NIST CRC Meta Report
smartnoise-synth

Report created on: January 31, 2024 19:54:31

Created with SDNIST v2.3.0

Motivation

SmartNoise Synthesizers is an open source, user-friendly python library for differentially private synthetic data; it's connected to the OpenDP framework.

SmartNoise organizes its synthesizers into two categories:
First, there are approaches that make specific statistical queries to the data (typically marginal queries), add noise to those query results, and then generate new data that aligns with the noisy statistics. In our archive, we have three examples from this category: mewm, mst and pac-synth.
Second, there are neural network approaches, which inject noise into each iteration of the machine learning training/optimization loop. In our archive we have one example from this category: patectgan.

You can find other synthesizers in the SmartNoise library here: [smartnoise-synth] And you can find the OpenDP platform itself here: [opendp]

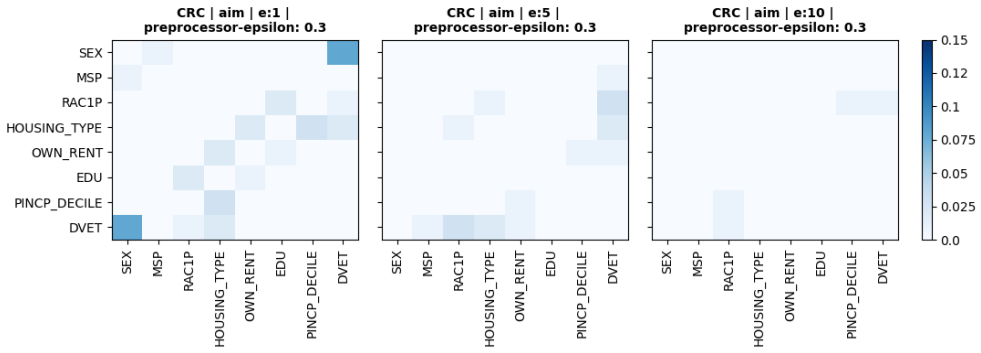
Comparisons

Correlation Comparison:

The Pearson Correlation difference was a popular utility metric during the HLG-MOS Synthetic Data Test Drive. Note that darker highlighting indicates pairs of features whose correlations were not well preserved by the deidentified data.

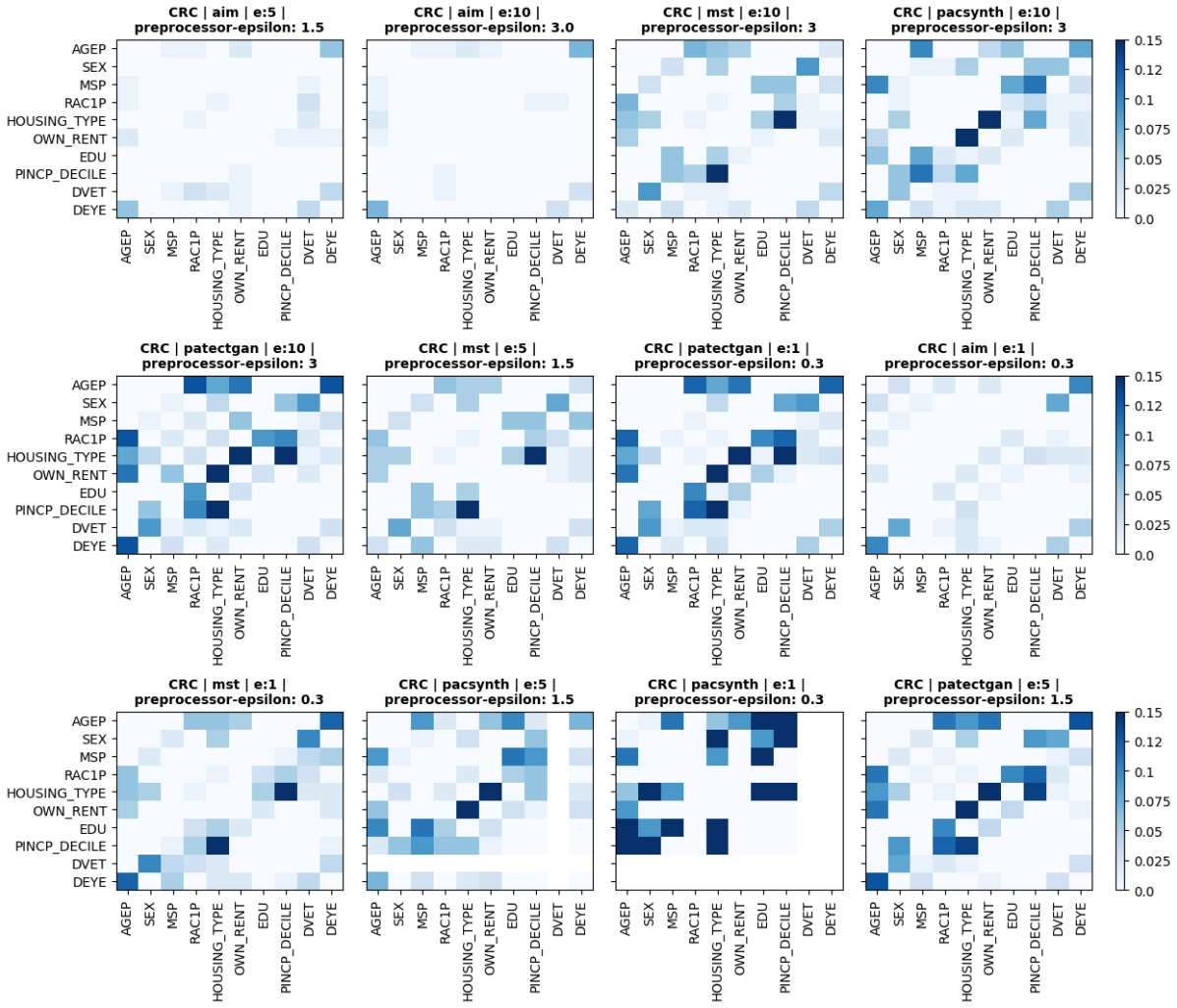
Feature Set: demographic-focused-except-AGEP-DEYE | Target Dataset: national2019:

Features: ['DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,135,134



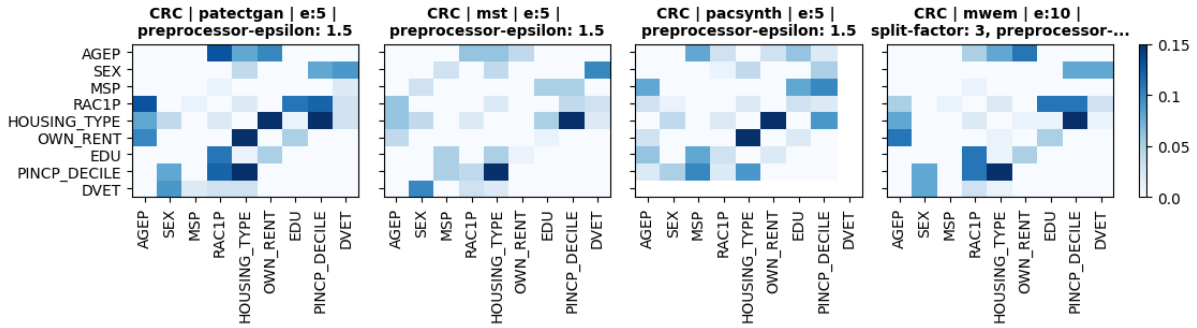
Feature Set: demographic-focused | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



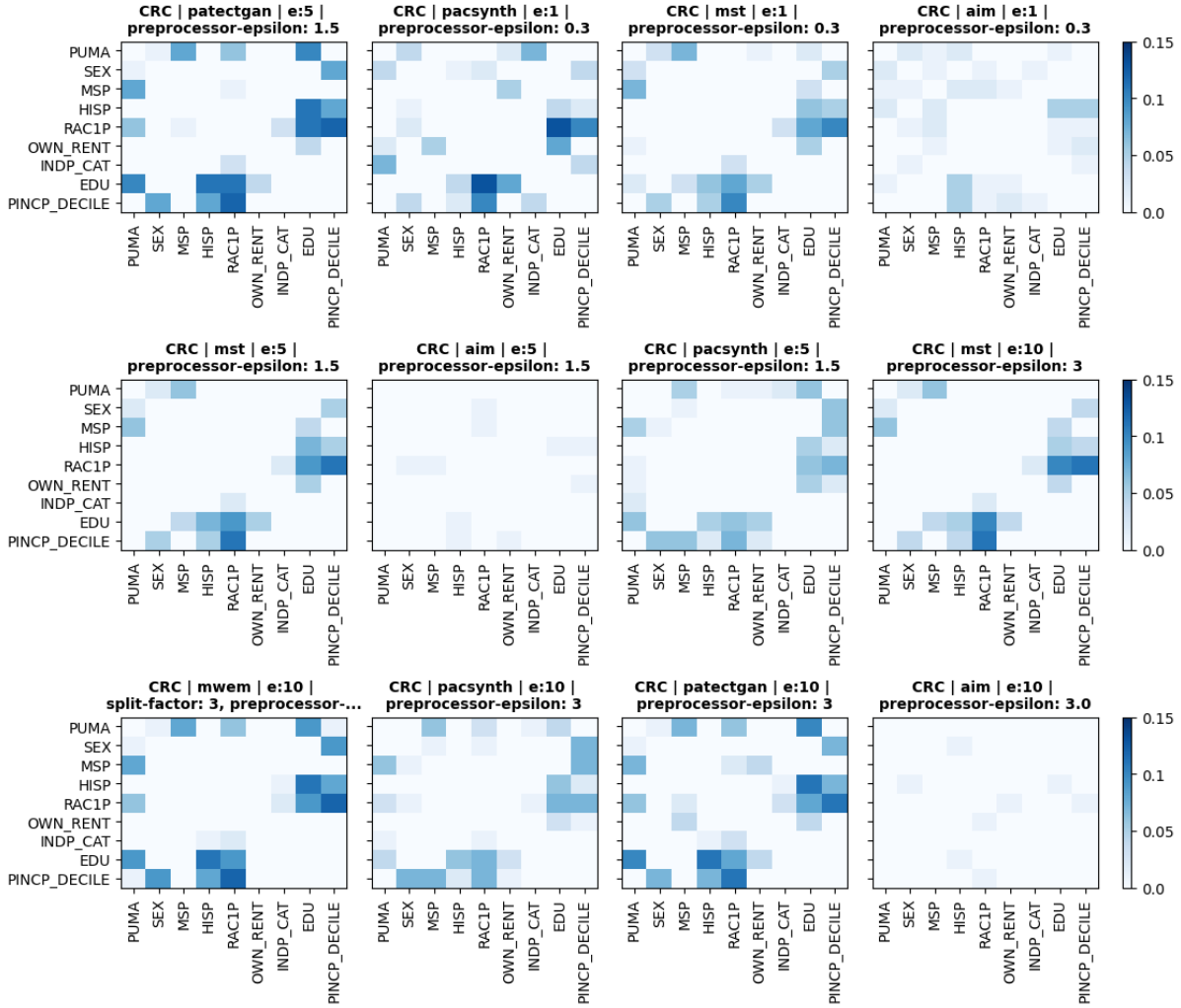
Feature Set: demographic-focused-except-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



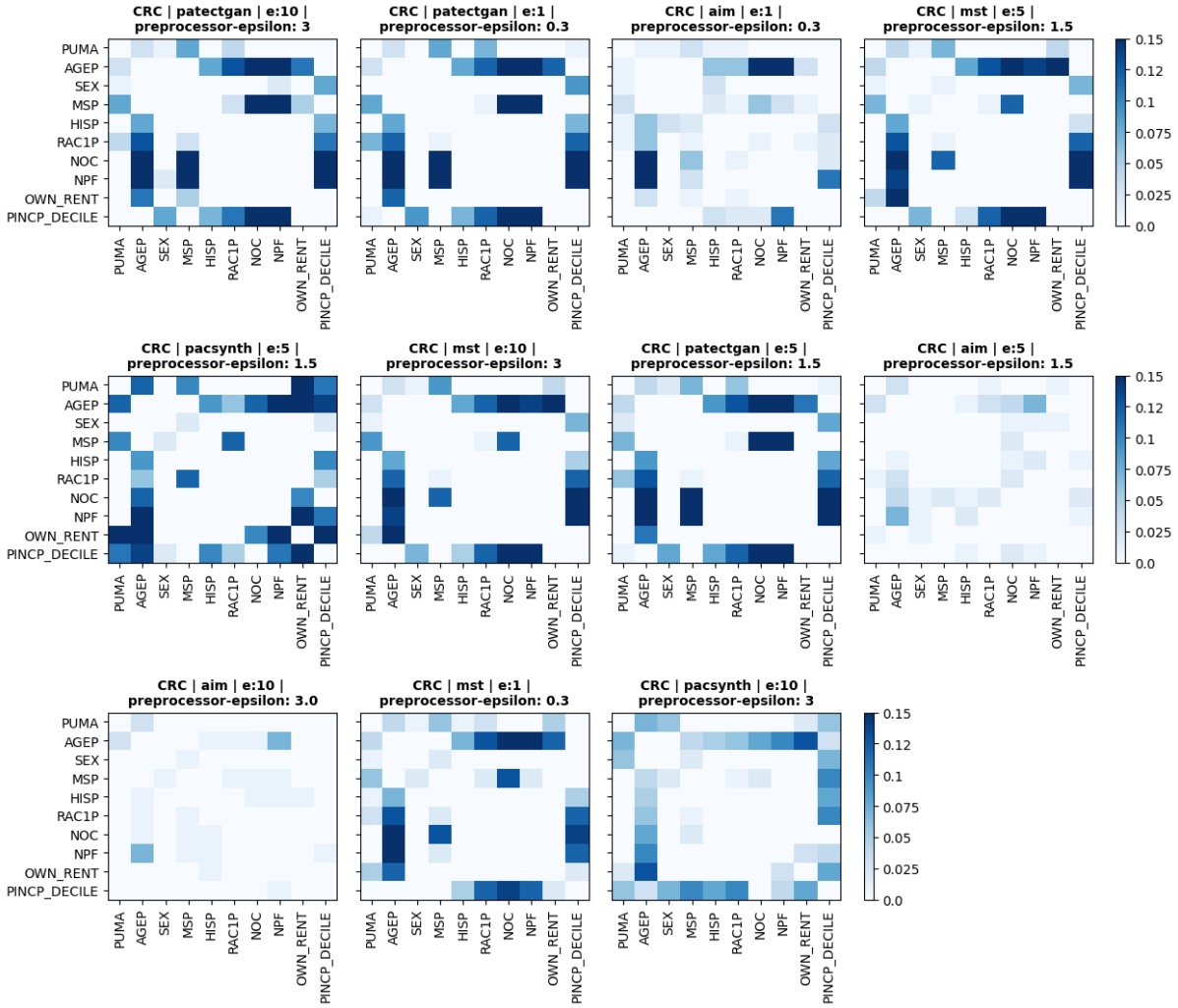
Feature Set: industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



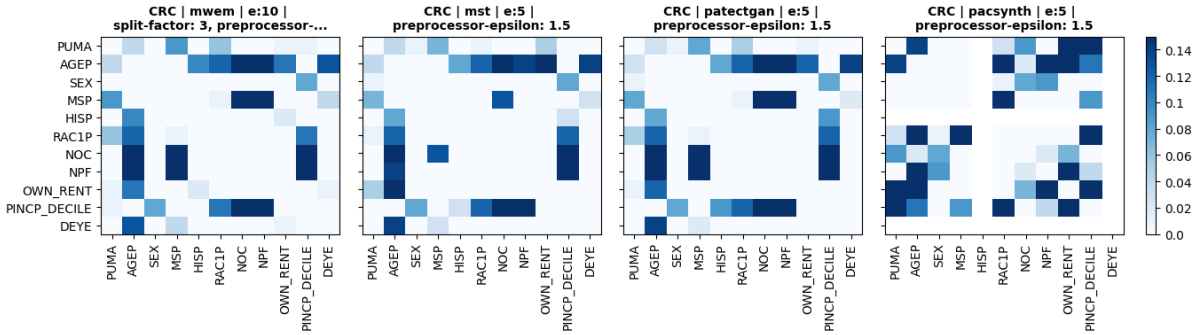
Feature Set: family-focused | Target Dataset: national2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



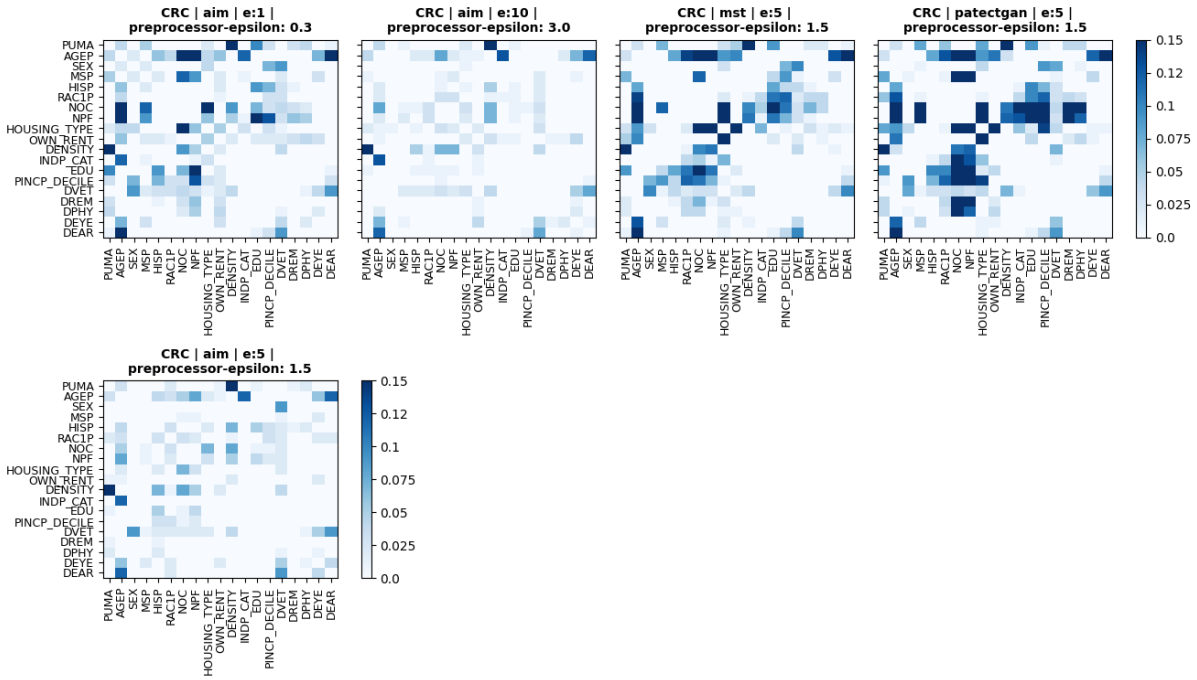
Feature Set: family-focused-with-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



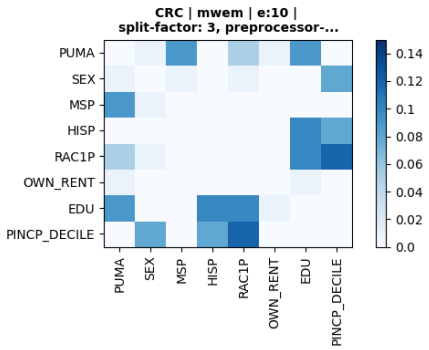
Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



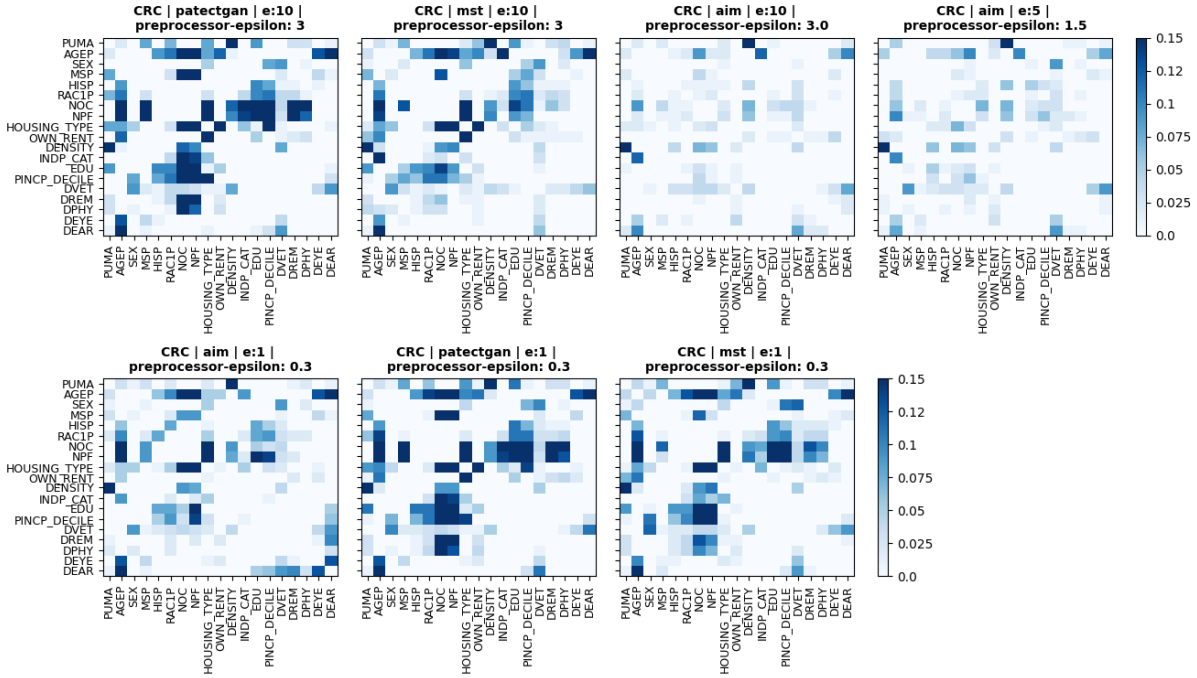
Feature Set: detailed-industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,427,025,600



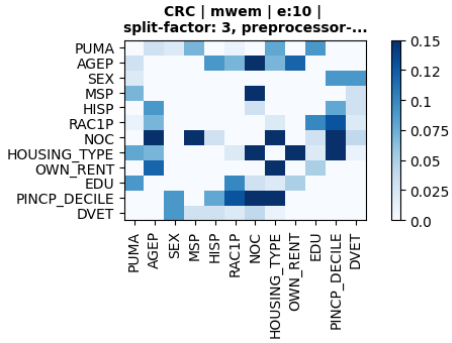
Feature Set: simple-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



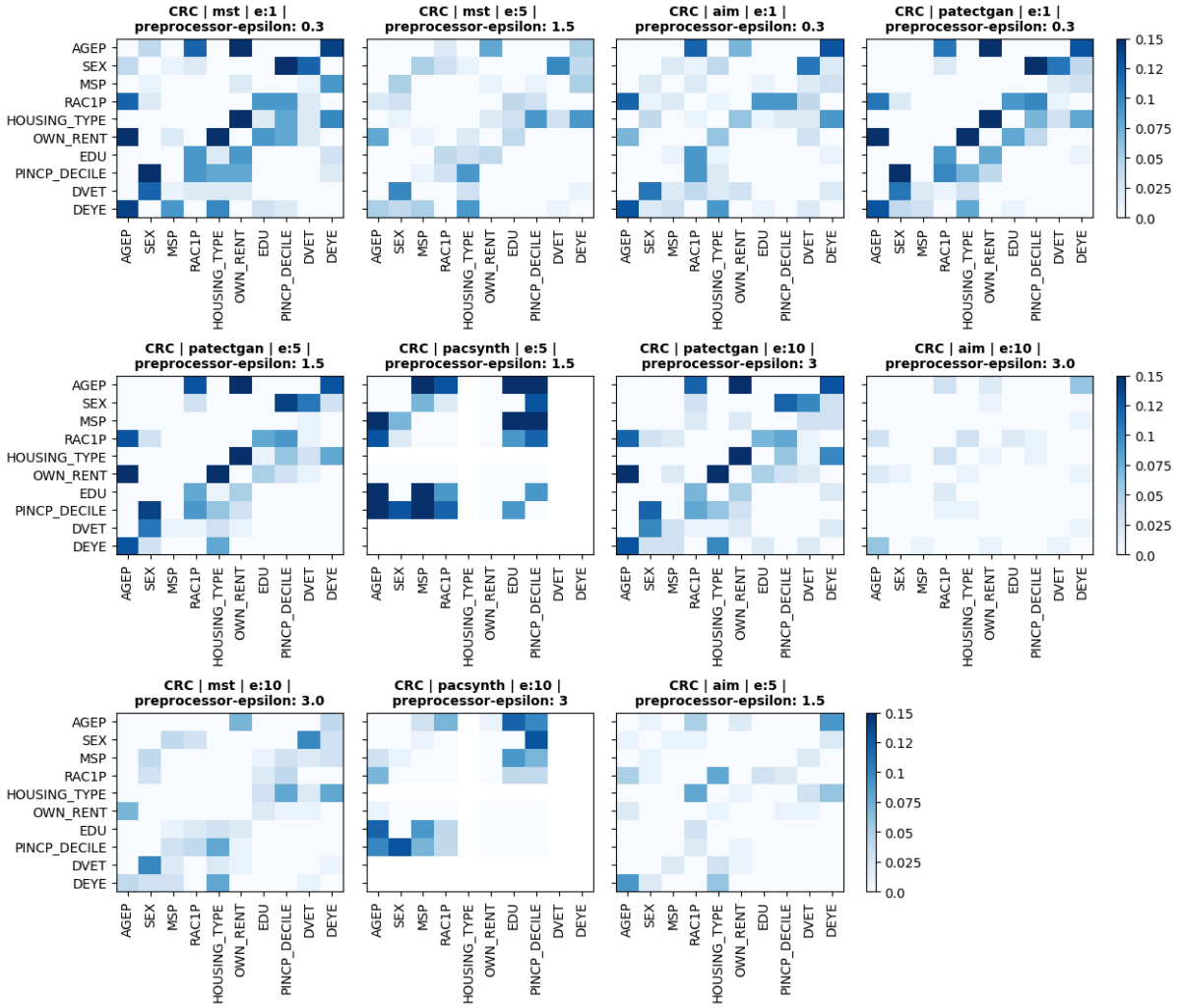
Feature Set: custom-features-12 | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'MSP', 'NOC', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400,000



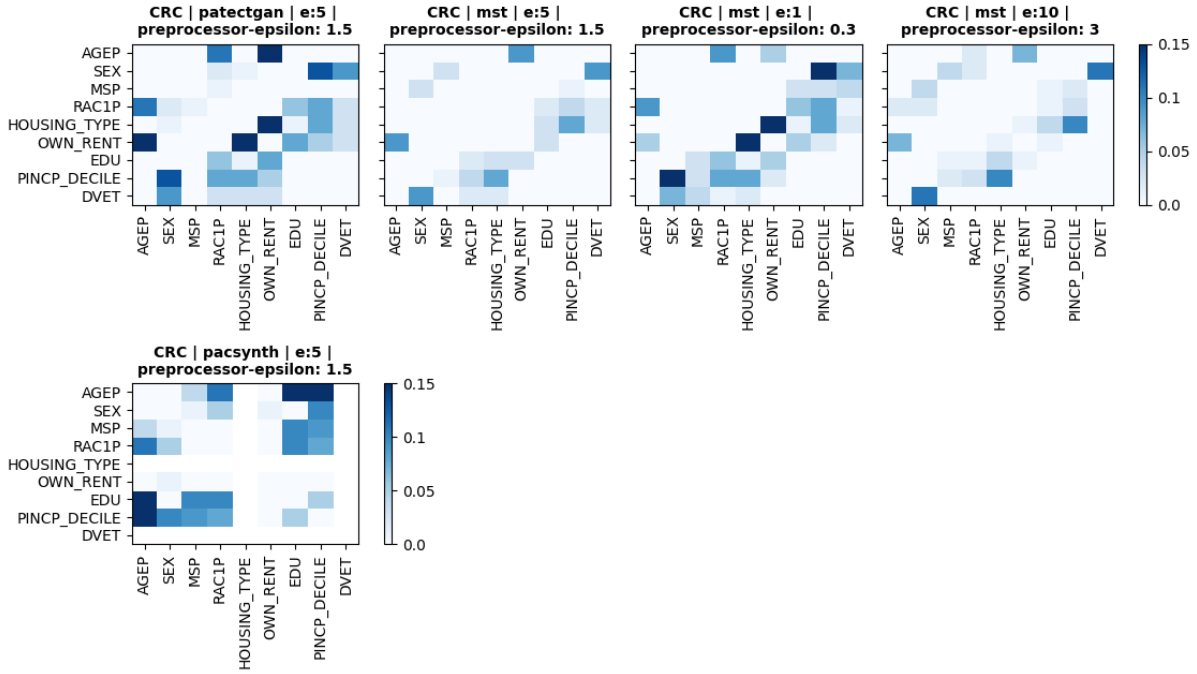
Feature Set: demographic-focused | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



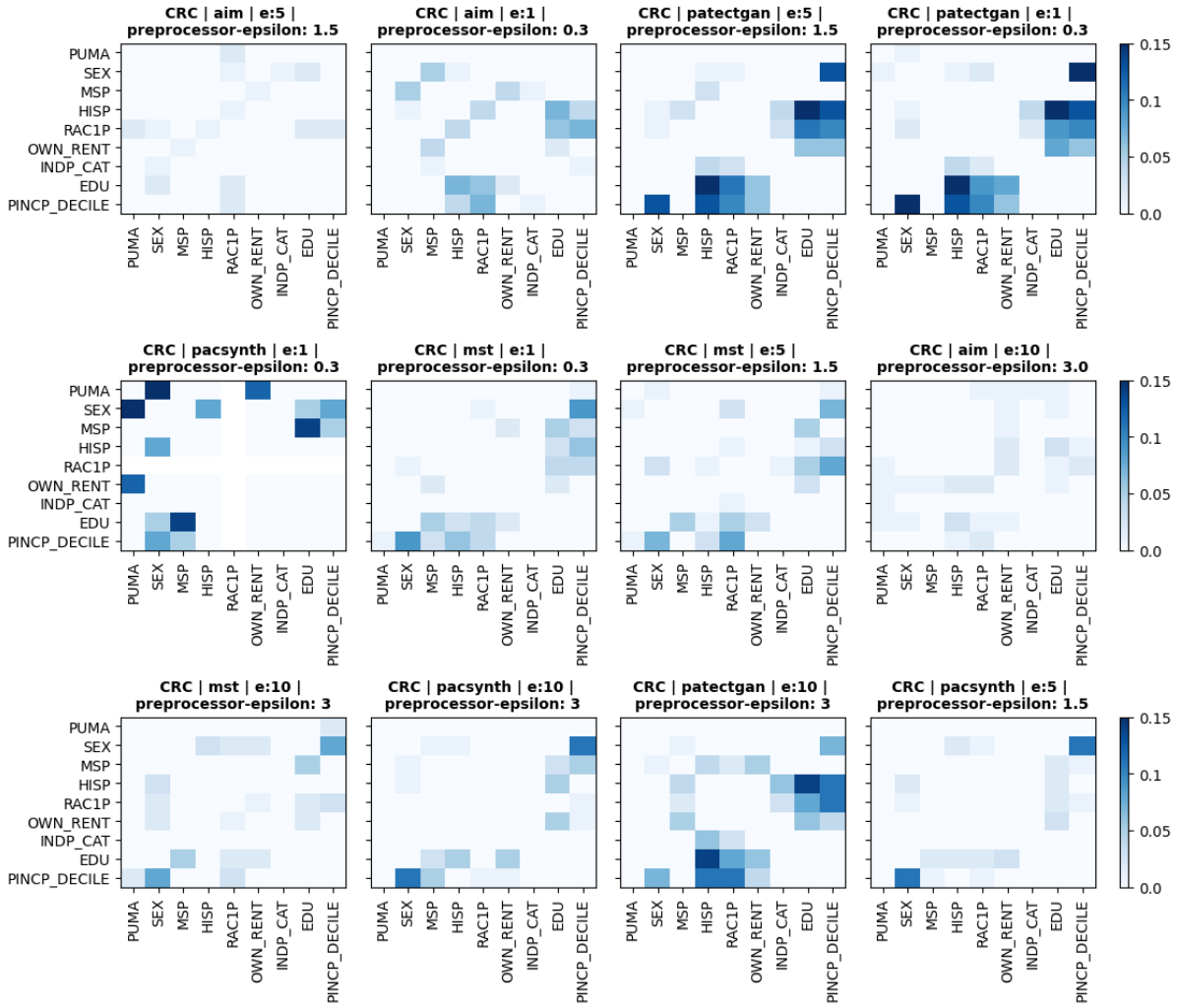
Feature Set: demographic-focused-except-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



