
  }
  imputation_list[[i]] = d
}
```

Diagnostics

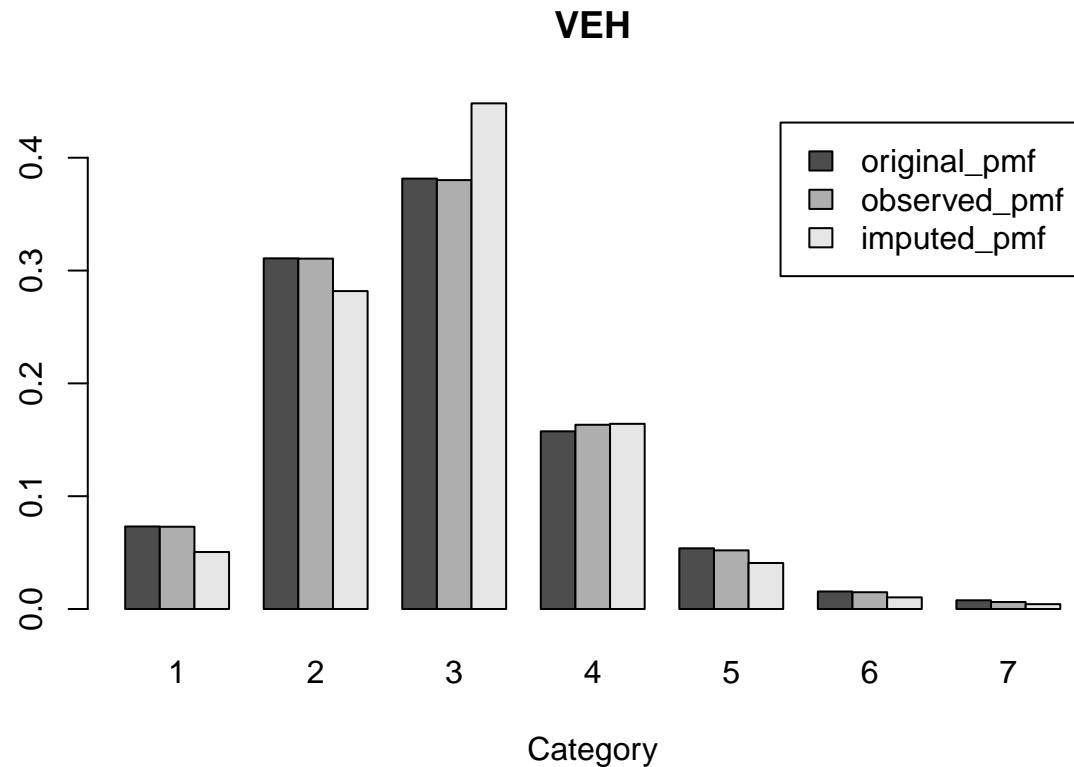
```
source("../utils/create_report.R")
create_report(imputation_list, max_nway=4, missing_col, df_observed)

