

Utility Check

Simulated Data - Simple Approach.

```
load(paste0(getwd(), "/results/results_sat.RData"))

## Warning in register(): Can't find generic `scale_type` in package ggplot2 to
## register S3 method.

library(corrplot)

## Warning: Paket 'corrplot' wurde unter R Version 4.1.2 erstellt

## corrplot 0.92 loaded

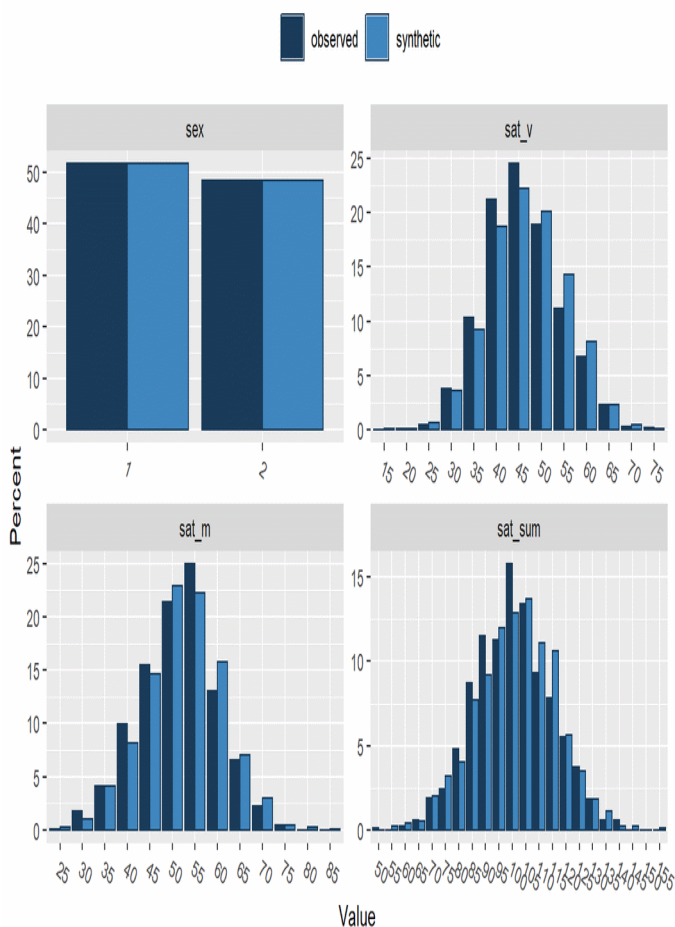
par(mfrow = c(1,2))

mean(results_syn_sat_mnorm_simple$ks > 0.05)

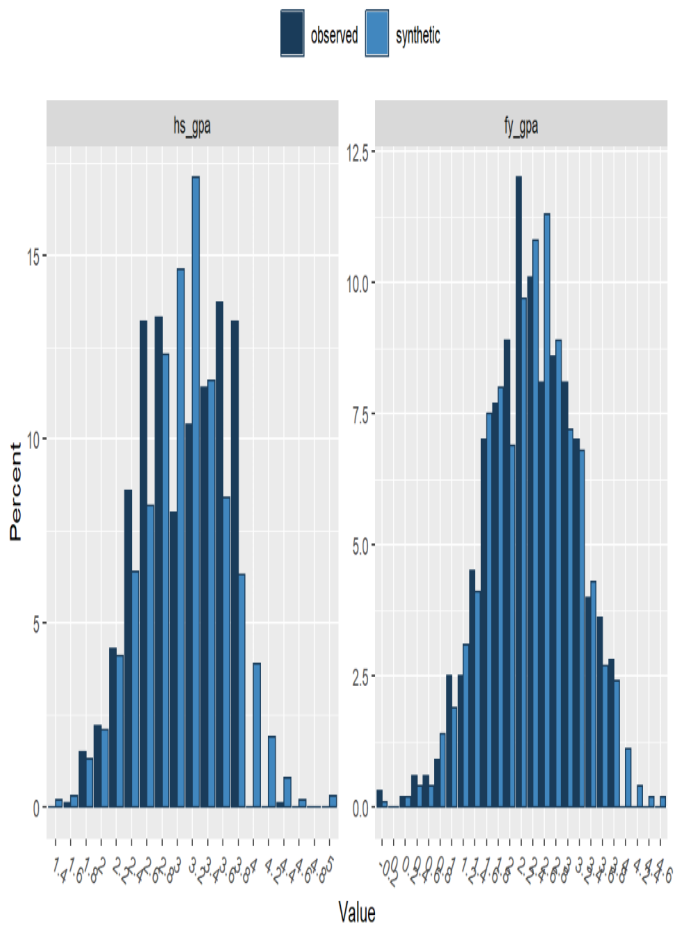
## [1] 0.8333333

results_syn_sat_mnorm_simple$comp

##
## Comparing percentages observed with synthetic
```



```
## Press return for next variable(s):
```



```
##
## Selected utility measures:
##           pMSE      S_pMSE df
## sex      0.000000  0.000000  1
## sat_v     0.000203  0.810003  4
## sat_m     0.001759  7.036278  4
## sat_sum   0.000381  1.525899  4
## hs_gpa    0.007564 30.257857  4
## fy_gpa    0.000243  0.970529  4
```

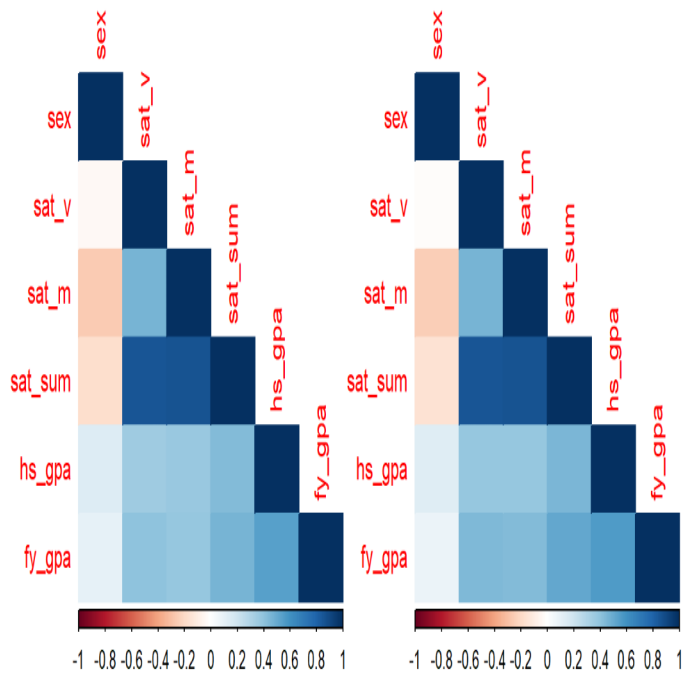
```
results_syn_sat_mnorm_simple$il
```

```
## [1] 0.6372805
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.5100000 0.8510521 0.8410040 0.9017978 0.3196433 0.4001858
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"
```

```
corrplot(results_syn_sat_mnorm_simple$cp1$corr, method = "color", type = "lower",
main = "Original")
corrplot(results_syn_sat_mnorm_simple$cp2$corr, method = "color", type = "lower",
main = "Synthetic")
```

