

# Quality Assurance Report

generated on 01 Feb 2022, 10:52AM  
for 300 target vs. 300 synthetic subjects

# Executive Summary

Overall Accuracy: 93.2%

Privacy Tests: 

Table	Columns	Target Data	Synthetic Data	Accuracy
NEWHIVDATA	1 numeric	300 rows	300 rows	univariate: 97.9%
	7 categorical			bivariate: 91.6%

# Privacy

# Privacy Tests

## Identical Match Share

(IMS)



### Holdout Data

IMS: 0.0% = 0 of 150 subjects have identical matches within the target data.

### Synthetic Data

IMS: 0.0% = 0 of 150 subjects have identical matches within the target data.

TEST PASSED! The share of subjects within the synthetic data that matches an actual subject from the target data is not significantly bigger than the share that is to be expected when analyzing the target data itself.

## Distance to Closest Record

(DCR)



### Holdout Data

5<sup>th</sup> percentile of the DCR = 0.37

### Synthetic Data

5<sup>th</sup> percentile of the DCR = 0.33

TEST PASSED! The normalized distance for synthetic subjects to their closest actual subject within the target data is not significantly closer than the distance that is to be expected when analyzing the target data itself.

## Nearest Neighbor Distance Ratio

(NNDR)



### Holdout Data

5<sup>th</sup> percentile of the NNDR = 0.53

### Synthetic Data

5<sup>th</sup> percentile of the NNDR = 0.66

TEST PASSED! The distance ratio between nearest and second-nearest record for synthetic subjects to their closest actual subject within the target data is not significantly closer than the ratio that is to be expected when analyzing the target data itself.

# Privacy Tests

## Identical Match Share

(IMS)



**Holdout Data: 0.0%**

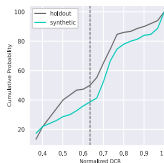
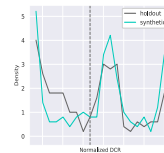
(0 of 150)

**Synthetic Data: 0.0%**

(0 of 150)

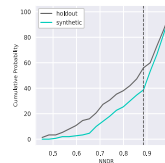
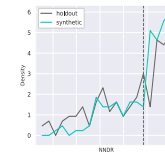
## Distance to Closest Record

(DCR)



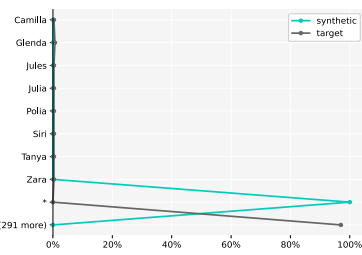
## Nearest Neighbor Distance Ratio

(NNDR)