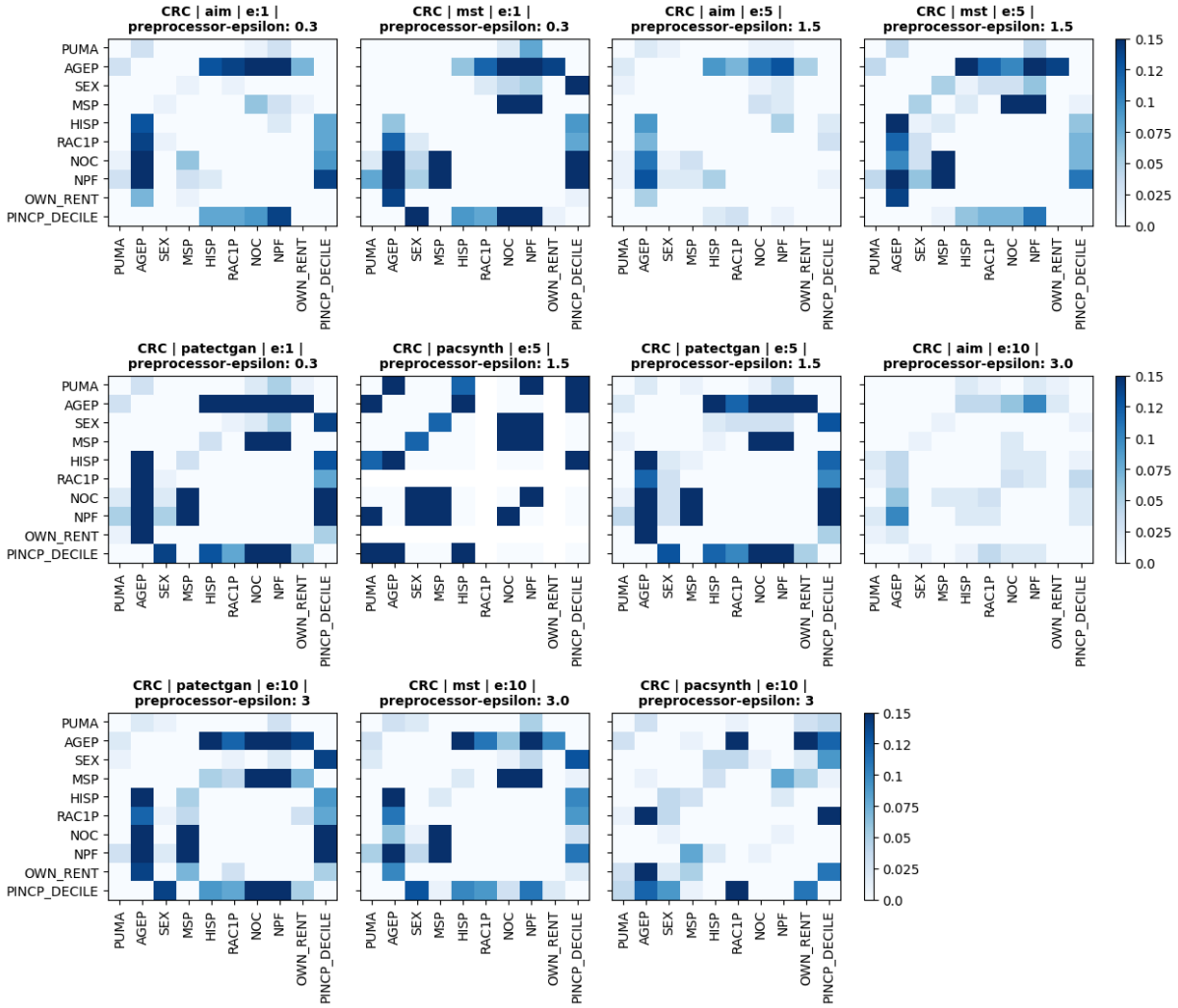
Feature Set: industry-focused | Target Dataset: tx2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



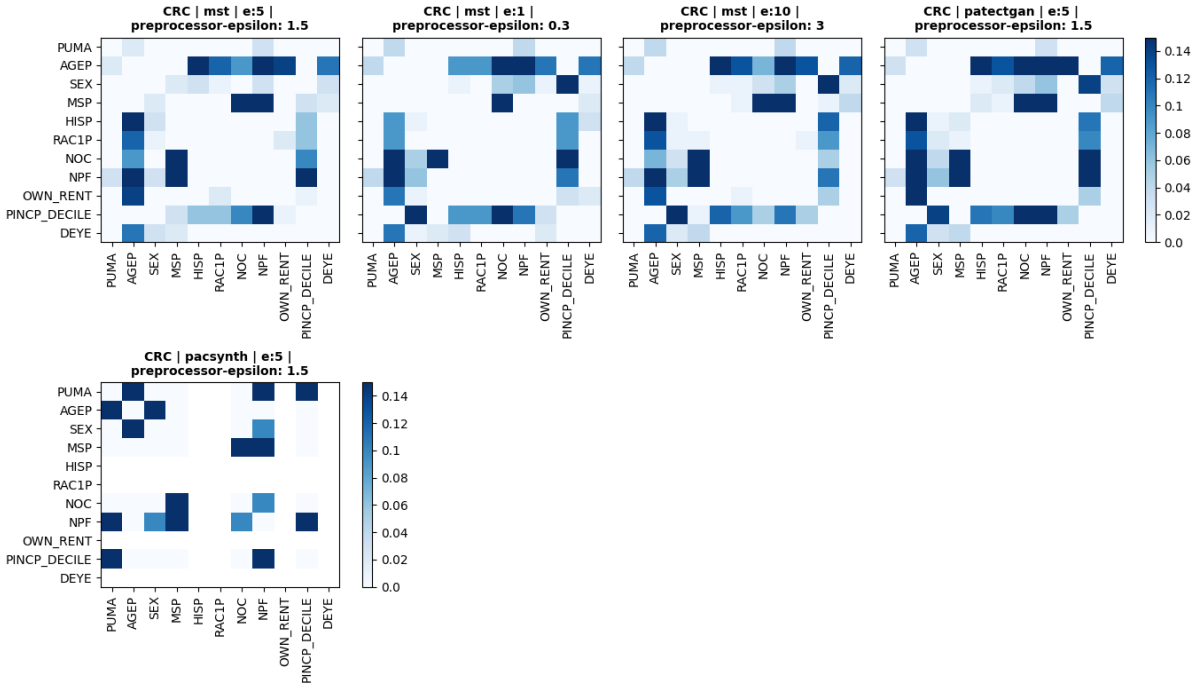
Feature Set: family-focused | Target Dataset: tx2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



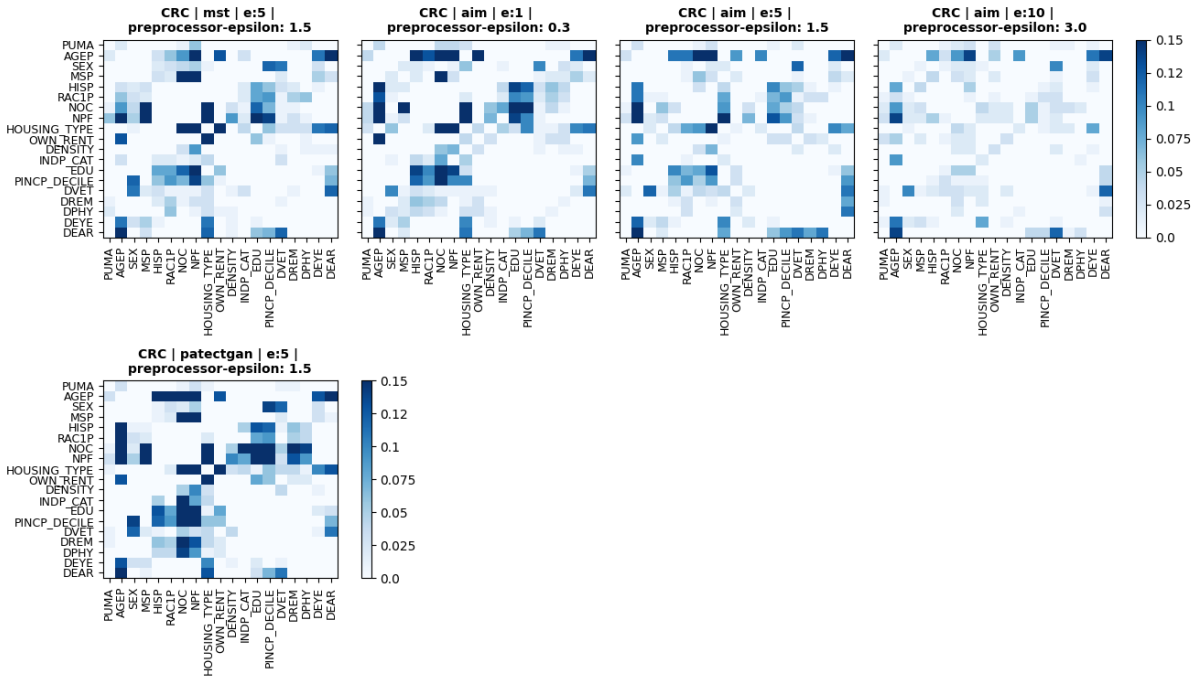
Feature Set: family-focused-with-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



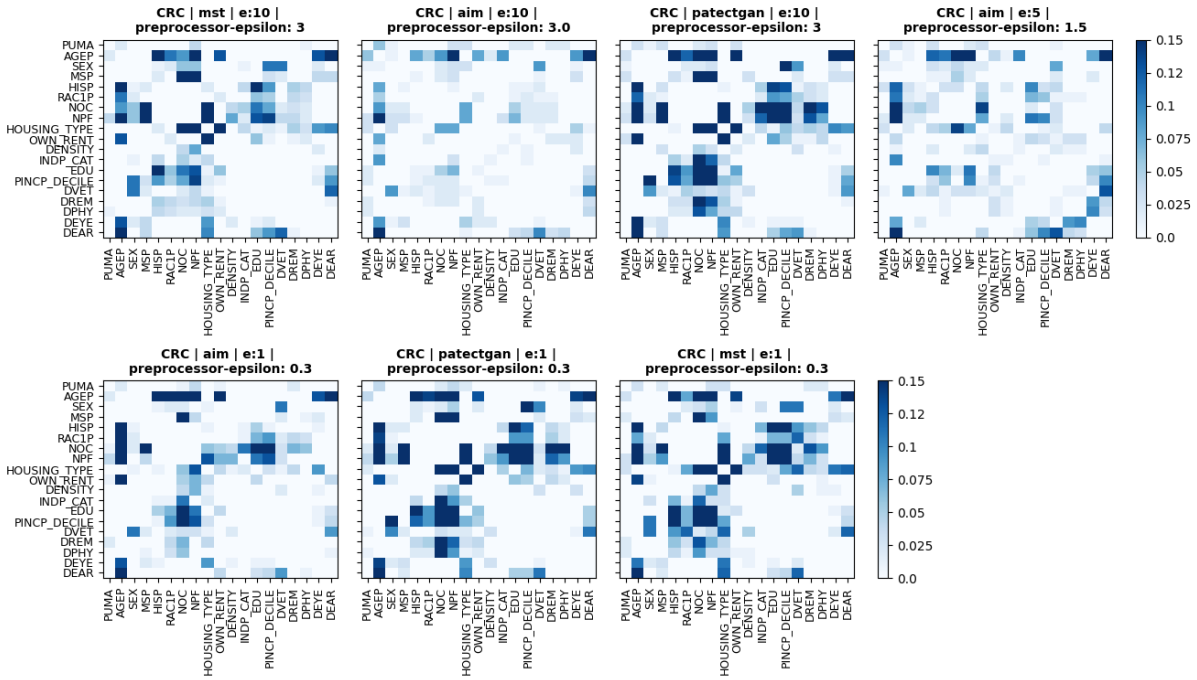
Feature Set: all-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



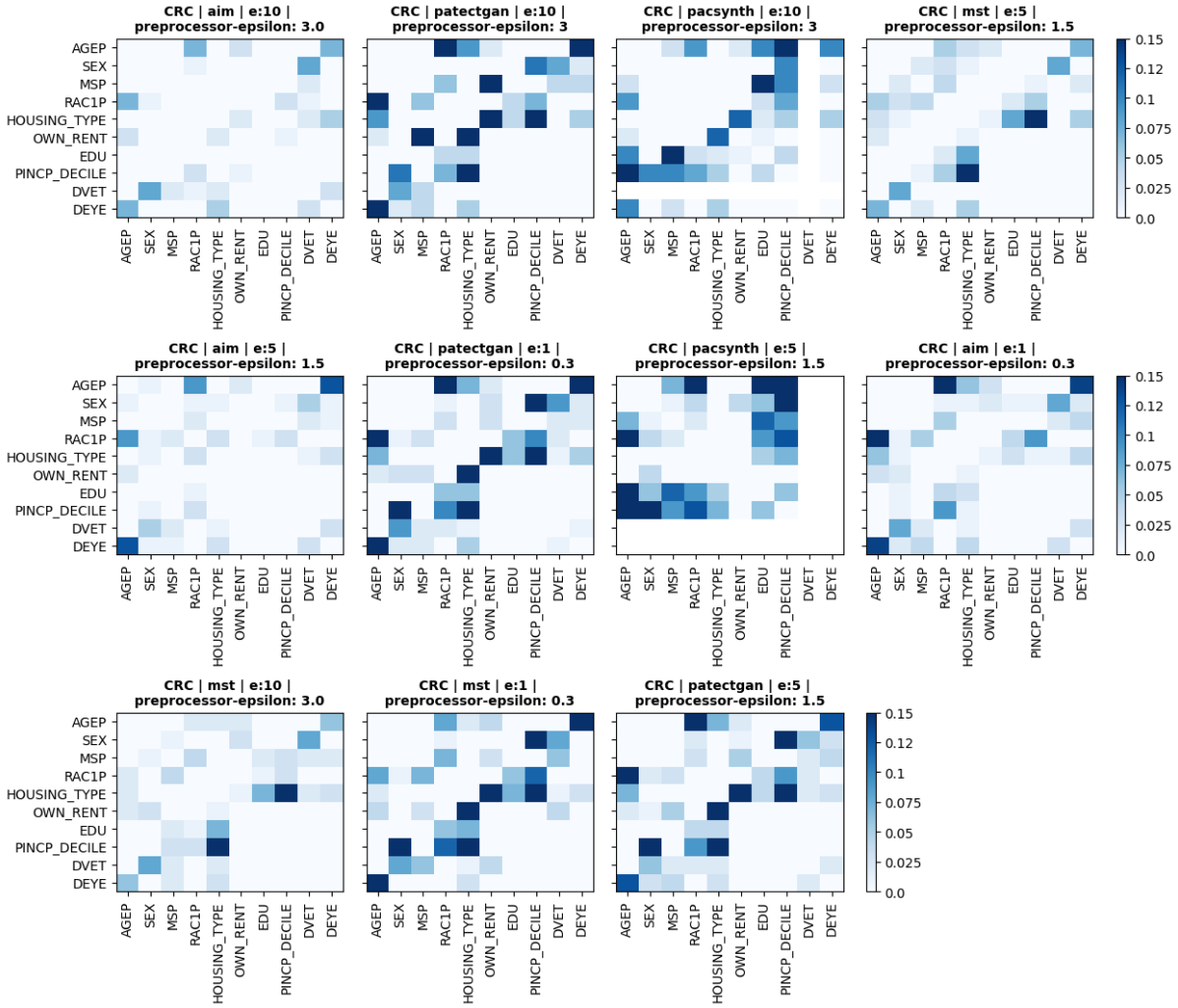
Feature Set: simple-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



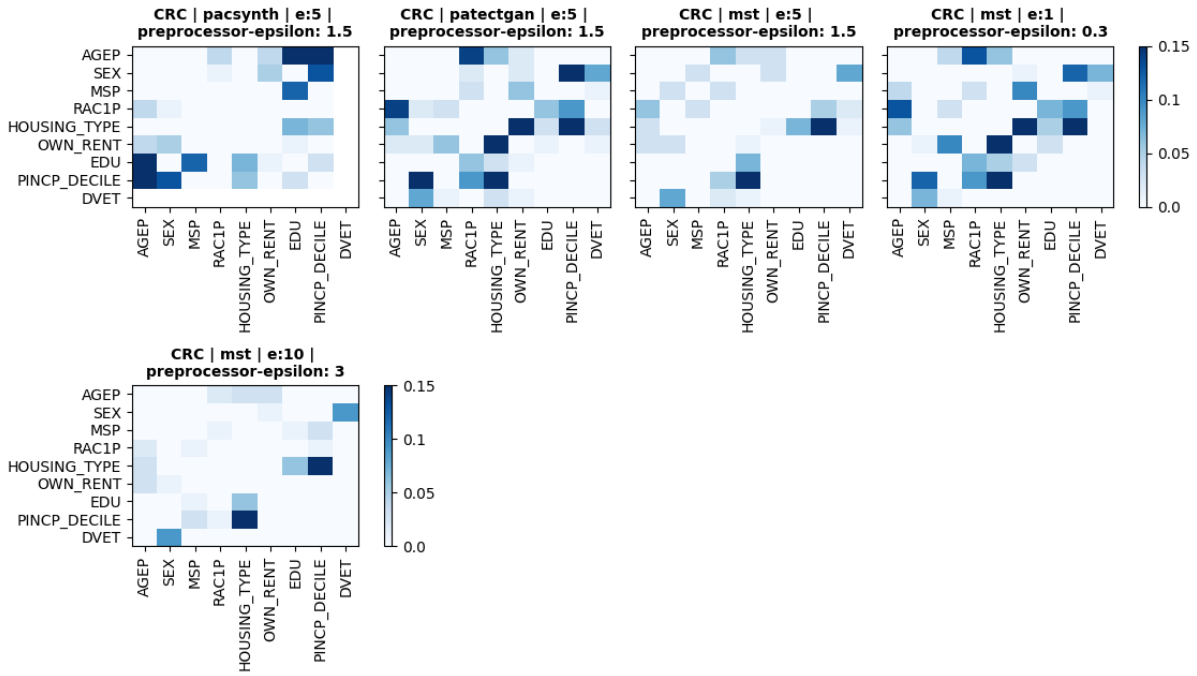
Feature Set: demographic-focused | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



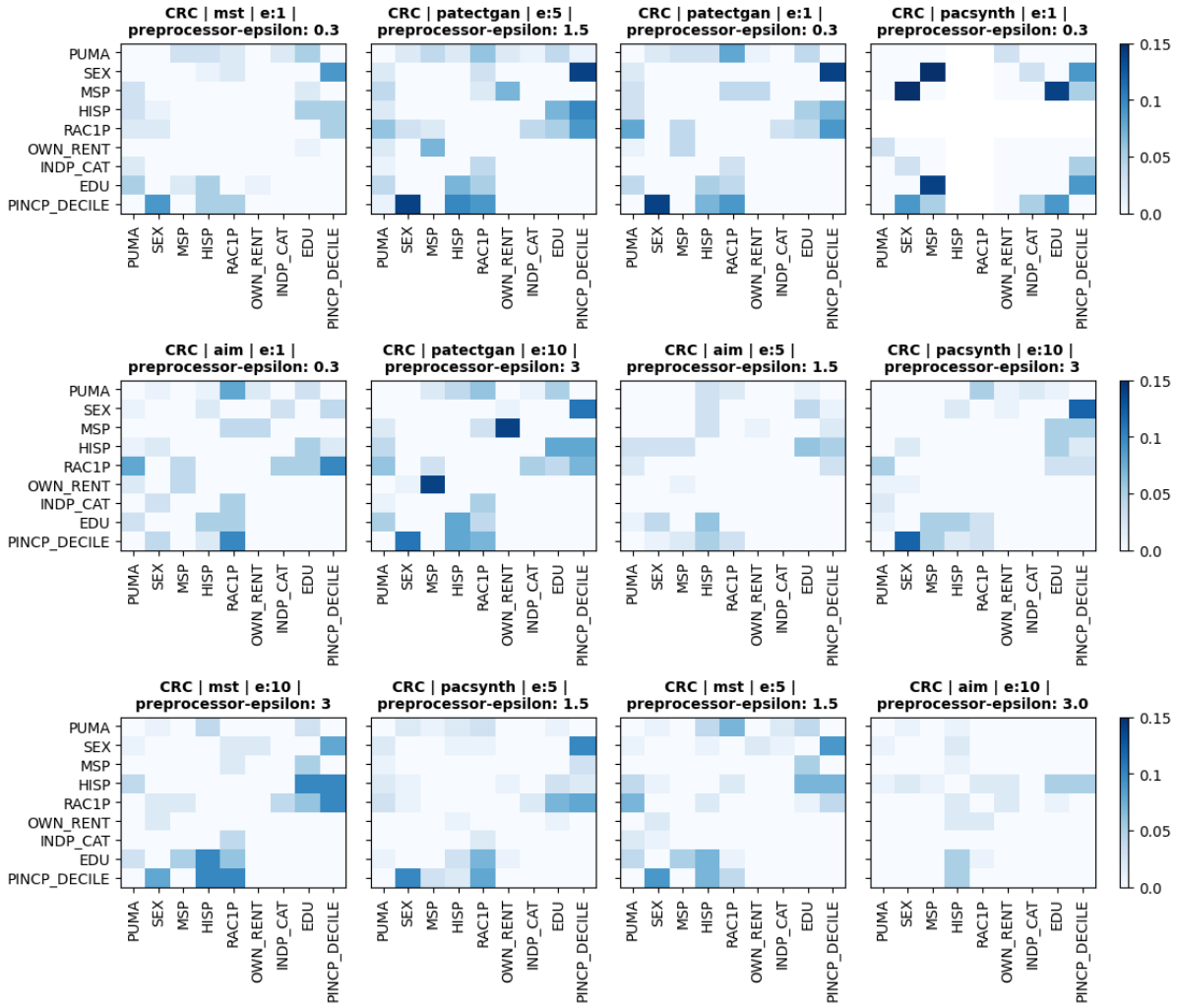
Feature Set: demographic-focused-except-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



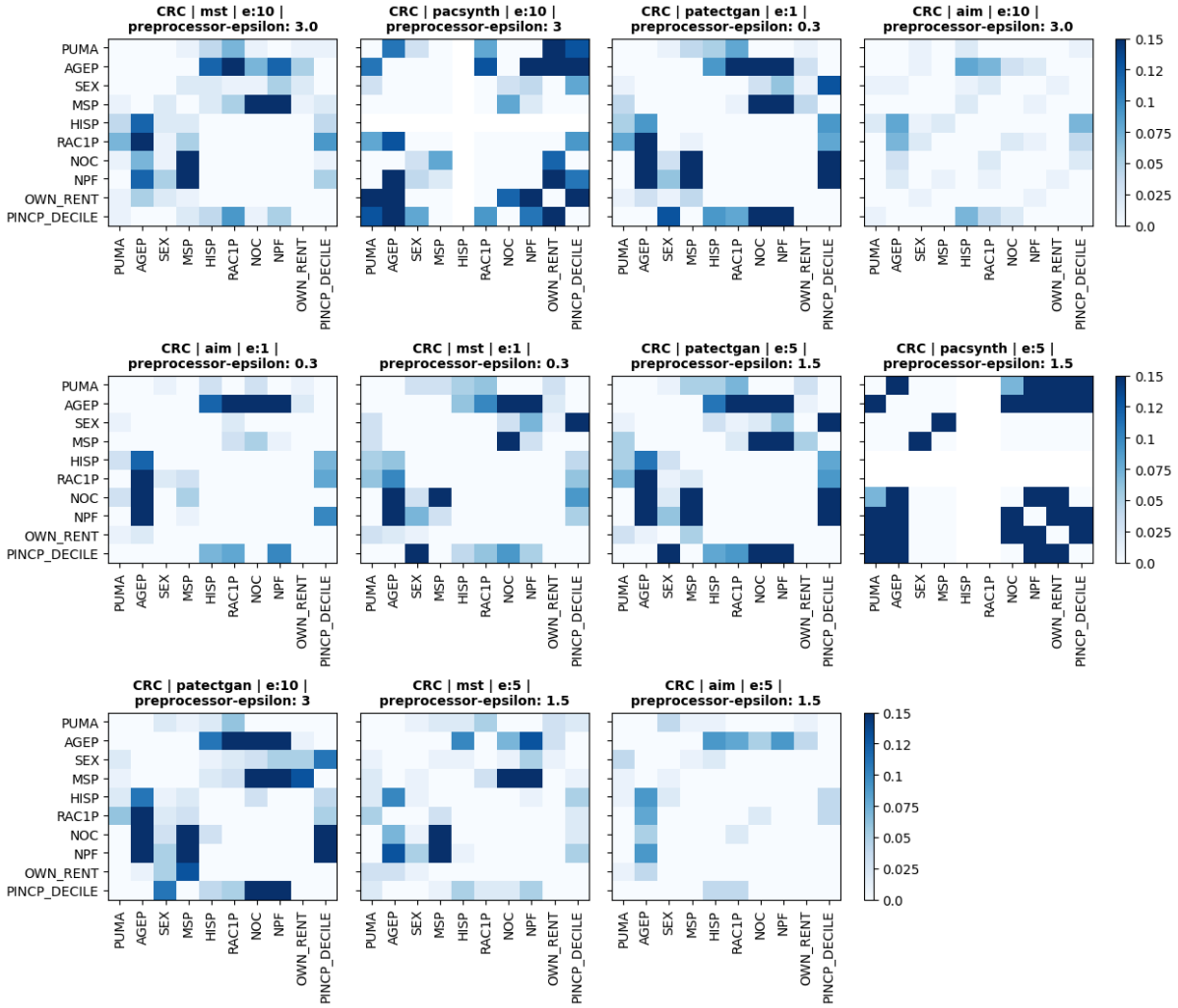
Feature Set: industry-focused | Target Dataset: ma2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



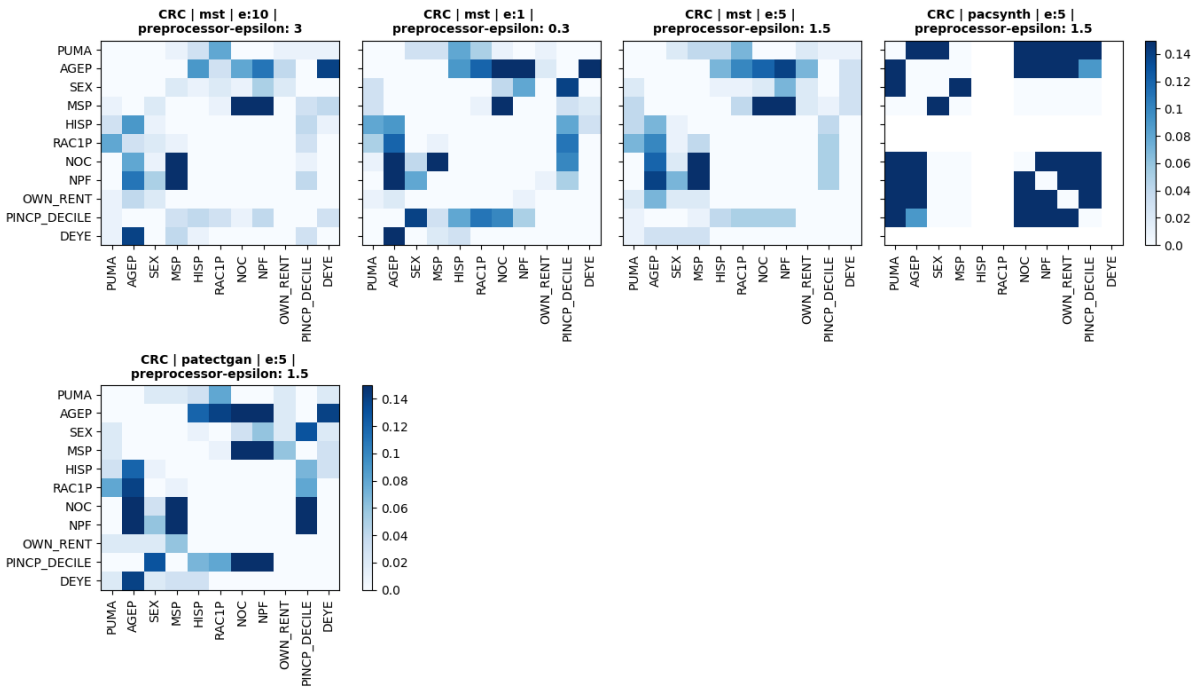
Feature Set: family-focused | Target Dataset: ma2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



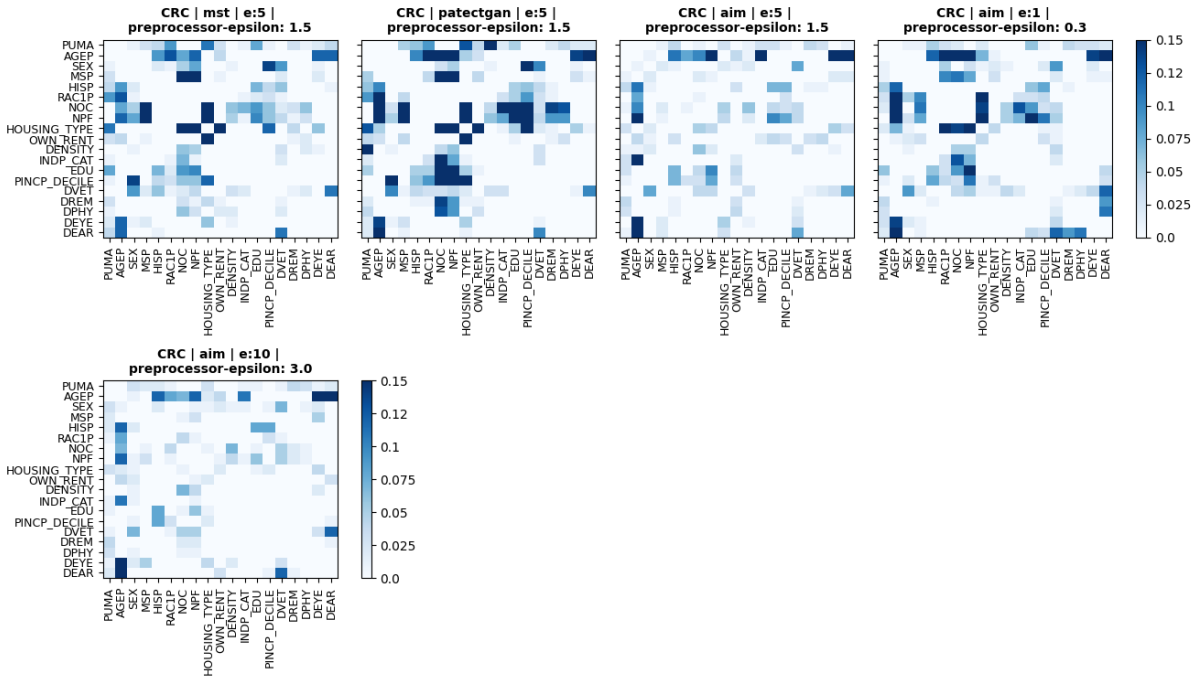
Feature Set: family-focused-with-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



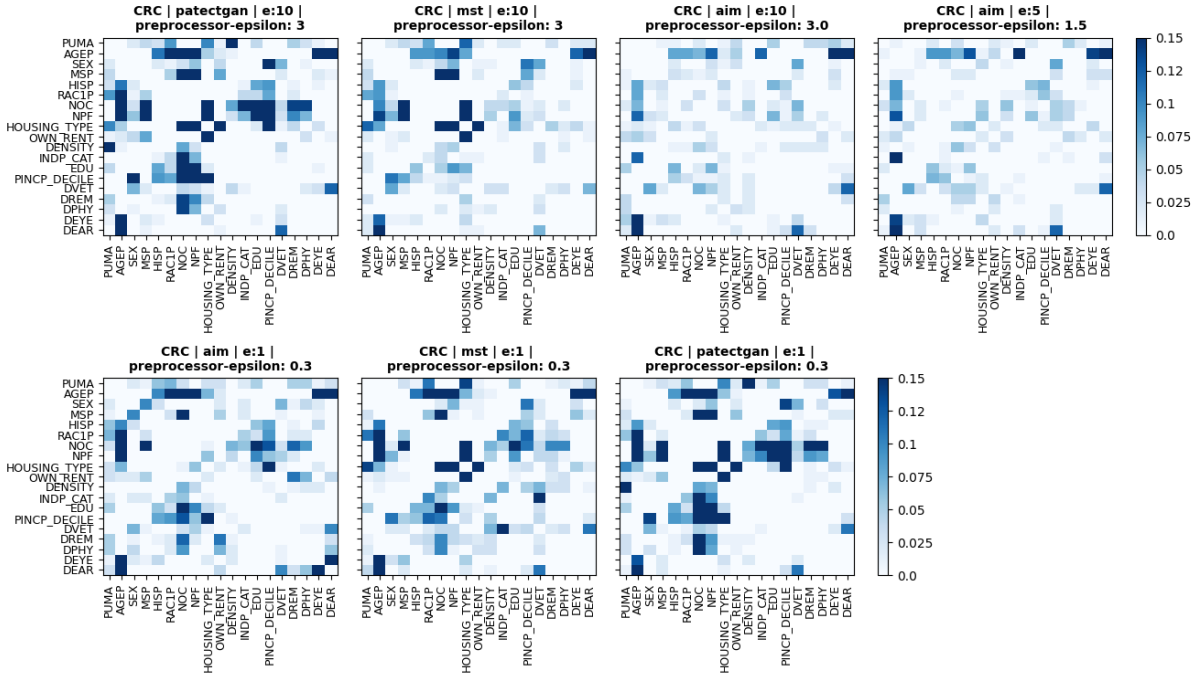
Feature Set: all-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



Feature Set: simple-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



Unique Exact Matches Comparison:

This is a count of unique records in the target data that were exactly reproduced in the deidentified data. Because these records were unique outliers in the target data, and they still appear unchanged in the deidentified data, they are potentially vulnerable to reidentification.