Original

Synthetic



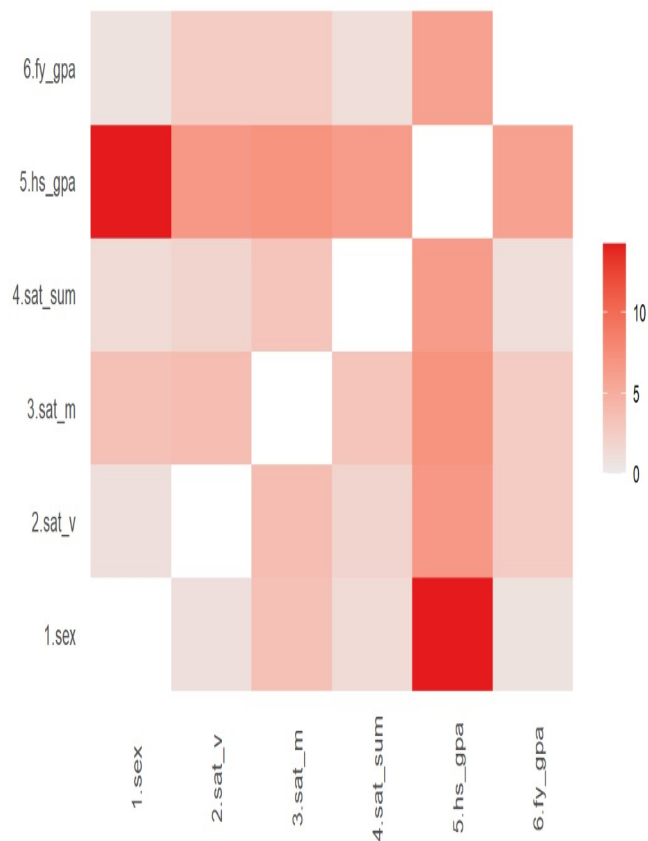
```
results_syn_sat_mnorm_simple$ug
```

```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = syn_sat_mnorm_simple, data =
## as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
##
## Selected utility measures
##      pMSE      S_pMSE
## 0.243853 5.563547
```

```
results_syn_sat_mnorm_simple$ut
```

```
##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:5.hs_gpa      3.sat_m:5.hs_gpa      2.sat_v:5.hs_gpa      4.sat_sum:5.hs_gpa
##              14.3265              7.0793              6.7851              6.3962
##      5.hs_gpa:6.fy_gpa
##              5.9794
```

Two-way utility: S_pMSE for pairs of variables



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE    0.0038  0.0106
## S_pMSE   3.1910 14.3265
## df      24.0000 24.0000
##
## For more details of all scores use print.tabs = TRUE.
```

Simulated Data - Complex Approach.

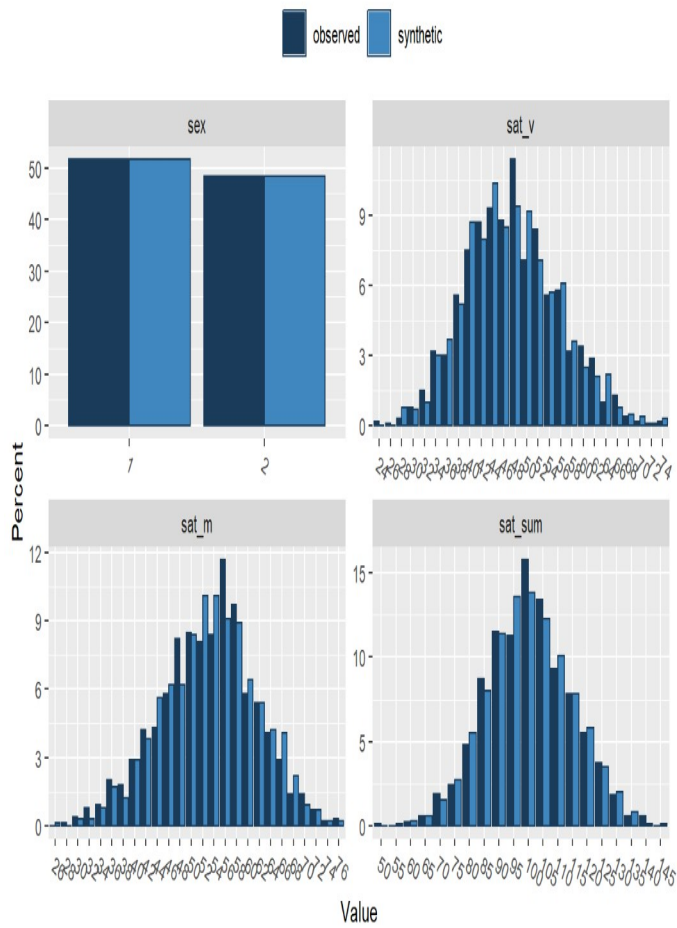
```
load(paste0(getwd(), "/results/results_sat.RData"))
library(corrplot)
par(mfrow = c(1,2))
```

```
mean(results_syn_sat_mnorm_complex$ks > 0.05)
```

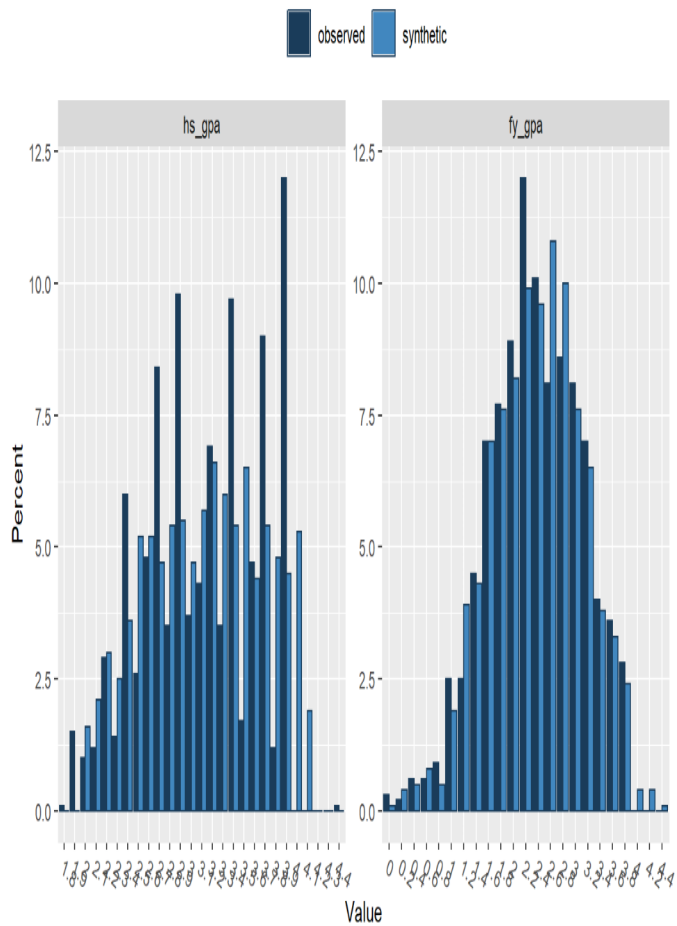
```
## [1] 0.5
```