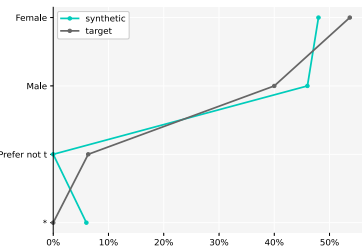


# Univariate Distributions

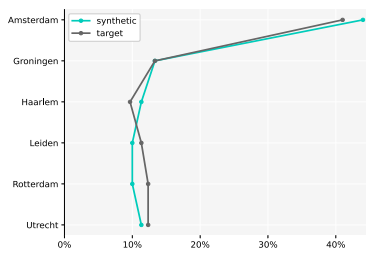
Name



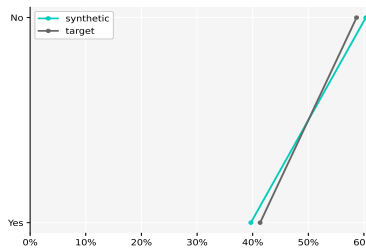
Gender



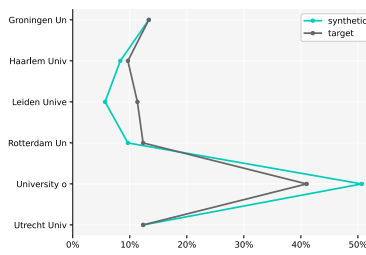
Location/City



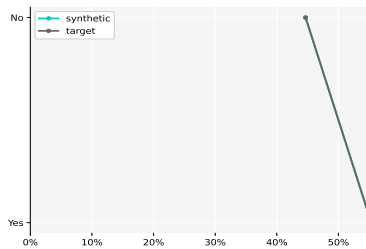
HIV Positive



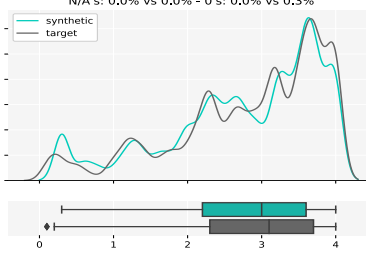
University