## ##### Coverage #####
## Coverage 1 way: 12.96 percent
## Coverage 2 way: 43.46 percent
## Coverage 3 way: 64.67 percent
```

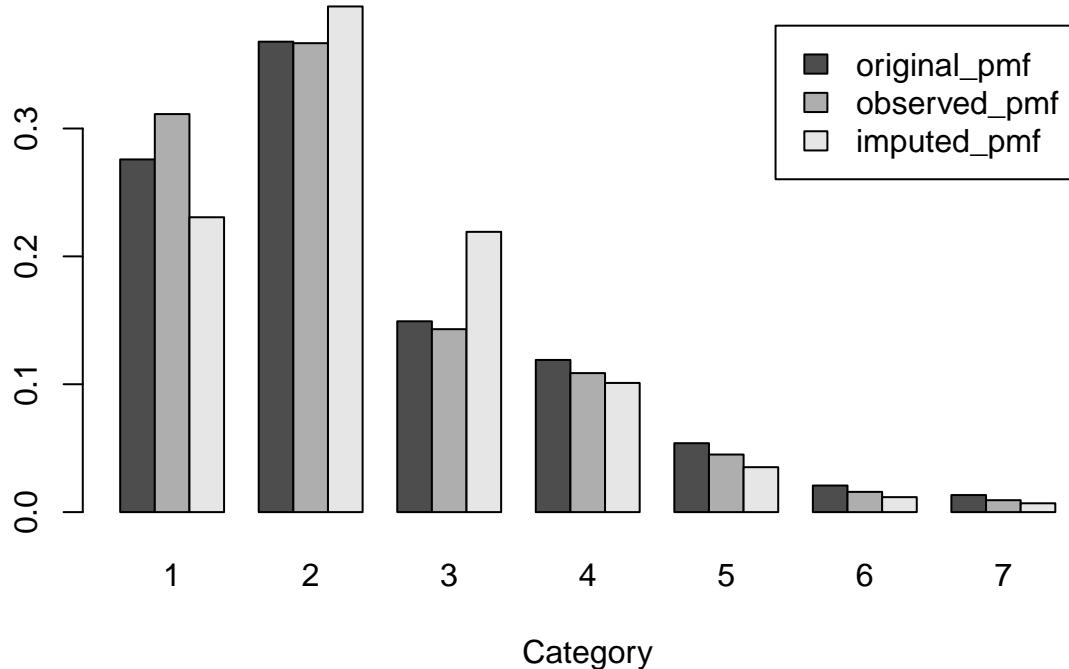