```
results_syn_sat_mnorm_complex$comp
```

```
##
## Comparing percentages observed with synthetic
```



Press return for next variable(s):



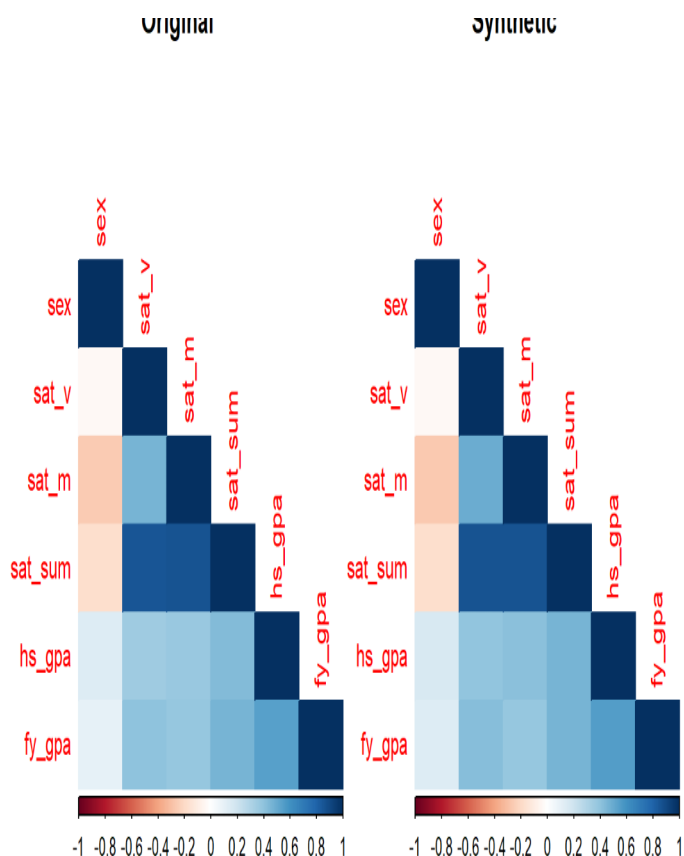
##

```
## Selected utility measures:
##           pMSE      S_pMSE df
## sex      0.000000  0.000000  1
## sat_v    0.001973  7.890132  4
## sat_m    0.002657 10.628605  4
## sat_sum  0.000404  1.614208  4
## hs_gpa   0.001424  5.694862  4
## fy_gpa   0.001199  4.796056  4
```

```
results_syn_sat_mnorm_complex$il
```

```
## [1] 0.6364775
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.5100000 0.8556561 0.8488076 0.8993197 0.3216012 0.3834804
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"
```

```
corrplot(results_syn_sat_mnorm_complex$cp1$corr, method = "color", type = "lower",
main = "Original")
corrplot(results_syn_sat_mnorm_complex$cp2$corr, method = "color", type = "lower",
main = "Synthetic")
```



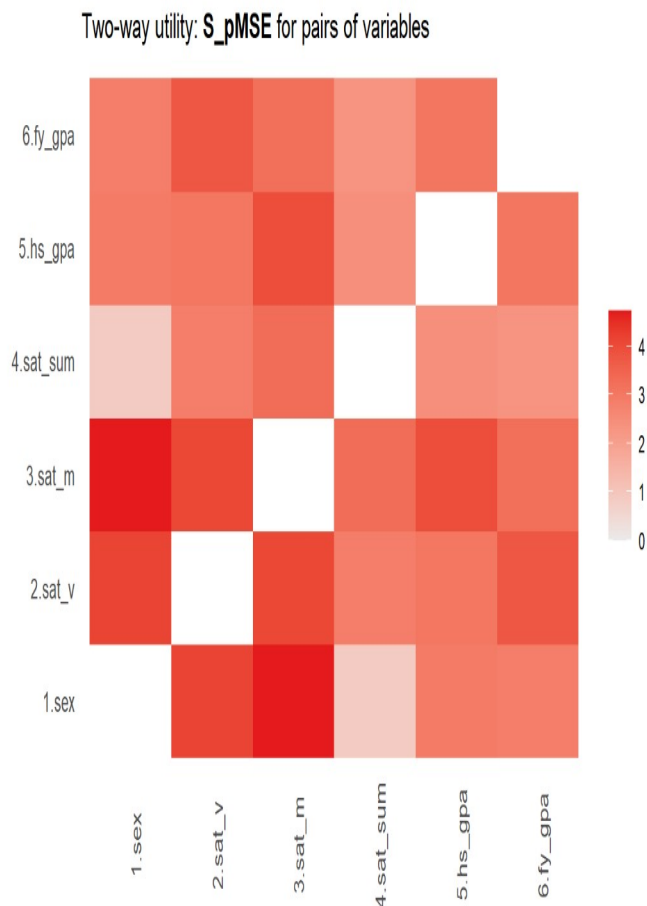
```
results_syn_sat_mnorm_complex$ug
```

```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = syn_sat_mnorm_complex, data =
## as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
```

```
##
## Selected utility measures
##      pMSE      S_pMSE
## 0.222253 5.098163

results_syn_sat_mnorm_complex$ut

##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:3.sat_m      1.sex:2.sat_v      2.sat_v:3.sat_m      3.sat_m:5.hs_gpa
##              4.7427              4.1772              4.1101              3.9939
##      2.sat_v:6.fy_gpa
##              3.7936
```



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE      0.004 0.0062
## S_pMSE    3.091 4.7427
## df        24.000 24.0000
##
## For more details of all scores use print.tabs = TRUE.
```

IPSO

```
load(paste0(getwd(), "/results/results_sat.RData"))
library(corrplot)
par(mfrow = c(1,2))

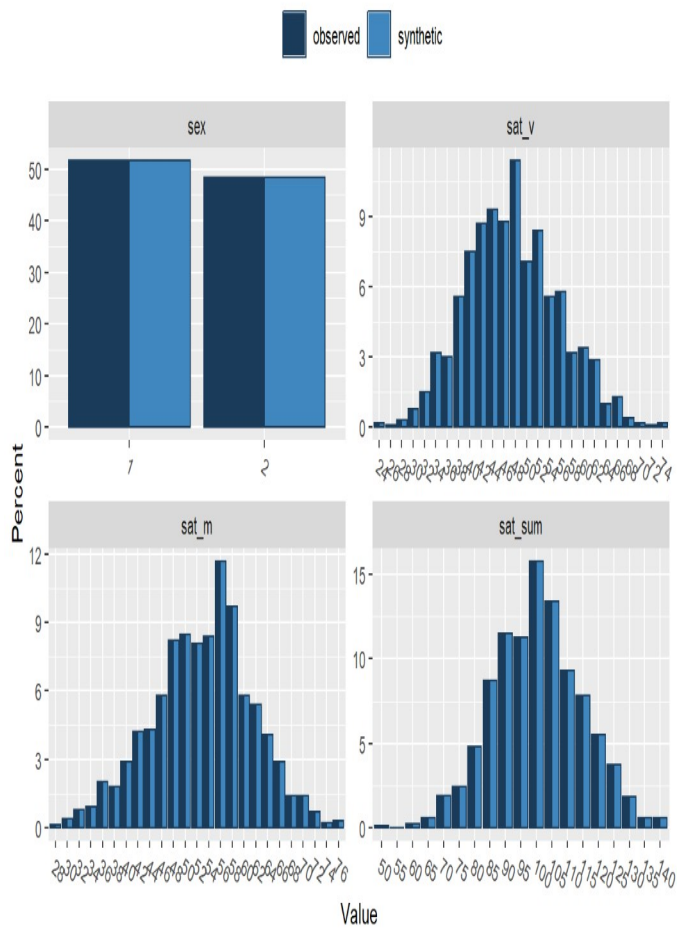
mean(results_sm_ipso_regsgdc_conf_hs_fy$ks > 0.05)

## [1] 0.8333333
```

