

# MCAR 45% missing - Generative Adversarial Imputation Nets (GAIN)

```
# 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']]
```

## Generative Adversarial Imputation Nets (GAIN)

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

```
# Load imputed dataset
d1 = read.csv('../GAIN/imputed_dataset/MCAR_45percent_1.csv', header = FALSE, sep = ',')
d2 = read.csv('../GAIN/imputed_dataset/MCAR_45percent_2.csv', header = FALSE, sep = ',')
d3 = read.csv('../GAIN/imputed_dataset/MCAR_45percent_3.csv', header = FALSE, sep = ',')
d4 = read.csv('../GAIN/imputed_dataset/MCAR_45percent_4.csv', header = FALSE, sep = ',')
d5 = read.csv('../GAIN/imputed_dataset/MCAR_45percent_5.csv', header = FALSE, sep = ',')

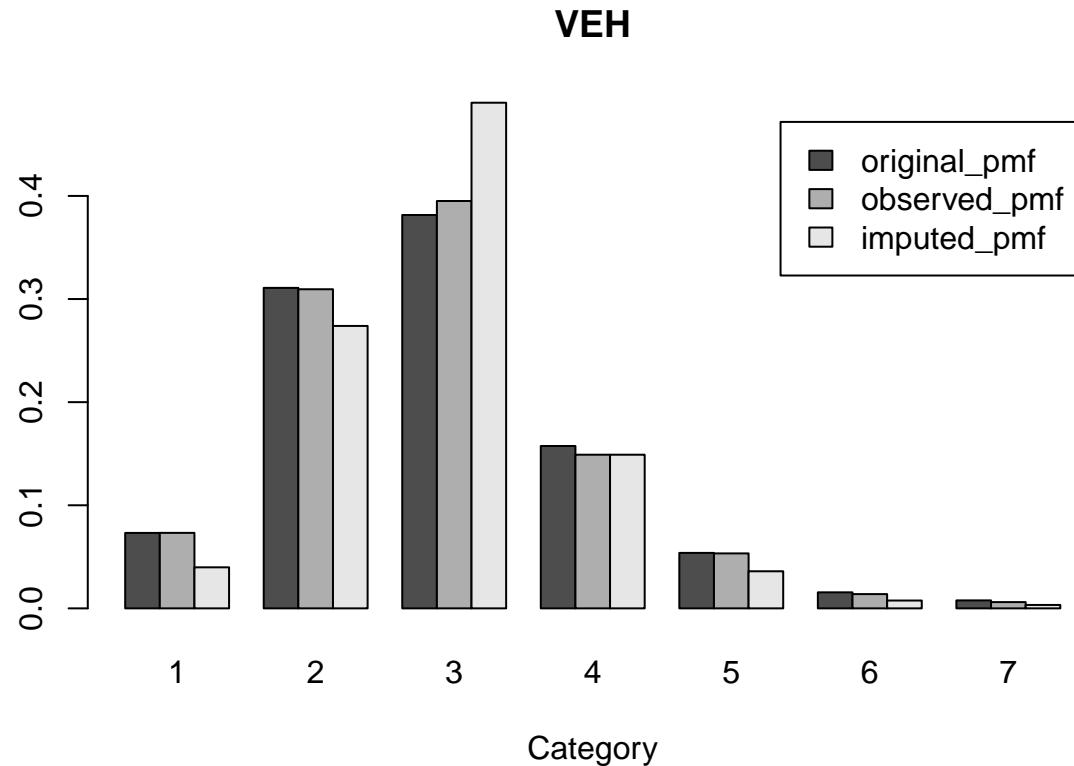
# Format imputed dataset into list
imputation_df = list(d1, d2, d3, d4, d5)
imputation_list = list()
levels = c(7,7,7,19,5,4,7,2,17,3,13)
for (i in 1:5) {
  # need to plus 1 here because the class index of DP function starts at 0
  d = imputation_df[[i]]
  colnames(d) = colnames(df_observed)