```

## Coverage 4 way: 77.52 percent
##
## ##### RMSE #####
## RMSE 1 way: 0.062257
## RMSE 2 way: 0.015964
## RMSE 3 way: 0.004101
## RMSE 4 way: 0.001048
##
## ##### MAE #####
## MAE 1 way: 0.033764
## MAE 2 way: 0.005184
## MAE 3 way: 0.000875
## MAE 4 way: 0.000158

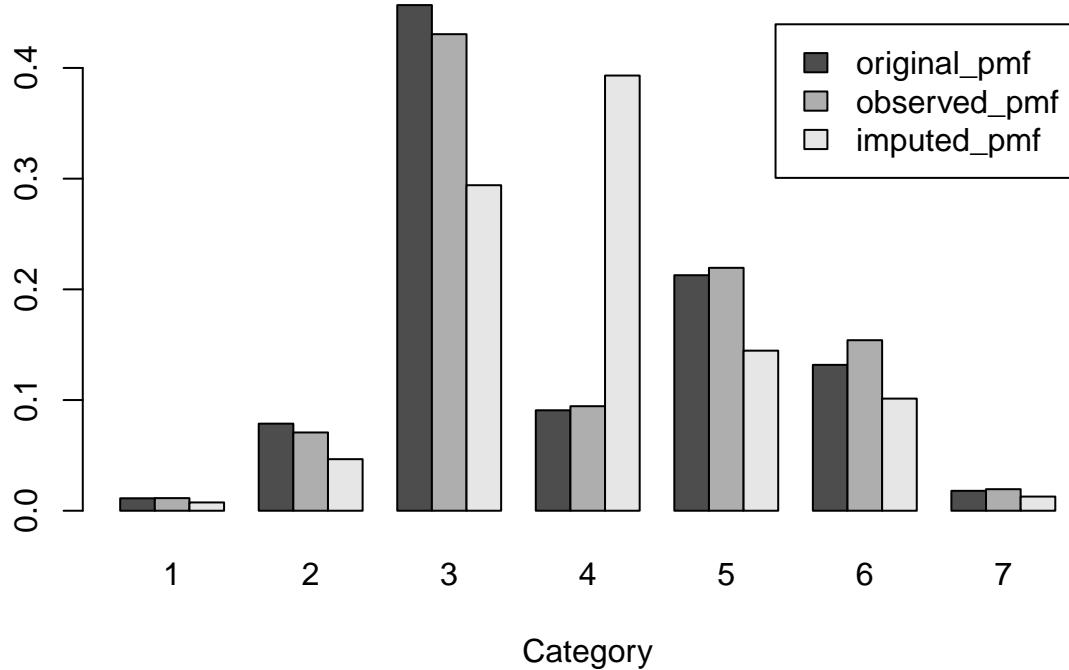
```



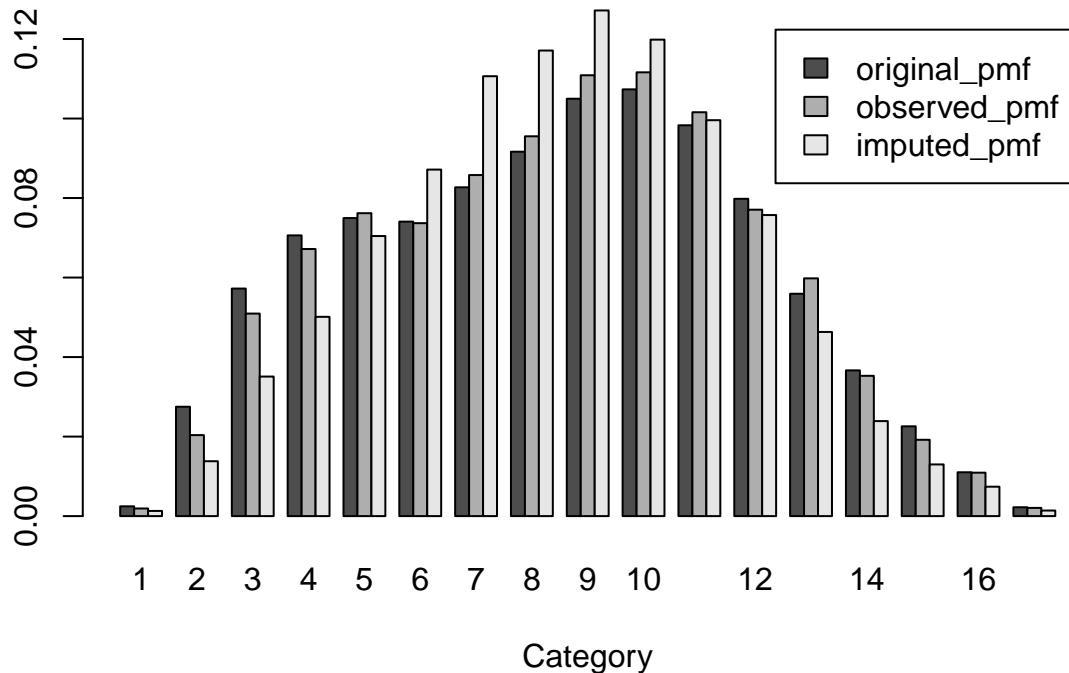