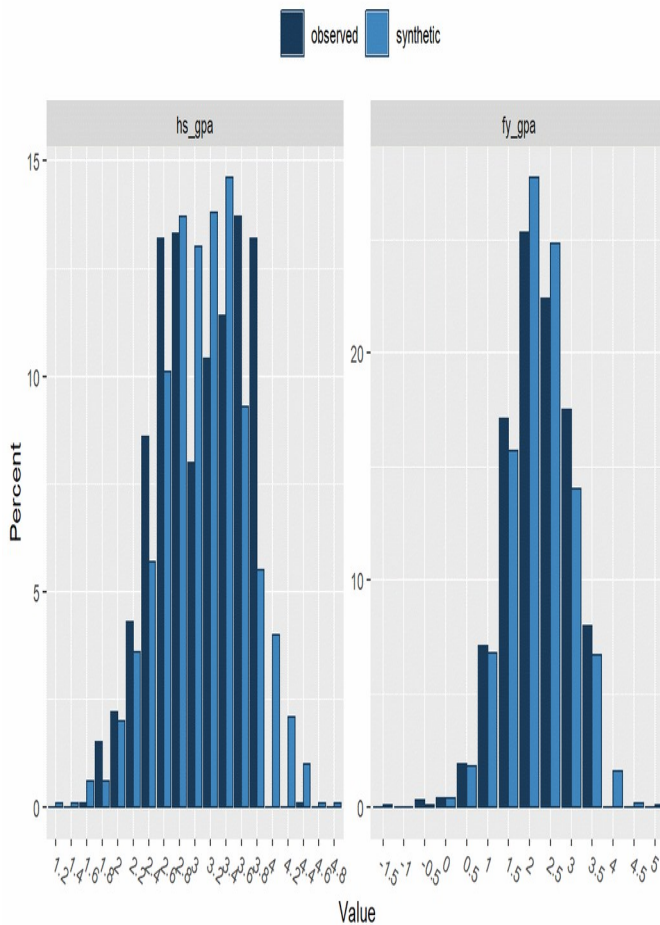
```
results_sm_ipso_regsd_conf_hs_fy$comp
```

```
##
```

```
## Comparing percentages observed with synthetic
```



```
## Press return for next variable(s):
```

```
##
## Selected utility measures:
##          pMSE      S_pMSE df
## sex      0.000000  0.000000  1
## sat_v     0.000000  0.000000  4
## sat_m     0.000000  0.000000  4
## sat_sum   0.000000  0.000000  4
## hs_gpa    0.007768 31.070045  4
## fy_gpa    0.000649  2.597249  4

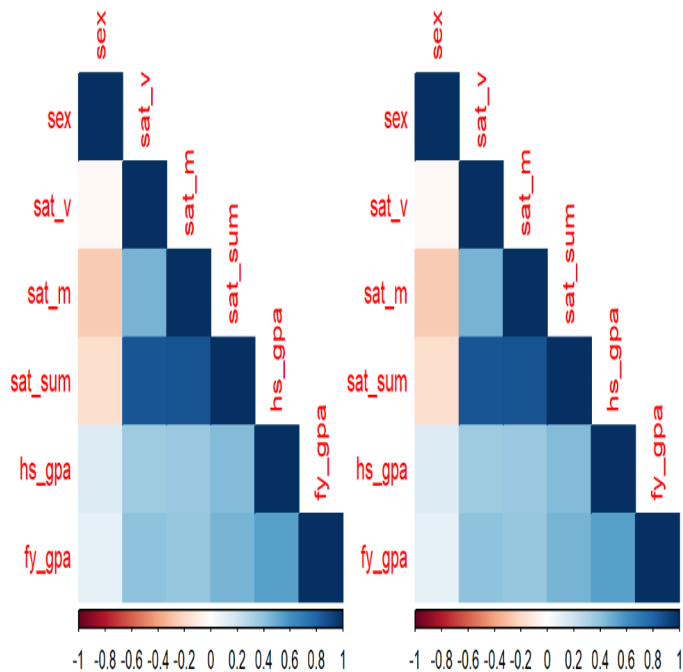
results_sm_ipso_regsdc_conf_hs_fy$il

## [1] 0.1058059
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.00000000 0.00000000 0.00000000 0.00000000 0.2843878 0.3504478
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"

corrplot(results_sm_ipso_regsdc_conf_hs_fy$cp1$corr, method = "color", type =
"lower", main = "Original")
corrplot(results_sm_ipso_regsdc_conf_hs_fy$cp2$corr, method = "color", type =
"lower", main = "Synthetic")
```

Original

Synthetic



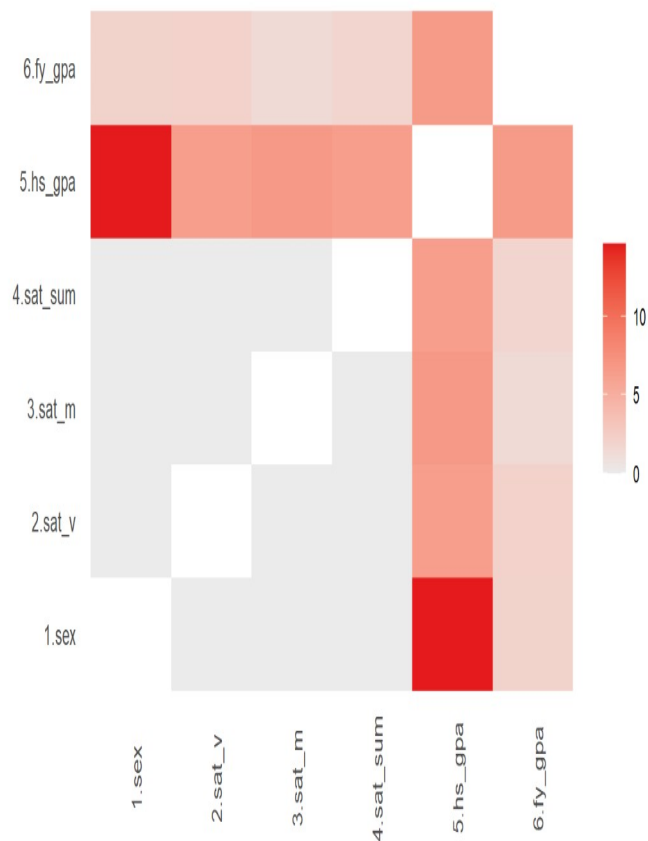
```
results_sm_ipso_regsdc_conf_hs_fy$ug
```

```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = sm_ipso_regsdc_conf_hs_fy, data =
## as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
##
## Selected utility measures
##      pMSE      S_pMSE
## 0.227840 5.735889
```

```
results_sm_ipso_regsdc_conf_hs_fy$ut
```

```
##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:5.hs_gpa      3.sat_m:5.hs_gpa      5.hs_gpa:6.fy_gpa      4.sat_sum:5.hs_gpa
##              14.6795              6.8080              6.6737              6.4191
##      2.sat_v:5.hs_gpa
##              6.4159
```