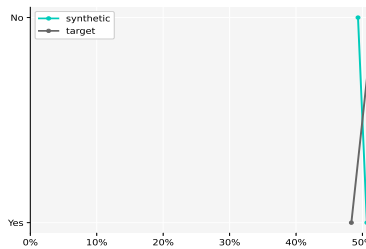
Other STDs



GPA



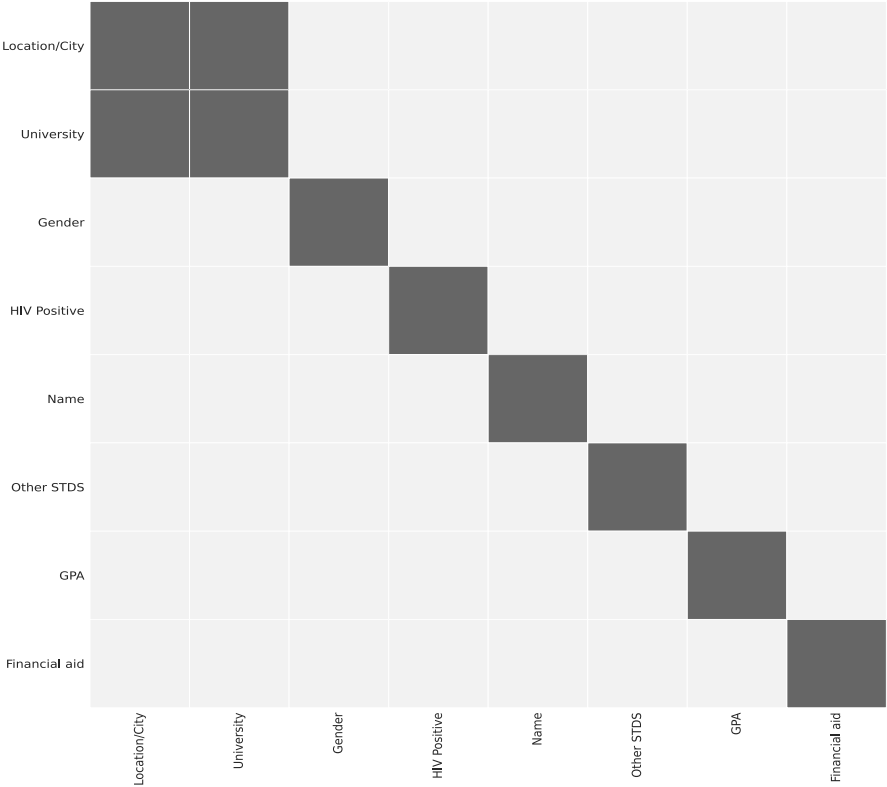
Financial aid



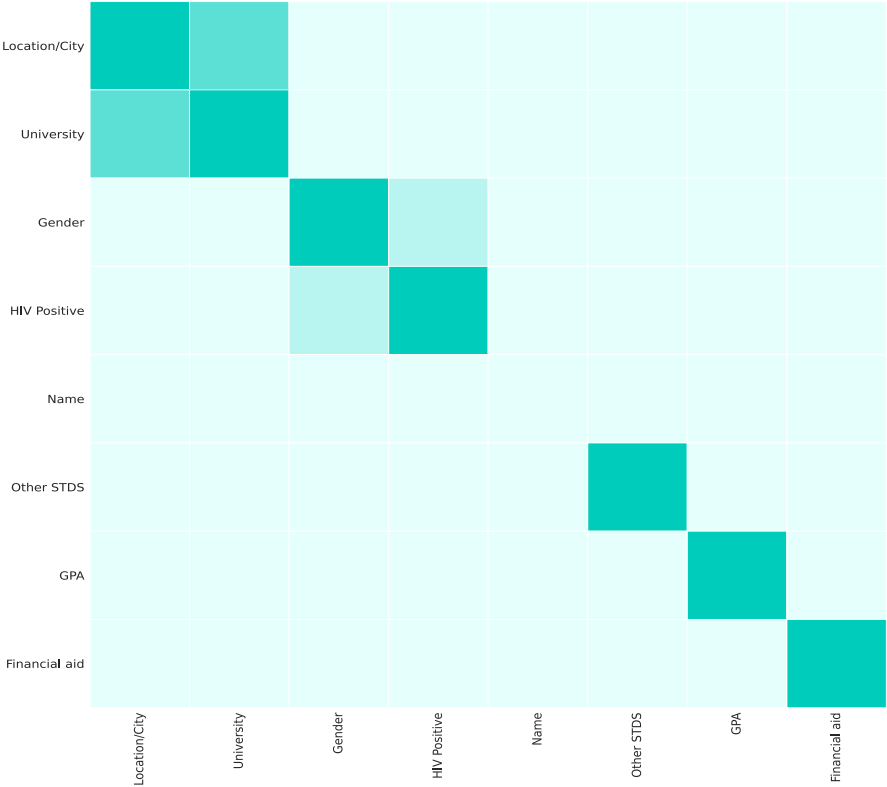
# Correlations



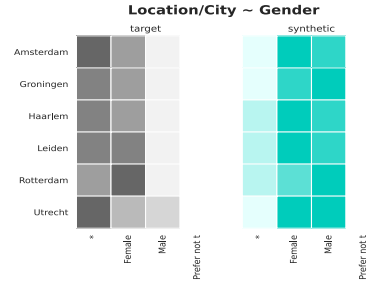
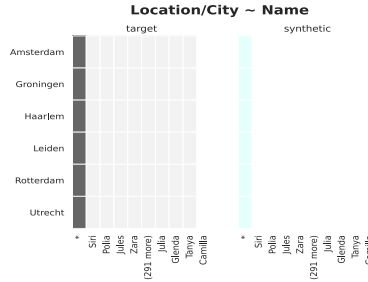
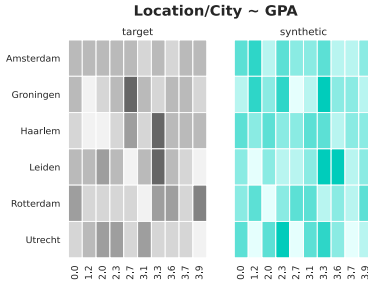
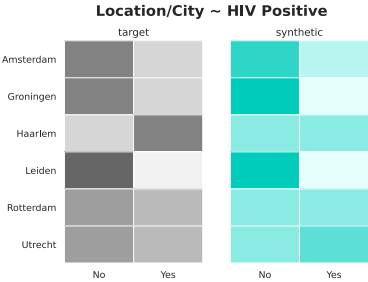