Feature Set: demographic-focused-except-AGEP-DEYE | Target Dataset: national2019:

Features: ['DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,135,134

Number of Unique Records in Target Data: 2773 (10.18%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:1 | preprocessor-epsilon: 0.3	959	34.58
CRC | aim | e:5 | preprocessor-epsilon: 0.3	1176	42.41
CRC | aim | e:10 | preprocessor-epsilon: 0.3	1214	43.78

Feature Set: demographic-focused | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800

Number of Unique Records in Target Data: 14918 (54.74%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:5 | preprocessor-epsilon: 1.5	2736	18.34
CRC | aim | e:10 | preprocessor-epsilon: 3.0	2789	18.7
CRC | mst | e:10 | preprocessor-epsilon: 3	2027	13.59
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	2013	13.49
CRC | patectgan | e:10 | preprocessor-epsilon: 3	1935	12.97
CRC | mst | e:5 | preprocessor-epsilon: 1.5	2030	13.61
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	5	0.03
CRC | aim | e:1 | preprocessor-epsilon: 0.3	2105	14.11
CRC | mst | e:1 | preprocessor-epsilon: 0.3	1860	12.47
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	1117	7.49
CRC | pacsynth | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	1158	7.76

Feature Set: demographic-focused-except-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400

Number of Unique Records in Target Data: 14708 (53.97%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	1295	8.8
CRC | mst | e:5 | preprocessor-epsilon: 1.5	2069	14.07
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	1521	10.34
CRC | mwem | e:10 | split-factor: 3, preprocessor-...	1517	10.31

Feature Set: industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000

Number of Unique Records in Target Data: 16132 (59.19%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	519	3.22
CRC | pacsynth | e:1 | preprocessor-epsilon: 0.3	165	1.02
CRC | mst | e:1 | preprocessor-epsilon: 0.3	1455	9.02
CRC | aim | e:1 | preprocessor-epsilon: 0.3	1886	11.69
CRC | mst | e:5 | preprocessor-epsilon: 1.5	1561	9.68
CRC | aim | e:5 | preprocessor-epsilon: 1.5	2422	15.01
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	1574	9.76
CRC | mst | e:10 | preprocessor-epsilon: 3	1534	9.51
CRC | mwem | e:10 | split-factor: 3, preprocessor-...	671	4.16
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	1579	9.79
CRC | patectgan | e:10 | preprocessor-epsilon: 3	595	3.69
CRC | aim | e:10 | preprocessor-epsilon: 3.0	2413	14.96

Feature Set: family-focused | Target Dataset: national2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000

Number of Unique Records in Target Data: 23908 (87.73%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:10 | preprocessor-epsilon: 3	16	0.07
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | aim | e:1 | preprocessor-epsilon: 0.3	366	1.53
CRC | mst | e:5 | preprocessor-epsilon: 1.5	243	1.02
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	91	0.38
CRC | mst | e:10 | preprocessor-epsilon: 3	261	1.09
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	2	0.01
CRC | aim | e:5 | preprocessor-epsilon: 1.5	648	2.71
CRC | aim | e:10 | preprocessor-epsilon: 3.0	694	2.9
CRC | mst | e:1 | preprocessor-epsilon: 0.3	332	1.39
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	208	0.87

Feature Set: family-focused-with-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000

Number of Unique Records in Target Data: 23945 (87.86%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mwem | e:10 | split-factor: 3, preprocessor-...	33	0.14
CRC | mst | e:5 | preprocessor-epsilon: 1.5	222	0.93
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	1	0.0
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	40	0.17

Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000

Number of Unique Records in Target Data: 27159 (99.66%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | aim | e:10 | preprocessor-epsilon: 3.0	0	0.0
CRC | mst | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | aim | e:5 | preprocessor-epsilon: 1.5	0	0.0

Feature Set: detailed-industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,427,025,600

Number of Unique Records in Target Data: 17944 (65.84%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mwem | e:10 | split-factor: 3, preprocessor-...	483	2.69

Feature Set: simple-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000

Number of Unique Records in Target Data: 25762 (94.53%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:10 | preprocessor-epsilon: 3	0	0.0
CRC | mst | e:10 | preprocessor-epsilon: 3	7	0.03
CRC | aim | e:10 | preprocessor-epsilon: 3.0	11	0.04
CRC | aim | e:5 | preprocessor-epsilon: 1.5	12	0.05
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | mst | e:1 | preprocessor-epsilon: 0.3	6	0.02

Feature Set: custom-features-12 | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'MSP', 'NOC', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400,000

Number of Unique Records in Target Data: 23169 (85.01%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mwem | e:10 | split-factor: 3, preprocessor-...	70	0.3

Feature Set: demographic-focused | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800

Number of Unique Records in Target Data: 5268 (56.79%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:1 | preprocessor-epsilon: 0.3	573	10.88
CRC | mst | e:5 | preprocessor-epsilon: 1.5	741	14.07
CRC | aim | e:1 | preprocessor-epsilon: 0.3	755	14.33
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	645	12.24
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	60	1.14
CRC | patectgan | e:10 | preprocessor-epsilon: 3	622	11.81
CRC | aim | e:10 | preprocessor-epsilon: 3.0	1066	20.24
CRC | mst | e:10 | preprocessor-epsilon: 3.0	815	15.47
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	359	6.81
CRC | aim | e:5 | preprocessor-epsilon: 1.5	893	16.95

Feature Set: demographic-focused-except-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400

Number of Unique Records in Target Data: 5205 (56.11%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	577	11.09
CRC | mst | e:5 | preprocessor-epsilon: 1.5	798	15.33
CRC | mst | e:1 | preprocessor-epsilon: 0.3	566	10.87
CRC | mst | e:10 | preprocessor-epsilon: 3	855	16.43
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	121	2.32

Feature Set: industry-focused | Target Dataset: tx2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000

Number of Unique Records in Target Data: 5504 (59.34%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:5 | preprocessor-epsilon: 1.5	838	15.23
CRC | aim | e:1 | preprocessor-epsilon: 0.3	698	12.68
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	441	8.01
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	5	0.09
CRC | pacsynth | e:1 | preprocessor-epsilon: 0.3	11	0.2
CRC | mst | e:1 | preprocessor-epsilon: 0.3	626	11.37
CRC | mst | e:5 | preprocessor-epsilon: 1.5	641	11.65
CRC | aim | e:10 | preprocessor-epsilon: 3.0	885	16.08
CRC | mst | e:10 | preprocessor-epsilon: 3	664	12.06
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	1002	18.2
CRC | patectgan | e:10 | preprocessor-epsilon: 3	530	9.63
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	792	14.39

Feature Set: family-focused | Target Dataset: tx2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000

Number of Unique Records in Target Data: 8601 (92.72%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:1 | preprocessor-epsilon: 0.3	95	1.1
CRC | mst | e:1 | preprocessor-epsilon: 0.3	85	0.99
CRC | aim | e:5 | preprocessor-epsilon: 1.5	173	2.01
CRC | mst | e:5 | preprocessor-epsilon: 1.5	176	2.05
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	5	0.06
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	3	0.03
CRC | aim | e:10 | preprocessor-epsilon: 3.0	248	2.88
CRC | patectgan | e:10 | preprocessor-epsilon: 3	96	1.12
CRC | mst | e:10 | preprocessor-epsilon: 3.0	137	1.59
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	81	0.94

Feature Set: family-focused-with-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000

Number of Unique Records in Target Data: 8615 (92.87%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:5 | preprocessor-epsilon: 1.5	131	1.52
CRC | mst | e:1 | preprocessor-epsilon: 0.3	79	0.92
CRC | mst | e:10 | preprocessor-epsilon: 3	138	1.6
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	3	0.03
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	1	0.01

Feature Set: all-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000

Number of Unique Records in Target Data: 9260 (99.83%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | aim | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | aim | e:10 | preprocessor-epsilon: 3.0	0	0.0
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	0	0.0

Feature Set: simple-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000

Number of Unique Records in Target Data: 9075 (97.83%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:10 | preprocessor-epsilon: 3	30	0.33
CRC | aim | e:10 | preprocessor-epsilon: 3.0	6	0.07
CRC | patectgan | e:10 | preprocessor-epsilon: 3	0	0.0
CRC | aim | e:5 | preprocessor-epsilon: 1.5	6	0.07
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | mst | e:1 | preprocessor-epsilon: 0.3	1	0.01

Feature Set: demographic-focused | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800

Number of Unique Records in Target Data: 4268 (55.91%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | aim | e:10 | preprocessor-epsilon: 3.0	856	20.06
CRC | patectgan | e:10 | preprocessor-epsilon: 3	559	13.1
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	213	4.99
CRC | mst | e:5 | preprocessor-epsilon: 1.5	621	14.55
CRC | aim | e:5 | preprocessor-epsilon: 1.5	816	19.12
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	27	0.63
CRC | aim | e:1 | preprocessor-epsilon: 0.3	429	10.05
CRC | mst | e:10 | preprocessor-epsilon: 3.0	659	15.44
CRC | mst | e:1 | preprocessor-epsilon: 0.3	342	8.01
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	392	9.18

Feature Set: demographic-focused-except-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400

Number of Unique Records in Target Data: 4219 (55.27%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	45	1.07
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	402	9.53
CRC | mst | e:5 | preprocessor-epsilon: 1.5	598	14.17
CRC | mst | e:1 | preprocessor-epsilon: 0.3	383	9.08
CRC | mst | e:10 | preprocessor-epsilon: 3	659	15.62

Feature Set: industry-focused | Target Dataset: ma2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000

Number of Unique Records in Target Data: 4265 (55.87%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:1 | preprocessor-epsilon: 0.3	595	13.95
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	463	10.86
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	4	0.09
CRC | pacsynth | e:1 | preprocessor-epsilon: 0.3	12	0.28
CRC | aim | e:1 | preprocessor-epsilon: 0.3	593	13.9
CRC | patectgan | e:10 | preprocessor-epsilon: 3	552	12.94
CRC | aim | e:5 | preprocessor-epsilon: 1.5	820	19.23
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	986	23.12
CRC | mst | e:10 | preprocessor-epsilon: 3	618	14.49
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	720	16.88
CRC | mst | e:5 | preprocessor-epsilon: 1.5	595	13.95
CRC | aim | e:10 | preprocessor-epsilon: 3.0	825	19.34

Feature Set: family-focused | Target Dataset: ma2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000

Number of Unique Records in Target Data: 6435 (84.29%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:10 | preprocessor-epsilon: 3.0	284	4.41
CRC | pacsynth | e:10 | preprocessor-epsilon: 3	124	1.93
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | aim | e:10 | preprocessor-epsilon: 3.0	437	6.79
CRC | aim | e:1 | preprocessor-epsilon: 0.3	95	1.48
CRC | mst | e:1 | preprocessor-epsilon: 0.3	134	2.08
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	4	0.06
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	4	0.06
CRC | patectgan | e:10 | preprocessor-epsilon: 3	169	2.63
CRC | mst | e:5 | preprocessor-epsilon: 1.5	238	3.7
CRC | aim | e:5 | preprocessor-epsilon: 1.5	328	5.1

Feature Set: family-focused-with-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000

Number of Unique Records in Target Data: 6456 (84.57%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:10 | preprocessor-epsilon: 3	265	4.1
CRC | mst | e:1 | preprocessor-epsilon: 0.3	117	1.81
CRC | mst | e:5 | preprocessor-epsilon: 1.5	216	3.35
CRC | pacsynth | e:5 | preprocessor-epsilon: 1.5	2	0.03
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	5	0.08

Feature Set: all-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000

Number of Unique Records in Target Data: 7626 (99.9%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | mst | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | patectgan | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | aim | e:5 | preprocessor-epsilon: 1.5	0	0.0
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | aim | e:10 | preprocessor-epsilon: 3.0	0	0.0

Feature Set: simple-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000

Number of Unique Records in Target Data: 7176 (94.0%)

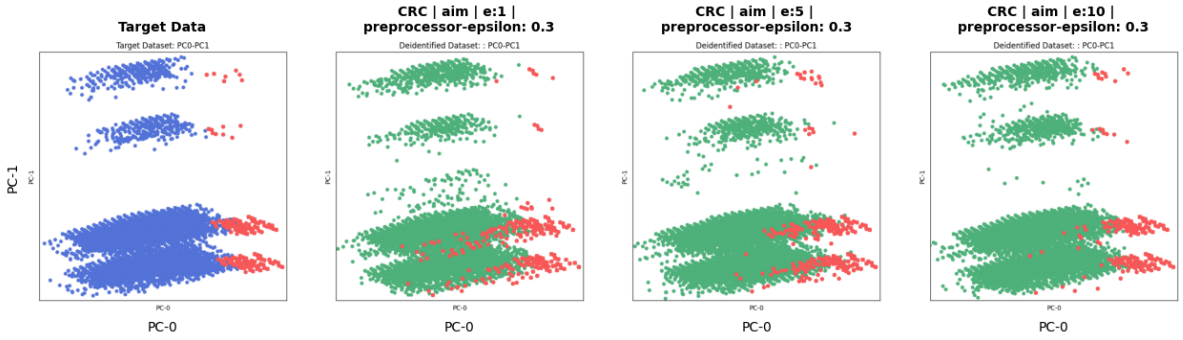
Variant	Records Matched In Target Data	Percent Records Matched In Target Data
CRC | patectgan | e:10 | preprocessor-epsilon: 3	0	0.0
CRC | mst | e:10 | preprocessor-epsilon: 3	31	0.43
CRC | aim | e:10 | preprocessor-epsilon: 3.0	20	0.28
CRC | aim | e:5 | preprocessor-epsilon: 1.5	8	0.11
CRC | aim | e:1 | preprocessor-epsilon: 0.3	0	0.0
CRC | mst | e:1 | preprocessor-epsilon: 0.3	4	0.06
CRC | patectgan | e:1 | preprocessor-epsilon: 0.3	0	0.0

PCA Comparison: (PC-0 & PC-1) with highlighted MSP-N (AGE < 15):

This is another approach for visualizing where the distribution of the deidentified data has shifted away from the target data. In this approach, we begin by using Principle Component Analysis to find a way of representing the target data in a lower dimensional space (in 5 dimensions rather than the full 22 dimensions of the original feature space). Descriptions of these new five dimensions (components) are given in the components table; the components will change depending on which target data set youâ€™re using. Five dimensions are better than 22, but we actually want to get down to two dimensions so we can plot the data on simple (x,y) axesâ€“ the plots below show the data across each possible pair combination of our five components. You can compare how the shapes change between the target data and the deidentified data, and consider what that might mean in light of the component definitions. This is a relatively new visualization metric that was introduced by the IPUMS International team during the HLG-MOS Synthetic Data Test Drive.

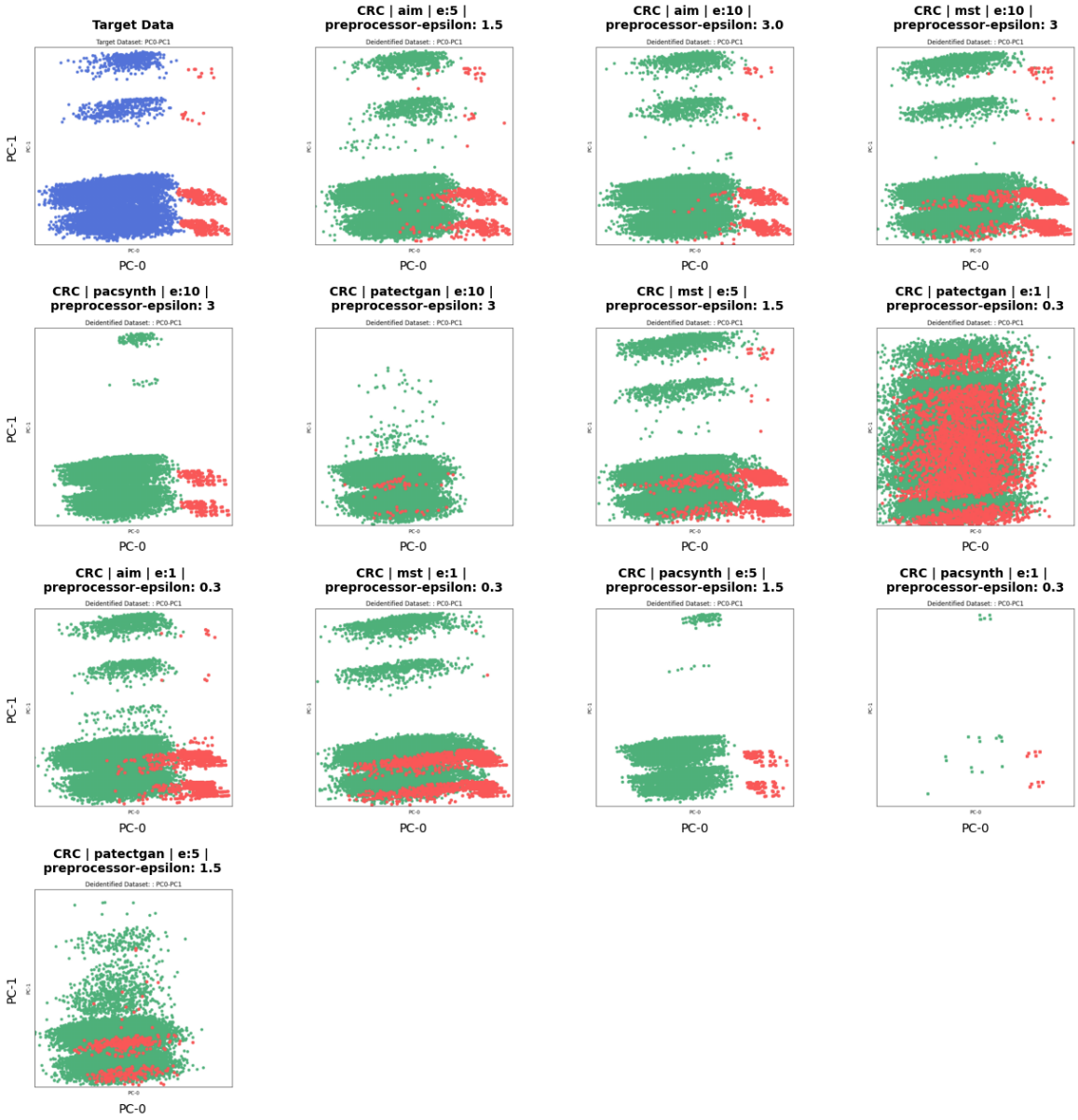
Feature Set: demographic-focused-except-AGEP-DEYE | Target Dataset: national2019:

Features: ['DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,135,134



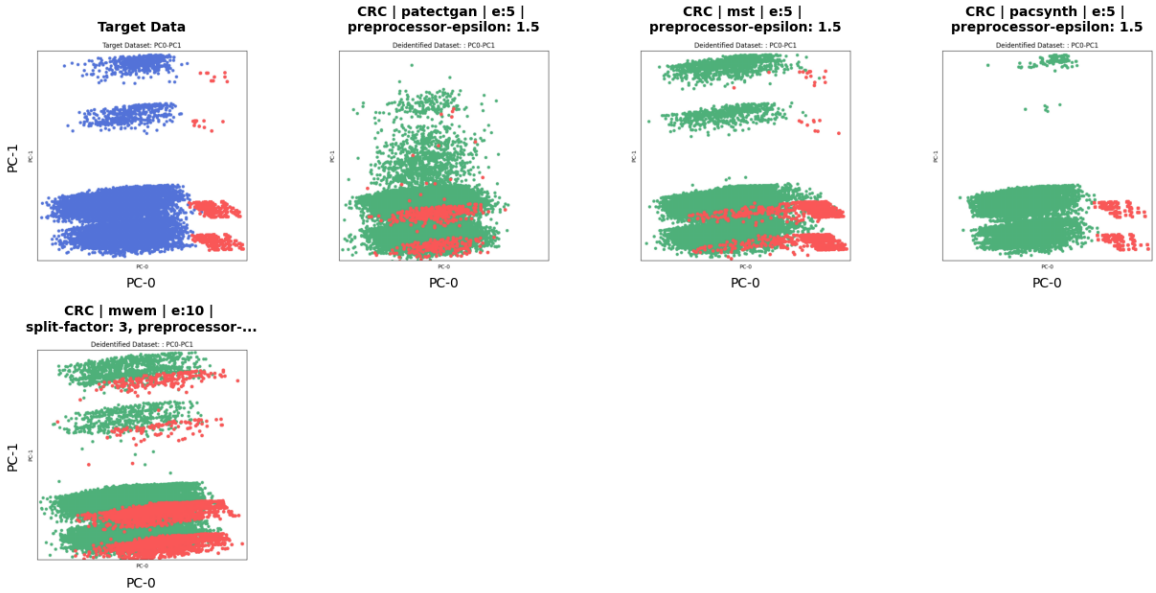
Feature Set: demographic-focused | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



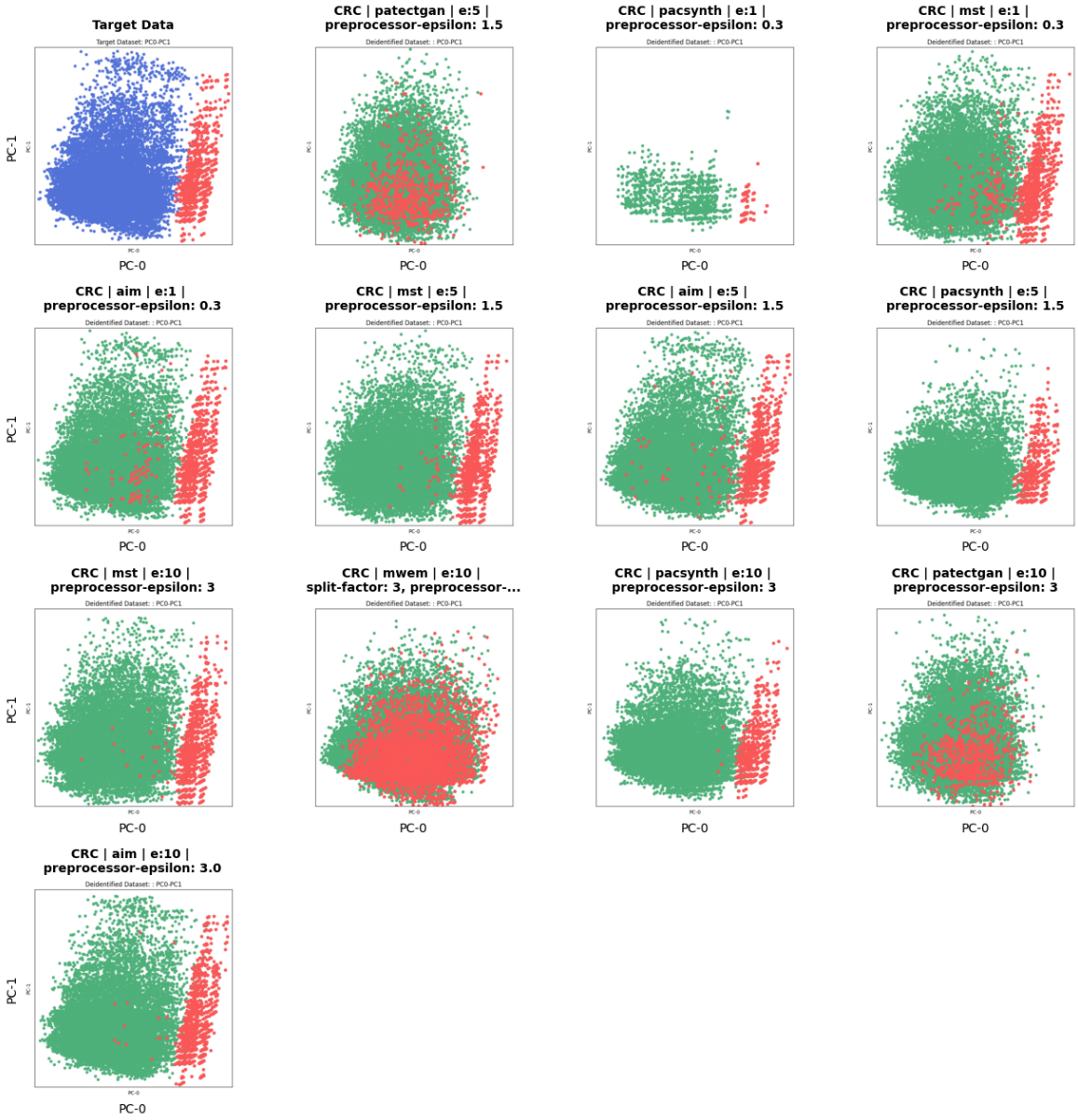
Feature Set: demographic-focused-except-DEYE I Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



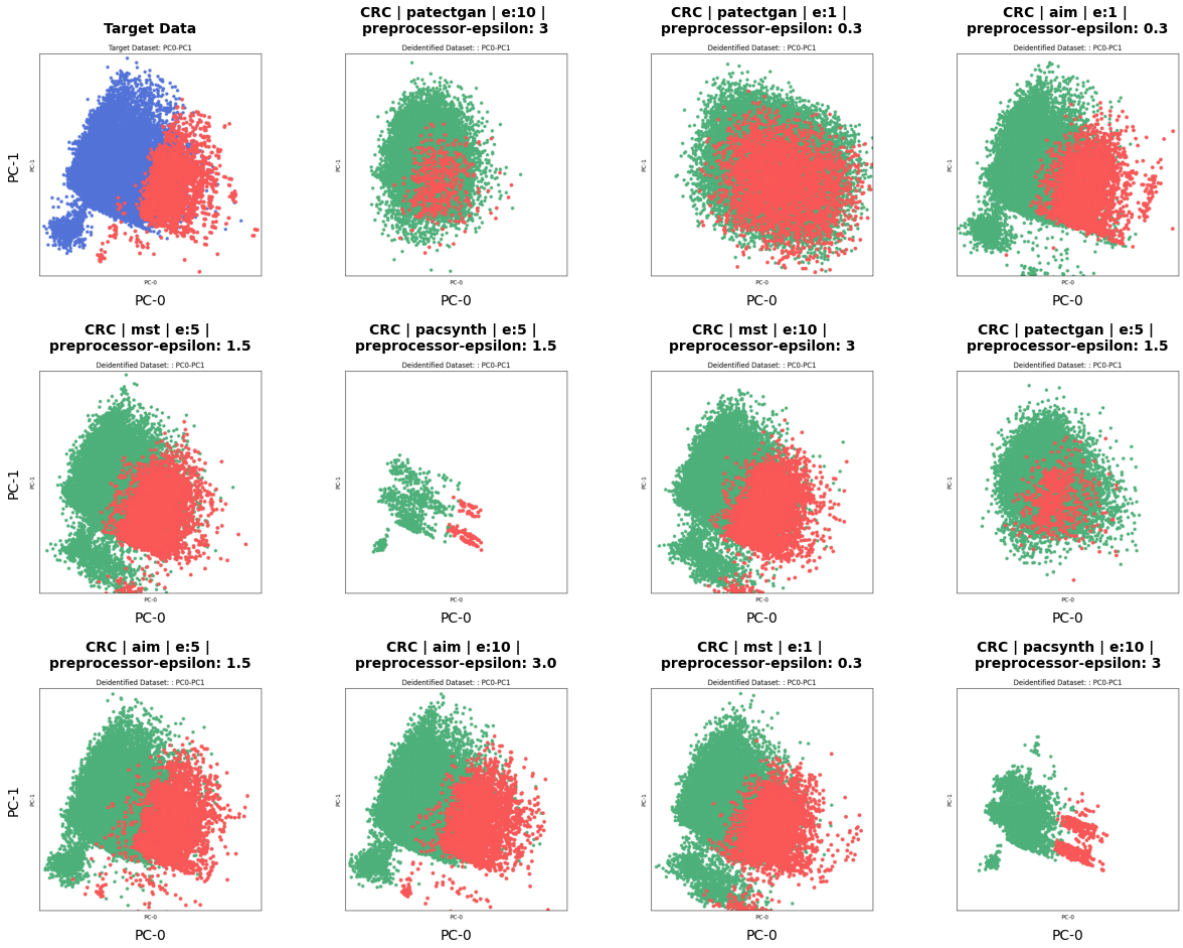
Feature Set: industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



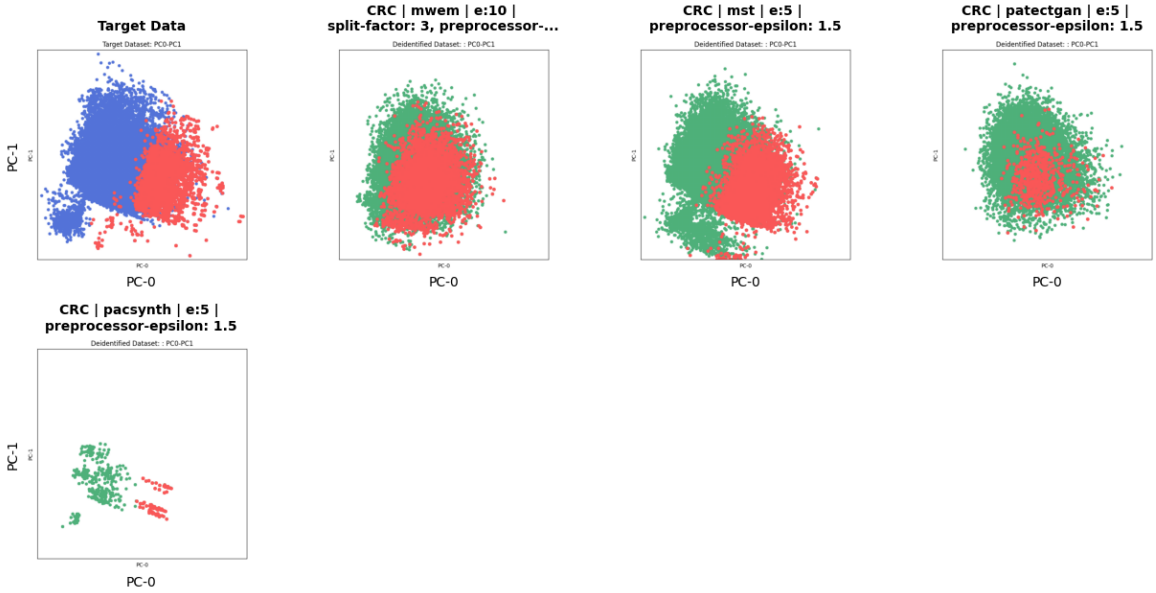
Feature Set: family-focused | Target Dataset: national2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