  # format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

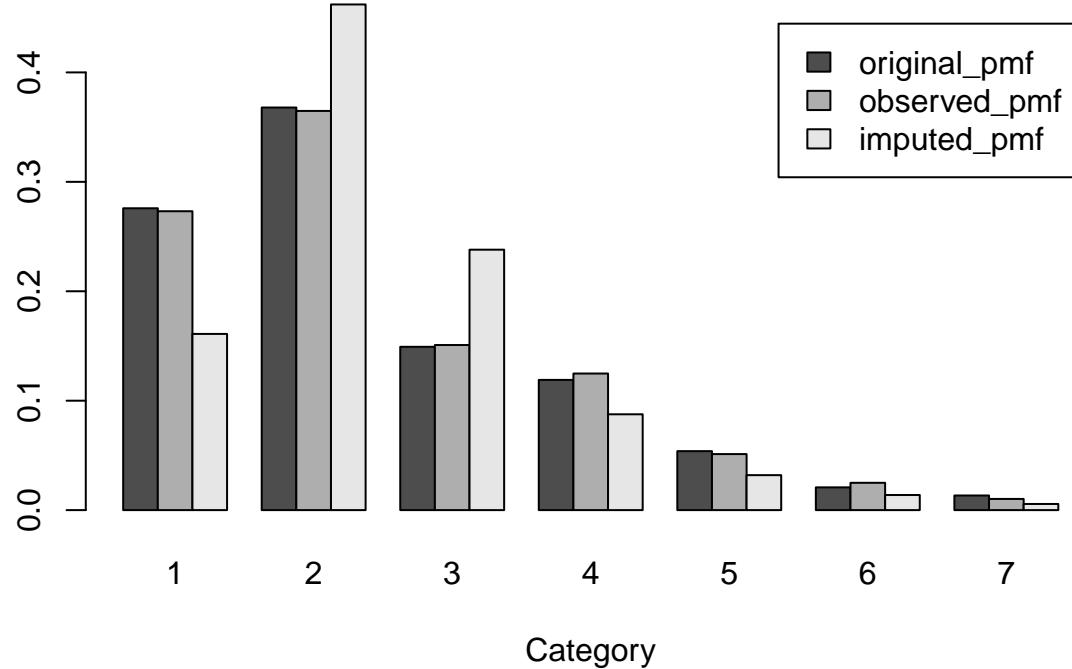
```
source("../utils/create_report.R")
create_report(imputation_list, max_nway=4, missing_col, df_observed)

## ##### Coverage #####
## Coverage 1 way: 11.11 percent
## Coverage 2 way: 37.32 percent
## Coverage 3 way: 62.07 percent
## Coverage 4 way: 76.64 percent
##
## ##### RMSE #####
```