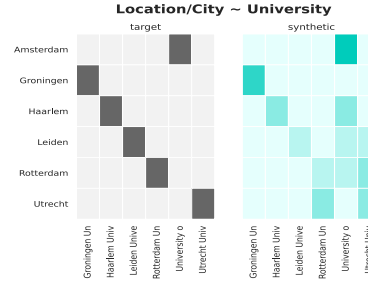
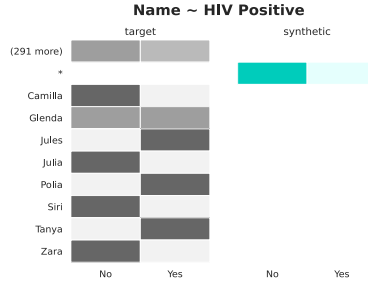
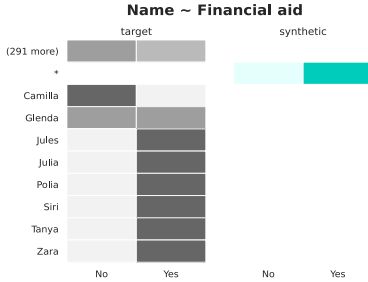
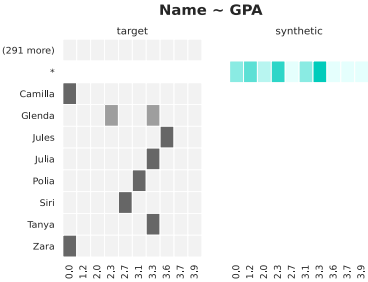
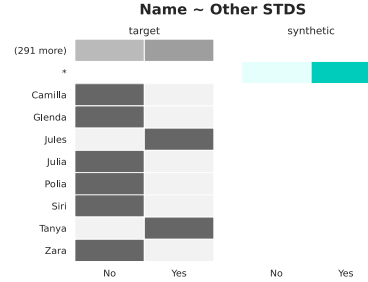
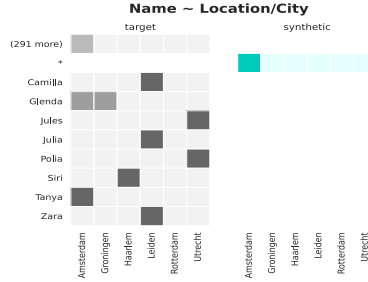
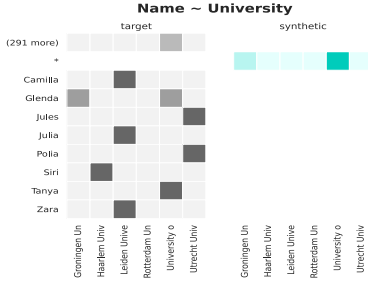
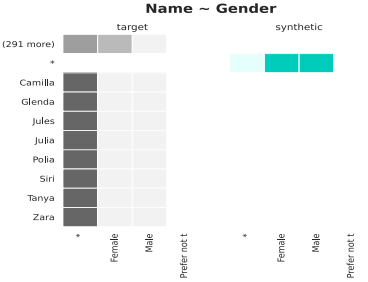
target