Two-way utility: S_pMSE for pairs of variables



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE    0.0021  0.0102
## S_pMSE   1.8209 14.6795
## df      24.0000 24.0000
##
## For more details of all scores use print.tabs = TRUE.
```

FCS CART

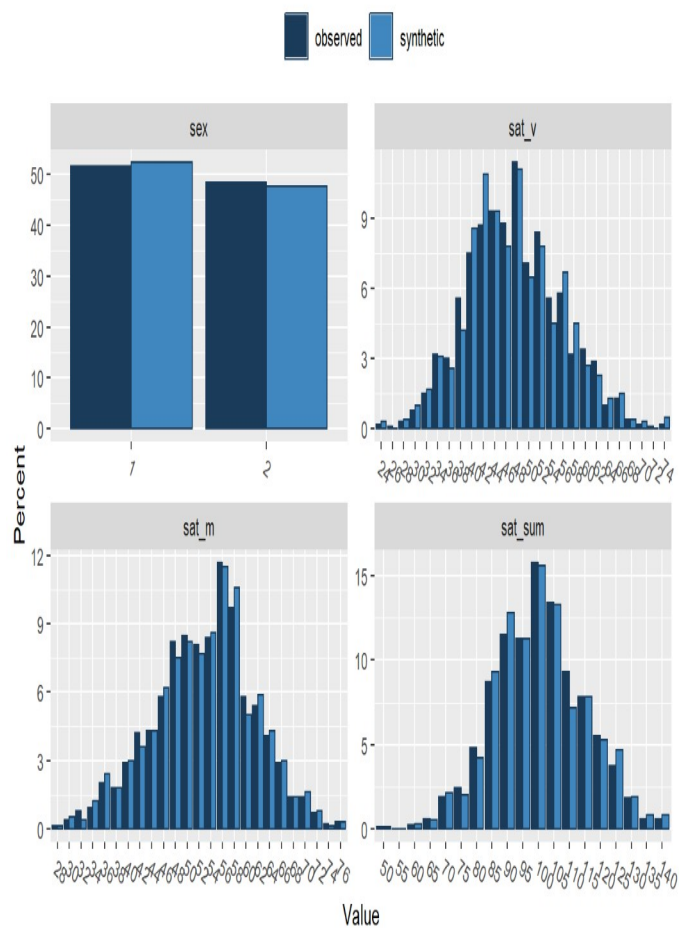
```
load(paste0(getwd(), "/results/results_sat.RData"))
library(corrplot)
par(mfrow = c(1,2))
```

```
mean(results_sm_sat_fcs_cart$ks > 0.05)
```

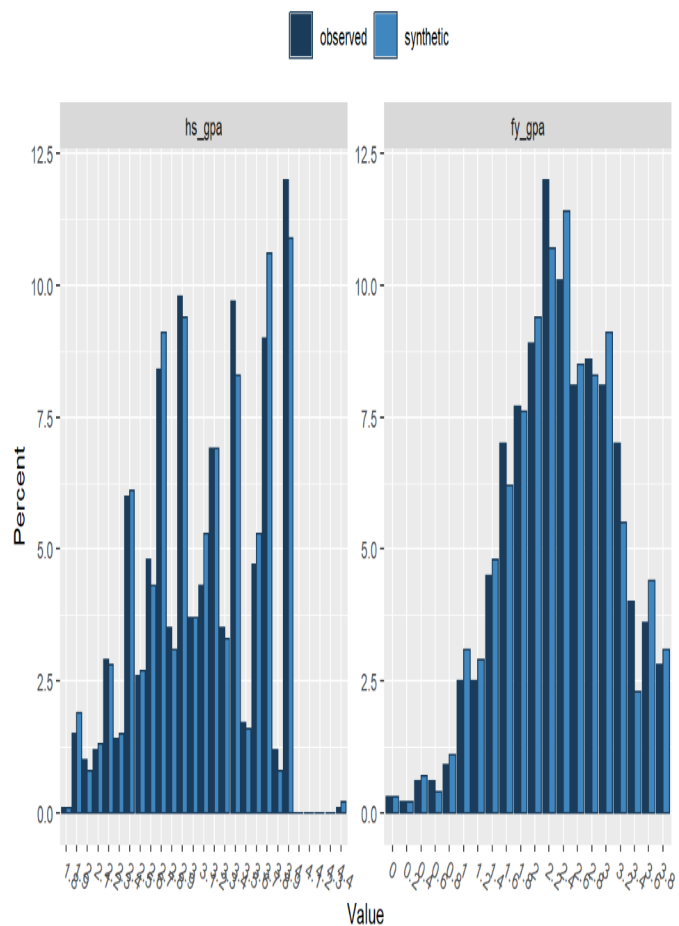
```
## [1] 1
```

```
results_sm_sat_fcs_cart$comp
```

```
##
## Comparing percentages observed with synthetic
```



Press return for next variable(s):



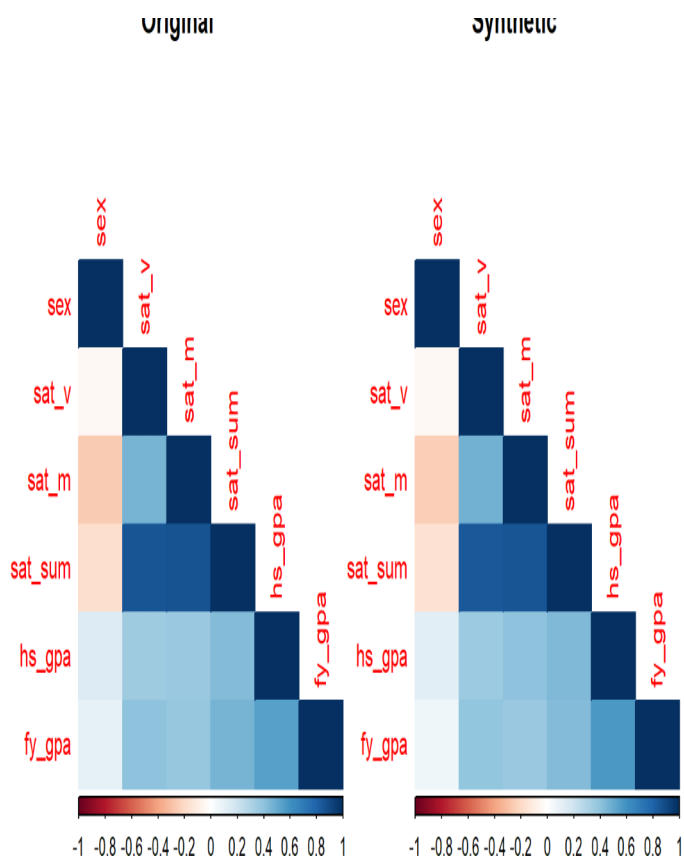
##

```
## Selected utility measures:
##           pMSE   S_pMSE df
## sex      0.000012 0.196299 1
## sat_v    0.000771 3.084445 4
## sat_m    0.000027 0.108885 4
## sat_sum  0.000277 1.109359 4
## hs_gpa   0.000031 0.122727 4
## fy_gpa   0.000205 0.821632 4
```

```
results_sm_sat_fcs_cart$il
```

```
## [1] 0.6254697
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.4790000 0.8378600 0.8433546 0.8951682 0.3124193 0.3850162
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"
```