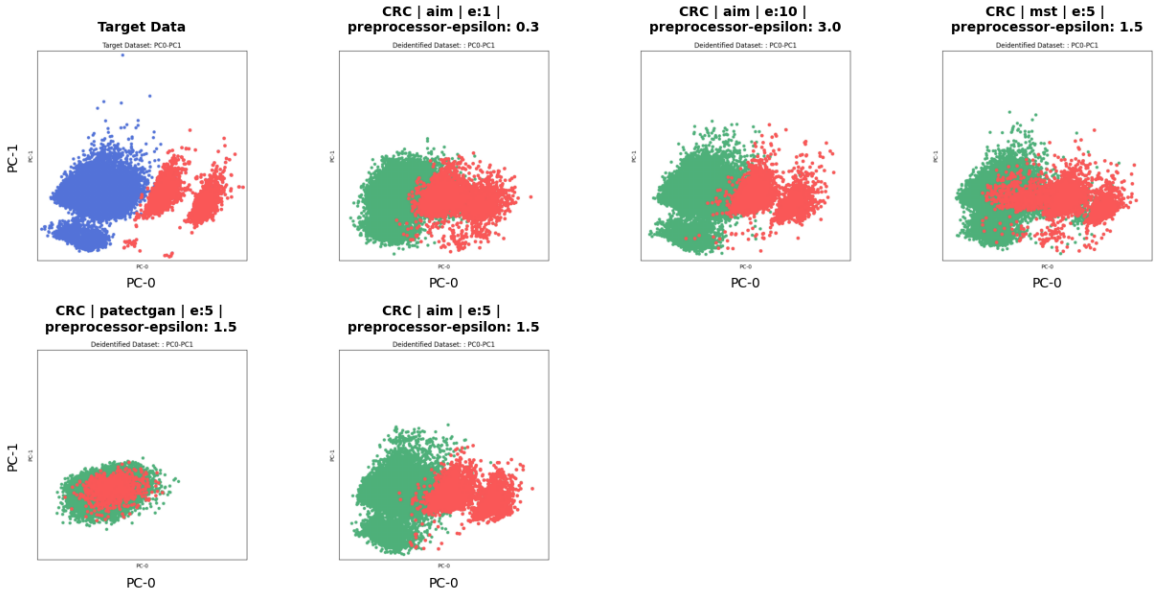
Feature Set: family-focused-with-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



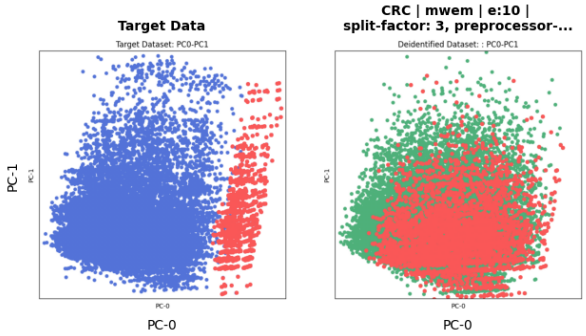
Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



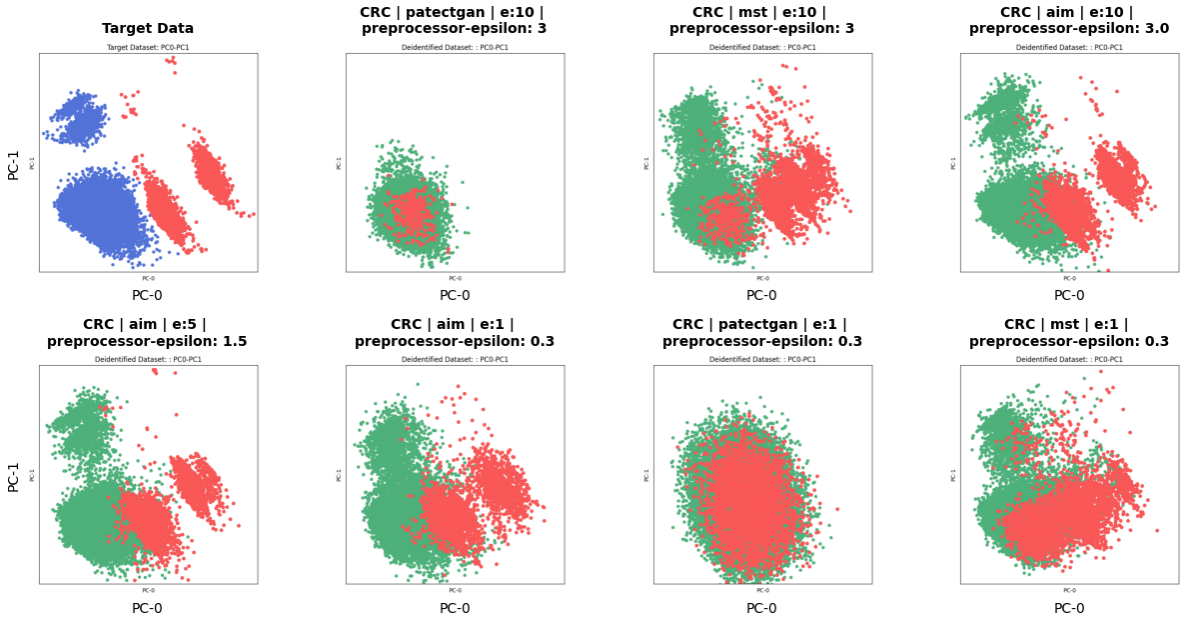
Feature Set: detailed-industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,427,025,600



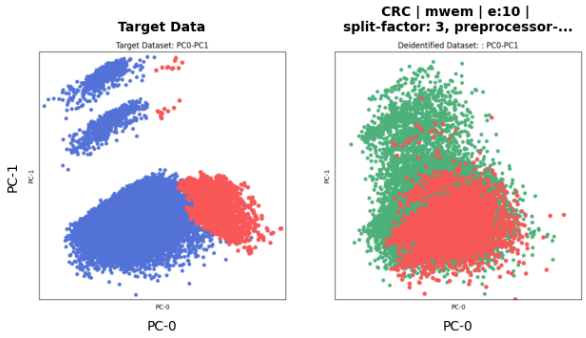
Feature Set: simple-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



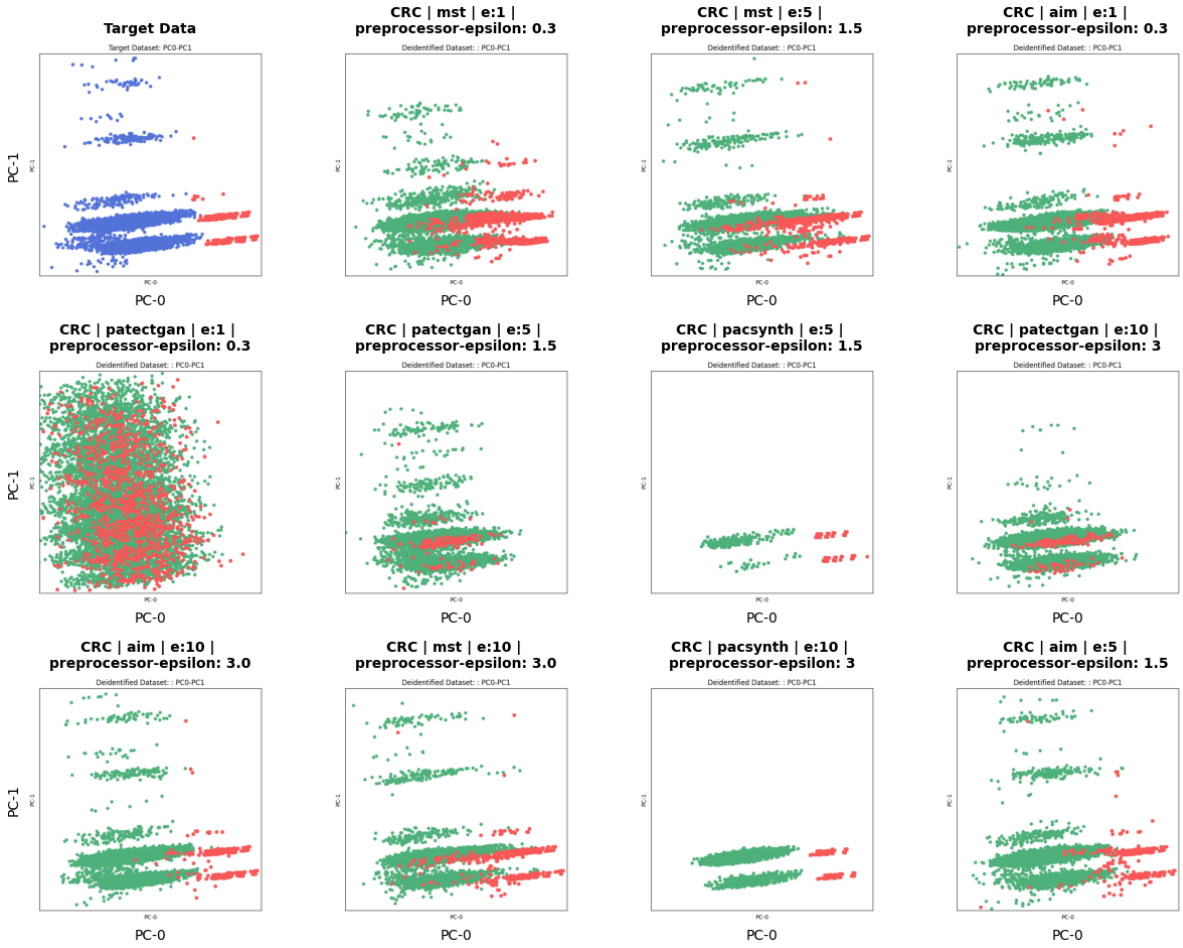
Feature Set: custom-features-12 | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'MSP', 'NOC', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400,000



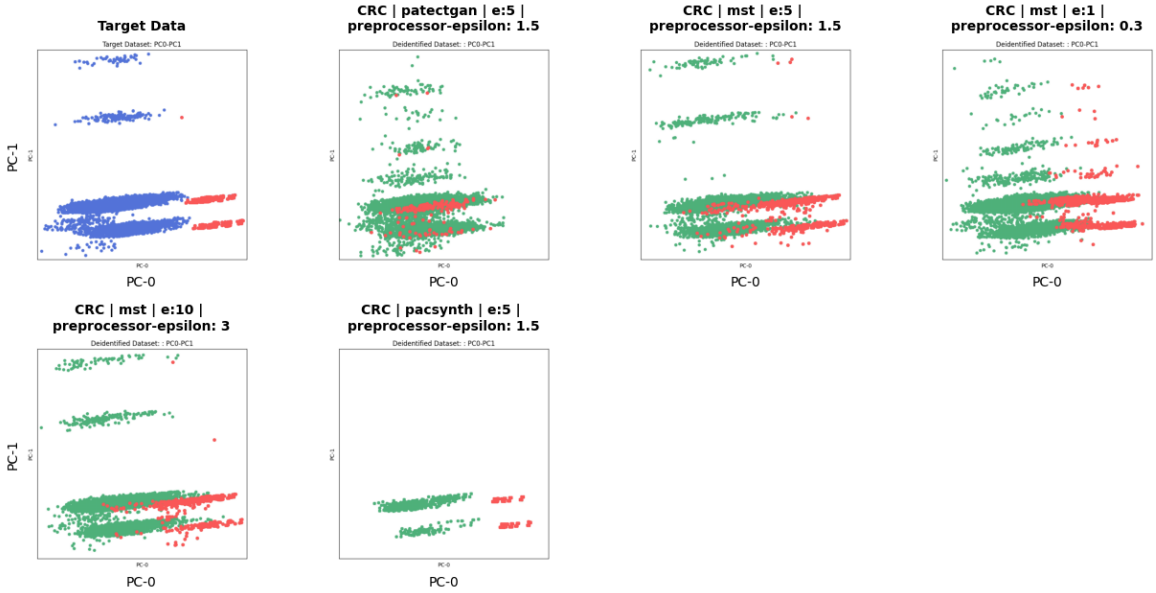
Feature Set: demographic-focused | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



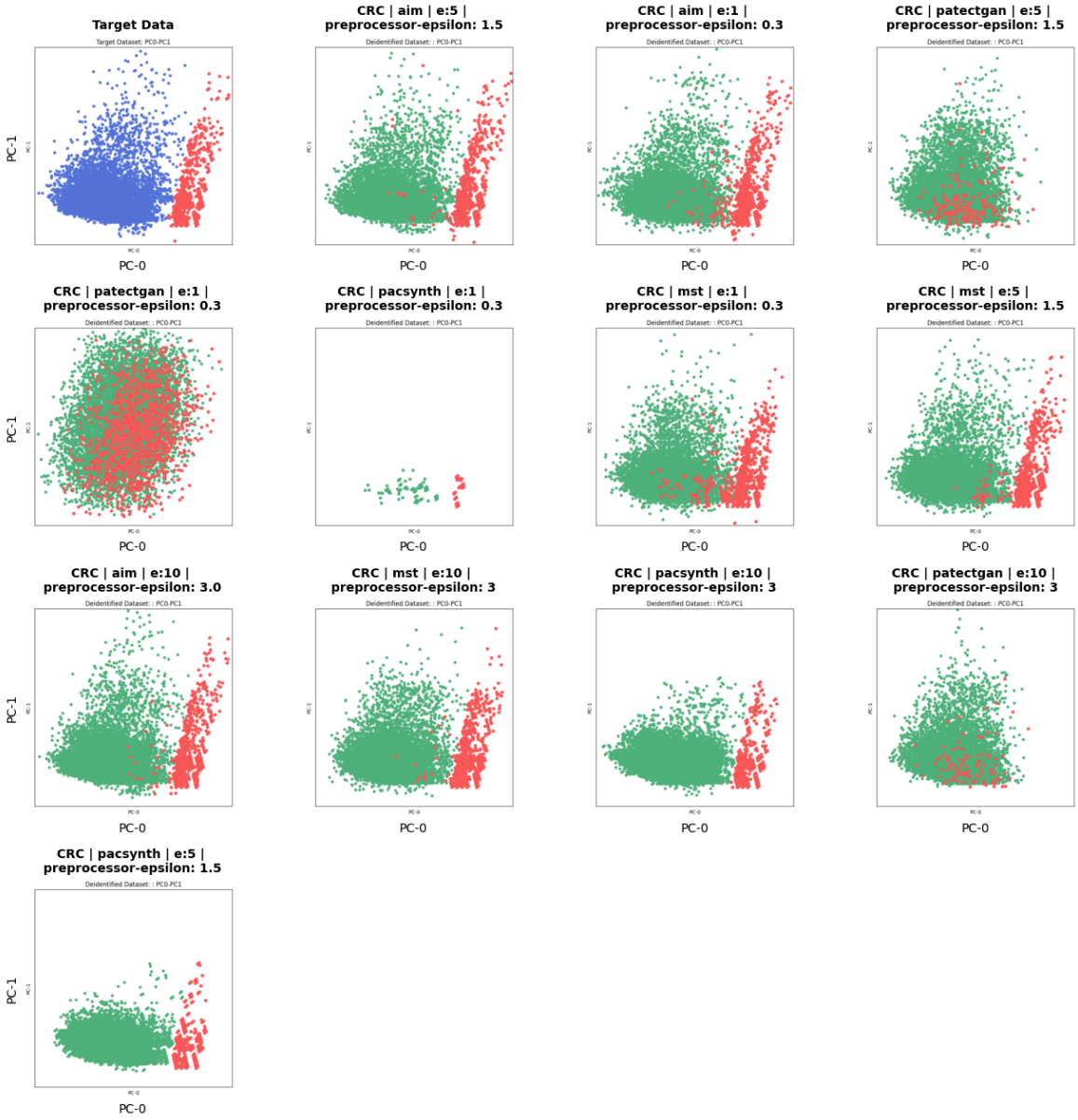
Feature Set: demographic-focused-except-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



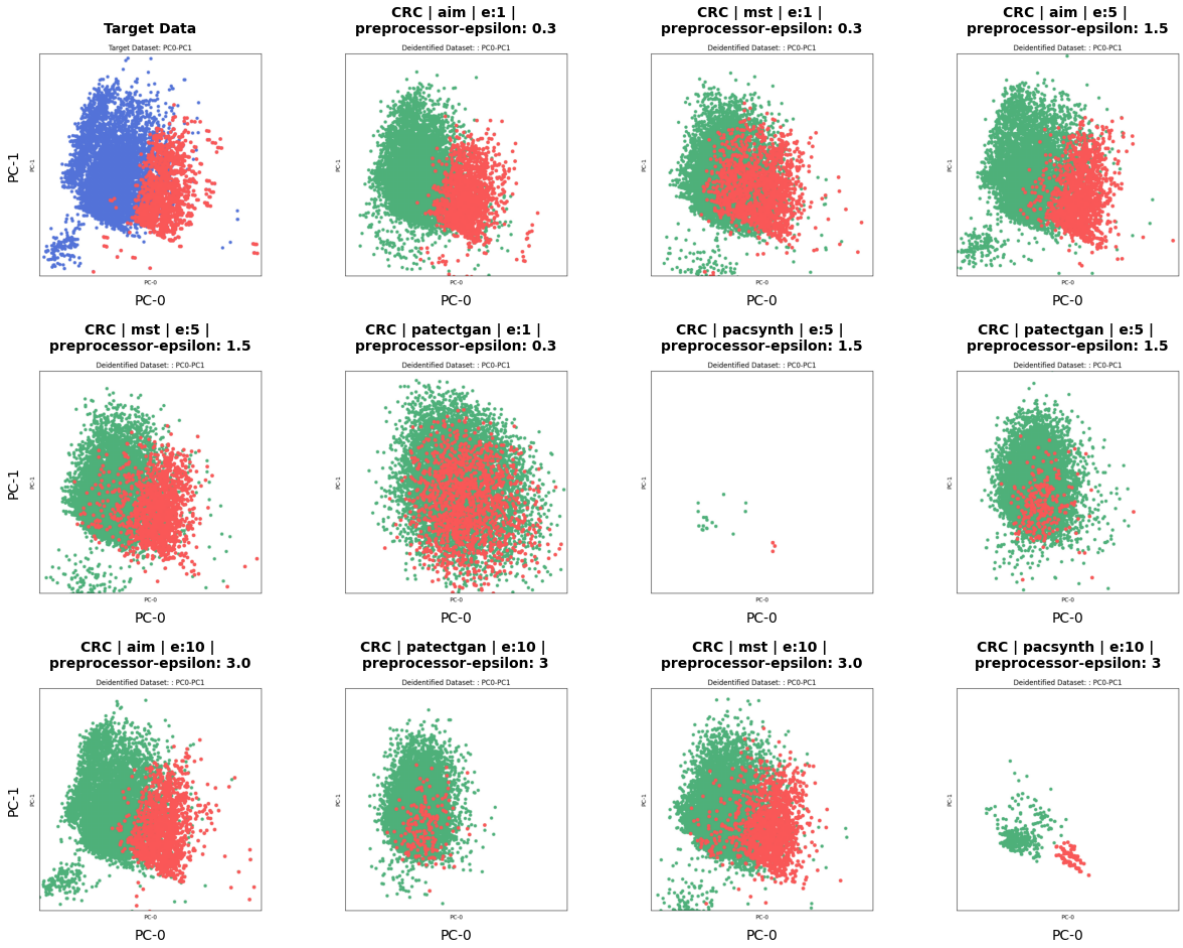
Feature Set: industry-focused | Target Dataset: tx2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



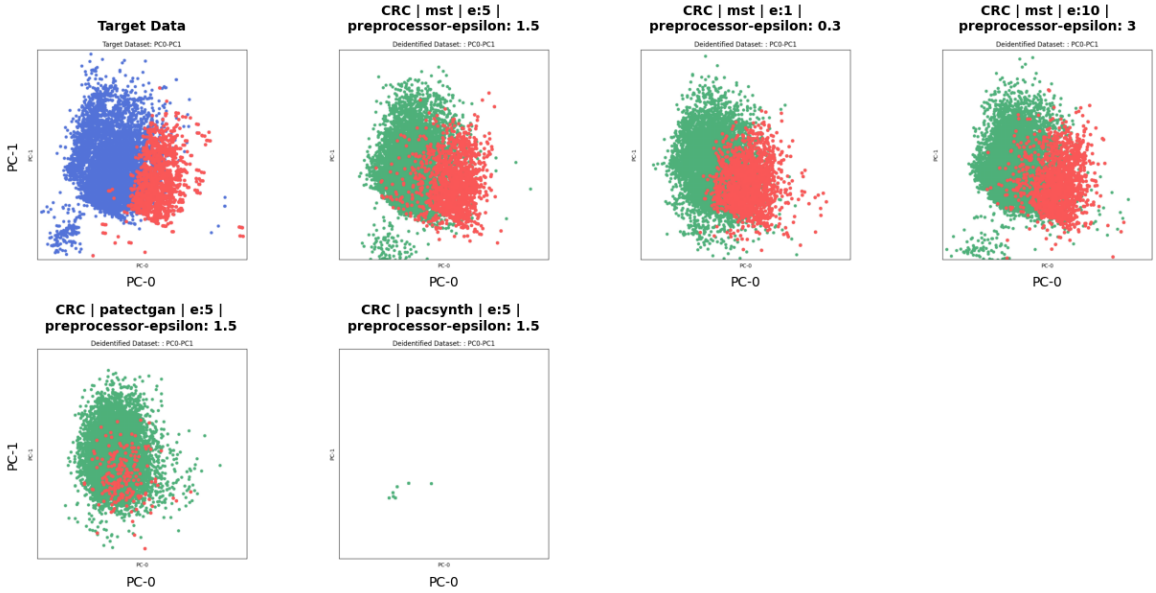
Feature Set: family-focused | Target Dataset: tx2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



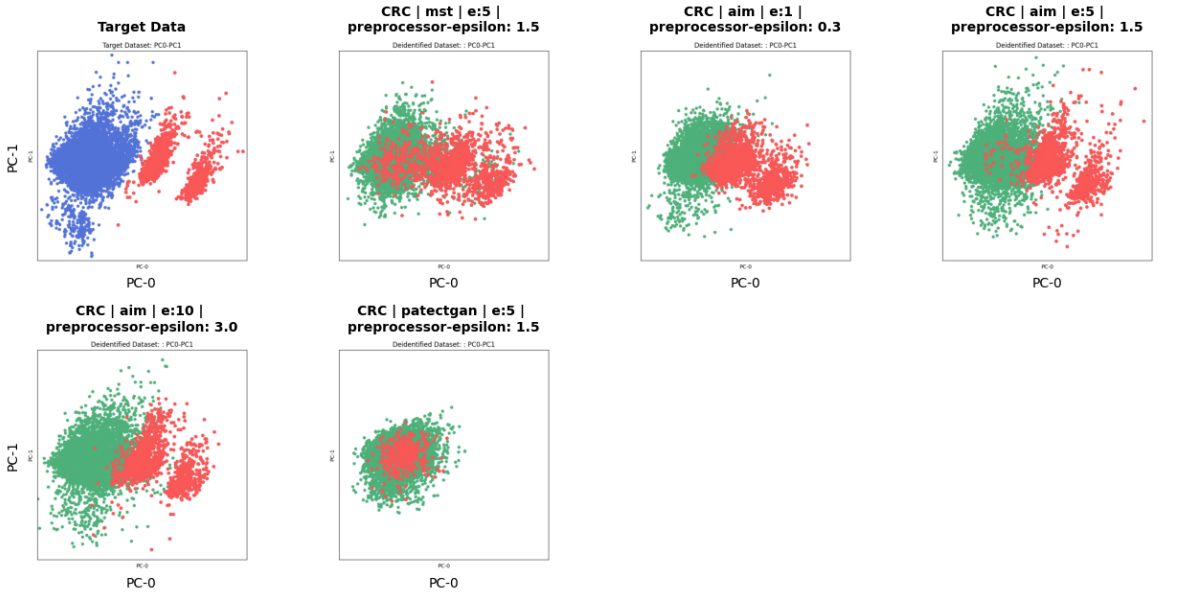
Feature Set: family-focused-with-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



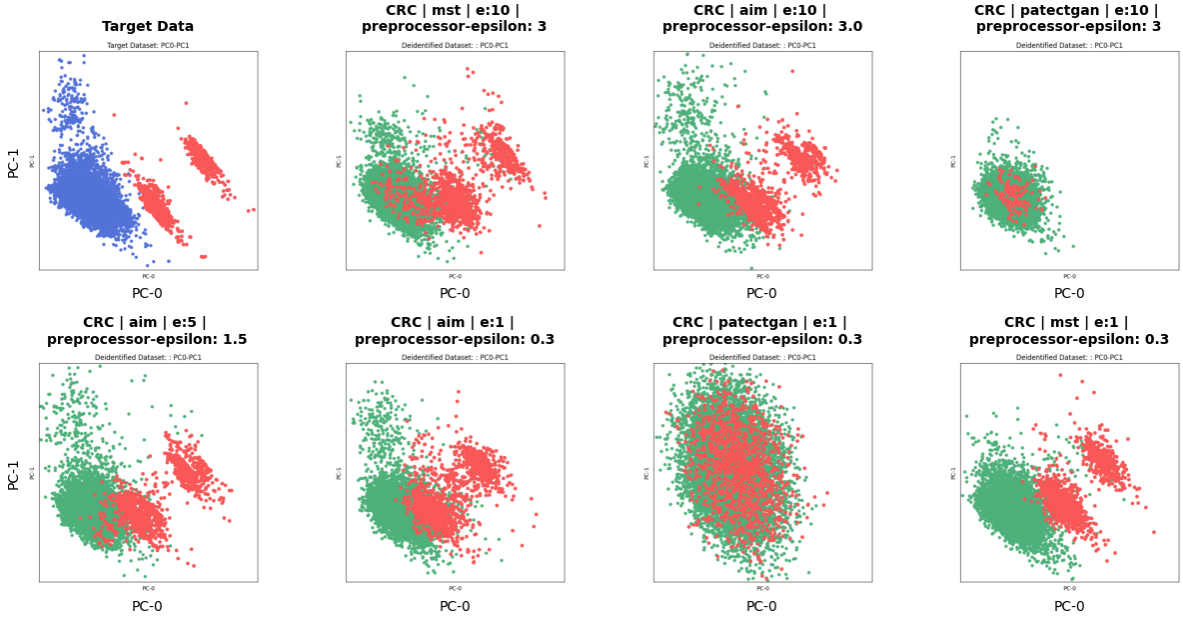
Feature Set: all-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



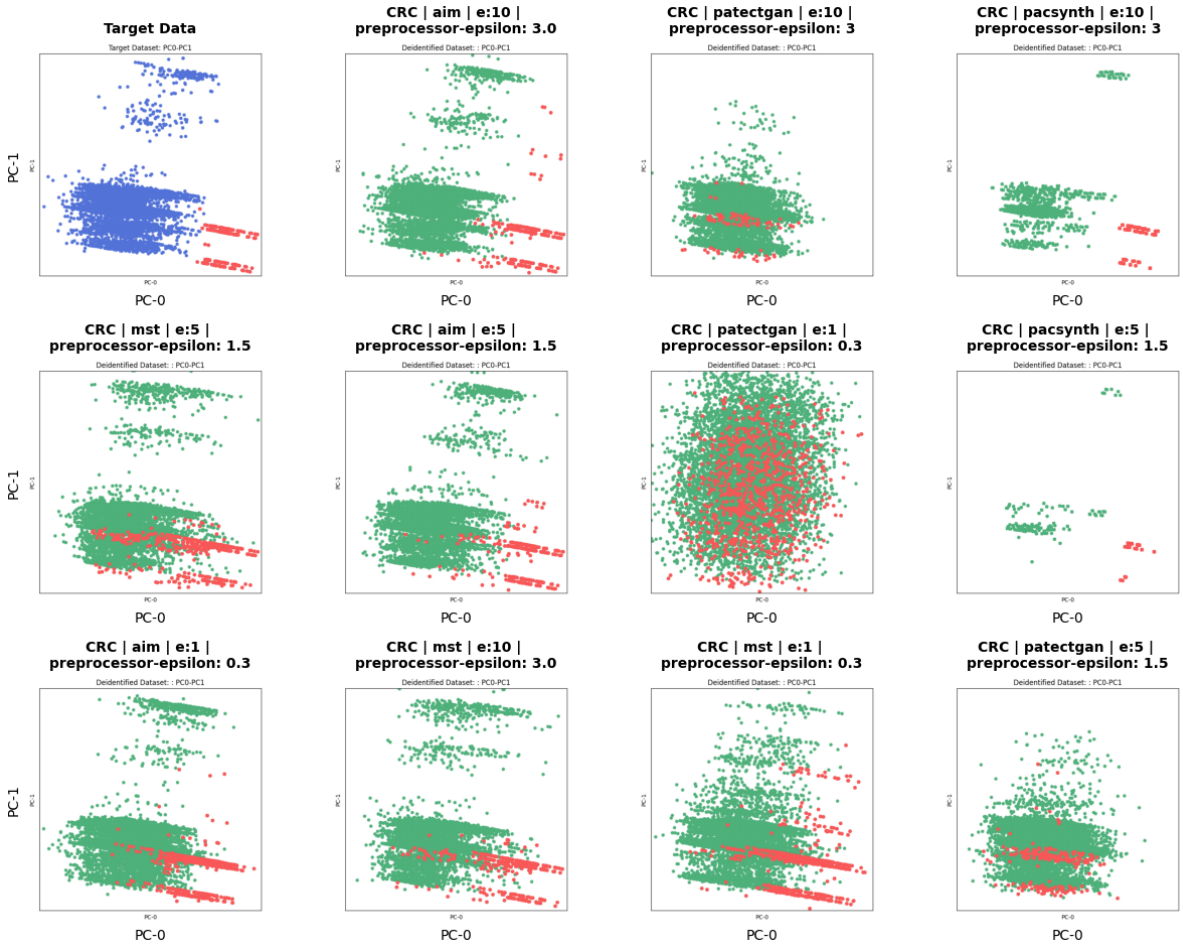
Feature Set: simple-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



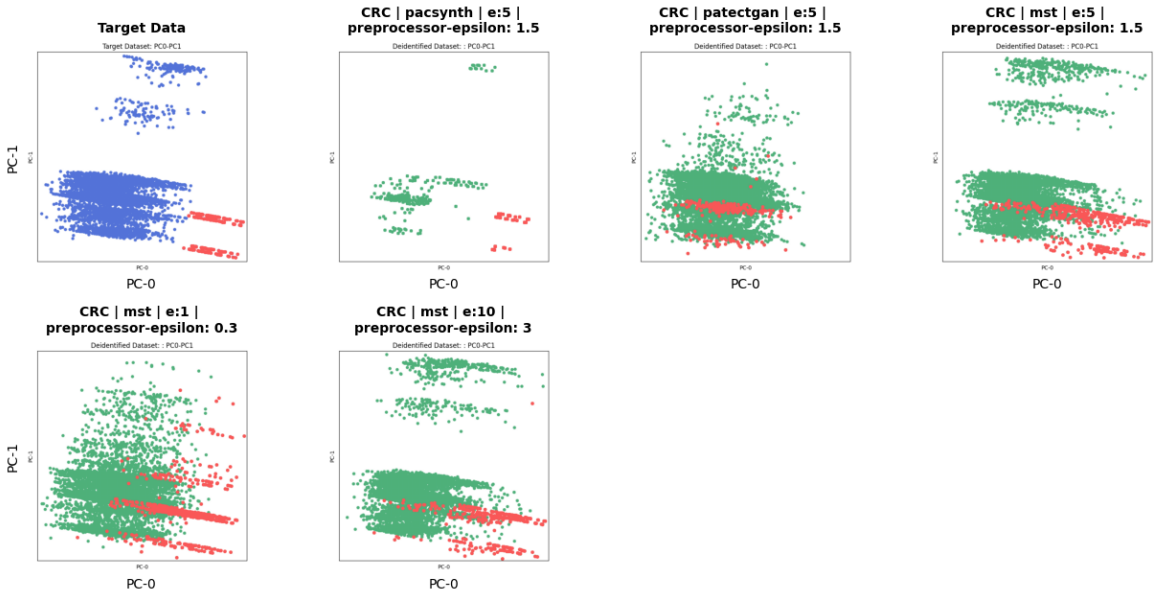
Feature Set: demographic-focused | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