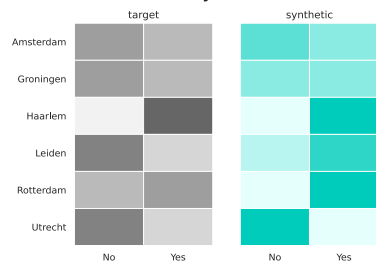
synthetic



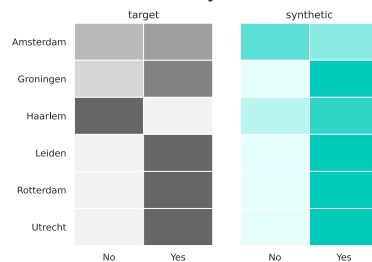
# Bivariate Distributions



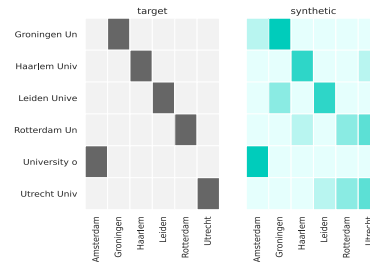
Location/City ~ Financial aid



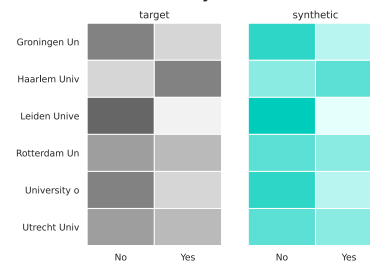
Location/City ~ Other STDs



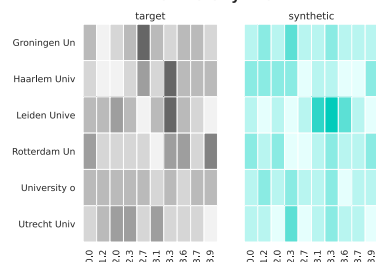
University ~ Location/City



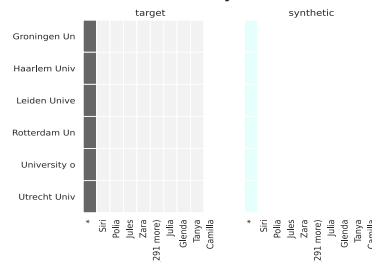
University ~ HIV Positive



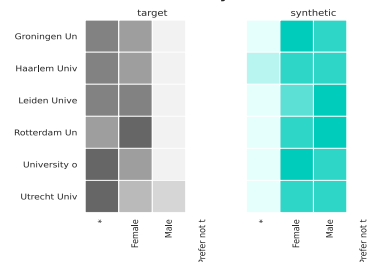
University ~ GPA



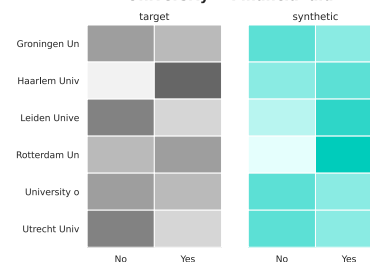
University ~ Name



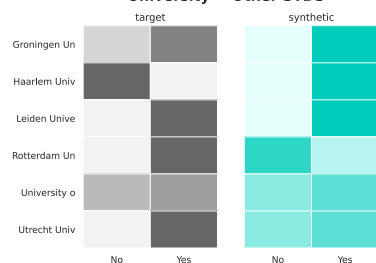
University ~ Gender



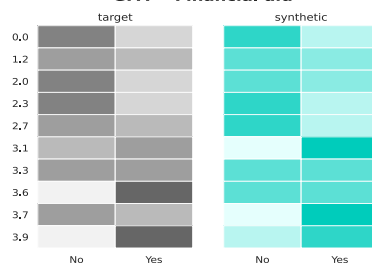
University ~ Financial aid



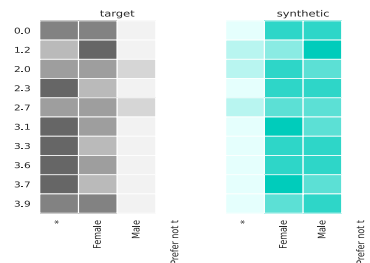
University ~ Other STDs



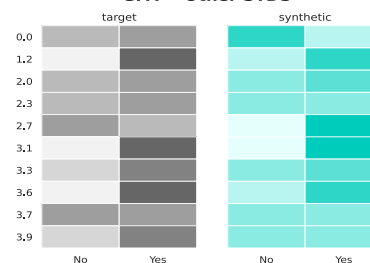
GPA ~ Financial aid



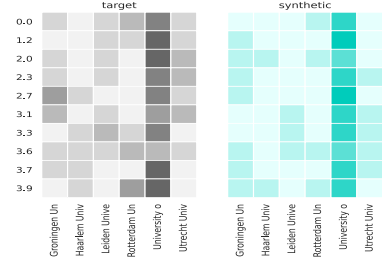
GPA ~ Gender



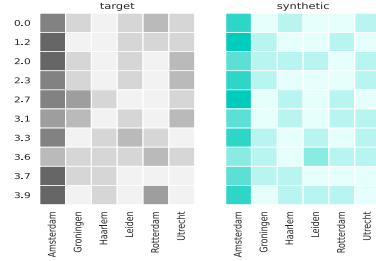