NP



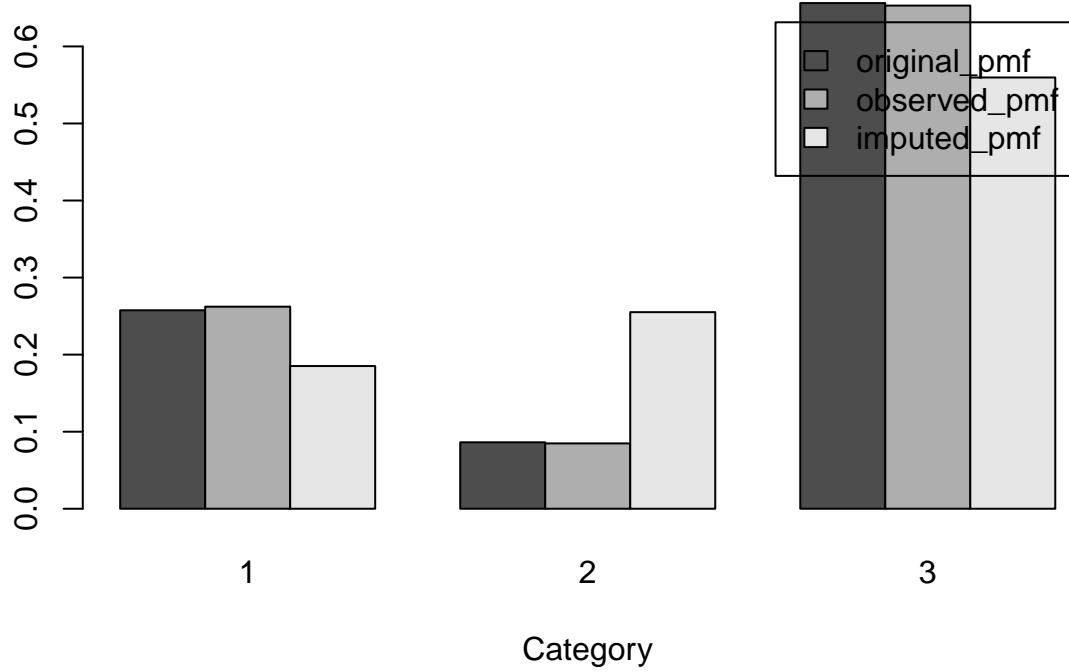
SCHL



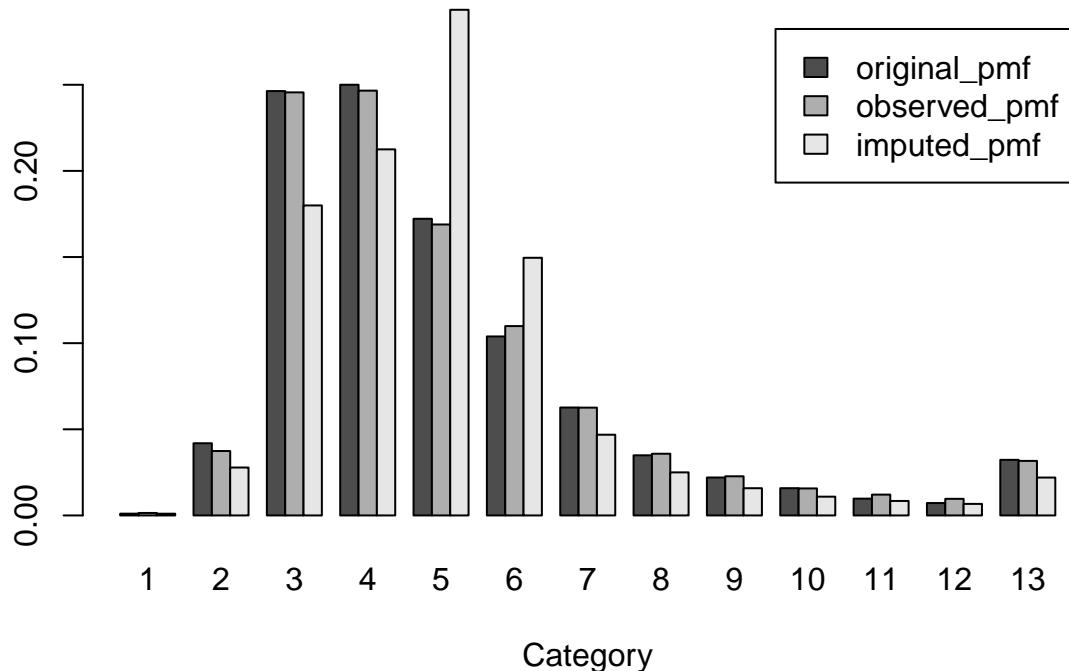
AGEP



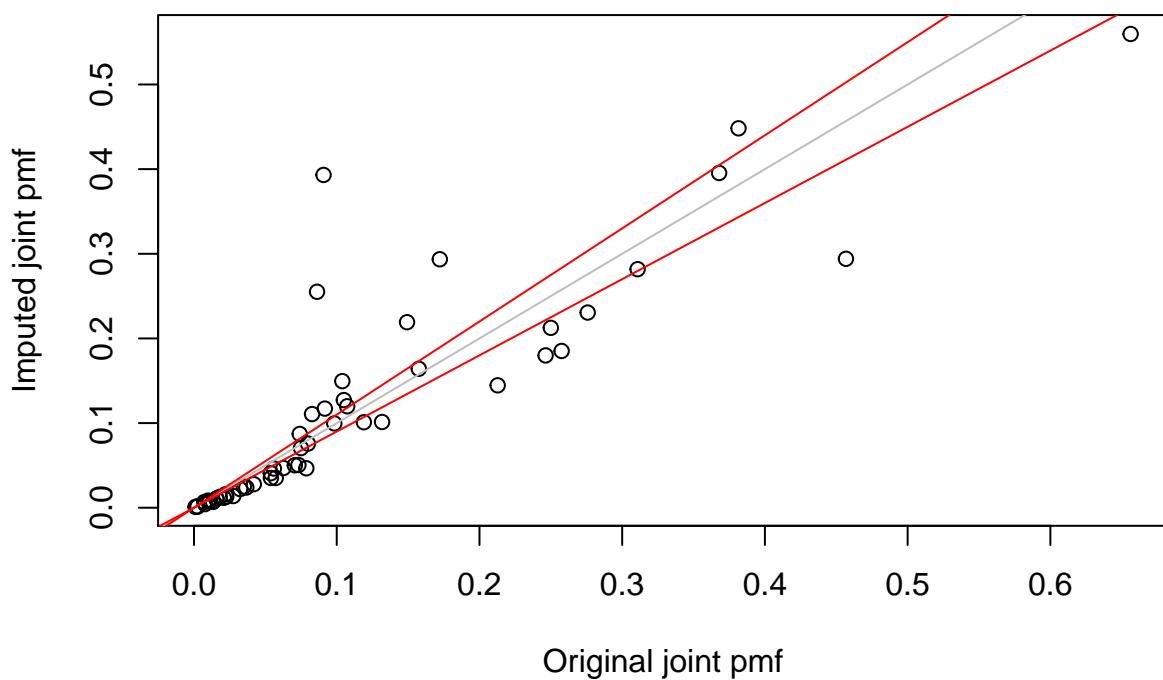
WKL



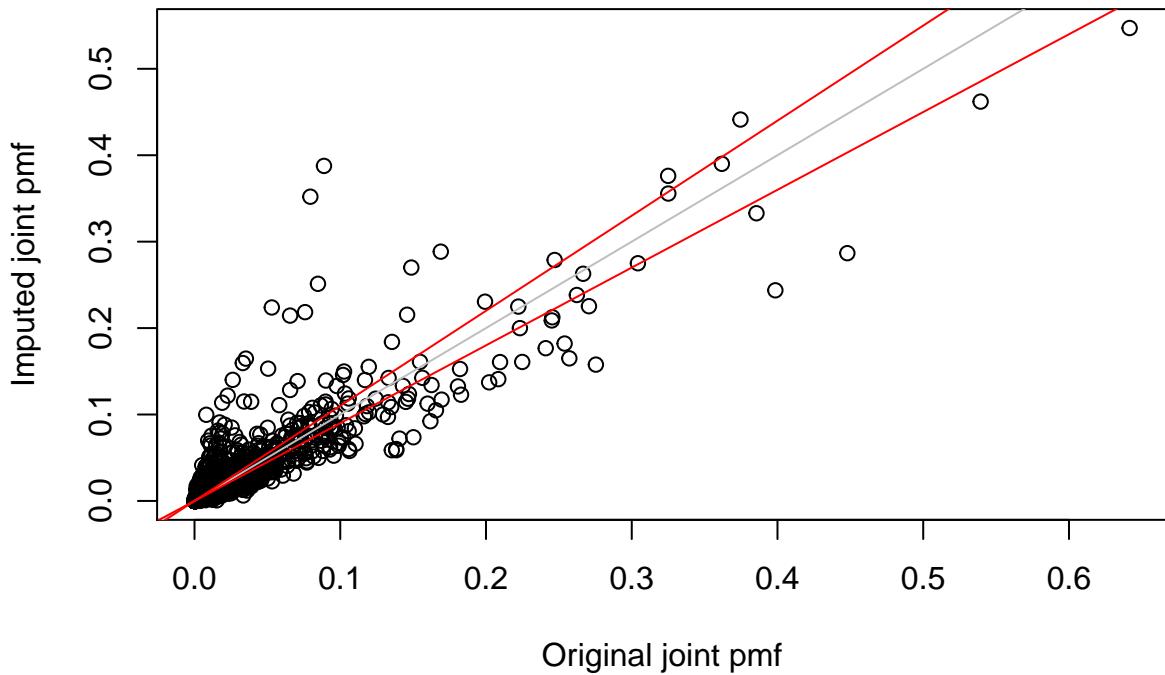