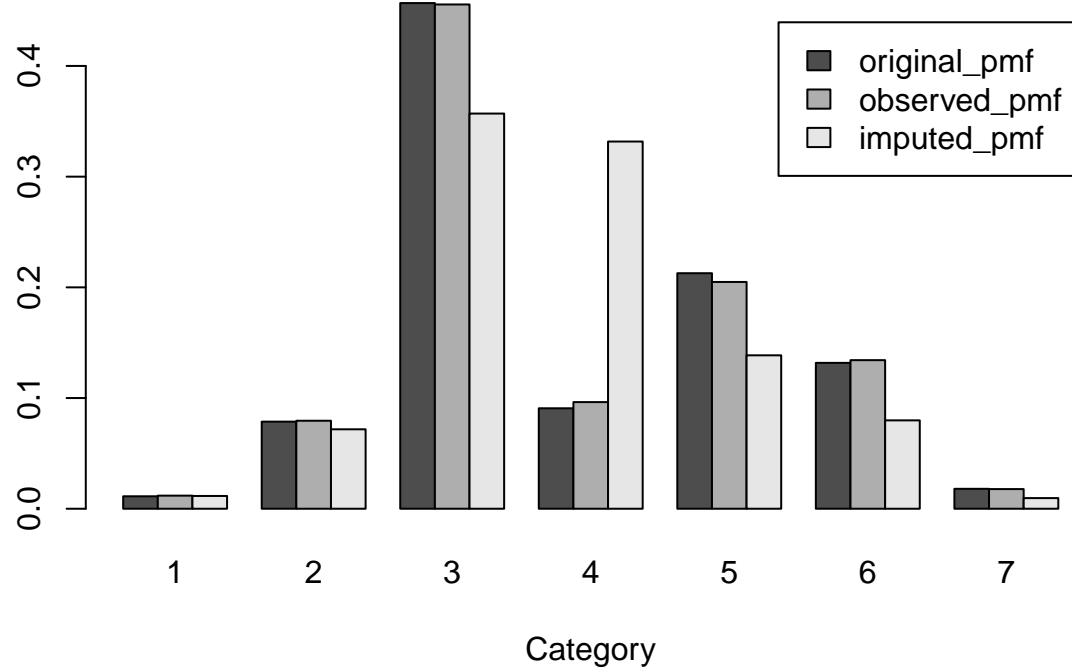
```
## RMSE 1 way: 0.059031
## RMSE 2 way: 0.01587
## RMSE 3 way: 0.004271
## RMSE 4 way: 0.001139
##
## ##### MAE #####
## MAE 1 way: 0.036622
## MAE 2 way: 0.005456
## MAE 3 way: 0.000889
## MAE 4 way: 0.000156
```



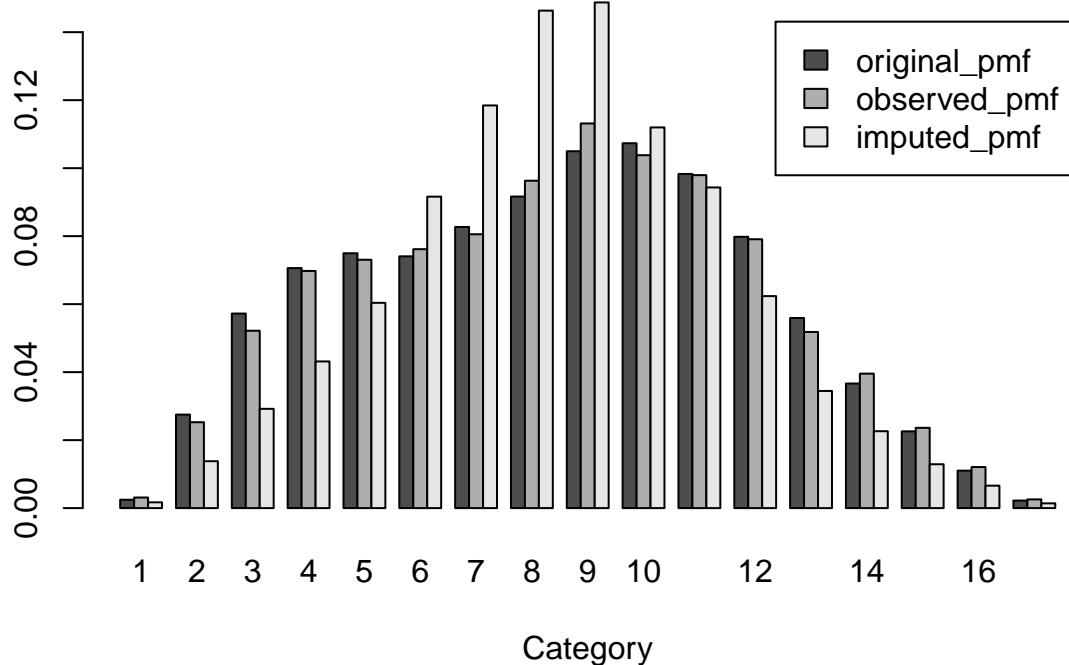
## NP



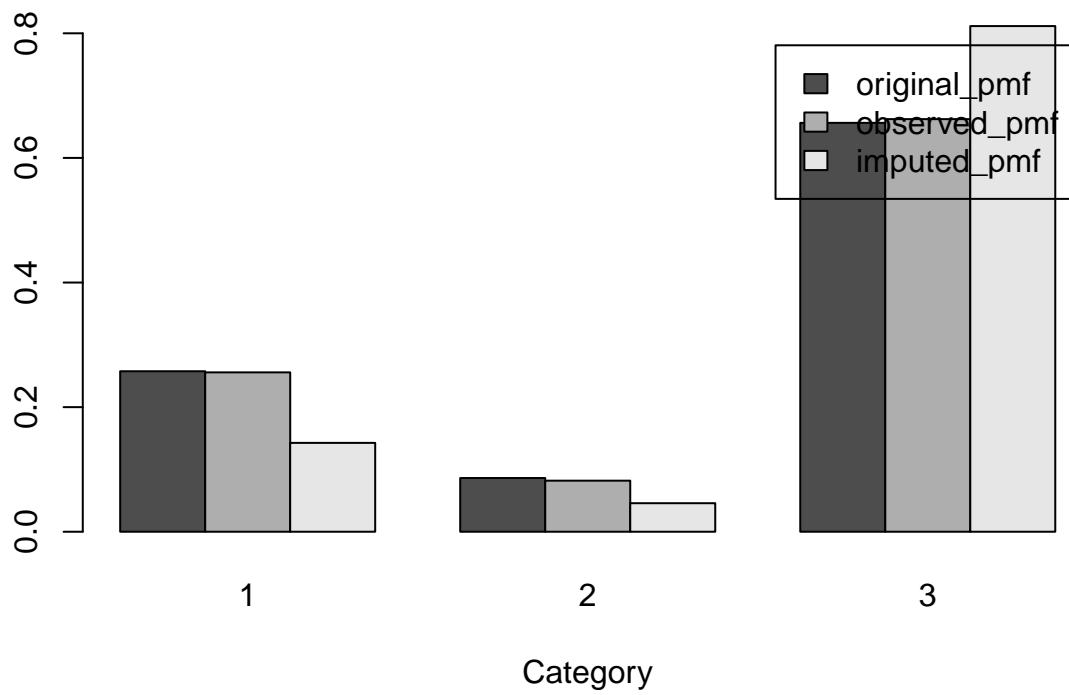