GPA ~ Other STDs



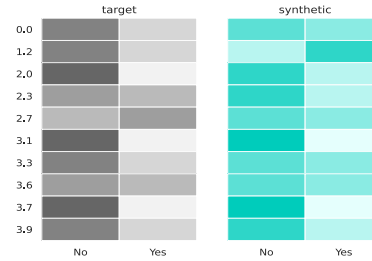
GPA ~ University



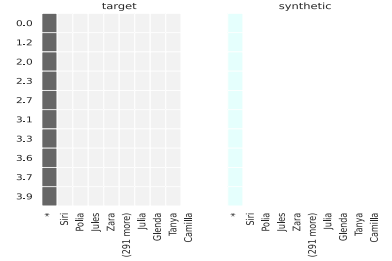
GPA ~ Location/City



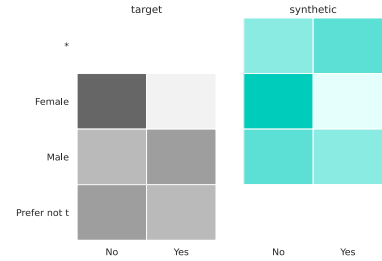
GPA ~ HIV Positive



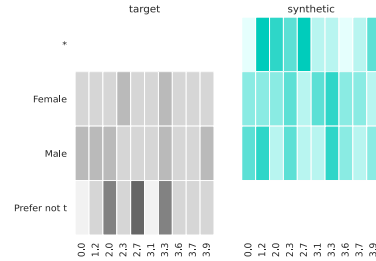
GPA ~ Name



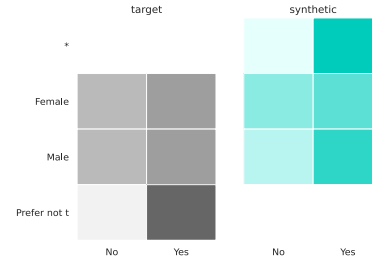
Gender ~ HIV Positive



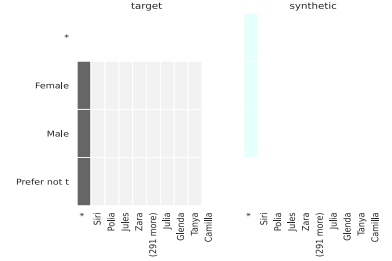
Gender ~ GPA



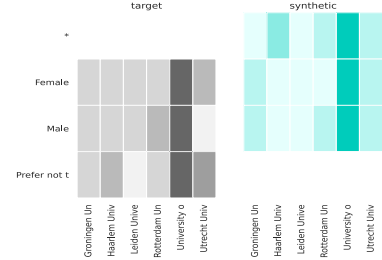
Gender ~ Other STDs



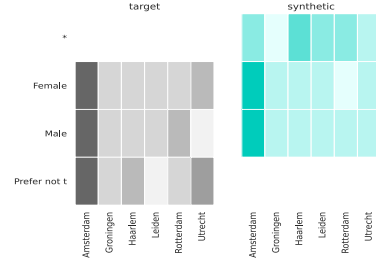
Gender ~ Name



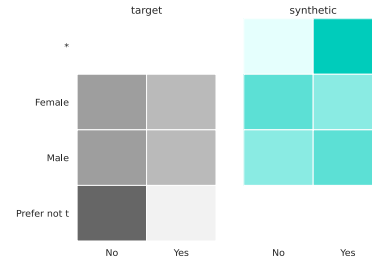
Gender ~ University



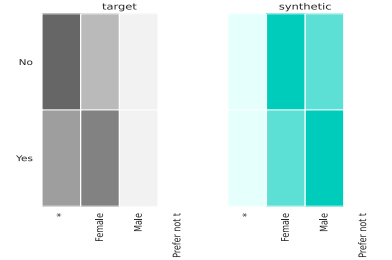
Gender ~ Location/City



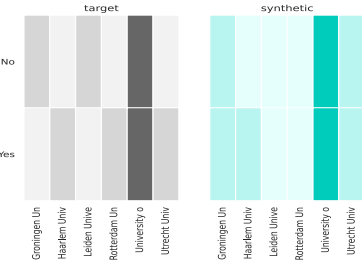
Gender ~ Financial aid



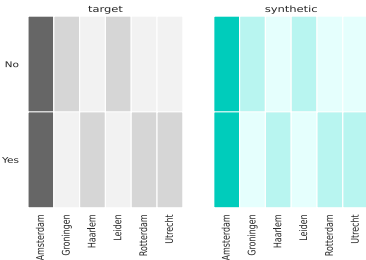
HIV Positive ~ Gender



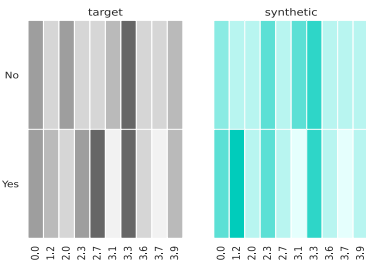
HIV Positive ~ University



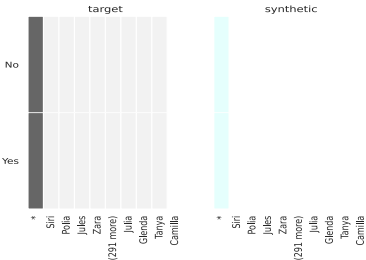
HIV Positive ~ Location/City