```
corrplot(results_sm_sat_fcs_cart$cp1$corr, method = "color", type = "lower", main =
"Original")
corrplot(results_sm_sat_fcs_cart$cp2$corr, method = "color", type = "lower", main =
"Synthetic")
```



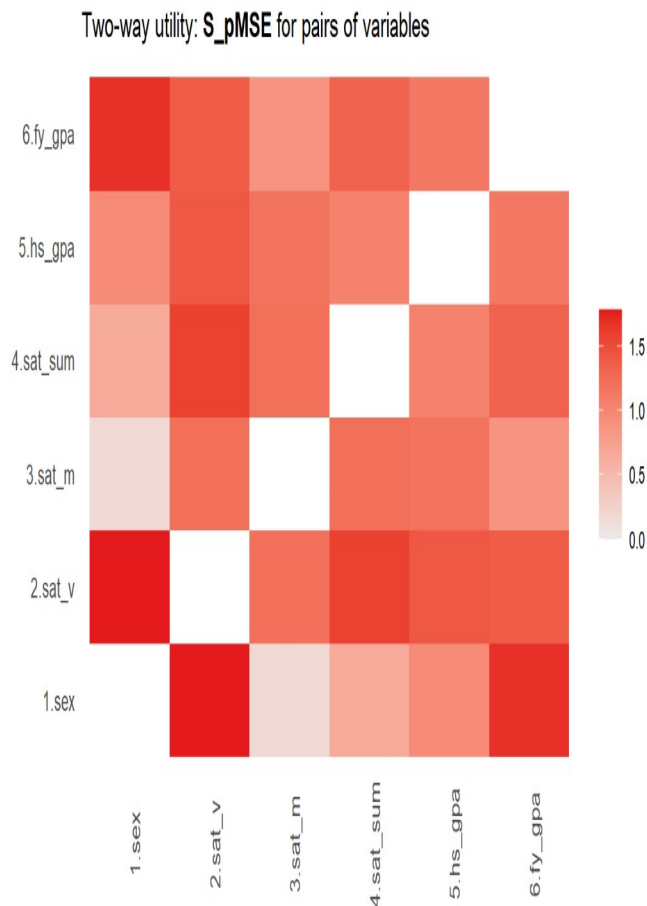
```
results_sm_sat_fcs_cart$ug
```

```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = sm_sat_fcs_cart, data = as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
##
```

```
## Selected utility measures
##      pMSE      S_pMSE
## 0.049752 1.454311
```

```
results_sm_sat_fcs_cart$ut
```

```
##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:2.sat_v      1.sex:6.fy_gpa 2.sat_v:4.sat_sum 2.sat_v:5.hs_gpa
##              1.7856              1.6913              1.5942              1.4261
##      2.sat_v:6.fy_gpa
##              1.3823
```



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE      0.0017 0.0022
## S_pMSE    1.2091 1.7856
## df        24.0000 24.0000
##
## For more details of all scores use print.tabs = TRUE.
```

GAN CTGAN

```
load(paste0(getwd(), "/results/results_sat.RData"))
library(corrplot)
par(mfrow = c(1,2))

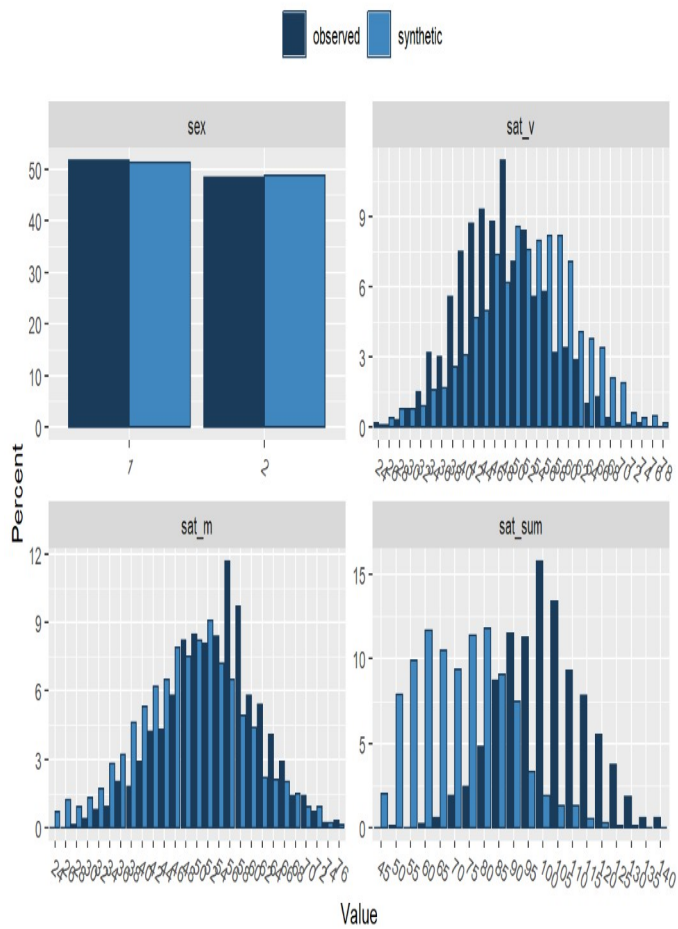
mean(results_sm_sat_gan_ctgan$ks > 0.05)

## [1] 0.1666667
```

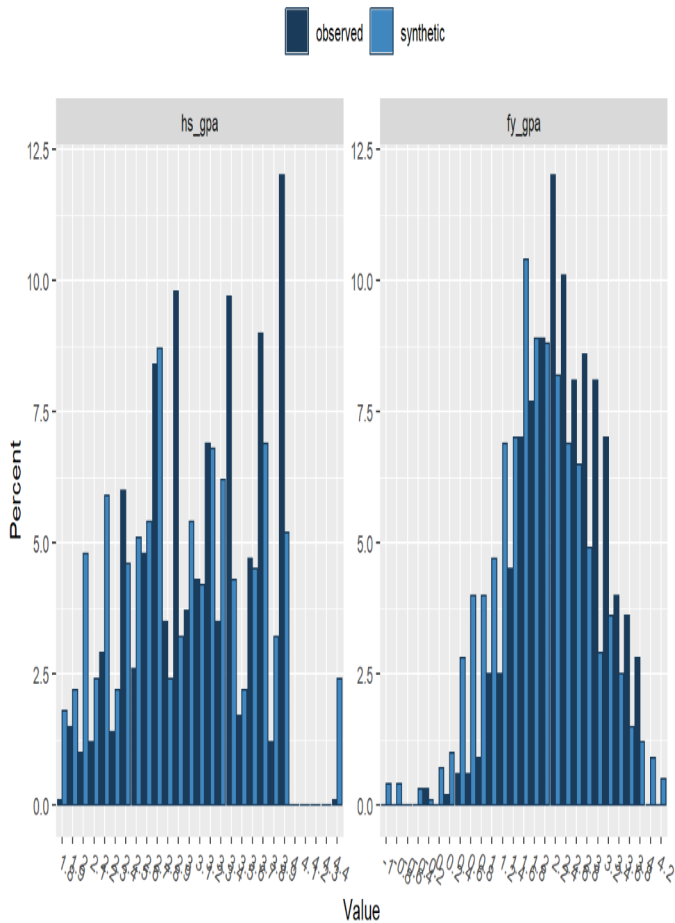
```
results_sm_sat_gan_ctgan$comp
```

```
##
```

```
## Comparing percentages observed with synthetic
```



```
## Press return for next variable(s):
```



```
##
## Selected utility measures:
##           pMSE      S_pMSE df
## sex      0.000004   0.06405  1
## sat_v     0.018286  73.14275  4
## sat_m     0.013173  52.69385  4
## sat_sum   0.124991 499.96473  4
## hs_gpa    0.004946  19.78465  4
## fy_gpa    0.021600  86.40012  4

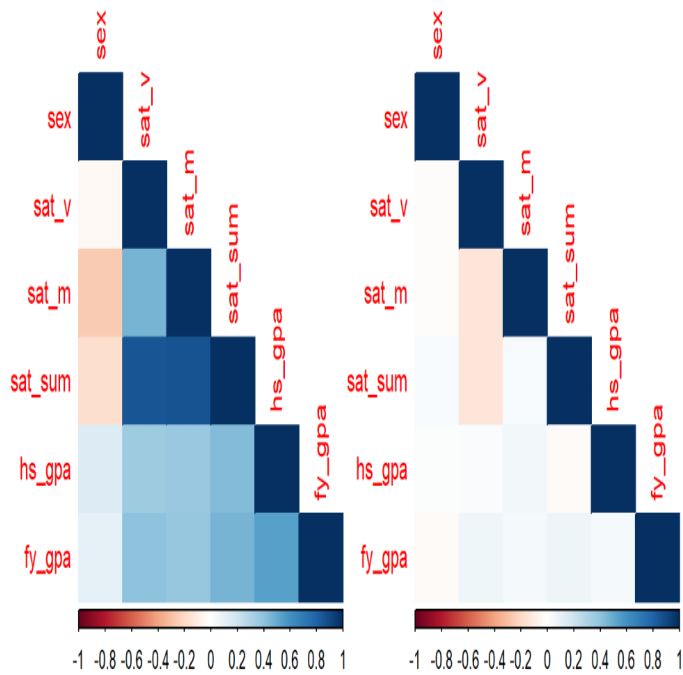
results_sm_sat_gan_ctgan$il

## [1] 0.6577558
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.5100000 0.8522727 0.8579451 0.9434975 0.3499452 0.4328742
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"

corrplot(results_sm_sat_gan_ctgan$cp1$corr, method = "color", type = "lower", main =
"Original")
corrplot(results_sm_sat_gan_ctgan$cp2$corr, method = "color", type = "lower", main =
"Synthetic")
```