PINCP



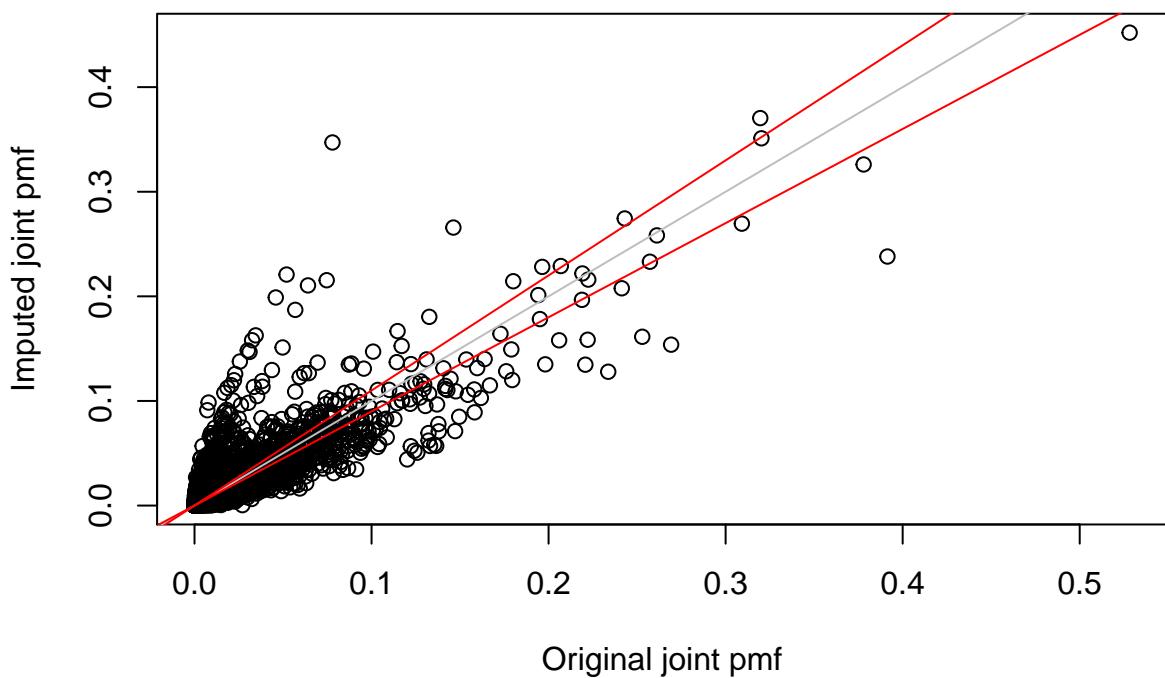
Assess imputed pmf: 1 way



Assess imputed pmf: 2 way



Assess imputed pmf: 3 way



Assess imputed pmf: 4 way