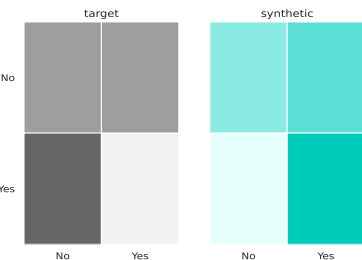
HIV Positive ~ GPA



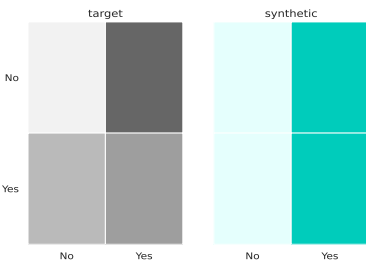
HIV Positive ~ Name



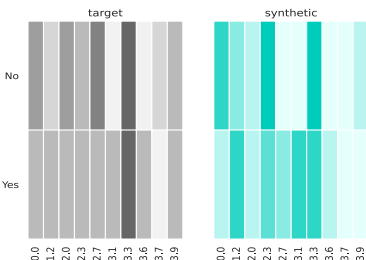
HIV Positive ~ Financial aid



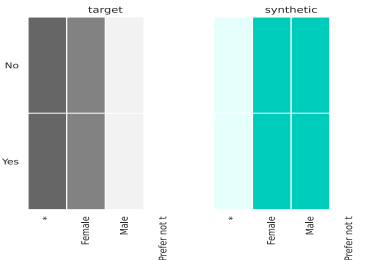
HIV Positive ~ Other STDs



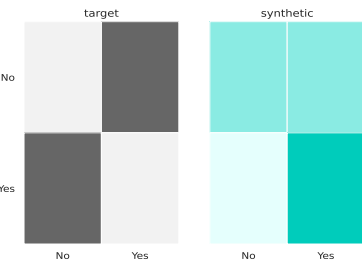
Other STDs ~ GPA



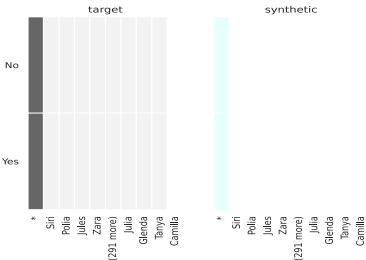
Other STDs ~ Gender



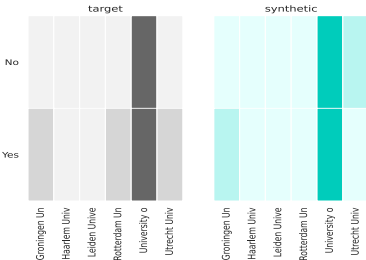
Other STDs ~ Financial aid



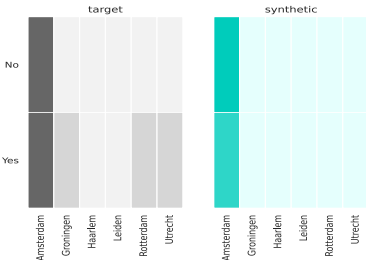
Other STDs ~ Name



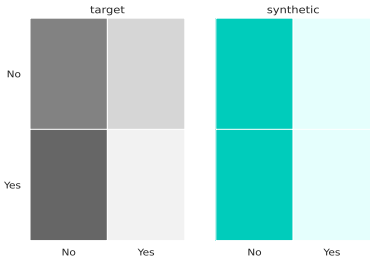
Other STDs ~ University



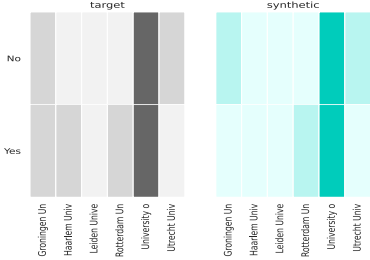
Other STDs ~ Location/City



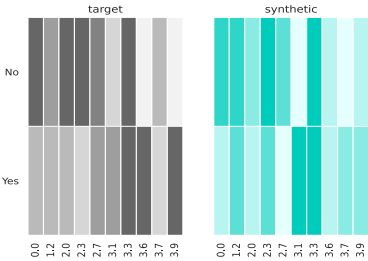
Other STDS ~ HIV Positive



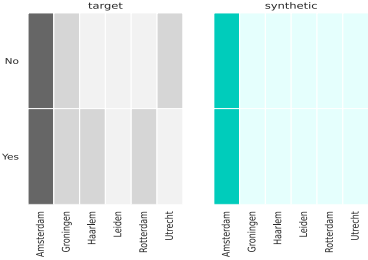
Financial aid ~ University



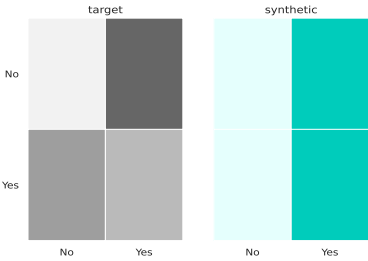
Financial aid ~ GPA



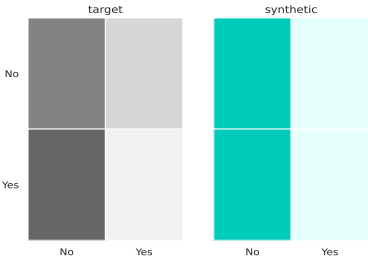