Original

Synthetic



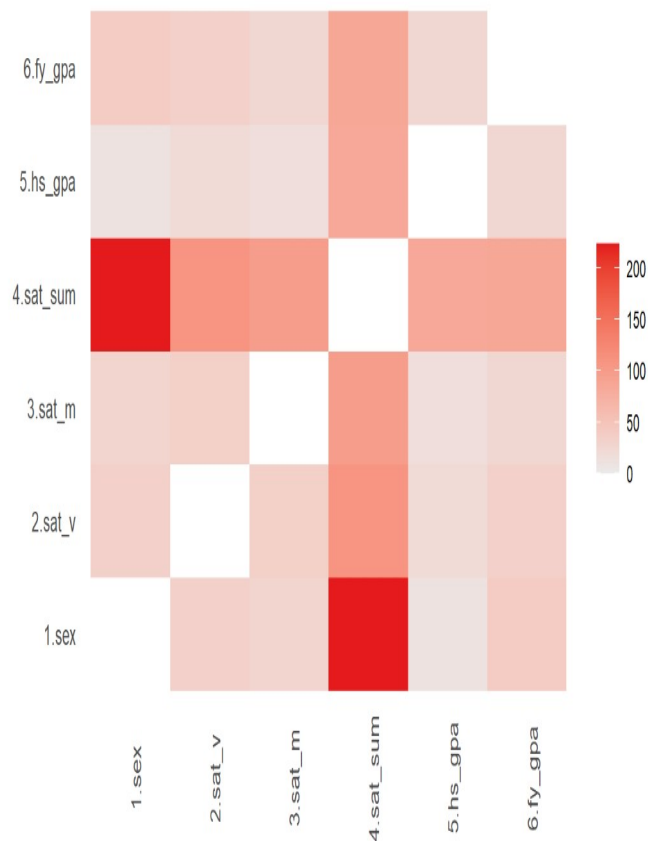
results_sm_sat_gan_ctgan\$ug

```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = sm_sat_gan_ctgan, data = as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
##
## Selected utility measures
##      pMSE      S_pMSE
## 0.206471 5.143291
```

results_sm_sat_gan_ctgan\$ut

```
##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:4.sat_sum  2.sat_v:4.sat_sum  3.sat_m:4.sat_sum  4.sat_sum:6.fy_gpa
##                224.5810                109.6788                99.9756                87.4208
##      4.sat_sum:5.hs_gpa
##                86.0208
```

Two-way utility: S_pMSE for pairs of variables



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE    0.0392  0.1645
## S_pMSE 33.3218 224.5810
## df      24.0000  24.0000
##
## For more details of all scores use print.tabs = TRUE.
```

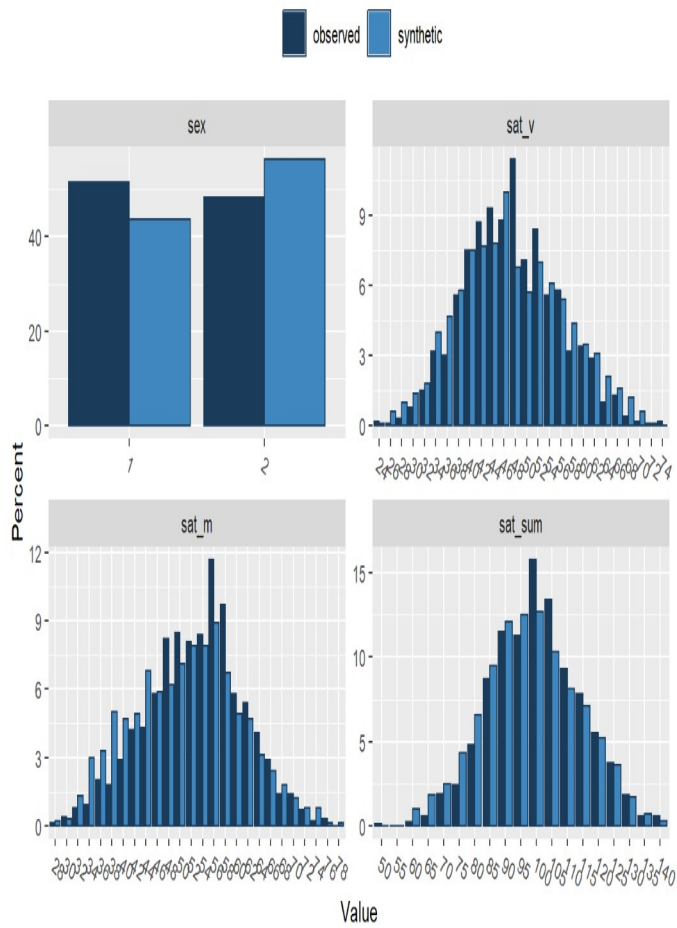
GAN EPOCH1000