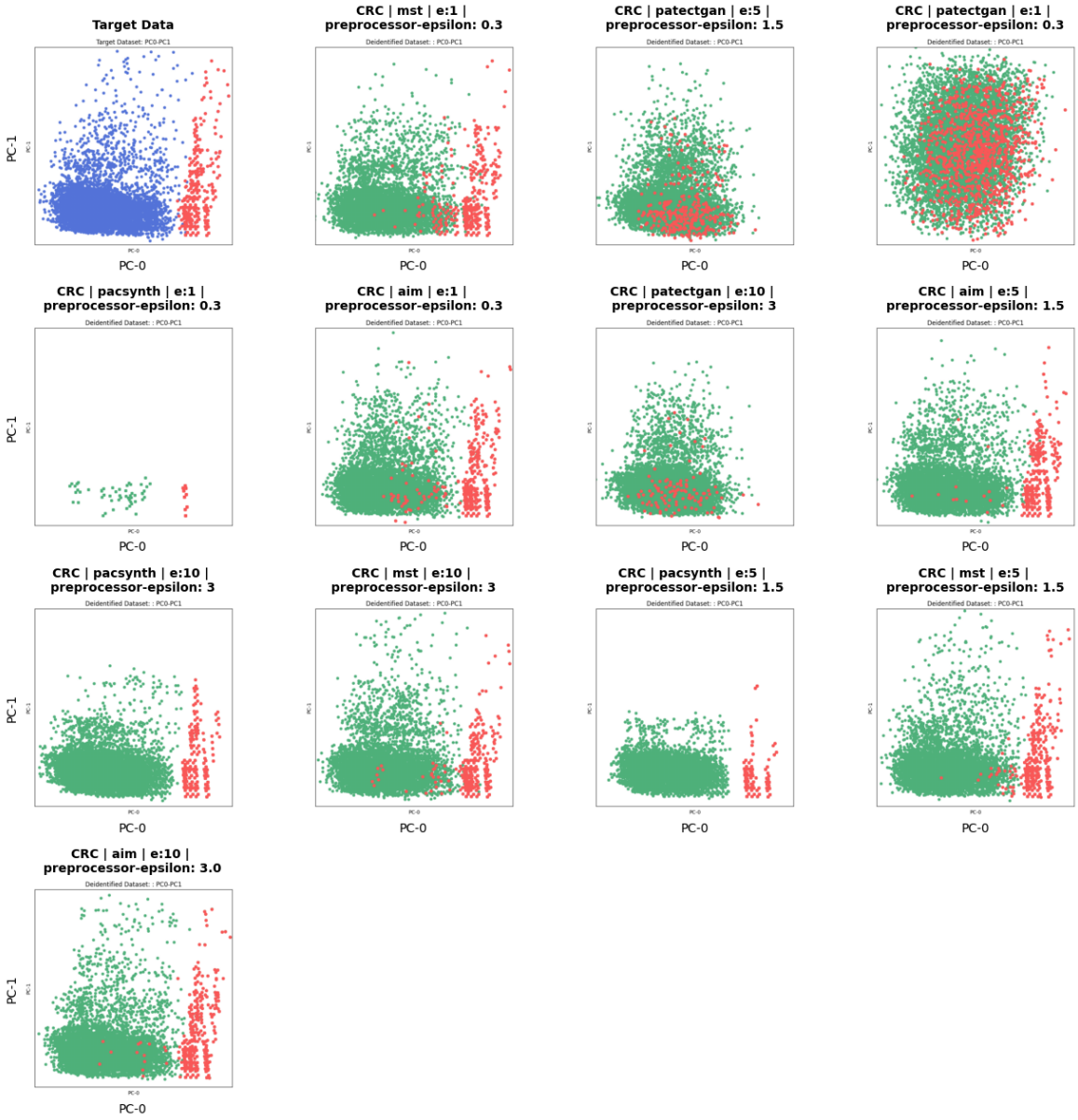
Feature Set: demographic-focused-except-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



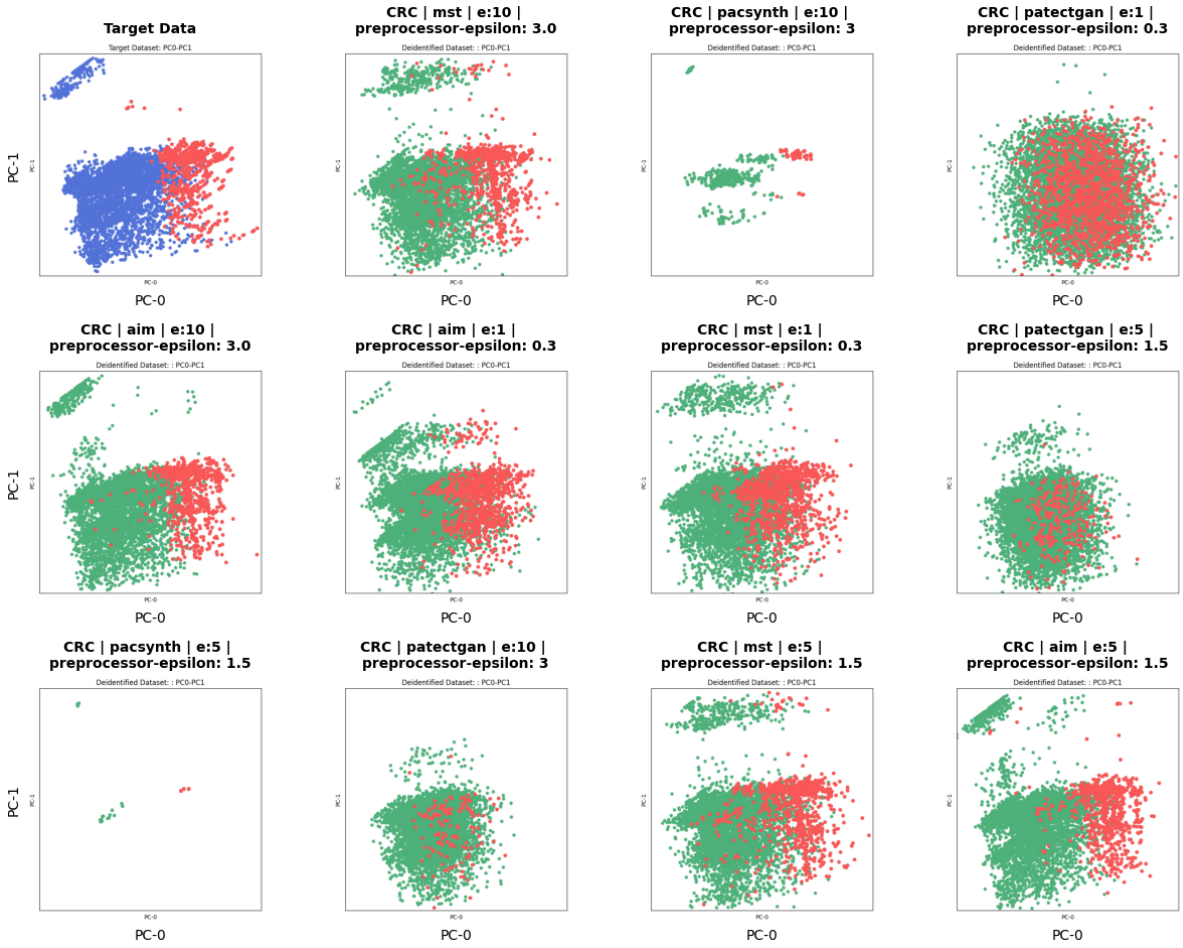
Feature Set: industry-focused | Target Dataset: ma2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



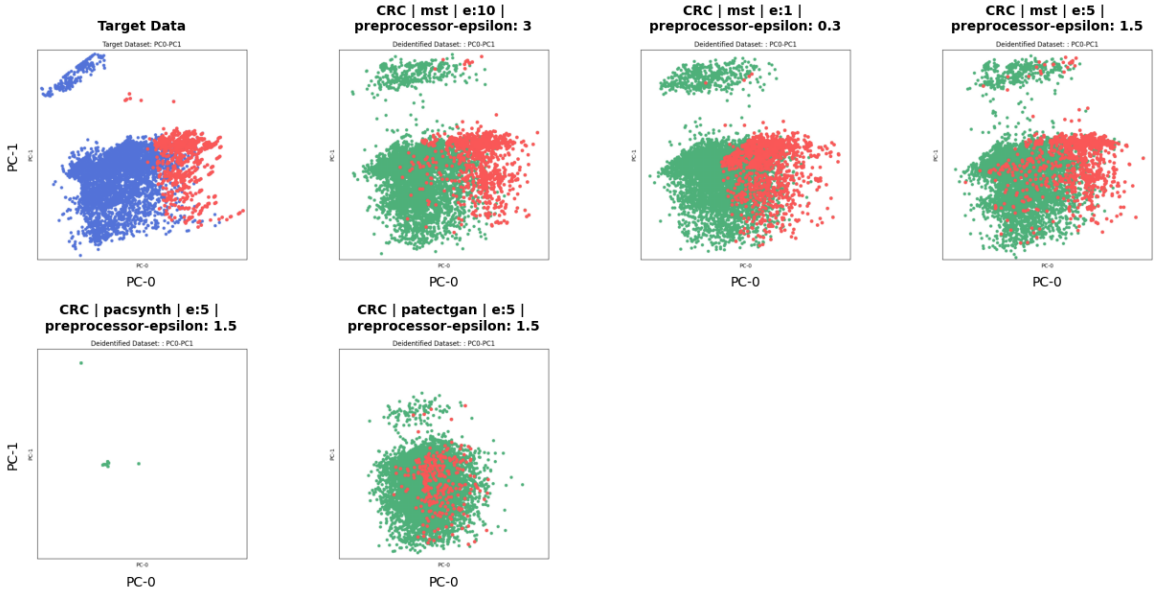
Feature Set: family-focused | Target Dataset: ma2019:

Features: ['AGEP', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 498,960,000,000



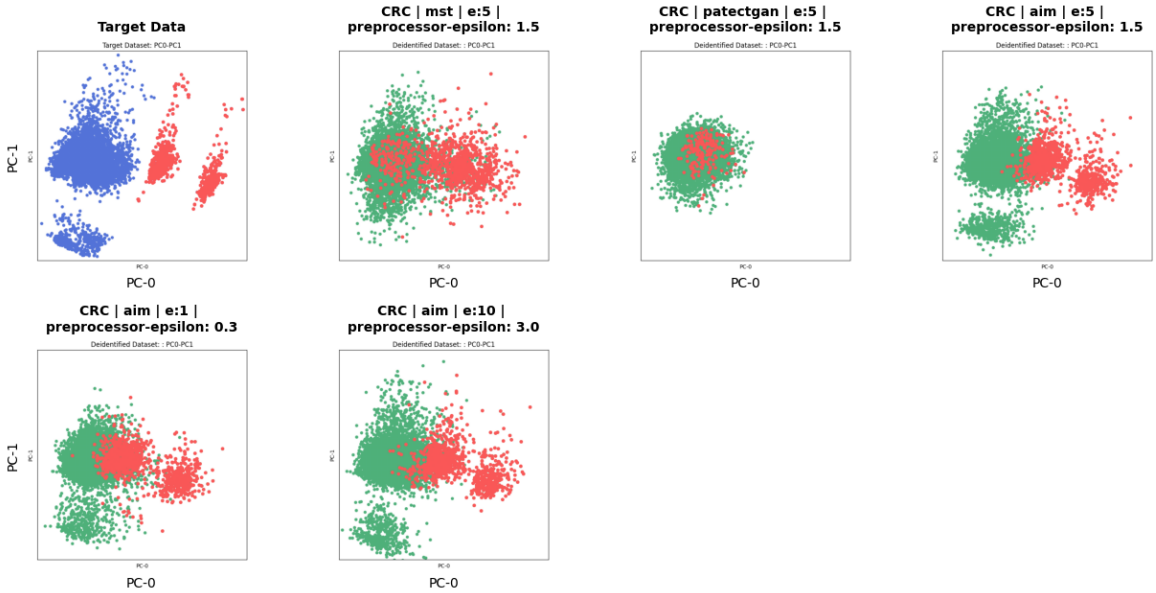
Feature Set: family-focused-with-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'HISP', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 997,920,000,000



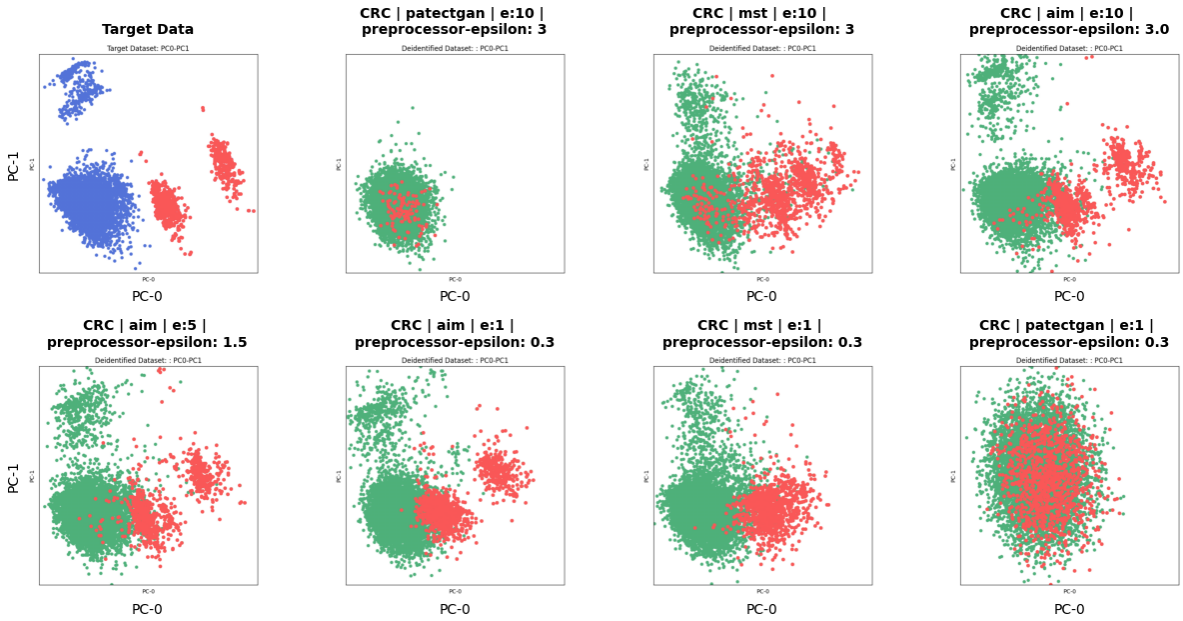
Feature Set: all-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



Feature Set: simple-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



Regression Comparison: White Men Data:

Linear regression is a fundamental data analysis technique that condenses a multi-dimensional data distribution down to a one dimensional (line) representation. It works by finding the line that sits in the 'middle' of the data, in some sense-- it minimizes the total distance between the points of the data and the line. There are more advanced forms of regression, but here we're focusing on the simplest case-- we fit a simple straight line to the data, getting the slope and y-intercept value of that line.

For this metric we're just looking at data from adults (AGEP > 15) and we're only considering the distribution of the data across two features:

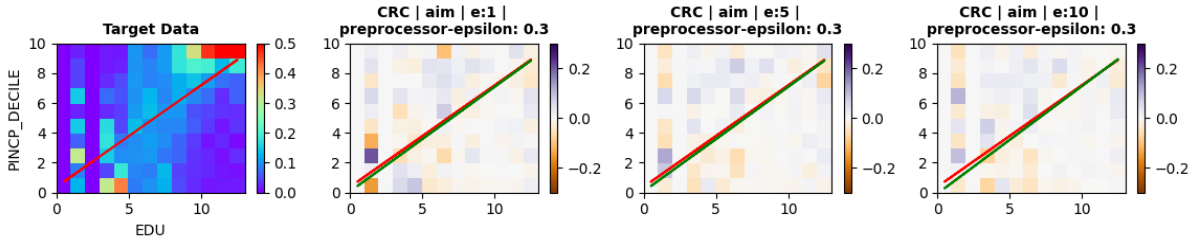
- EDU: The highest education level this individual has attained, ranging from 1 (elementary school) to 12 (PhD). See Appendix of this report for the full list of code values.
- PINCP_DECILE: The individual's income decile relative to their PUMA. This helps us account for differences in cost of living across the country. If an individual makes a moderate income but lives in a very low income area, they may have a high value for PINCP_DECILE indicating that they have a high income for their PUMA).

The basic idea is that higher values of EDU should lead to higher values of PINCP_DECILE, and this is broadly true. However, it is known that the relationship between EDU and PINCP_DECILE is different for different demographic subgroups. The heatmaps in the left column below show the density distribution of the true data for each subgroup, normalized by education category (so the density values in each column sum to 1; note that when a cell in the heatmap contains too few people (< 20), it is left blank; its not expected that the deidentified data will match the original distribution precisely). The regression line is drawn in red over the heatmap, so you can see the relationship between the target data distribution and its linear regression analysis. In the right column for each subgroup we show how the deidentified data's regression line compares to the target data's regression line, along with a heatmap of the density differences between the two distributions. Redder areas are where the deidentified data has created too many people, bluer areas are where it's created too few people.

We've broken this metric down into demographic subgroups so we can see not only how well the privacy techniques preserve the overall relationship between these features, but also whether they preserve how that overall relationship is built up from the different relationships that hold at each major demographic subgroup. It's important that deidentification techniques preserve these distinct subgroup patterns for analysis.

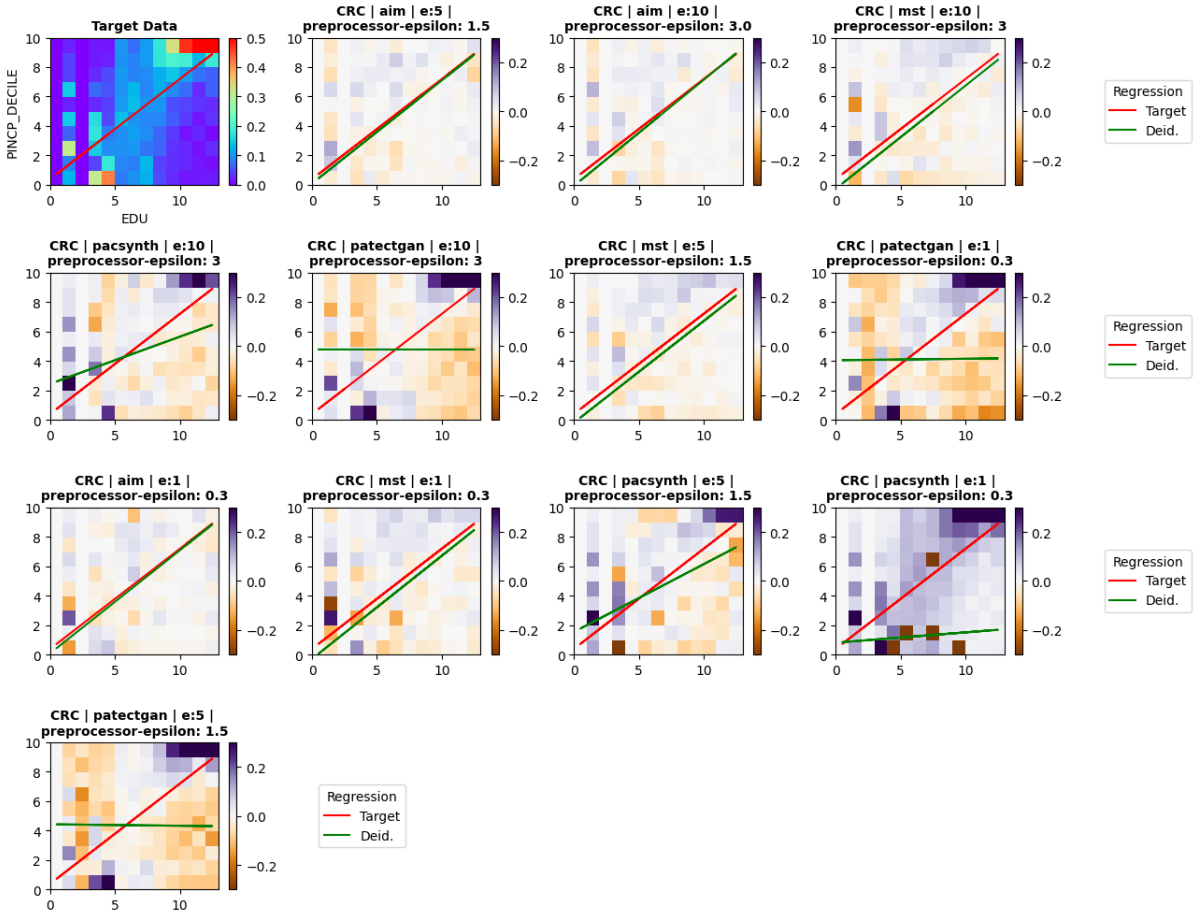
Feature Set: demographic-focused-except-AGEP-DEYE | Target Dataset: national2019:

Features: ['DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,135,134



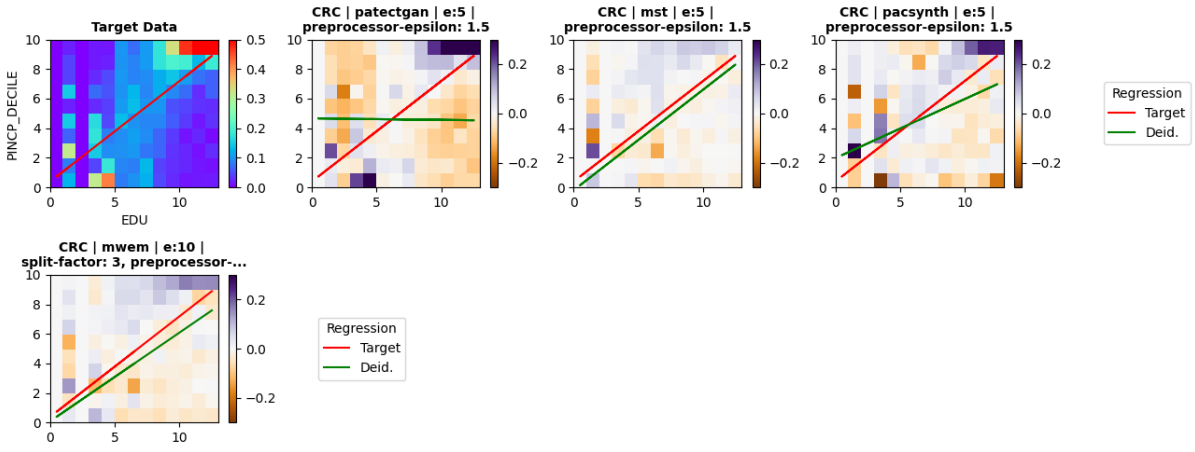
Feature Set: demographic-focused | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



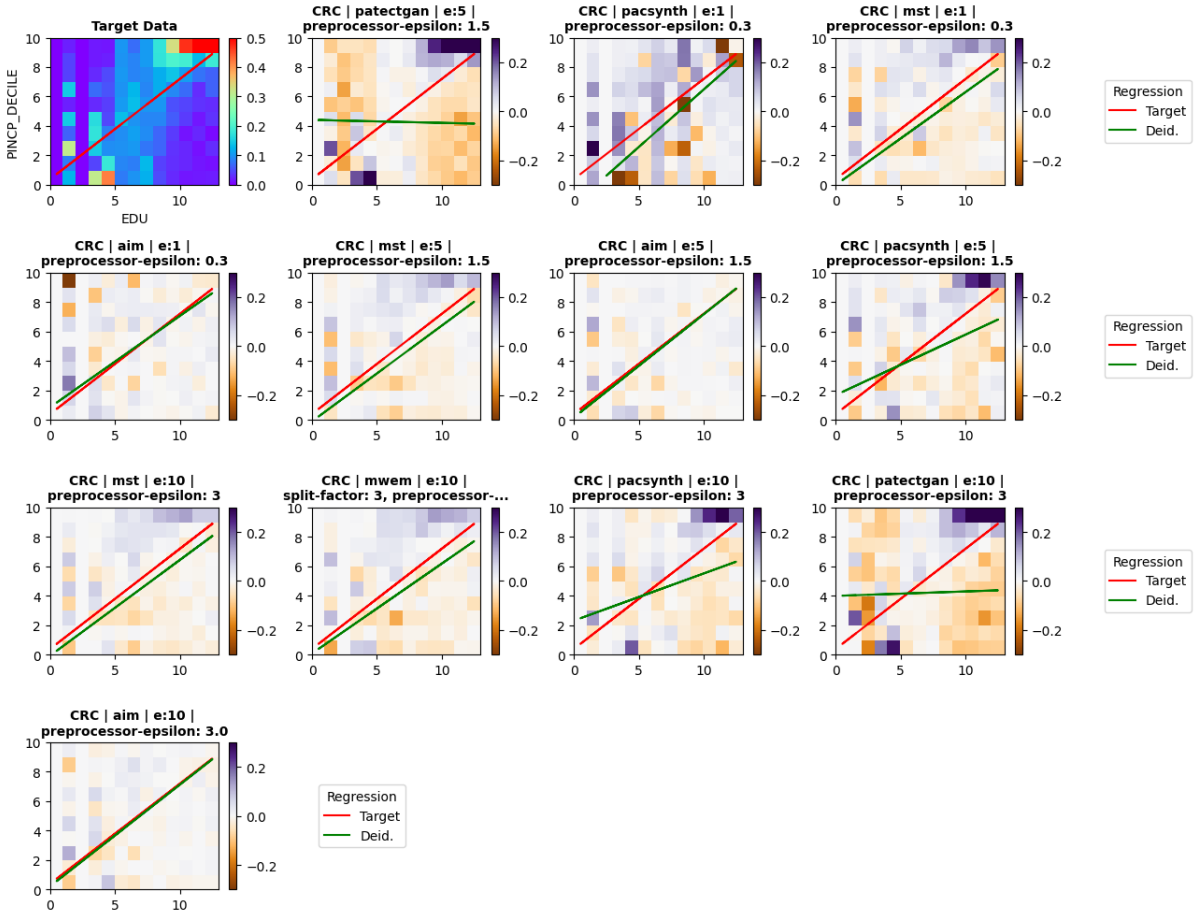
Feature Set: demographic-focused-except-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



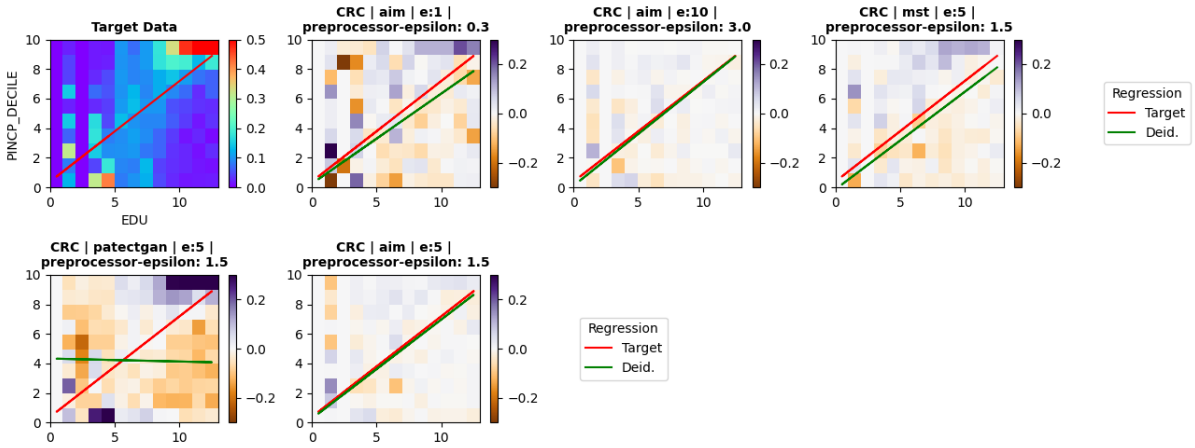
Feature Set: industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



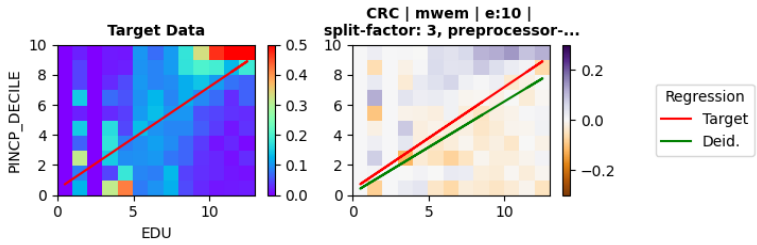
Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



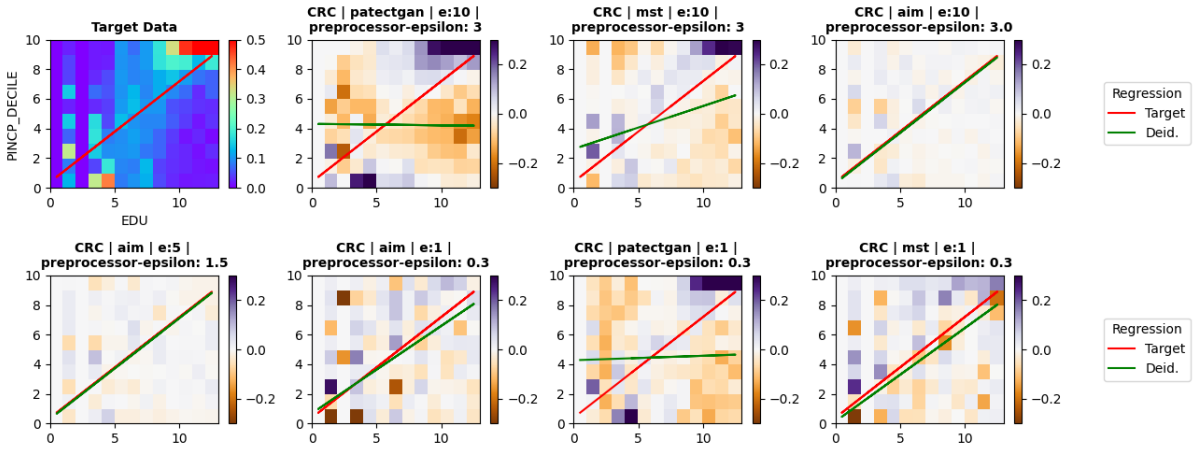
Feature Set: detailed-industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,427,025,600



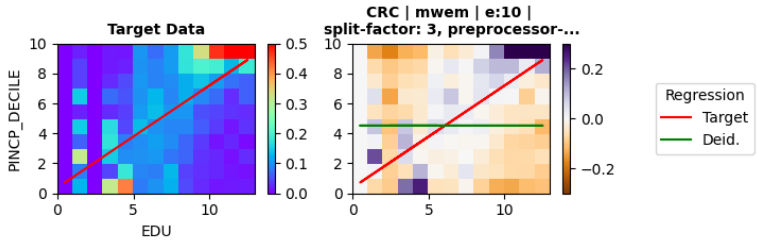
Feature Set: simple-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



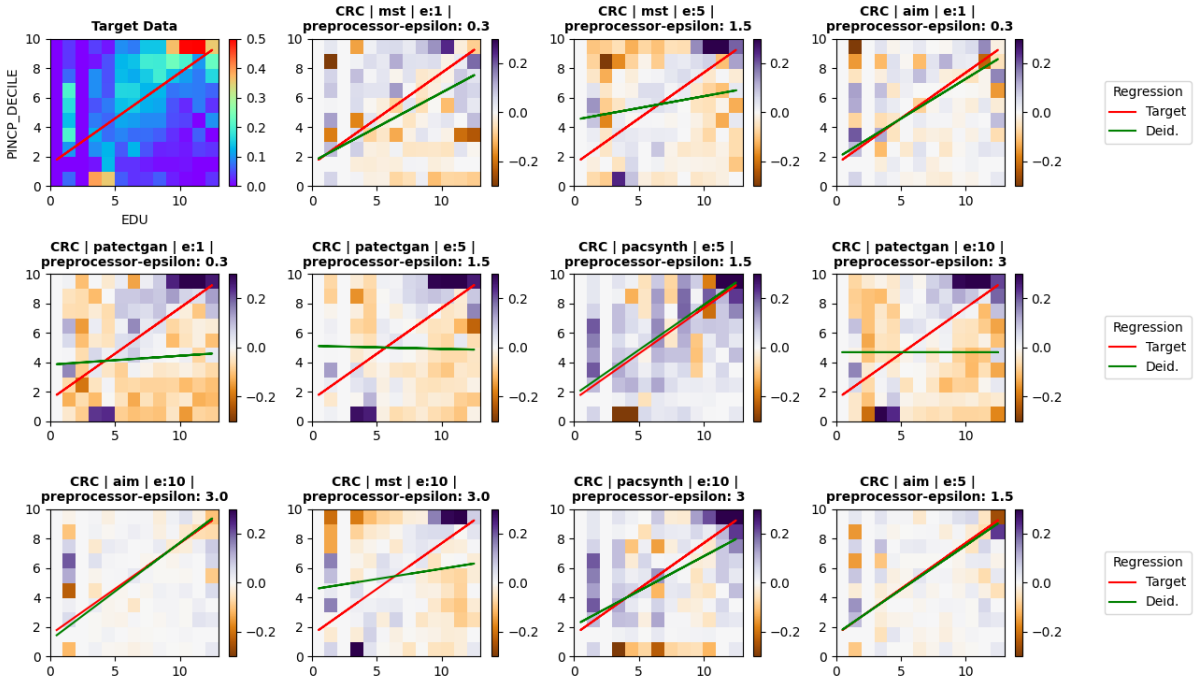
Feature Set: custom-features-12 | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'MSP', 'NOC', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400,000