Financial aid ~ Location/City



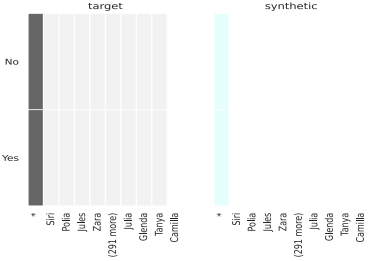
Financial aid ~ Other STDS



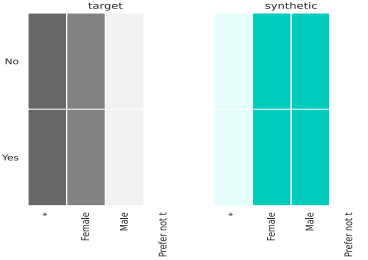
Financial aid ~ HIV Positive



Financial aid ~ Name



Financial aid ~ Gender



# Accuracy



**Accuracy 'target ~ synthetic': 93.22% [40.33%, 100.0%]**

Accuracy is defined as '100% - (Maximum Deviation in Probability)'

Name	88.9	57.0	57.0	84.3	46.3	40.3	42.3	46.3
Location/City	57.0	90.4	63.3	90.4	88.3	95.0	96.0	96.2
University	57.0	63.3	96.8	90.7	89.7	97.4	93.2	94.2
GPA	84.3	90.4	90.7	98.2	87.6	92.5	93.4	98.2
Gender	46.3	88.3	89.7	87.6	94.0	93.3	94.4	95.0
HIV Positive	40.3	95.0	97.4	92.5	93.3	98.3	96.8	97.7
Other STDS	42.3	96.0	95.2	93.4	94.4	96.8	100.0	93.7
Financial aid	46.3	96.2	94.2	90.2	95.0	97.7	93.7	97.7
	Name	Location/City	University	GPA	Gender	HIV Positive	Other STDS	Financial aid

# MOSTLY·AI

© <https://mostly.ai>