```
library(corrplot)
par(mfrow = c(1,2))
load(paste0(getwd(), "/results/results_sat.RData"))
mean(results_sm_sat_gan_ctgan_epoch1000$ks > 0.05)
```

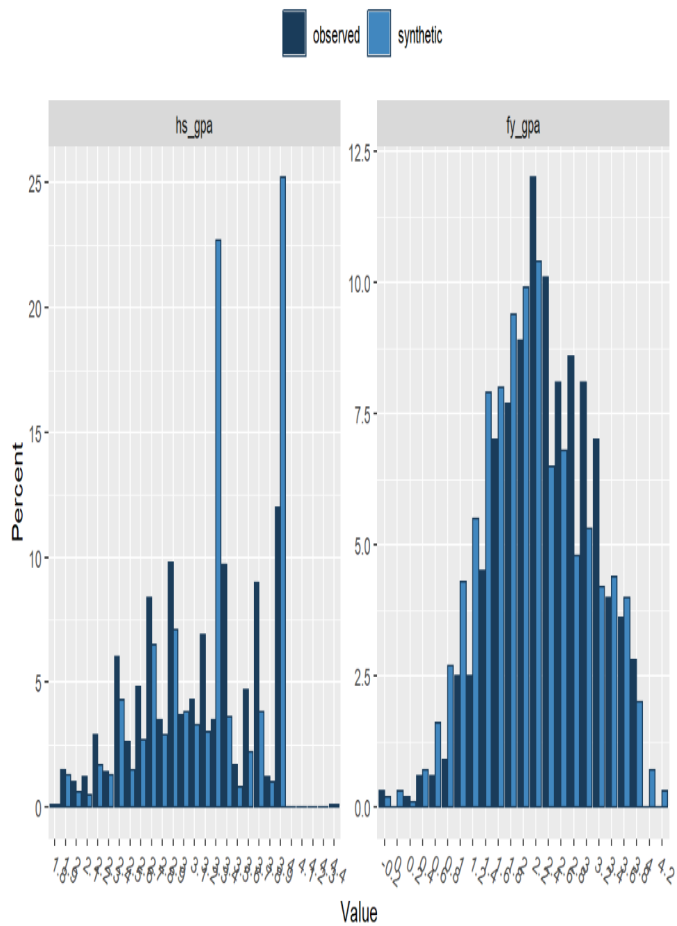
```
## [1] 0.1666667
```

```
results_sm_sat_gan_ctgan_epoch1000$comp
```

```
##
## Comparing percentages observed with synthetic
```



Press return for next variable(s):



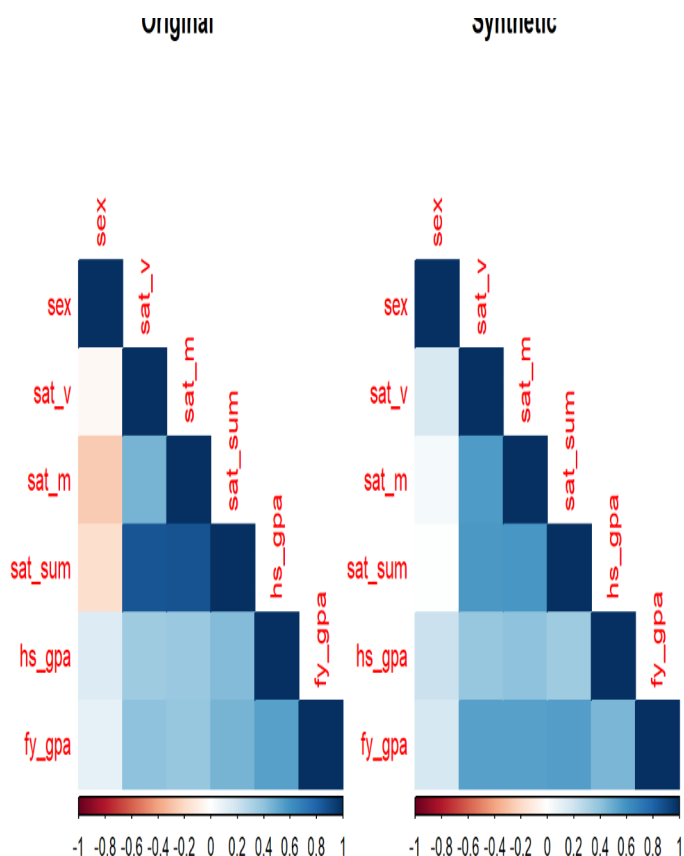
##

```
## Selected utility measures:
##           pMSE      S_pMSE df
## sex      0.001564 25.019268  1
## sat_v     0.001935  7.738043  4
## sat_m     0.004628 18.510251  4
## sat_sum   0.002794 11.176962  4
## hs_gpa    0.008166 32.663548  4
## fy_gpa    0.006987 27.950000  4
```

```
results_sm_sat_gan_ctgan_epoch1000$il
```

```
## [1] 0.6363493
## attr("indiv_distances")
##      sex      sat_v      sat_m      sat_sum      hs_gpa      fy_gpa
## 0.4750000 0.8464092 0.8570983 0.9093859 0.3230122 0.4071904
## attr("n")
## [1] 1000
## attr("class")
## [1] "il_variables"
```

```
corrplot(results_sm_sat_gan_ctgan_epoch1000$cp1$corr, method = "color", type =
"lower", main = "Original")
corrplot(results_sm_sat_gan_ctgan_epoch1000$cp2$corr, method = "color", type =
"lower", main = "Synthetic")
```



```
results_sm_sat_gan_ctgan_epoch1000$ug
```

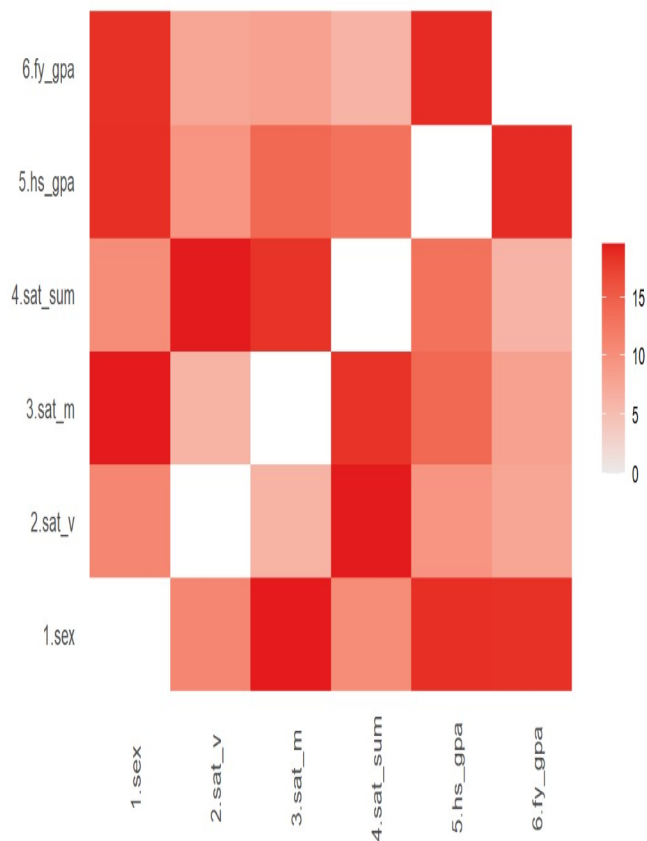
```
##
## Utility score calculated by method: cart
##
## Call:
## utility.gen.data.frame(object = sm_sat_gan_ctgan_epoch1000, data =
as.data.frame(orig_sat))
##
## Null utilities simulated from a permutation test with 50 replications.
```

```
##
## Selected utility measures
##      pMSE      S_pMSE
## 0.118700 3.079181

results_sm_sat_gan_ctgan_epoch1000$ut

##
## Two-way utility: S_pMSE value plotted for 15 pairs of variables.
##
## Variable combinations with worst 5 utility scores (S_pMSE):
##      1.sex:3.sat_m 2.sat_v:4.sat_sum 5.hs_gpa:6.fy_gpa      1.sex:5.hs_gpa
##              19.5920              19.5505              18.8641              18.6489
##      1.sex:6.fy_gpa
##              18.5117
```

Two-way utility: S_pMSE for pairs of variables



```
##
## Medians and maxima of selected utility measures for all tables compared
##      Medians  Maxima
## pMSE      0.0117 0.0293
## S_pMSE    13.2194 19.5920
## df        24.0000 24.0000
##
## For more details of all scores use print.tabs = TRUE.
```