## SCHL



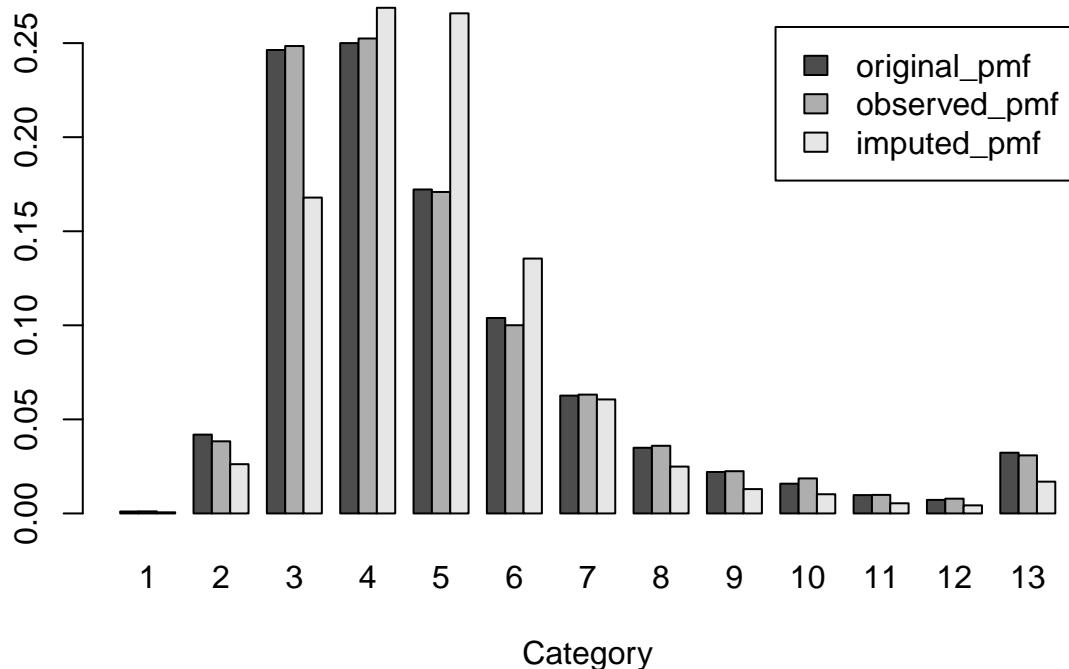
## AGEP



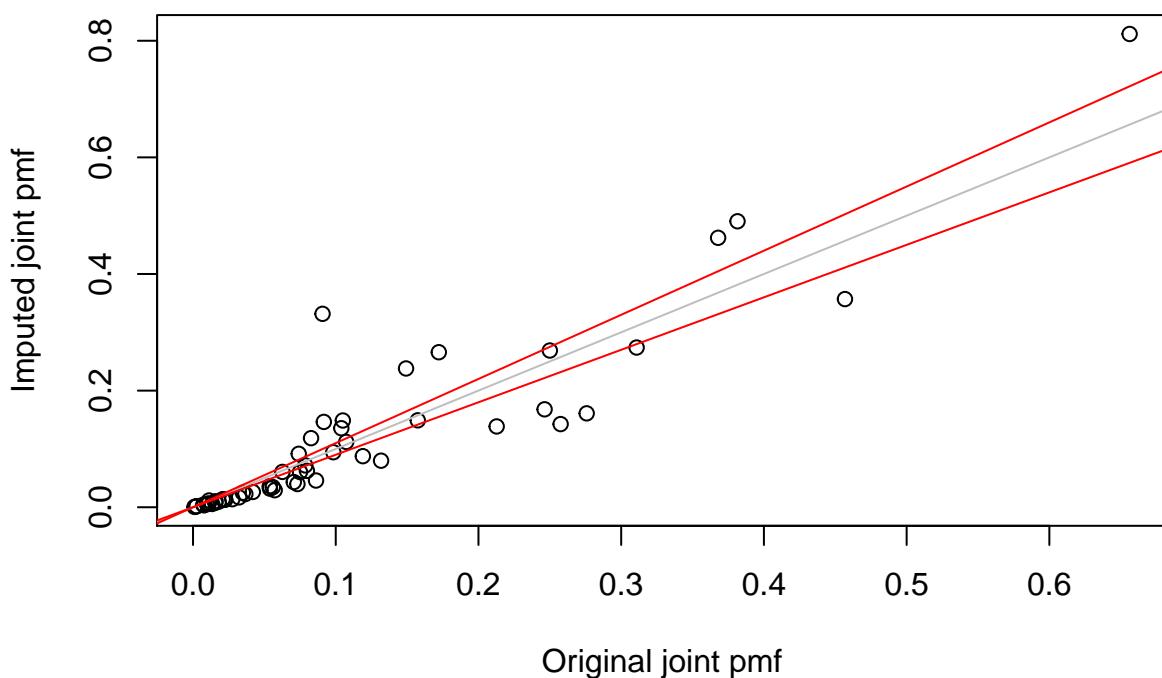
## WKL



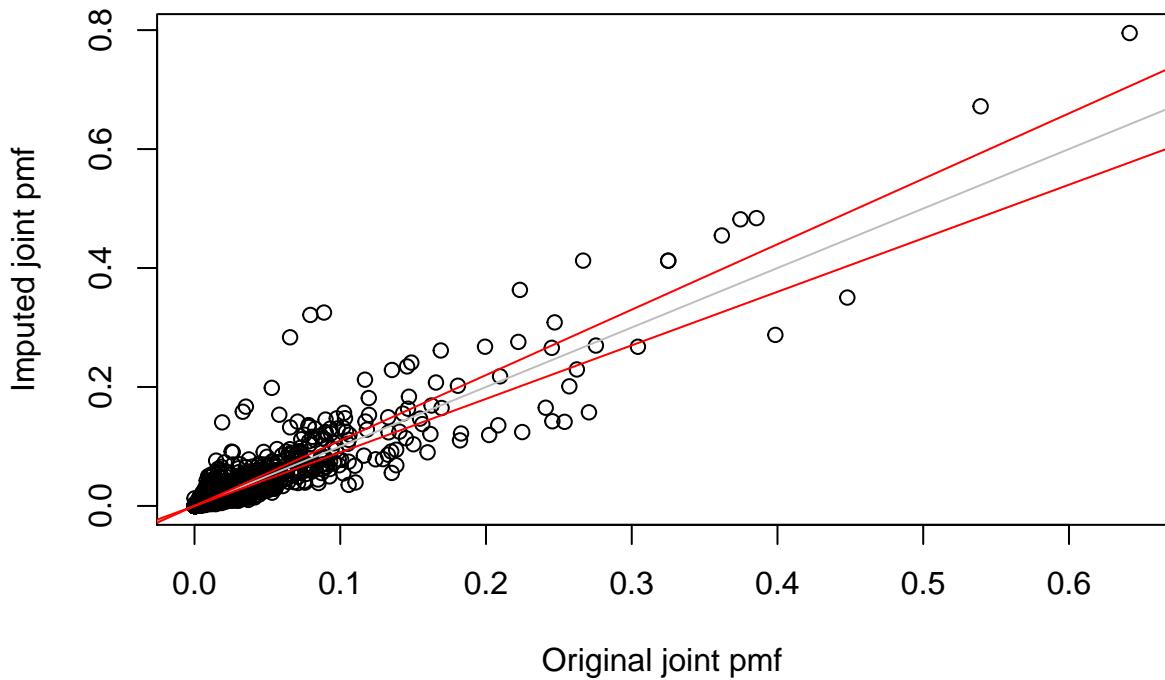
## PINCP



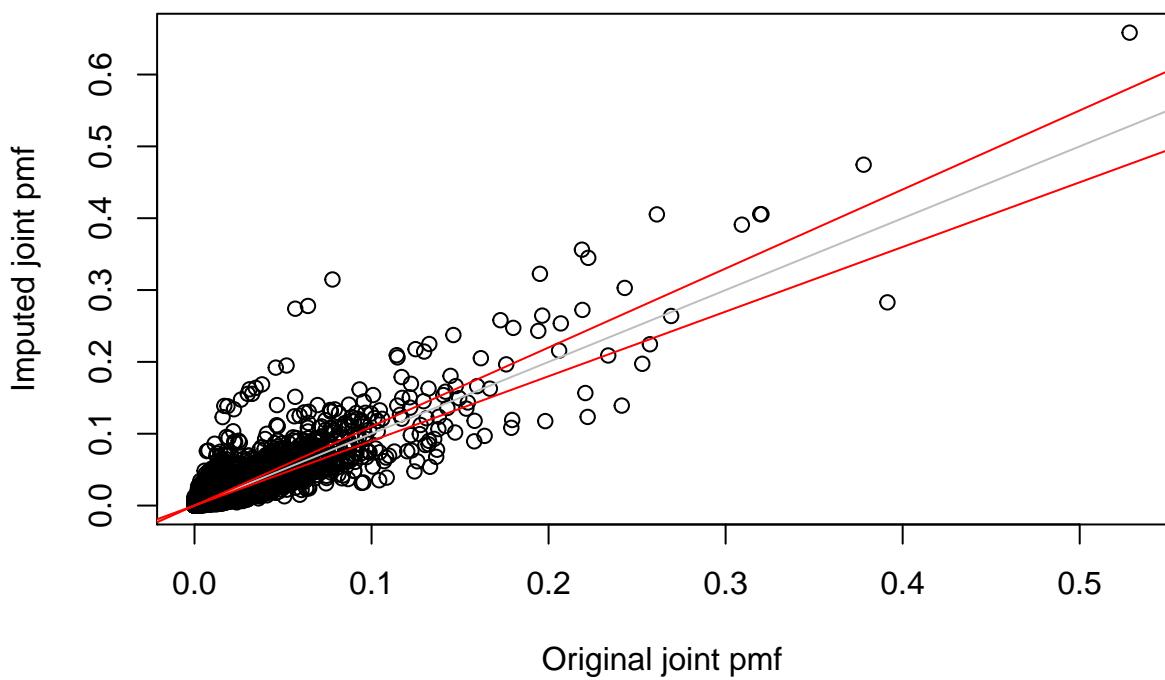
## Assess imputed pmf: 1 way



### Assess imputed pmf: 2 way



### Assess imputed pmf: 3 way



### Assess imputed pmf: 4 way