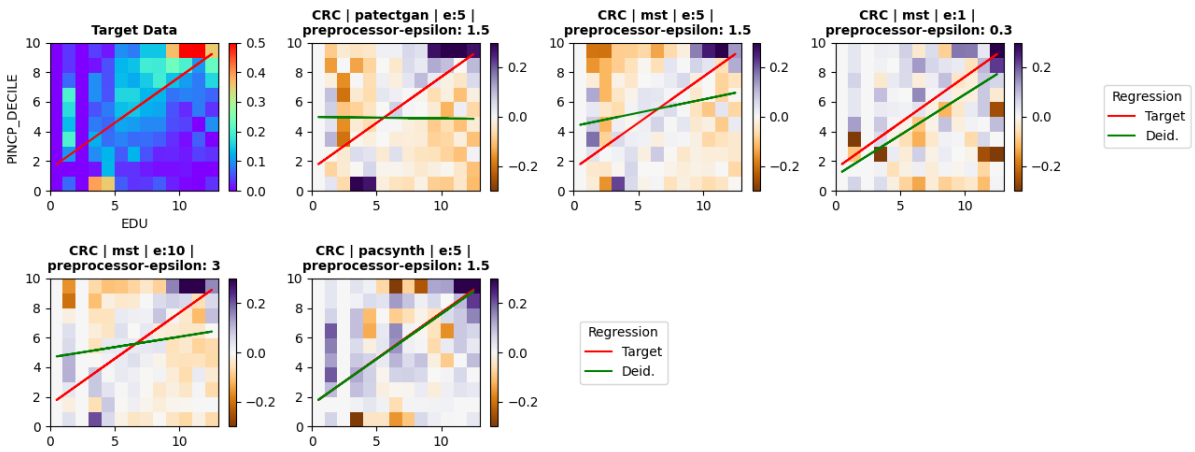
Feature Set: demographic-focused | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



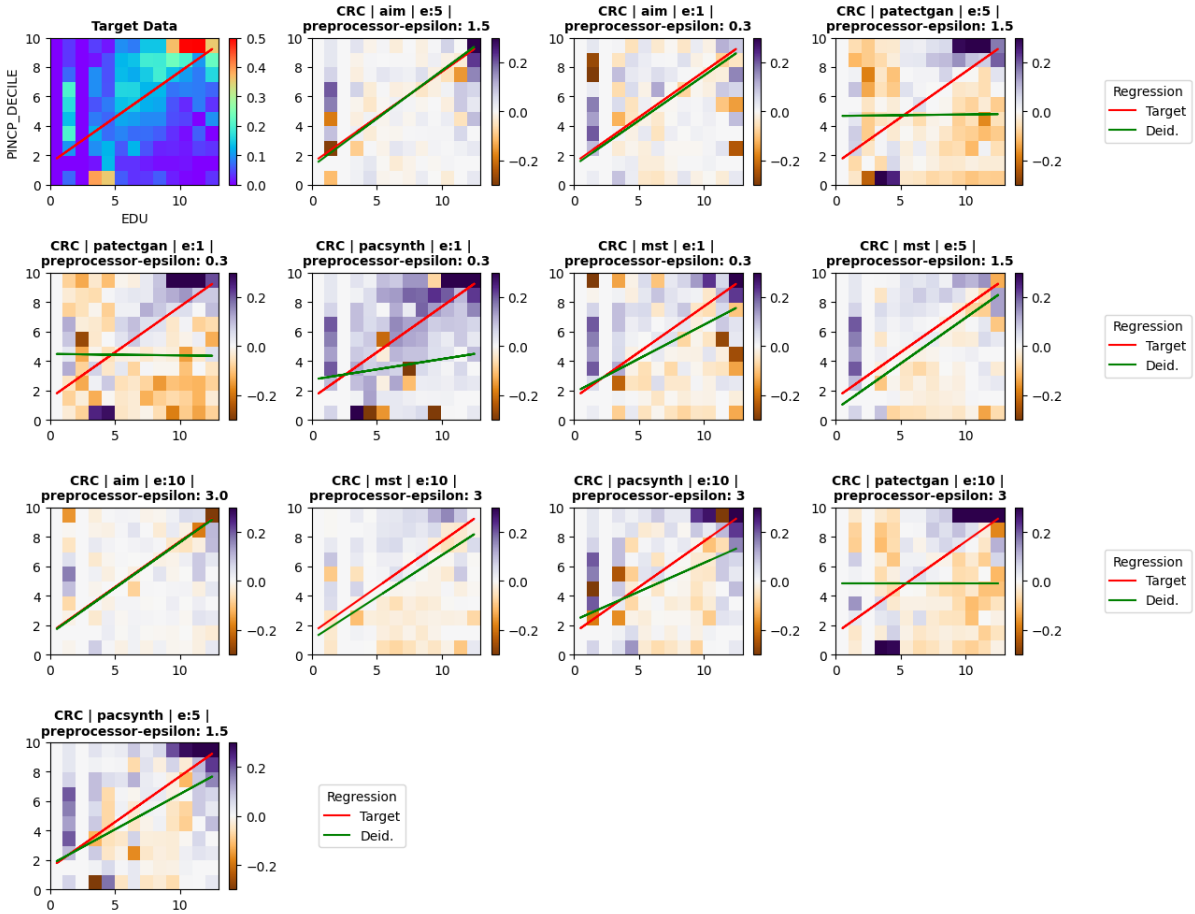
Feature Set: demographic-focused-except-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



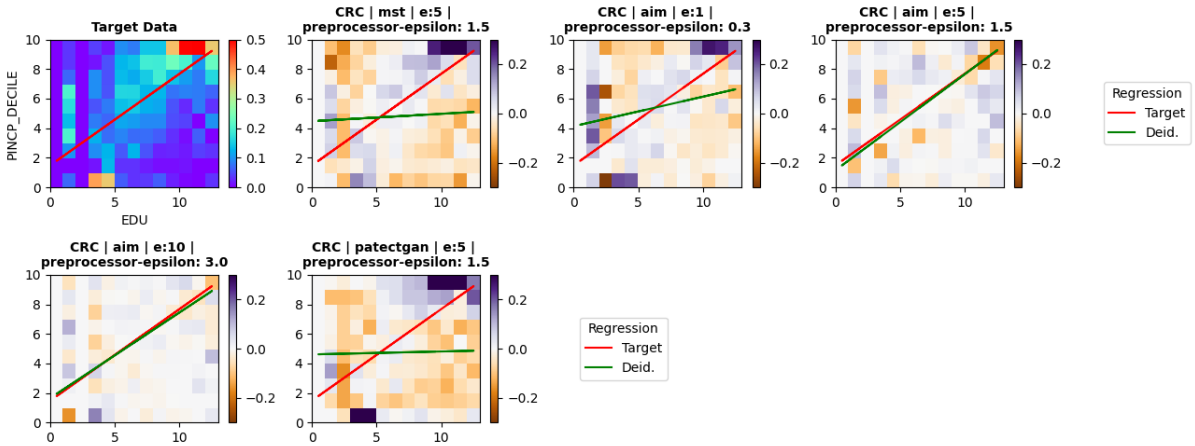
Feature Set: industry-focused | Target Dataset: tx2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



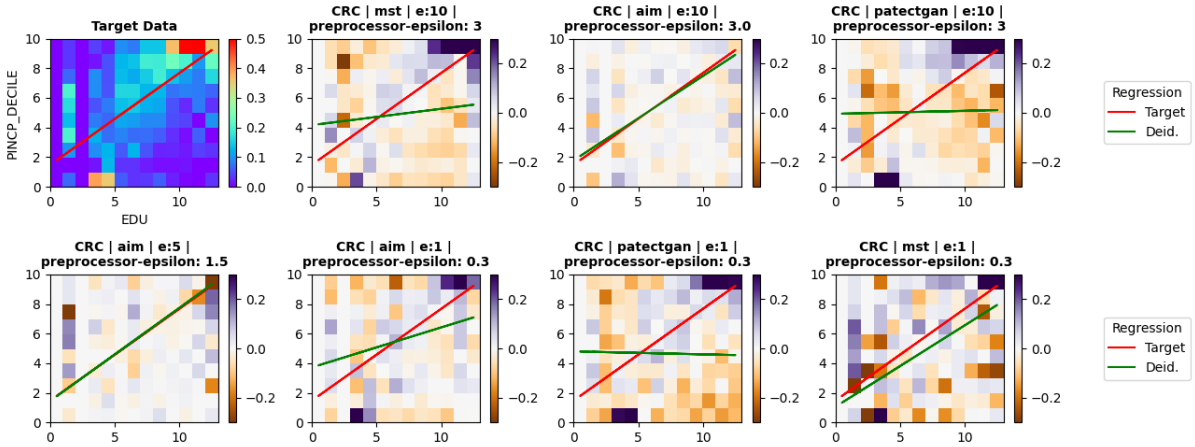
Feature Set: all-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



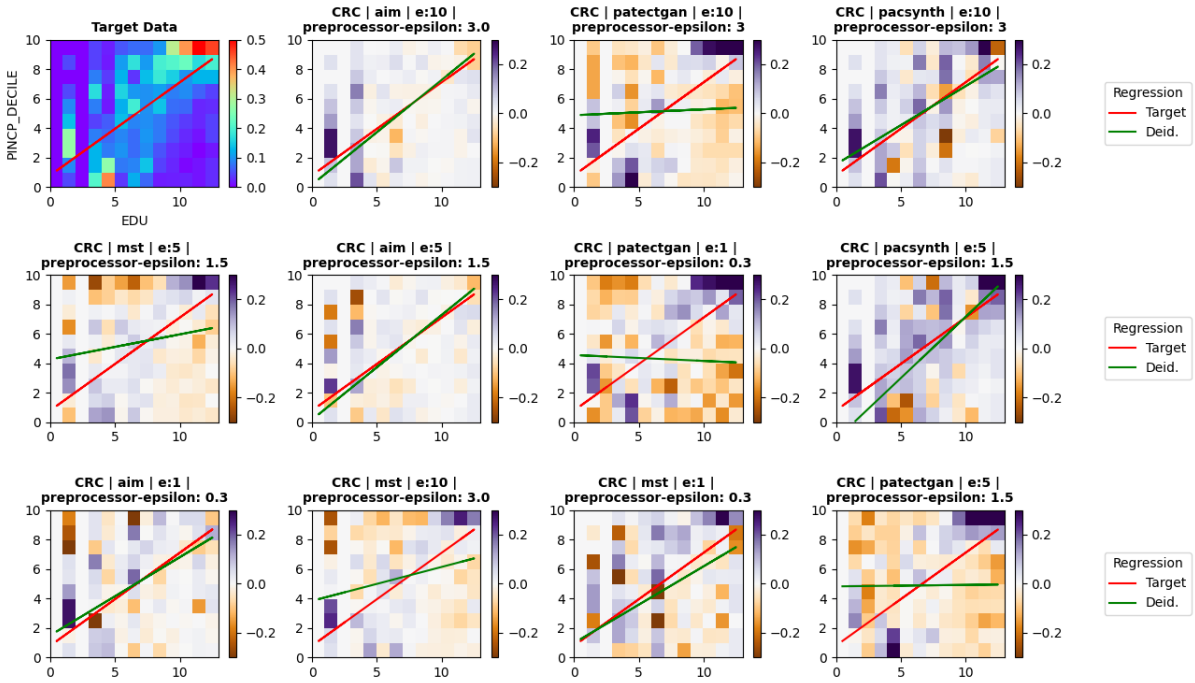
Feature Set: simple-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



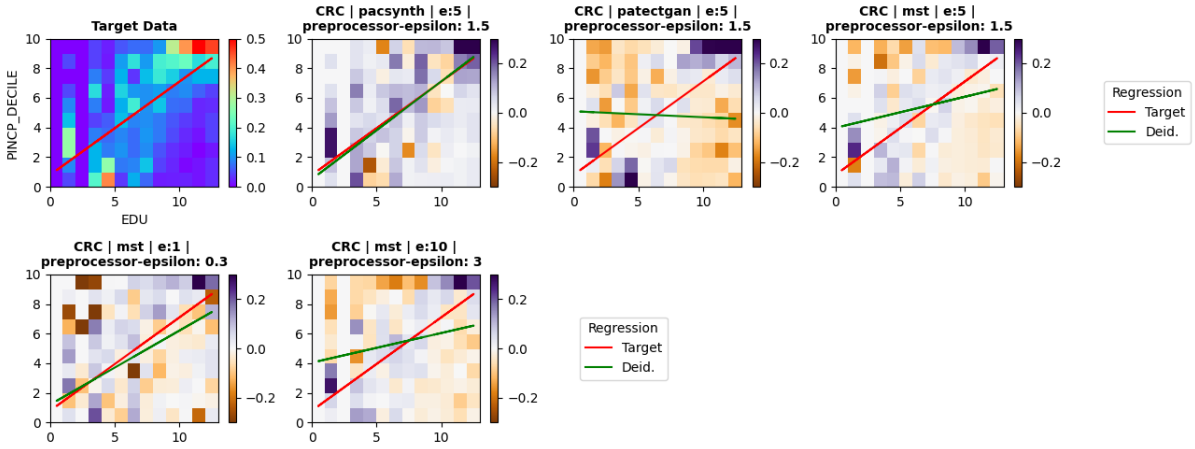
Feature Set: demographic-focused | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



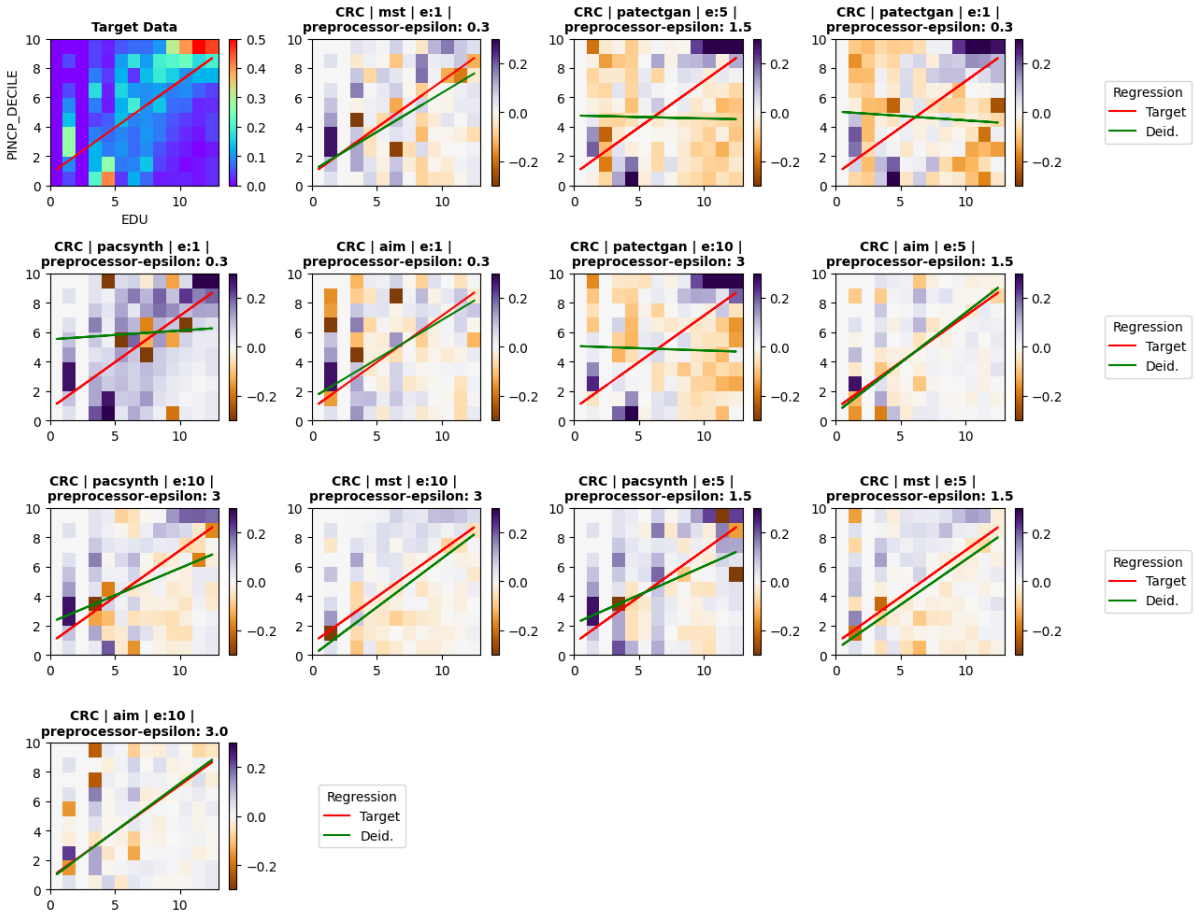
Feature Set: demographic-focused-except-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



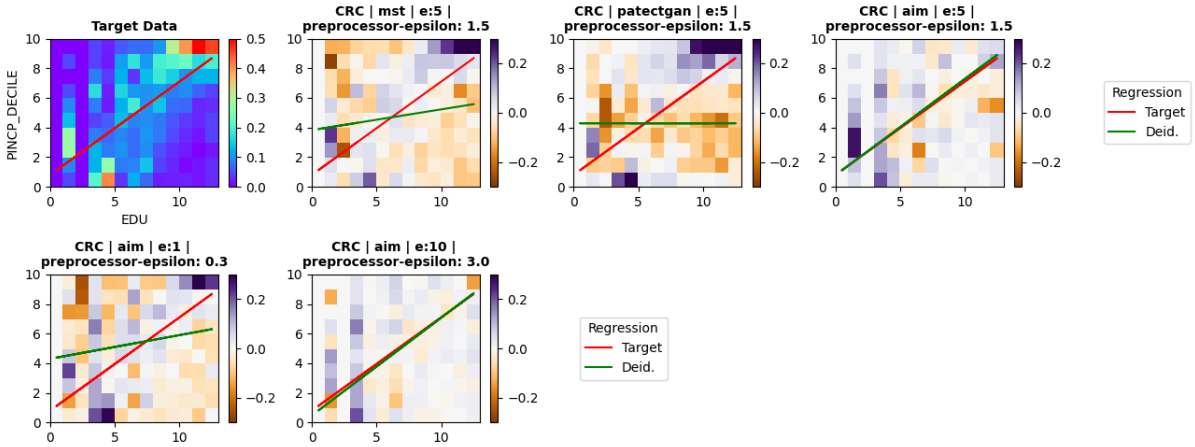
Feature Set: industry-focused | Target Dataset: ma2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



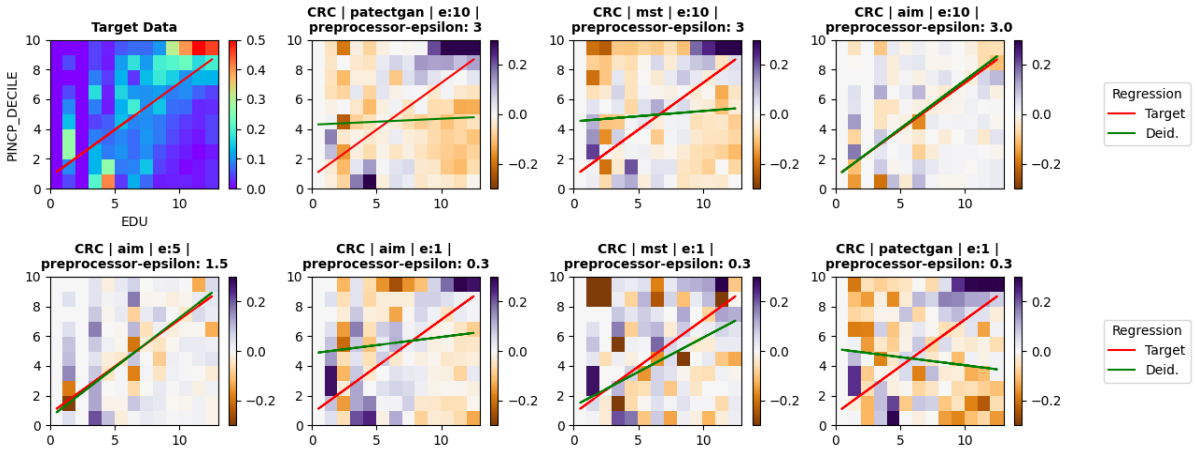
Feature Set: all-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



Feature Set: simple-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



Regression Comparison: Black Women Data:

Linear regression is a fundamental data analysis technique that condenses a multi-dimensional data distribution down to a one dimensional (line) representation. It works by finding the line that sits in the 'middle' of the data, in some sense-- it minimizes the total distance between the points of the data and the line. There are more advanced forms of regression, but here we're focusing on the simplest case-- we fit a simple straight line to the data, getting the slope and y-intercept value of that line.

For this metric we're just looking at data from adults (AGEP > 15) and we're only considering the distribution of the data across two features:

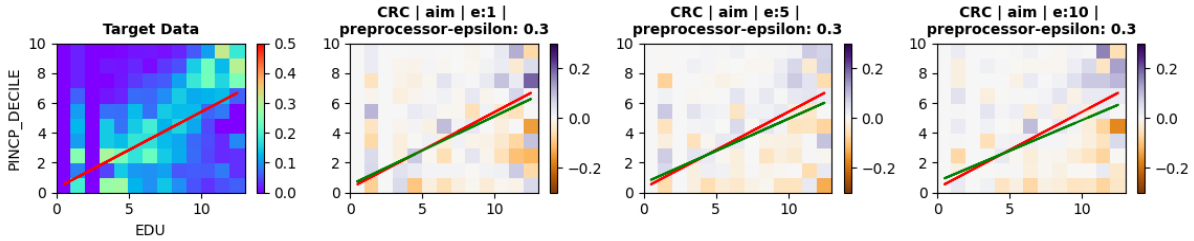
- EDU: The highest education level this individual has attained, ranging from 1 (elementary school) to 12 (PhD). See Appendix of this report for the full list of code values.
- PINCP_DECILE: The individual's income decile relative to their PUMA. This helps us account for differences in cost of living across the country. If an individual makes a moderate income but lives in a very low income area, they may have a high value for PINCP_DECILE indicating that they have a high income for their PUMA).

The basic idea is that higher values of EDU should lead to higher values of PINCP_DECILE, and this is broadly true. However, it is known that the relationship between EDU and PINCP_DECILE is different for different demographic subgroups. The heatmaps in the left column below show the density distribution of the true data for each subgroup, normalized by education category (so the density values in each column sum to 1; note that when a cell in the heatmap contains too few people (< 20), it is left blank; its not expected that the deidentified data will match the original distribution precisely). The regression line is drawn in red over the heatmap, so you can see the relationship between the target data distribution and its linear regression analysis. In the right column for each subgroup we show how the deidentified data's regression line compares to the target data's regression line, along with a heatmap of the density differences between the two distributions. Redder areas are where the deidentified data has created too many people, bluer areas are where it's created too few people.

We've broken this metric down into demographic subgroups so we can see not only how well the privacy techniques preserve the overall relationship between these features, but also whether they preserve how that overall relationship is built up from the different relationships that hold at each major demographic subgroup. It's important that deidentification techniques preserve these distinct subgroup patterns for analysis.

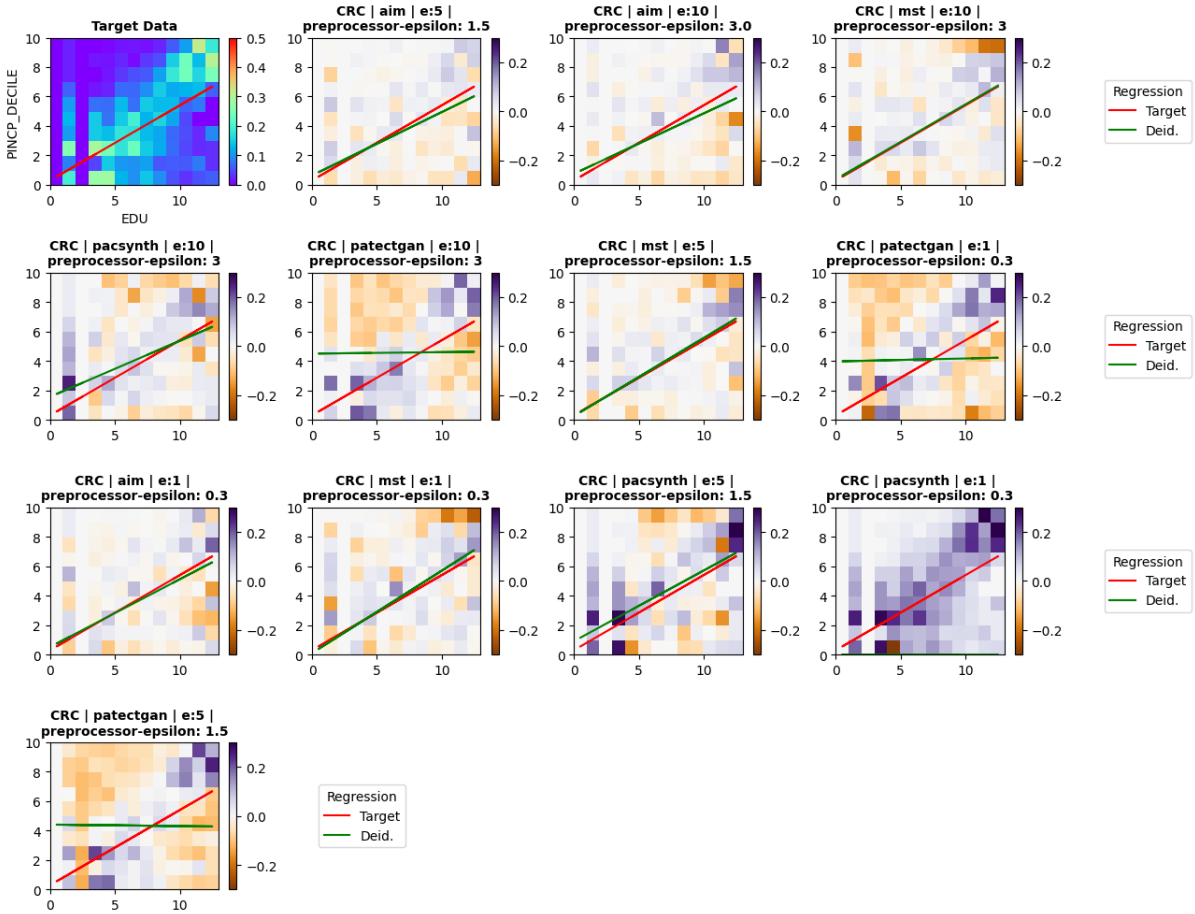
Feature Set: demographic-focused-except-AGEP-DEYE | Target Dataset: national2019:

Features: ['DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,135,134



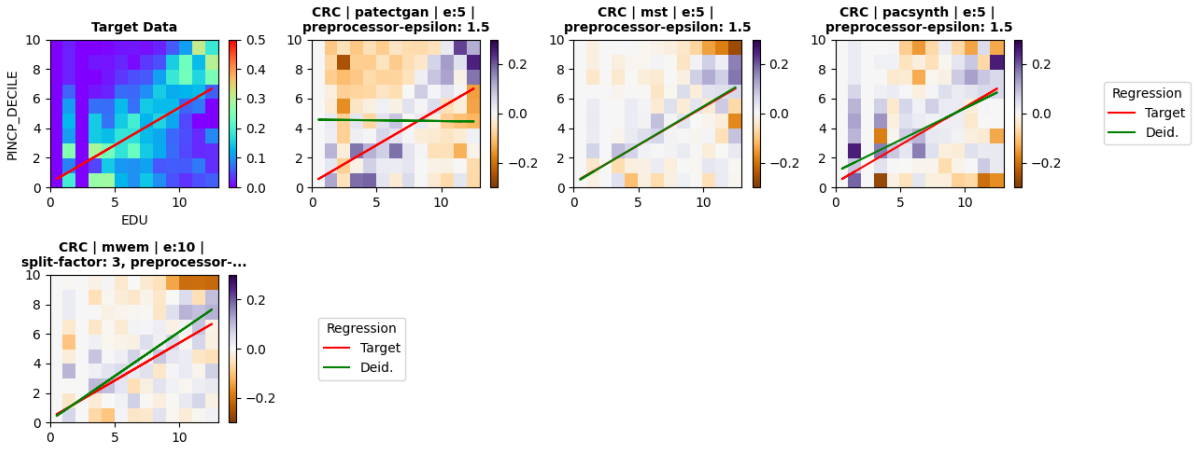
Feature Set: demographic-focused | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



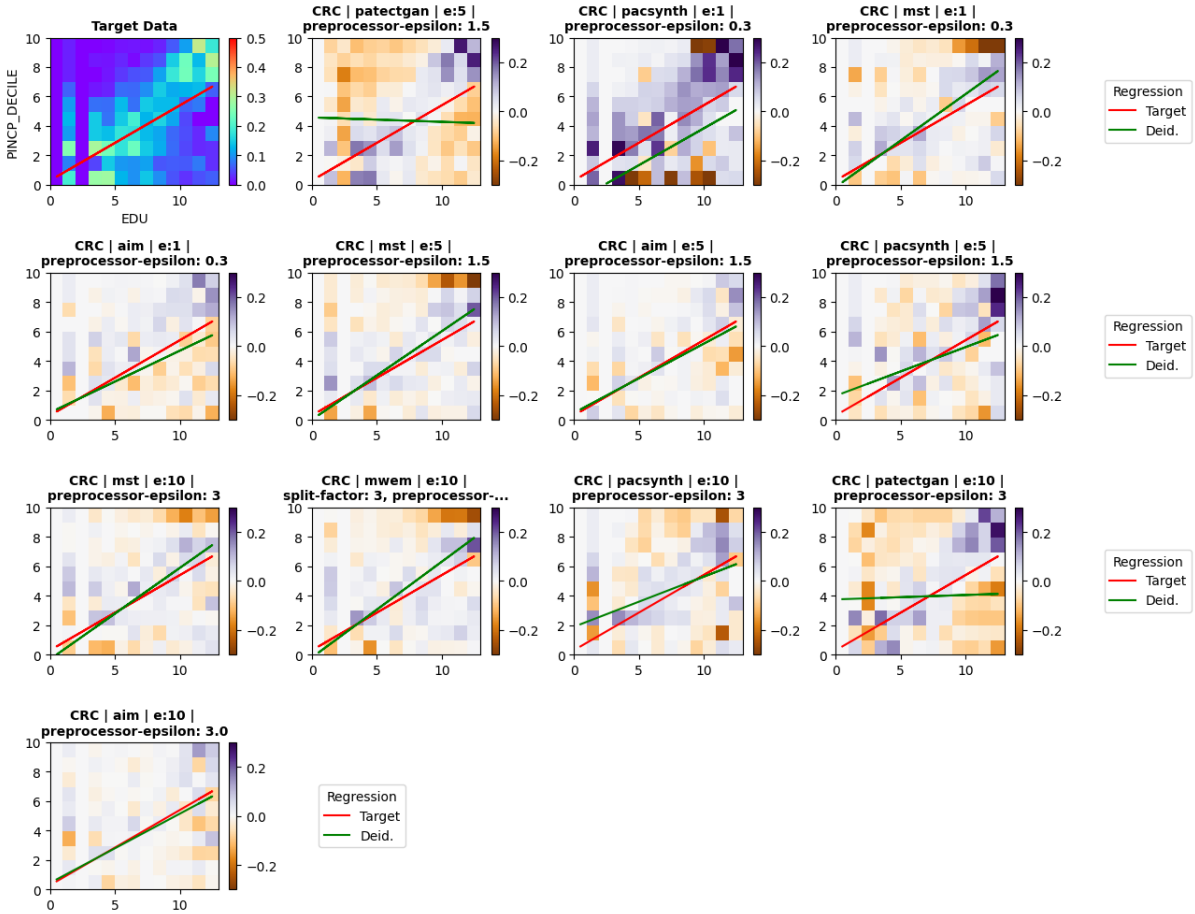
Feature Set: demographic-focused-except-DEYE | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



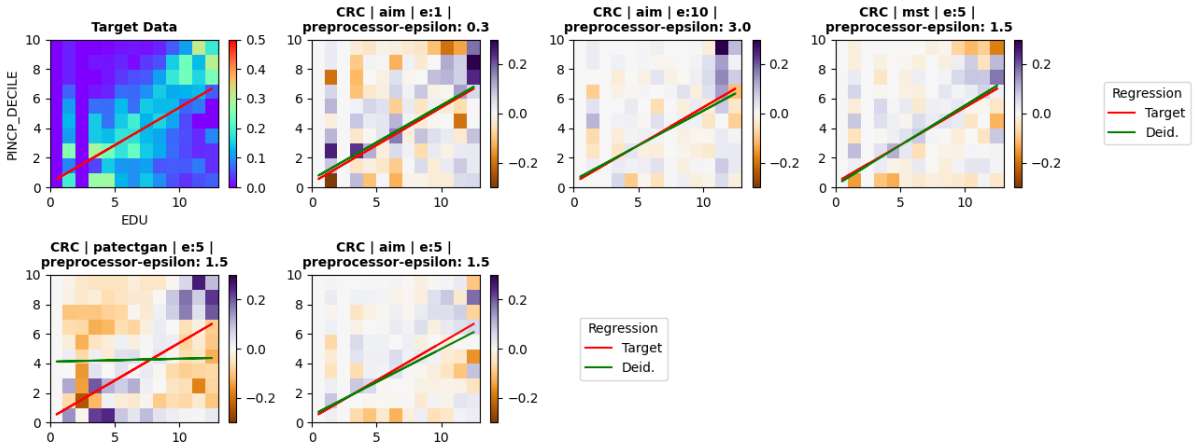
Feature Set: industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



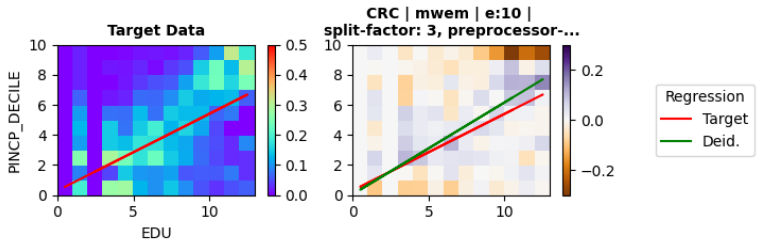
Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



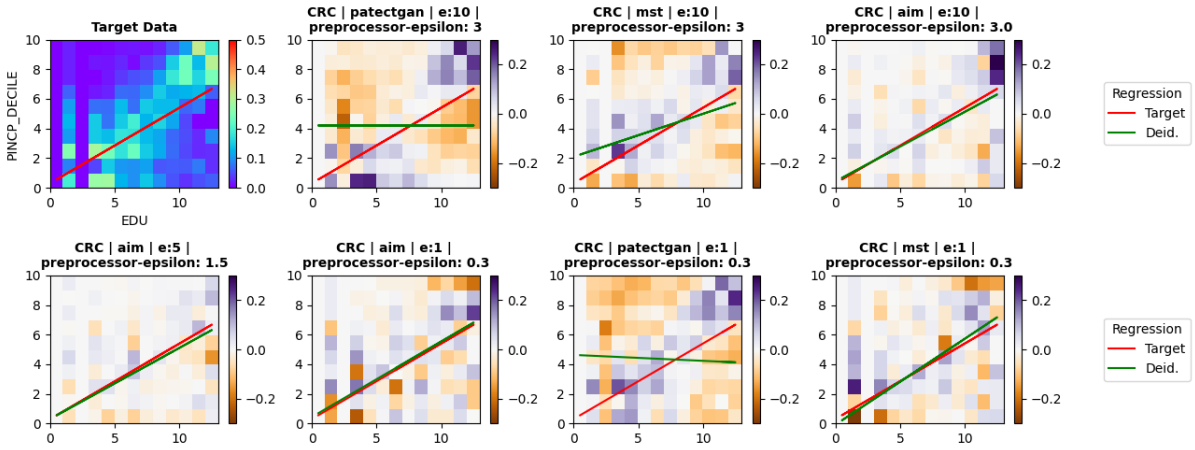
Feature Set: detailed-industry-focused | Target Dataset: national2019:

Features: ['EDU', 'HISP', 'INDP', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 1,427,025,600



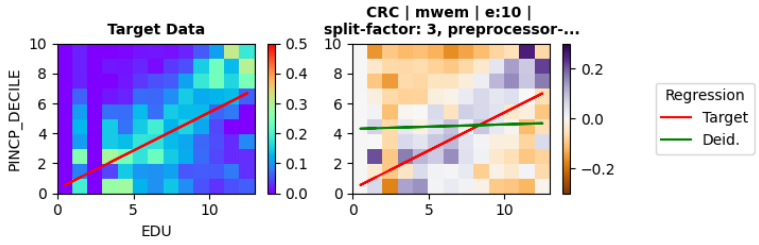
Feature Set: simple-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



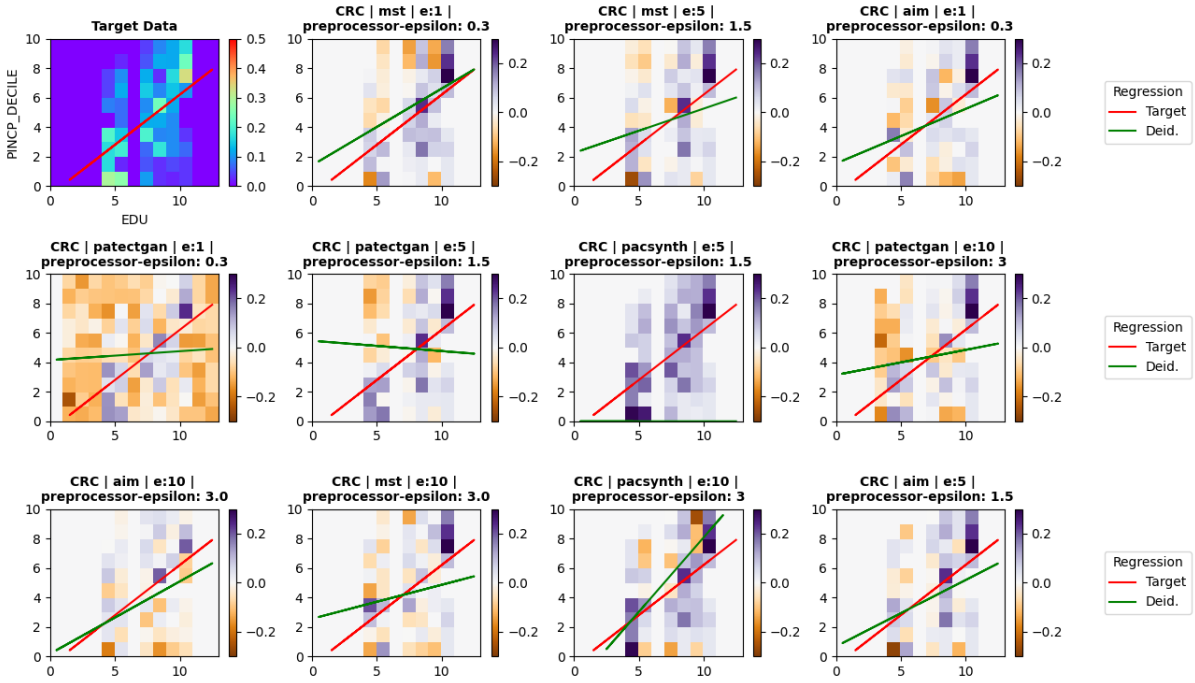
Feature Set: custom-features-12 | Target Dataset: national2019:

Features: ['AGEP', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'MSP', 'NOC', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400,000



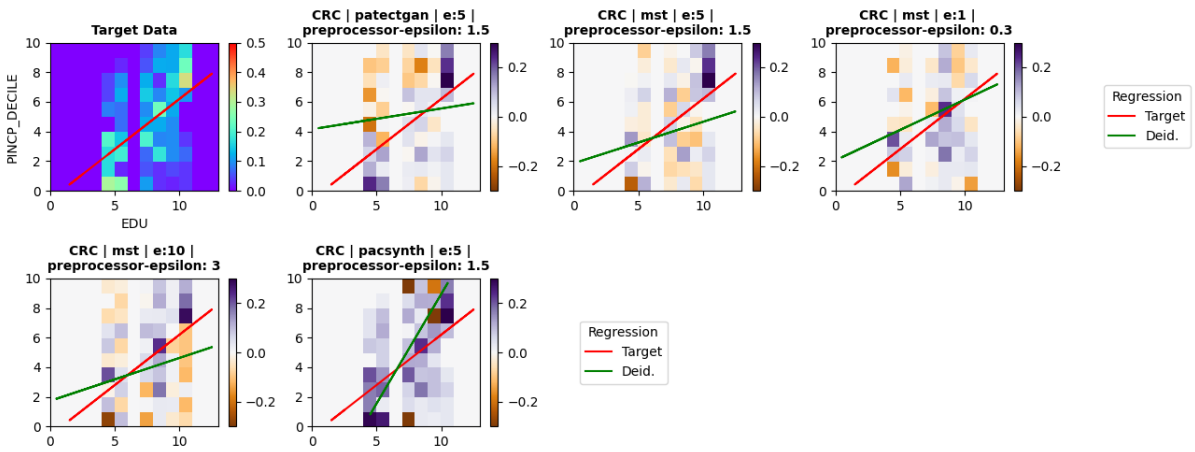
Feature Set: demographic-focused | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



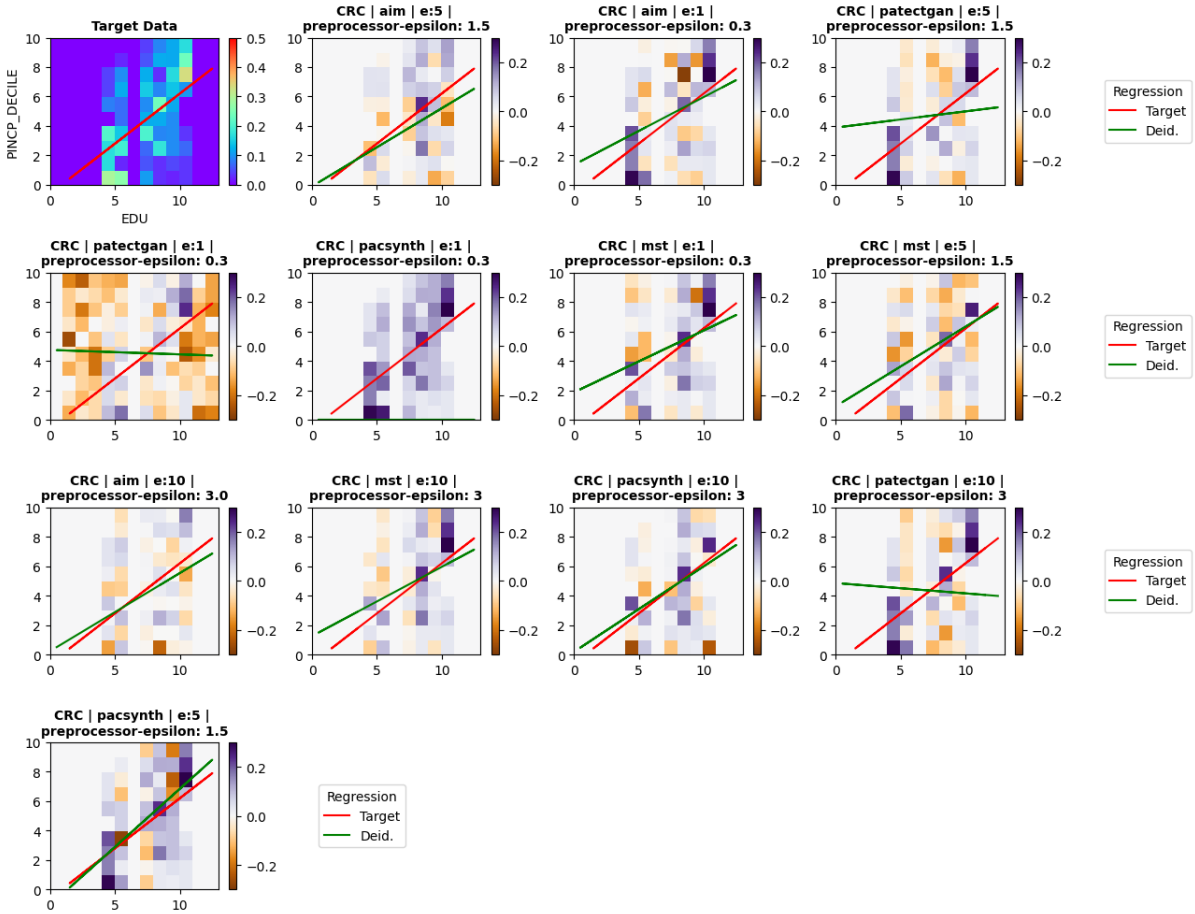
Feature Set: demographic-focused-except-DEYE | Target Dataset: tx2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



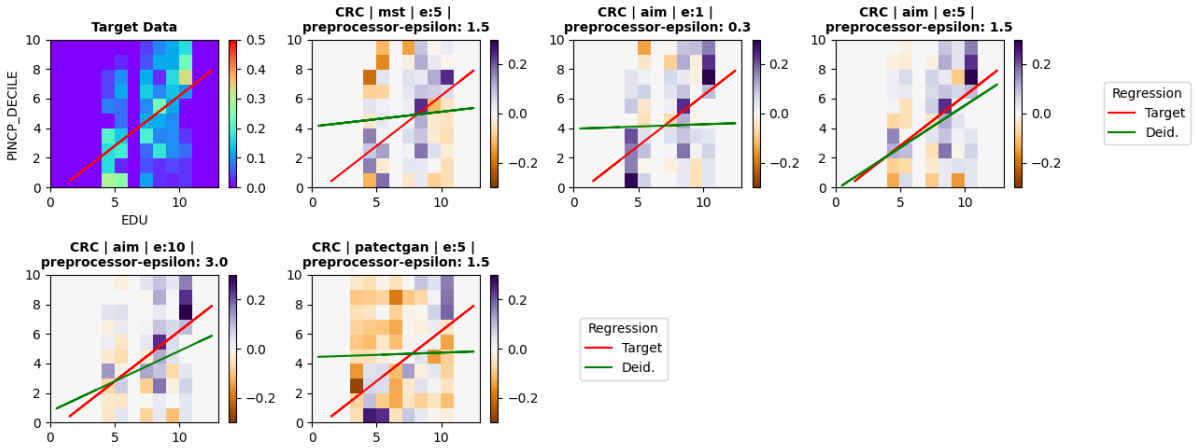
Feature Set: industry-focused | Target Dataset: tx2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



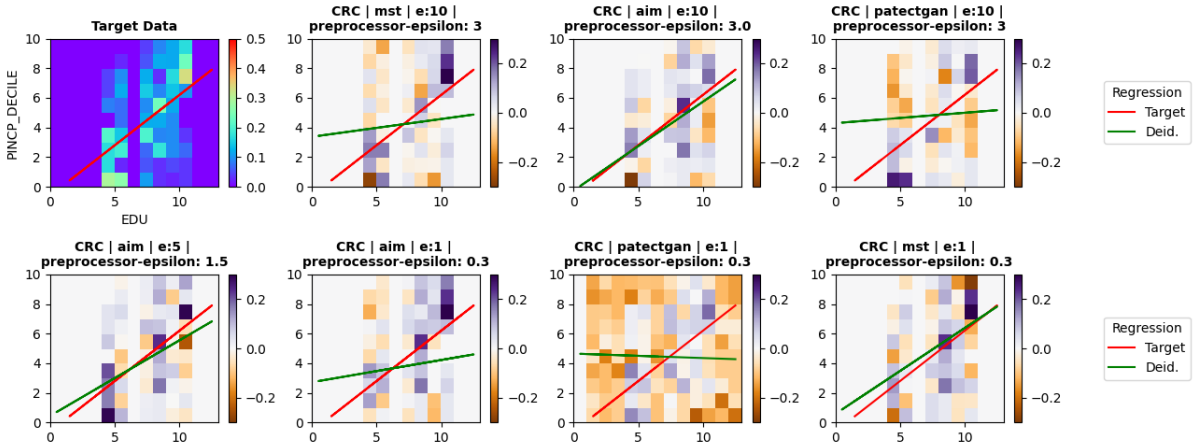
Feature Set: all-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



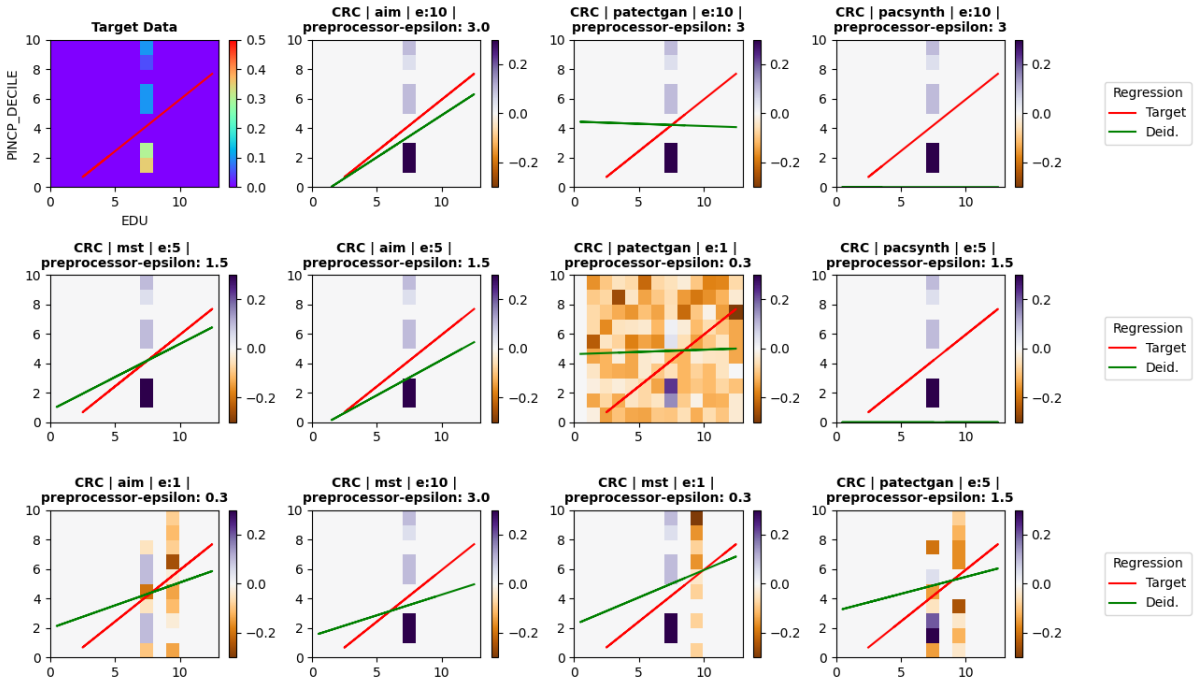
Feature Set: simple-features | Target Dataset: tx2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



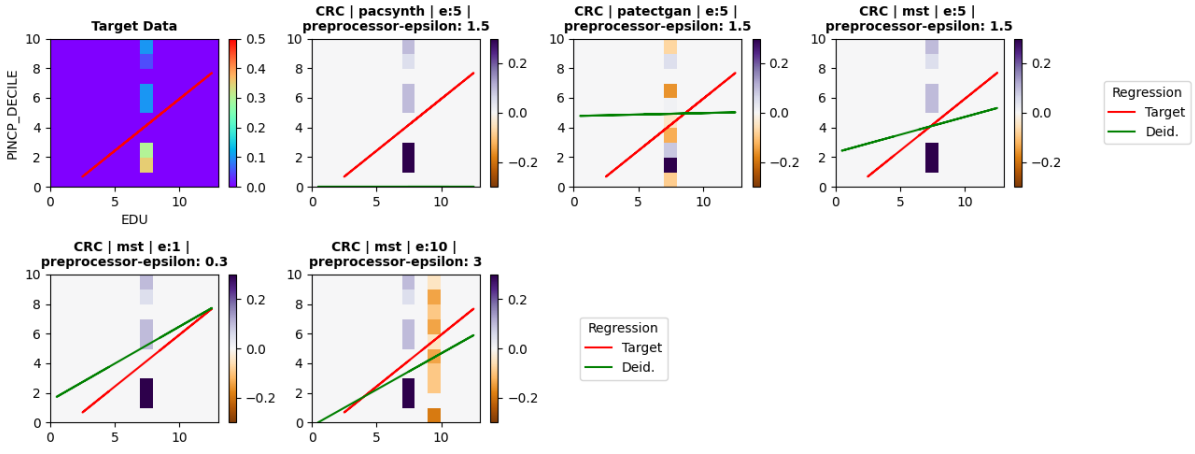
Feature Set: demographic-focused | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 227,026,800



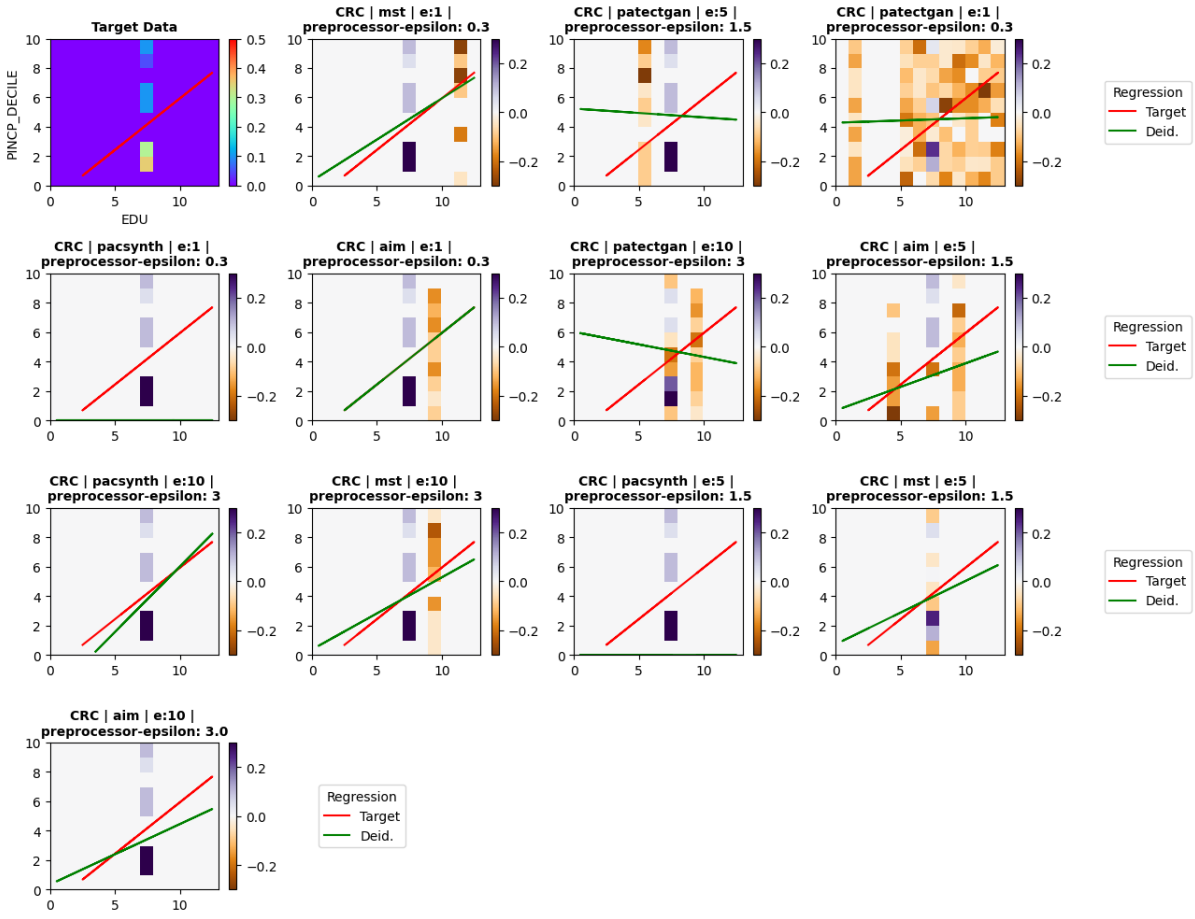
Feature Set: demographic-focused-except-DEYE | Target Dataset: ma2019:

Features: ['AGEP', 'DVET', 'EDU', 'HOUSING_TYPE', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'RAC1P', 'SEX']
Feature Space (possible combinations): 113,513,400



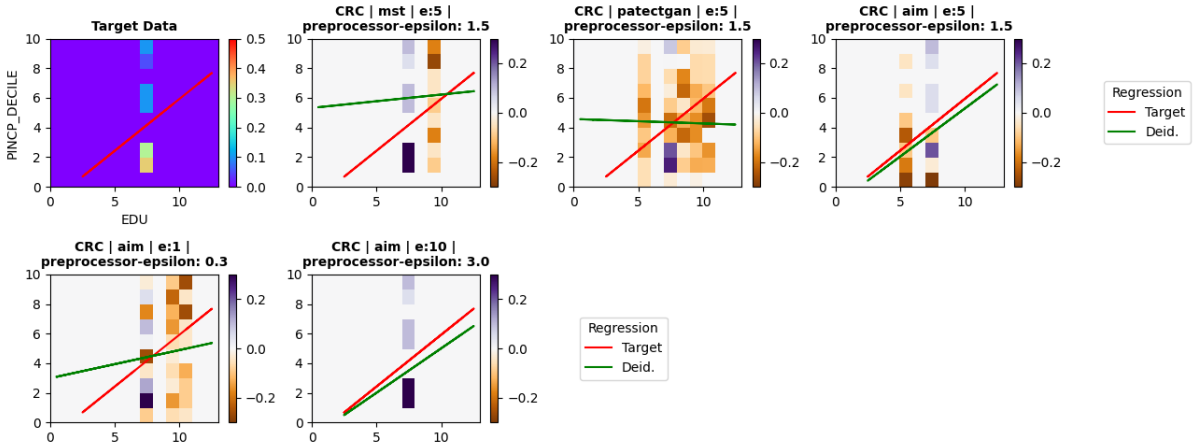
Feature Set: industry-focused | Target Dataset: ma2019:

Features: ['EDU', 'HISP', 'INDP_CAT', 'MSP', 'OWN_RENT', 'PINCP_DECILE', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 108,108,000



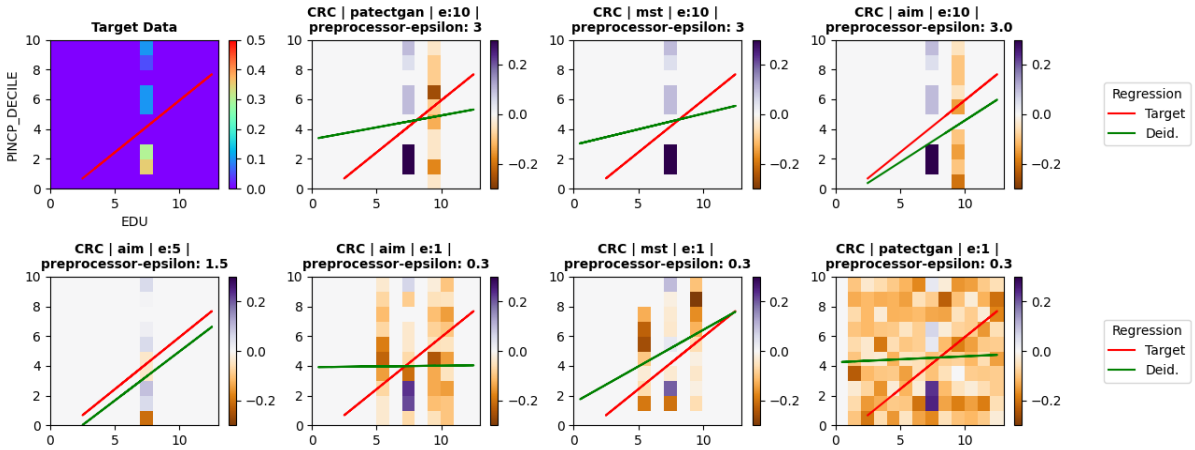
Feature Set: all-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'PWGTP', 'RAC1P', 'SEX', 'WGTP']
Feature Space (possible combinations): 517,839,049,728,000,000,000,000,000



Feature Set: simple-features | Target Dataset: ma2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING_TYPE', 'INDP_CAT', 'MSP', 'NOC', 'NPF', 'OWN_RENT', 'PINCP', 'PINCP_DECILE', 'POVPIP', 'PUMA', 'RAC1P', 'SEX']
Feature Space (possible combinations): 196,151,155,200,000,000,000



Observations

How do the two categories of approaches (marginal vs. neural network) differ from each other? Which one seems to perform the best?

Epsilon 10 is actually a very weak privacy guarantee. Which algorithms seem to have better or worse privacy in practice (ie, unique exact match metric) at the epsilon = 10 parameter setting?

Are there any properties or structures in the data that seem to cause difficulty for multiple different approaches? How do different feature sets affect performance? Can you recognize any 'artifacts' in the deidentified data (strange, structured deviations from the target distribution).

Which target data set seems to be the most difficult? Does that differ between different algorithms?

Data Description

Deidentified (Deid.) Datasets:

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-AGEP-DEYE
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677
Features List	SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Deid Data Id	efc2fa104f2454dbd556307f0bb591001dfa241b

Property	Value
Filename	aim_e_1_cf8
Records	27253
Features	8

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-AGEP-DEYE
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677
Features List	SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Deid Data Id	fb6965e304123a0e42b64c47c4b32ad874cbceff

Property	Value
Filename	aim_e_5_cf8
Records	27253
Features	8

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-AGEP-DEYE
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677
Features List	SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Deid Data Id	4e1828d11ca182a97f57ec6ac467c33419c7b581

Property	Value
Filename	aim_e_10_cf8
Records	27253
Features	8

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	b1dc323244c8ce6e503430f7055f83bebf883970
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_1_demographic
Records	9276
Features	10

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	8031989cee2202434abf2c38b30c9028943675e5
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_demographic
Records	9276
Features	10

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	0499dd8c3fad848a0827bed3e2c03e696bdfe642
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_demographic
Records	9276
Features	10

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	ec11a1e2afc50abefa651bf381ff48974248ba4f
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_demographic_focused_tx2019
Records	9276
Features	10

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	9b8fc36de40bb588b7bba567aaccf7c6a42db02f
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_demographic
Records	27253
Features	10

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	ad485b1eeb135222e94321c5534da9f5277b38a3
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_demographic
Records	27253
Features	10

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	4785c6e6a765361293d70f34a166b7f070fd6902
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_demographic
Records	9276
Features	10

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	3222405fc925dceee93cf323ea9a8f538937fcd3
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_demographic
Records	3319
Features	10

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	ace4ca9a5db6c2b9b7d6bc7a5a45abdcd4409782
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_demographic_focused_tx2019
Records	9276
Features	10

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	94b5f35973b02890deaafa766883e6024dd0d053
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_demographic
Records	7634
Features	10

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	31f694caa18638878c335973e10f019b27f2f2c1
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_demographic
Records	9276
Features	10

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	8d2686be13eab9890512ea526d886c6a6414eb9e
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_demographic_focused_ma2019
Records	7634
Features	10

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	506a9dc57d04140f26510caaa6e3aec47e0407b0
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e10_demographic_focused_na2019
Records	27253
Features	10

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	f74bbada21e12ce5905272fa5dde283d3b660d94
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_demographic_ma2019
Records	8344
Features	10

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	2c80801236f3504a9bb7fdb4930d099df448b81e
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_10_demographic_focused_na2019
Records	27472
Features	10

| mst | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	a7304d9fc5bef11c3305a693139360586620ffa7
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_10_demographic
Records	9276
Features	10

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	bae078d27f74a6f7743165d5230ca0b7b9b60653
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_demographic_tx2019
Records	10626
Features	10

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	502abdd38e2a95c49305ed6cfed5eb816ead0bfa
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_demographic_focused_na2019
Records	27253
Features	10

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	46bb7492cd13843cfd309086bff581a87c6aadc3
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_demographic
Records	27253
Features	10

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	0fa17c3a42f3fa8563a6bb292552a17074df6a8a
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_demographic
Records	7634
Features	10

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	ba6d88711205841da7e57e02752936db736a359b
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_demographic
Records	7634
Features	10

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	fc19e2743bc0e905611266506efedf2ed859b65c
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_demographic
Records	9276
Features	10

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	neural net
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	903da32c5136aa180647b8d64f9601fc0a3b7465
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_demographic
Records	27253
Features	10

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	cb7ae88ac0d8ac9091831c4cb12e5c99f07fd9d3
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_demographic
Records	27253
Features	10

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	7fddbec8634086c42dec03ac30f3ac7bc8a08e9a
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_demographic_focused_ma2019
Records	7634
Features	10

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	96225249f2c9e4efd22e96734715b1c2c2acc592
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_demographic_na2019
Records	27253
Features	10

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	a90c96b24751a715f1f3fa9d7982d4afe638aac6
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_demographic
Records	2101
Features	10

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	868ec48cd249f692191d5dc33d94e9d7104332bf
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_demographic
Records	7634
Features	10

| mst | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	ce992cc36a37938d34ac8668d435d608c22e0bdc
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_10_demographic
Records	7634
Features	10

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	1db2c49587abae371c12872c319ef9b9ec339ae3
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_1_demographic
Records	7634
Features	10

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	c38d28c27b6ebfd99fcd4845379c0201239ad1c2
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_demographic
Records	18372
Features	10

| pacsynth | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Privacy Category	dp
Deid Data Id	8e560a74f5748dbb17872f48907931476703ee8f
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_1_demographic_na2019
Records	1502
Features	10

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	2ae1b5e9b639d44c926d0a16cb3fe079b92339e8
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_demographic
Records	27253
Features	10

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	demographic-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET, DEYE
Deid Data Id	8662b2c370e82e18e51b2232b9ad2a01bf4ba0b0
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_demographic
Records	7634
Features	10

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	b64c6dbdb24b0779b84e13d40072afcda07f162e
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_demographic_except_DEYE
Records	9276
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	3381b70fbc90f9c42046da37afabbe7598a3ae29
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_demographic_except_DEYE
Records	5168
Features	9

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	d9007a56b0b86d69774149a8c80c049184975c48
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_demographic_except_DEYE
Records	7634
Features	9

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	6bfd2a6eec2c527becb3228d188114f0db7a5a3d
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_demographic_except_DEYE
Records	27253
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	185fd015c97fb873313347f115adc804ac261336
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_demographic_except_DEYE
Records	27253
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	d6305c272c28a4e82f3412e4ea4baaeb4ba4cad7
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_demographic_except_DEYE
Records	24845
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	006dca8fbdd0a530d06dfba14f16ca8adce0812d
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_demographic_except_DEYE
Records	7634
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	e41ec25108620d7965bf6f82916b64de73103a8f
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_demographic_except_DEYE
Records	9276
Features	9

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	0d546c0b8c876457c3300f85103a90a05b52fa99
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_demographic_except_DEYE_ma2019
Records	7634
Features	9

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	53ab8dcdab222dc8e51e124c4e9b8faaef2d8d8e
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_demographic_except_DEYE_tx2019
Records	9276
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	7cef749f52b3e39b3113211aa29848706a2f3866
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_demographic_except_DEYE_ma2019
Records	7634
Features	9

| mwem | | split-factor: 3, preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mwem
Target Dataset	national2019
Epsilon	10
Variant Label	split-factor: 3, preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	09fee0d92885042bc2e702c6d53e98a8544c1546
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Multiplicative Weights Exponential Mechanism. From â€œA Simple and Practical Algorithm for Differentially Private Data Releaseâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.1012.4763

Property	Value
Filename	mwem_e_10_c9_na2019
Records	27253
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	971c6b24ddbbad3a55ef378951cdecb4986e0cc4
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_demographic_except_DEYE_tx2019
Records	9276
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	demographic-focused-except-DEYE
Privacy Category	dp
Deid Data Id	42c61cb6c3caa78388f2f86214a7a5c3d80d7466
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_demographic_except_DEYE
Records	7973
Features	9

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	44cbd85b84434c87d57fd4dd98932a5a7bfdfc0f
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_industry
Records	9276
Features	9

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	5bec159744a0cf96c13b716d963dc78c28a13d5a
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_industry
Records	27253
Features	9

| pacsynth | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	2e501f20d48f35b5e92bb403249ad8d0c2e06a99
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_1_industry_na2019
Records	8513
Features	9

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	9bbb72f9941883832f67f51f483a7fbe6d29a080
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_industry
Records	9276
Features	9

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	95f827acc71472a99b9a733f6241c0a5ce0a69d4
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_industry
Records	9276
Features	9

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	7a502dfb87d78a96e4e50122e70a6322ff4ea6bd
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_industry_na2019
Records	27253
Features	9

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	1bc98a67e6f19b15e0101d2a516c27eaa262b859
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_industry
Records	27253
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	9d3c376ae70f799f15ca910f4a2a4c767c99b3a6
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_industry
Records	27253
Features	9

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	4ab593e9d112b42cd07c30adc18eec805c1e7f54
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_industry_ma2019
Records	7634
Features	9

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	c18d7b9fa03c57e0bc3c1685471a83c27cce2064
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_industry
Records	7634
Features	9

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	71a8cb7c3586cff24da333ec8b48a8a5a11acddf
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_industry_focused_ma2019
Records	7634
Features	9

| pacsynth | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	711f451cc6ab3e289c72c6fe76e34a00fa9958a0
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_1_industry_ma2019
Records	3240
Features	9

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	343f10b6b0553605b58a5514630a916c59c46b9a
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_industry
Records	7634
Features	9

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	462dbe0f2c6da27ca04790fa5224065d6bc5fb45
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_industry_focused_tx2019
Records	9276
Features	9

| pacsynth | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	00083665bddad26c4ea569f7787651e6868bb235
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_1_industry_tx2019
Records	4056
Features	9

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	04431d376506f728162e5c1235c621d4eb8f075c
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_industry_tx2019
Records	9276
Features	9

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	082b1e56292d799af0cc1ef700c5090bc6945cf3
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_industry
Records	27253
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	d70925821ab449edab2c790c5f44d0b4ec434611
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_industry
Records	9276
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	d683cc919395675a03652c2cdcb40b704fb0dd11
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_industry
Records	26612
Features	9

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	082ca80abb820674f5fb9723659fde68ea0b5448
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_industry_focused_ma2019
Records	7634
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	5c58e58b9649aa91a1b1081c2f9e424d0e333288
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e10_industry_focused_na2019
Records	27253
Features	9

| mwem | | split-factor: 3, preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mwem
Target Dataset	national2019
Epsilon	10
Variant Label	split-factor: 3, preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	2f81ed9a75beb1aa49f4f734e814debb7918da15
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Multiplicative Weights Exponential Mechanism. From â€œA Simple and Practical Algorithm for Differentially Private Data Releaseâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.1012.4763

Property	Value
Filename	mwem_e_10_indp_focused_na2019
Records	27253
Features	9

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	5ea31566b4eb19a6ee64a21fa8da4e485769af6a
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_industry
Records	7634
Features	9

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	b0a133c78c4b15eb14a47fc0267ab5926227648d
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_industry_focused_na2019
Records	29537
Features	9

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	09428317010945b07c1eb4a21e84bc9bf8377a7a
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_industry_focused_na2019
Records	27253
Features	9

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	20acbfafb860d34b3e3aaea46c1e6383a02bff32
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_industry
Records	9276
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	8436f92b9610d55b75a5cd8d421b12bfe40ded66
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_industry_tx2019
Records	9276
Features	9

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	26fa4a2ad25638cc5b325711f33d2918f8310eef
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_industry_ma2019
Records	13221
Features	9

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	fe8915d3c09a91131deb9e2143641c811405f330
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_industry_tx2019
Records	14855
Features	9

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	5583b71fcd41cb52b1116bb5cf8030a8ec3371b7
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_industry_focused_tx2019
Records	9276
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	0b3e9c1791f88cdc7f619ac57a23e328f38e655d
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_industry_ma2019
Records	7634
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	e45608d99ae95674826fee7c8437dd92f9691443
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_industry
Records	11131
Features	9

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	9498c4ced7ddb2dd5f6ec1460ad3401c0a747c52
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_industry
Records	7634
Features	9

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Privacy Category	dp
Deid Data Id	f84f79b22b0463f356204e8c7ab8ed1d6312b8b3
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_industry
Records	13230
Features	9

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	c3390d42d0fbff280f0f91e756dff504161374c9
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_industry
Records	7634
Features	9

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	industry-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP_CAT, EDU, PINCP_DECILE
Deid Data Id	dfac272b083233a9014c3b76b286f4fbd8eaf4f0
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_industry
Records	27253
Features	9

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	aad6974f832fb329f33fdff38c305ad3d8c699c4
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_family
Records	9276
Features	11

| mst | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	47309c163a2a15323a8b9e57eb54cdb90949ea99
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_10_family
Records	7634
Features	11

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	0a69bd584368244cdcc9d16e620d9059315a0a0c
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_family_focused_na2019
Records	27253
Features	11

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	neural net
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	74d75b85feeb12df1411ad4ae88bcfa719d86f06
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_family
Records	27253
Features	11

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	e9d4bb87e8daa30d288f2529f38233939f127678
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_family_ma2019
Records	2424
Features	11

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	209141e079a3561e4a1bd141b1acdf0156413f2d
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_1_family
Records	9276
Features	11

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	a7aef3f19b2fe6ae390097780baba8fbf66b2c62
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_family
Records	27253
Features	11

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	7e1d2c880b13b6f55bcc7b48201dc835c274b626
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_family_focused_ma2019
Records	7634
Features	11

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	02174de8579367c6c35f6c2e5fedb4349044880a
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_family
Records	27253
Features	11

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	a11416c6fa9863ade968f4bc266a9785c8e3f3c8
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_family
Records	9276
Features	11

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	e819cef32901c059fd00d73d26ee103ecabe1370
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_family
Records	1814
Features	11

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	cfbd9853673cafd44b4580f1fd6a182c7eedb5da
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_family
Records	7634
Features	11

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	c64f7bd6ef1942d88d7c4b687299bab1e40ed930
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_family
Records	9276
Features	11

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	bcd14617fabccf9a18e4389ded5266cbc45a67b7
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_family_focused_tx2019
Records	9276
Features	11

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	054138cd0b67906c163f8099e35654d9265cdafb
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_family_focused_na2019
Records	27253
Features	11

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	84b79aa7d5cb5520048b0b846bee9173a627daf3
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_family
Records	27253
Features	11

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	c67f4ce95e576dd04d1ae82a38461cb052142450
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_family
Records	27253
Features	11

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	08b818aec3732a76e8ff3aa5de06559607d70c8a
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_family
Records	27253
Features	11

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	11271bd92f5aec0d368b49d050b4ead89f407395
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_family
Records	7634
Features	11

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	3fd446badbabea3eb16c48d0597dbebce0b72842
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_family
Records	75
Features	11

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	805718ececeaa8e9ee2ffd008db3134fedd4a59a
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_family
Records	9276
Features	11

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	61971f4c39ef09012b2e62fc22a220b223ee27dd
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_1_family
Records	7634
Features	11

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	934acab8390c3741299bbd42e40f1d2c84810e9b
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_family_na2019
Records	27253
Features	11

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	68609f8dfc357dc4a436cece1b186e0059f74d59
Team	CRC
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	patectgan_e_5_family
Records	7634
Features	11

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	d1c505204c355d0920659bc51fd34d490ef2a63e
Team	CRC
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	pacsynth_e_5_family
Records	197
Features	11

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	e4e5ab974fa99c35bdd0228005c49d7766b97491
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_family_focused_ma2019
Records	7634
Features	11

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	4df2fe6efebb5bc3f7b6c944813e5de167aa5d28
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_5_family
Records	7634
Features	11

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	44e2092e155fe2b9b97c2286bc404ed42a8f0cd8
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_family_focused_na2019
Records	4579
Features	11

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	7f5181aecd2ced788b5a6dc8f9afd3176bb2f718
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_family
Records	9276
Features	11

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	34f077eec0ce71557336f9a50eeeeab1caf8b120
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_family_focused_tx2019
Records	9276
Features	11

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	3f0ddadba24689e2eff8145d40472bd44fd0699f
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_family
Records	7634
Features	11

| mst | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	mst
Library Name	smartnoise-synth
Feature Set Name	family-focused
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Deid Data Id	682f872f4ef8cbfd41d434210a95673b38a20ded
Team	CRC
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Research Papers	https://doi.org/10.48550/arXiv.2108.04978
Submission Timestamp	9/5/2023 00:00:00

Property	Value
Filename	mst_e_10_family
Records	9276
Features	11

| pacsynth | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused
Privacy Category	dp
Deid Data Id	8495bd84e21ae5f48582dc24dfd121f67c4b6fbb
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pac_synth_e_10_family_tx2019
Records	1479
Features	11

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	58e13f95cc69a9179a7ca24716c0d9e0d4aaa565
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_family_with_DEYE
Records	9276
Features	12

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	0003eee13499f4ae7510431525f7e7d9b3462cf3
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_family_with_DEYE_tx2019
Records	9276
Features	12

| mwem | | split-factor: 3, preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mwem
Target Dataset	national2019
Epsilon	10
Variant Label	split-factor: 3, preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	3300d7626304270473241cad81f4fb631a605187
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Multiplicative Weights Exponential Mechanism. From â€œA Simple and Practical Algorithm for Differentially Private Data Releaseâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.1012.4763

Property	Value
Filename	mwem_e_10_fam_with_DEYE_na2019
Records	27253
Features	12

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	aa08d8ea2ed75ebb27d5f7c05043349cb2e91883
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_family_with_DEYE_ma2019
Records	7634
Features	12

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	b94dacf7f27f2328d9090785e3f8c524adec198b
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_family_with_DEYE_ma2019
Records	7634
Features	12

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	aa4b53e0cacfc333ab033eaa1ca9e39f0b09c3e3
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_family_with_DEYE
Records	7634
Features	12

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	ddd80cb0418fceda76ccde5619297f5738a120f5
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_family_with_DEYE
Records	187
Features	12

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	a7f8557c1253e20ad078543aaf9d42ba184742e3
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_family_with_DEYE_tx2019
Records	9276
Features	12

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	2e4e4c82594dc80e1cb7c54cedcd5cdd39dc238d
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_family_with_DEYE
Records	7634
Features	12

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	c298000ccf0e13d74817c950b511dbd2c1cc00f8
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_family_with_DEYE
Records	27253
Features	12

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	dc469a8c3adfe54134141de9d4063087af891b33
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_family_with_DEYE
Records	27253
Features	12

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	c4558d1b8273d7c0de0932f996bbaaec0116f9bd
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_family_with_DEYE
Records	9276
Features	12

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	8a52ec3da67016c5c2393f85d5d051d9c41f9a9c
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_family_with_DEYE
Records	96
Features	12

| pacsynth | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	pacsynth
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	family-focused-with-DEYE
Privacy Category	dp
Deid Data Id	c31c4bd3fa38be629c5dfc825326aa456bb21026
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, OWN_RENT, PINCP_DECILE, POVPIP, DEYE
Privacy Label Detail	The pac-synth synthesizer will suppress marginal combinations that could uniquely fingerprint individuals (similar to k-anonymity). pac-synth is a differentially-private synthesizer that computes differentially private marginals to build synthetic data.
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://github.com/microsoft/synthetic-data-showcase/tree/main/docs/dp

Property	Value
Filename	pacsynth_e_5_family_with_DEYE
Records	1057
Features	12

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	9b3e1e4b5a9e02a7841552f386d7f438031d53d4
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_all
Records	9276
Features	24

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	1bf69ee3868bb34d680b2ca73a44407cce604e60
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_all
Records	27253
Features	24

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	c88cfa5762e113b08b9015b8b8bb31503d74e9c3
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_all
Records	9276
Features	24

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	f2bc3757f7960c90733c187b754d016cdb65694e
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_all
Records	9276
Features	24

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	51547e42438af33f574890429bd3f6074eb6507d
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_all
Records	9276
Features	24

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	2205aa483ba9174a6dffc4bbf6872d23046f8212
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_all
Records	27253
Features	24

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	c90f4014d1e3b67ca4c1932a820edf75364912ea
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_all
Records	27253
Features	24

| mst | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	47d86ee19426f71e1e24ed9108ecf82577e36827
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_5_all
Records	7634
Features	24

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	9475d3b11b5f843a41b249626a93859c19491538
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_all
Records	7634
Features	24

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	0b71f55b35376ccc0ddd9f2306d5be08f83319f1
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_all
Records	7634
Features	24

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	ae683726072017edbae5e8be7b4f8b0d67384b2e
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_all
Records	27253
Features	24

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	ec3a4517ad57e5005b056455675e08ca05cd80aa
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_all
Records	27253
Features	24

| patectgan | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	all-features
Privacy Category	dp
Deid Data Id	d193f5780998c7e8c631e659230a1ee26635f390
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_5_all
Records	9276
Features	24

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	58b33e5a776d6c5b01fc76ebd86785424ded8534
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_all
Records	7634
Features	24

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	all-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	e7d1a6fdd69b27bc54a5ffcb56b4333ab97e5e96
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_all
Records	7634
Features	24

| mwem | | split-factor: 3, preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mwem
Target Dataset	national2019
Epsilon	10
Variant Label	split-factor: 3, preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	detailed-industry-focused
Privacy Category	dp
Deid Data Id	7356236b049baebf88d6ea081242a33b972964ba
Features List	PUMA, SEX, MSP, HISP, RAC1P, OWN_RENT, INDP, EDU, PINCP_DECILE
Privacy Label Detail	Multiplicative Weights Exponential Mechanism. From â€œA Simple and Practical Algorithm for Differentially Private Data Releaseâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.1012.4763

Property	Value
Filename	mwem_e_10_det_indp_focused_na2019
Records	27253
Features	9

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	43580c159faf094b7535c561d24636bb72286d5d
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_simple_tx2019
Records	9276
Features	21

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	595813734f27d11874a1412cbe1964ccd1e7c711
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_simple
Records	9276
Features	21

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	343edd37a4cd2177fe380bab198ecfb2b41b2735
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_simple_features_na2019
Records	27253
Features	21

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	ea02977d2beea53fec7c7229d311748fdbdfab88
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_simple_na2019
Records	27253
Features	21

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	c4230a41d5ac8a0017553587701d3da8af25106a
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_simple
Records	27253
Features	21

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	18d870198862e25ee994964c5dcd452093772b3b
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_simple_features_ma2019
Records	7634
Features	21

| mst | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	ce36740d1636cb149dd081e4e01615ca3ded7978
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_10_simple_ma2019
Records	7634
Features	21

| aim | | preprocessor-epsilon: 3.0:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 3.0
Epsilon	10
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	c8932ec4c63f0ed1783009799e824c3876b440e9
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_10_simple
Records	7634
Features	21

| patectgan | | preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	10
Variant Label	preprocessor-epsilon: 3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	a22fd858c15643efff27cf6a0f984a75a16adadb
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_10_simple_features_tx2019
Records	9276
Features	21

| mwem | | split-factor: 3, preprocessor-epsilon: 3:

Label Name	Label Value
Algorithm Name	mwem
Target Dataset	national2019
Epsilon	10
Variant Label	split-factor: 3, preprocessor-epsilon: 3
Algorithm Type	query matching
Library Name	smartnoise-synth
Feature Set Name	custom-features-12
Privacy Category	dp
Deid Data Id	297bae0ce01d90dc7a58cb2f08adb1d54051d256
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, HOUSING_TYPE, OWN_RENT, EDU, PINCP_DECILE, DVET
Privacy Label Detail	Multiplicative Weights Exponential Mechanism. From â€œA Simple and Practical Algorithm for Differentially Private Data Releaseâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.1012.4763

Property	Value
Filename	mwem_e_10_cf12_na2019
Records	27253
Features	12

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	408ef3ae623efcb135195a4eabb70d5ec06d364d
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_simple
Records	9276
Features	21

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	7532b0e396ec551dc9651bbafeddb22bb6a2b7cc
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_simple
Records	27253
Features	21

| aim | | preprocessor-epsilon: 1.5:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 1.5
Epsilon	5
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	97701073872cbcf1dd9598f9c795c54eb5f63a2b
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_5_simple
Records	7634
Features	21

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	tx2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	dcbdb11cdb53869947fe469e89f74e147723f5b0
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_simple
Records	9276
Features	21

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	06e8de7b0a5cdef6e79c276f96f59f5ac6f08dfc
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_simple_features_tx2019
Records	9276
Features	21

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	tx2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	31cde42525a8f74539b1f535a2b6d0ed18ec63a0
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_simple_tx2019
Records	9276
Features	21

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	national2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	02e738d10970a5c45853e6d4276f7ba065eff981
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_simple
Records	27253
Features	21

| aim | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	aim
Library Name	smartnoise-synth
Feature Set Name	simple-features
Target Dataset	ma2019
Variant Label	preprocessor-epsilon: 0.3
Epsilon	1
Algorithm Type	stat model
Privacy Category	dp
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Deid Data Id	c7e37f89a95828ddf3ea674f19951d3b79b8b880
Team	CRC
Submission Timestamp	9/5/2023 00:00:00
Privacy Label Detail	AIM is a workload-adaptive algorithm that first selects a set of queries, then measures those queries with added noise to satisfy differential privacy, and finally generates synthetic data from the noisy measurements. It uses a set of innovative features to iteratively select the most useful measurements, reflecting both their relevance to the workload and their value in approximating the input data. We also provide analytic expressions to bound per-query error with high probability.
Research Papers	https://doi.org/10.48550/arXiv.2201.12677

Property	Value
Filename	aim_e_1_simple
Records	7634
Features	21

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	87a56494d7e7918df5941908739738426 91eba27
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_simple_ma2019
Records	7634
Features	21

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	d2bae376f0c8d58bb3109007ec52d84b2a9b4a55
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_simple_features_na2019
Records	27253
Features	21

| mst | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	mst
Target Dataset	national2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	stat model
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	93f0a6cf2eda899c0059dbc2ff90427ff6d3759c
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Differentially private synthetic data. From McKenna et al. â€œWinning the NIST Contest: A scalable and general approach to differentially private synthetic dataâ€
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://doi.org/10.48550/arXiv.2108.04978

Property	Value
Filename	mst_e_1_simple_na2019
Records	27253
Features	21

| patectgan | | preprocessor-epsilon: 0.3:

Label Name	Label Value
Algorithm Name	patectgan
Target Dataset	ma2019
Epsilon	1
Variant Label	preprocessor-epsilon: 0.3
Algorithm Type	neural net
Library Name	smartnoise-synth
Feature Set Name	simple-features
Privacy Category	dp
Deid Data Id	4659211f8e39de76647b33dd962a4ab669ca6fec
Features List	PUMA, AGEP, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPIP, DVET, DREM, DPHY, DEYE, DEAR
Privacy Label Detail	Conditional tabular GAN using Private Aggregation of Teacher Ensembles. From â€œPATE-GAN: Generating Synthetic Data with Differential Privacy Guaranteesâ€ and â€œModeling Tabular data using Conditional GANâ€ .
Submission Timestamp	5/20/2023 00:00:00
Team	CRC
Research Papers	https://openreview.net/forum?id=S1zk9iRqF7

Property	Value
Filename	patectgan_e_1_simple_features_ma2019
Records	7634
Features	21

Appendix

Data Dictionary:

PUMA: Public use microdata area code:

PUMA Code	Code Description
25-00503	Middlesex County--Waltham City, Lexington, Burlington, Bedford & Lincoln Towns
25-00703	Essex County (East)--Salem, Beverly, Gloucester & Newburyport Cities
25-01000	Peabody City, Danvers, Reading, North Reading & Lynnfield Towns
25-01300	Billerica, Andover, Tewksbury & Wilmington Towns
25-02800	Woburn, Melrose Cities, Saugus, Wakefield & Stoneham Towns
48-02510	Tarrant County (North)--North Richland Hills (North) & Keller Cities
48-02102	Johnson County
48-02101	Ellis County
48-02515	Tarrant County (West)--Fort Worth City (West)
48-02507	Tarrant County (East)--Arlington City (West)--South of I-30 & East of Loop I-820
48-02516	Tarrant County (Southwest)--Fort Worth (Southwest) & Benbrook Cities
01-01301	Birmingham City (West)
06-07502	San Francisco County (North & East)--North Beach & Chinatown
06-08507	Santa Clara County (Southwest)--Cupertino, Saratoga Cities & Los Gatos Town
08-00803	Boulder County (Central)--Boulder City
13-04600	Atlanta Regional Commission--Fulton County (Central)--Atlanta City (Central)
17-03529	Chicago City (South)--South Shore, Hyde Park, Woodlawn, Grand Boulevard & Douglas
17-03531	Chicago City (South)--Auburn Gresham, Roseland, Chatham, Avalon Park & Burnside
19-01700	Des Moines City
24-01004	Montgomery County (South)--Bethesda, Potomac & North Bethesda
26-02702	Washtenaw County (East Central)--Ann Arbor City Area
28-01100	Central Region--Jackson City (East & Central)
29-01901	St. Louis City (North)
30-00600	East Montana (Outside Billings City)
32-00405	Las Vegas City (Southeast)
36-03710	NYC-Bronx Community District 1 & 2--Hunts Point, Longwood & Melrose
36-04010	NYC-Brooklyn Community District 17--East Flatbush, Farragut & Rugby
38-00100	West North Dakota--Minot City
40-00200	Cherokee, Sequoyah & Adair Counties
51-01301	Arlington County (North)
51-51255	Alexandria City

AGEP: Person's age:

AGEP Code	Code Description
min	0
max	99

SEX: Person's gender:

SEX Code	Code Description
1	Male
2	Female

MSP: Marital Status:

MSP Code	Code Description
N	N/A (age less than 15 years)
1	Now married, spouse present
2	Now Married, spouse absent
3	Widowed
4	Divorced
5	Separated
6	Never married

HISP: Hispanic origin:

HISP Code	Code Description
0	Not Spanish/Hispanic/Latino
1	Mexican
2	Puerto Rican
3	Cuban
4	All other Spanish/Hispanic/Latino

RAC1P: Person's Race:

RAC1P Code	Code Description
1	White alone
2	Black or African American alone
3	American Indian alone
4	Alaska Native alone
5	American Indian and Alaska Native tribes specified; or American Indian or Alaska Native, not specified and no other races
6	Asian alone
7	Native Hawaiian and Other Pacific Islander alone
8	Some Other Race alone
9	Two or More Races

NOC: Number of own children in household (unweighted):

NOC Code	Code Description
N	N/A (GQ/vacant)
0	No own children
min	1
max	19

NPF: Number of persons in family (unweighted):

NPF Code	Code Description
N	N/A (GQ/vacant/non-family household
min	2
max	20

HOUSING_TYPE: Housing unit or group quarters:

HOUSING_TYPE Code	Code Description
1	Housing Unit
2	Institutional Group Quarters
3	Non-institutional Group Quarters

OWN_RENT: Housing unit rented or owned:

OWN_RENT Code	Code Description
0	Group quarters
1	Own housing unit
2	Rent housing unit

DENSITY: Population density among residents of each PUMA:

DENSITY Code	Code Description
min	16.3
max	52864.7

Density Bin: 0 | Bin Range: (0, 150]

PUMA	DENSITY	PUMA NAME
30-00600	16.3	East Montana (Outside Billings City)
38-00100	73.0	West North Dakota--Minot City
40-00200	90.7	Cherokee, Sequoyah & Adair Counties

Density Bin: 2 | Bin Range: (309.67, 475.62]

PUMA	DENSITY	PUMA NAME
48-02101	357.4	Ellis County
48-02102	450.9	Johnson County

Density Bin: 5 | Bin Range: (1121.99, 1723.27]

PUMA	DENSITY	PUMA NAME
25-01300	1457.2	Billerica, Andover, Tewksbury & Wilmington Towns
48-02516	1338.4	Tarrant County (Southwest)--Fort Worth (Southwest) & Benbrook Cities

Density Bin: 6 | Bin Range: (1723.27, 2646.76]

PUMA	DENSITY	PUMA NAME
25-00703	2195.3	Essex County (East)--Salem, Beverly, Gloucester & Newburyport Cities
25-01000	2447.1	Peabody City, Danvers, Reading, North Reading & Lynnfield Towns
48-02515	2134.8	Tarrant County (West)--Fort Worth City (West)

Density Bin: 7 | Bin Range: (2646.76, 4065.16]

PUMA	DENSITY	PUMA NAME
01-01301	2731.2	Birmingham City (West)
06-08507	3305.1	Santa Clara County (Southwest)--Cupertino, Saratoga Cities & Los Gatos Town
08-00803	3393.2	Boulder County (Central)--Boulder City
13-04600	3670.4	Atlanta Regional Commission--Fulton County (Central)--Atlanta City (Central)
19-01700	3572.3	Des Moines City
25-00503	2872.7	Middlesex County--Waltham City, Lexington, Burlington, Bedford & Lincoln Towns
25-02800	3683.9	Woburn, Melrose Cities, Saugus, Wakefield & Stoneham Towns
28-01100	2674.3	Central Region--Jackson City (East & Central)
48-02507	3731.1	Tarrant County (East)--Arlington City (West)--South of I-30 & East of Loop I-820
48-02510	3092.4	Tarrant County (North)--North Richland Hills (North) & Keller Cities

Density Bin: 8 | Bin Range: (4065.16, 6243.68]

PUMA	DENSITY	PUMA NAME
24-01004	4187.9	Montgomery County (South)--Bethesda, Potomac & North Bethesda
26-02702	4817.2	Washtenaw County (East Central)--Ann Arbor City Area
29-01901	5434.8	St. Louis City (North)

Density Bin: 9 | Bin Range: (6243.68, 9589.66]

PUMA	DENSITY	PUMA NAME
32-00405	7990.5	Las Vegas City (Southeast)

Density Bin: 10 | Bin Range: (9589.66, 14728.75]

PUMA	DENSITY	PUMA NAME
17-03531	11171.6	Chicago City (South)--Auburn Gresham, Roseland, Chatham, Avalon Park & Burnside
51-01301	11162.8	Arlington County (North)
51-51255	11224.3	Alexandria City

Density Bin: 11 | Bin Range: (14728.75, 22621.88]

PUMA	DENSITY	PUMA NAME
17-03529	15097.5	Chicago City (South)--South Shore, Hyde Park, Woodlawn, Grand Boulevard & Douglas

Density Bin: 12 | Bin Range: (22621.88, 34744.92]

PUMA	DENSITY	PUMA NAME
06-07502	33632.6	San Francisco County (North & East)--North Beach & Chinatown

Density Bin: 13 | Bin Range: (34744.92, 53364.7]

PUMA	DENSITY	PUMA NAME
36-03710	52864.7	NYC-Bronx Community District 1 & 2--Hunts Point, Longwood & Melrose
36-04010	50441.6	NYC-Brooklyn Community District 17--East Flatbush, Farragut & Rugby

INDP: Industry codes:

See codes in ACS data dictionary. Find codes by searching the string: INDP, in the ACS data dictionary

INDP_CAT: Industry categories:

INDP_CAT Code	Code Description
N	N/A (less than 16 years old, or last worked more than 5 years ago, or never worked)
0	AGR: Agriculture, Forestry, Fishing and Hunting
1	EXT: Mining, Quarrying, and Oil and Gas Extraction
2	UTL: Utilities
3	CON: Construction
4	MFG: Manufacturing
5	WHL: Wholesale Trade
6	RET: Retail Trade
7	TRN: Transportation and Warehousing
8	INF: Information
9	FIN: Finance, Insurance, Real Estate
10	PRF: Professional, Scientific and Technical Services
11	EDU: Educational Services
12	MED: Health Care
13	SCA: Social Assistance
14	ENT: Arts, Entertainment, Accommodation, Food Services and Recreation
15	SRV: Other Services
16	ADM: Government, Public Administration
17	MIL: Military
18	UNEMPLOYED

EDU: Educational attainment:

EDU Code	Code Description
N	N/A (less than 3 years old)
1	No schooling completed
2	Nursery school, Preschool, or Kindergarten
3	Grade 1 to grade 8
4	Grade 9 to grade 12, no diploma
5	High School diploma
6	GED
7	Some College, no degree
8	Associate degree
9	Bachelors degree
10	Masters degree
11	Professional degree
12	Doctorate degree

PINCP: Person's total income in dollars:

PINCP Code	Code Description
N	N/A (less than 15 years old)
min	-9000
max	1341000

PINCP_DECILE: Person's total income rank (with respect to their state) discretized into 10% bins.:

PINCP_DECILE Code	Code Description
N	N/A (less than 15 years old
9	90th percentile
8	80th percentile
7	70th percentile
6	60th percentile
5	50th percentile
4	40th percentile
3	30th percentile
2	20th percentile
1	10th percentile
0	0th percentile

POVPIP: Income-to-poverty ratio (ex: 250 = 2.5 x poverty line):

POVPIP Code	Code Description
N	N/A
min	0
max	500
501	income above 5 x poverty line

DVET: Veteran service connected disability rating (percentage):

DVET Code	Code Description
N	N/A (No service-connected disability/never served in military
1	0 percent
2	10 or 20 percent
3	30 or 40 percent
4	50 or 60 percent
5	70, 80, 90 or 100 percent
6	Not reported

DREM: Cognitive difficulty:

DREM Code	Code Description
N	N/A (Less than 5 years old)
1	Yes
2	No

DPHY: Ambulatory (walking) difficulty:

DPHY Code	Code Description
N	N/A (Less than 5 years old)
1	Yes
2	No

DEYE: Vision difficulty:

DEYE Code	Code Description
1	Yes
2	No

DEAR: Hearing difficulty:

DEAR Code	Code Description
1	Yes
2	No

WGTP: Housing unit sampling weight:

See description of weights.

WGTP Code	Code Description
0	Group quarters place holder record
min	1
max	9999

PWGTP: Person's sampling weight:

See description of weights.

PWGTP Code	Code Description
min	1
max	9999