

# NIST CRC Meta Report

## YData SDK

Report created on: January 31, 2024 19:41:07

Created with [SDNIST v2.3.0](#)

## Motivation

We have deidentified data samples from the following techniques in this library:

Algorithm Name: **YData Fabric Synthesizers**

*Proprietary ensemble of Deep NN and Machine Learning generators. The process is automated with semantic analysis for the identification of data types, encoding selection, and data processing.*

[More Info on CRC Website](#)

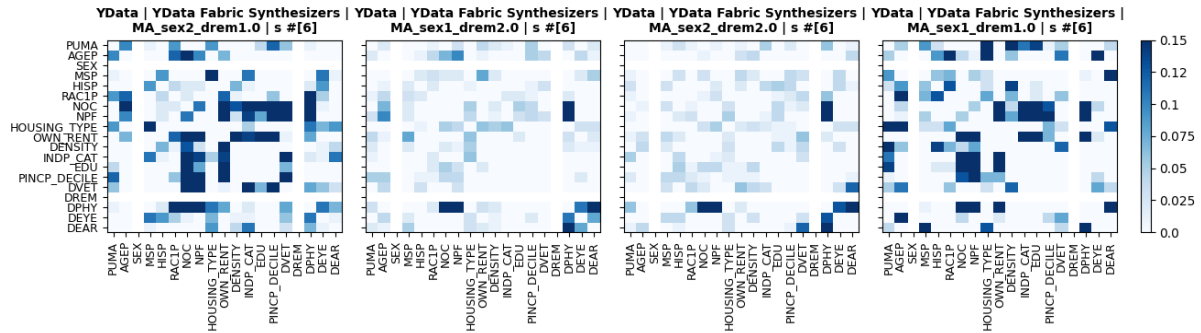
# Comparisons

## 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.

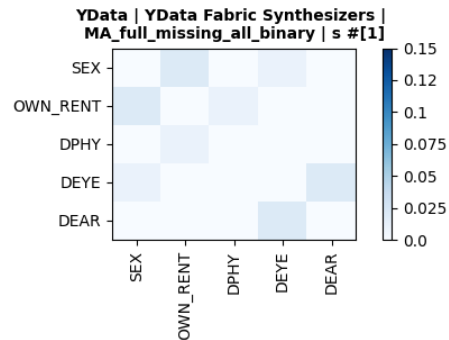
### Feature Set: all-features I Target Dataset: ma2019:

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): 13,330,105,588,800,000,000,000,000



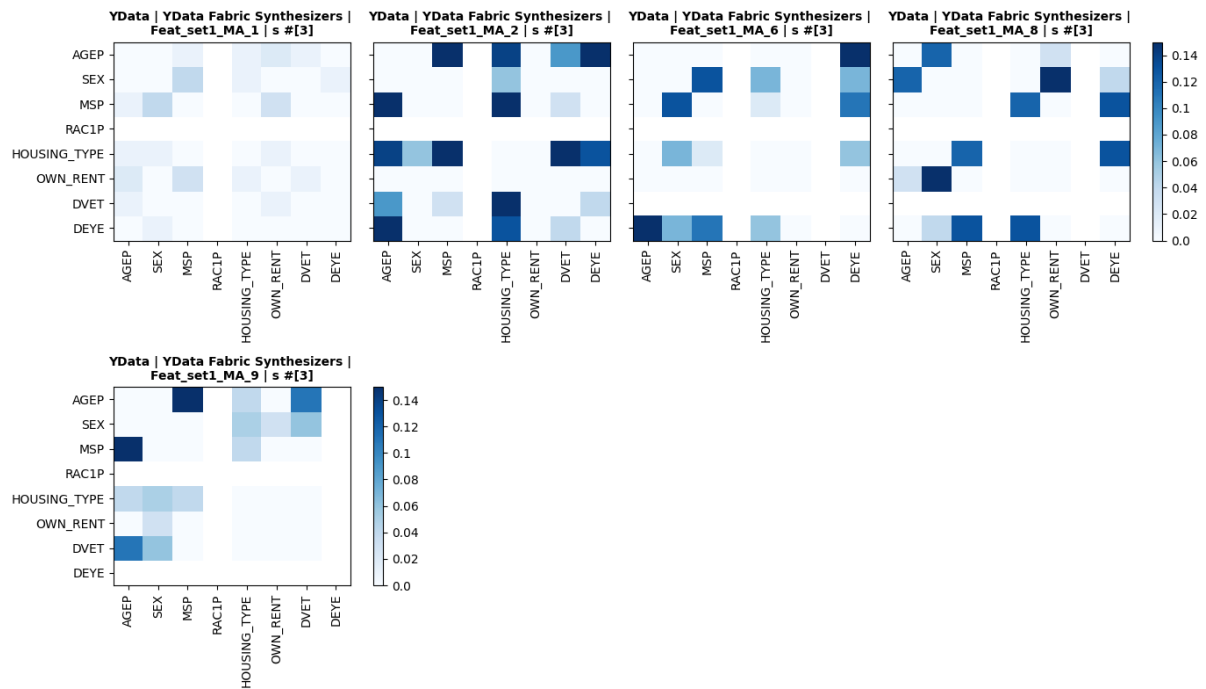
### Feature Set: custom-features-5 I Target Dataset: ma2019:

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']  
Feature Space (possible combinations): 72



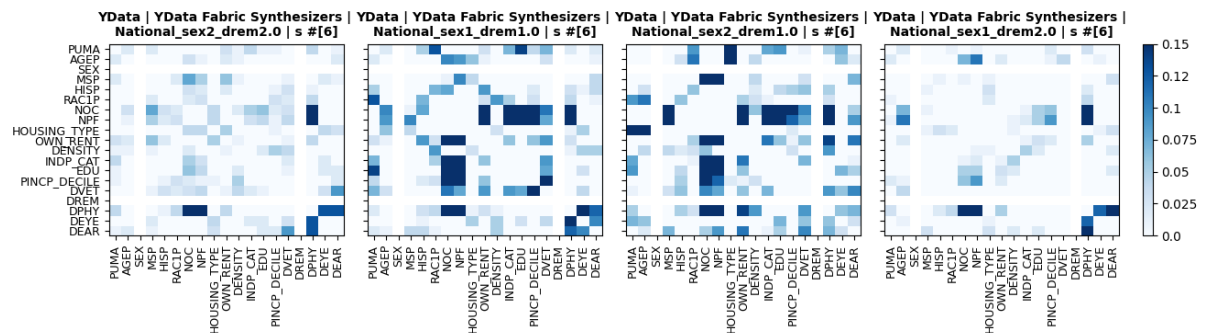
## Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600



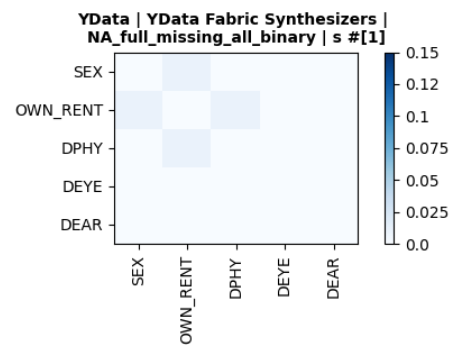
## Feature Set: all-features | Target Dataset: national2019:

Features: ['AGEP', 'DEAR', 'DENSITY', 'DEYE', 'DPHY', 'DREM', 'DVET', 'EDU', 'HISP', 'HOUSING\_TYPE', 'INDP', 'INDP\_CAT', 'MSP', 'NOC', 'NPF', 'OWN\_RENT', 'PINCP', 'PINCP\_DECILE', 'POVPIPI', 'PUMA', 'PWGTGP', 'RAC1P', 'SEX', 'WGTP']  
Feature Space (possible combinations): 13,330,105,588,800,000,000,000,000



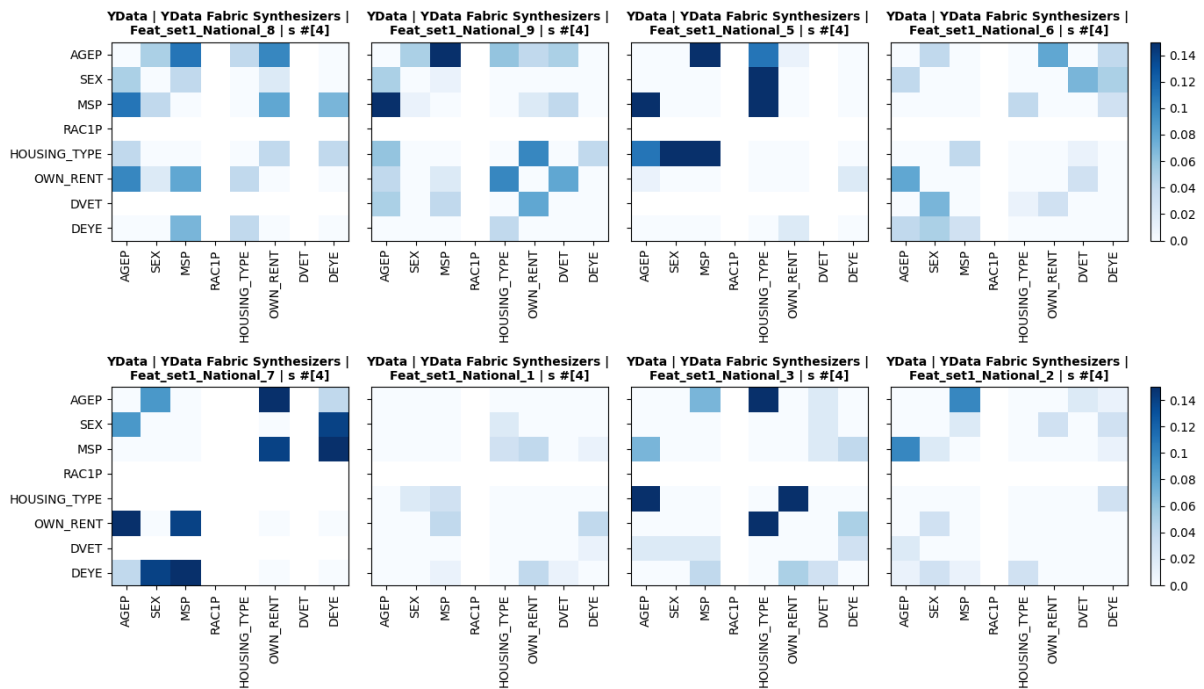
## Feature Set: custom-features-5 | Target Dataset: national2019:

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']  
Feature Space (possible combinations): 72



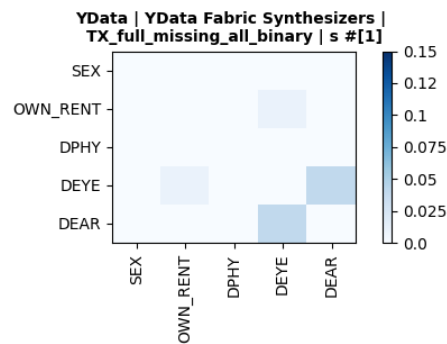
Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600



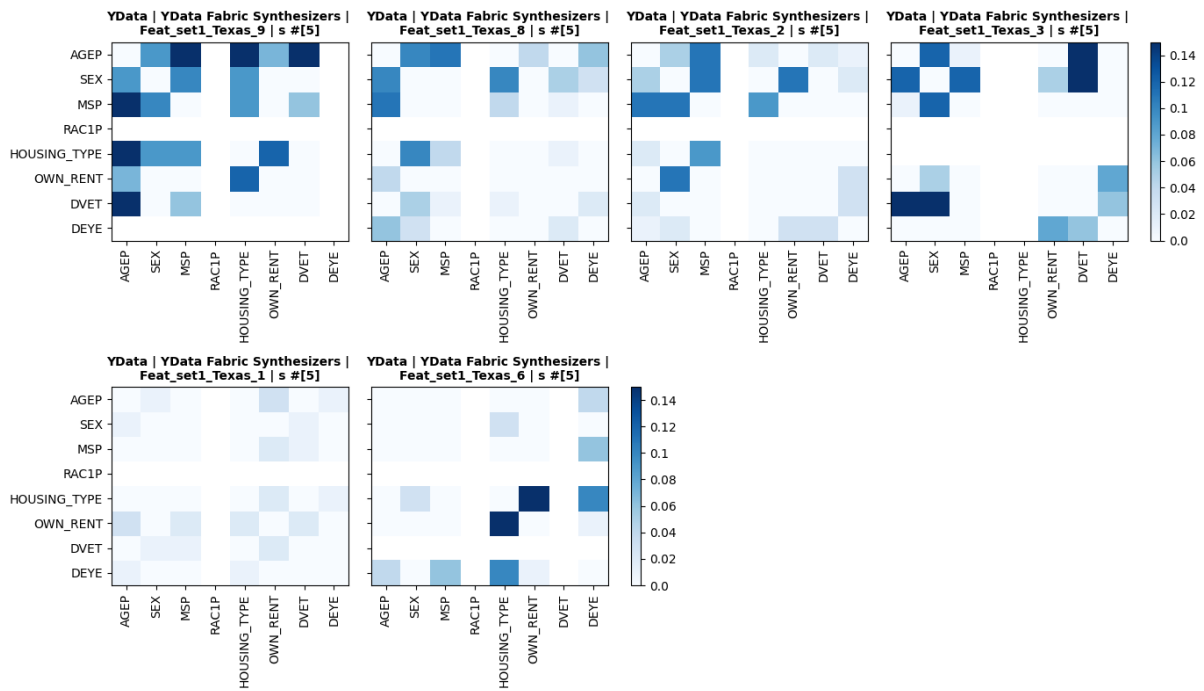
Feature Set: custom-features-5 | Target Dataset: tx2019:

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']  
Feature Space (possible combinations): 72



Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600



## 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: 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): 13,330,105,588,800,000,000,000,000

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

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   MA_sex2_drem1.0   s #[6]	3	0.04
YData   YData Fabric Synthesizers   MA_sex1_drem2.0   s #[6]	141	1.85
YData   YData Fabric Synthesizers   MA_sex2_drem2.0   s #[6]	180	2.36
YData   YData Fabric Synthesizers   MA_sex1_drem1.0   s #[6]	0	0.0

### Feature Set: custom-features-3 | Target Dataset: ma2019:

Features: ['AGEP', 'PINCP', 'POVPIP']

Feature Space (possible combinations): 1,000,000

Number of Unique Records in Target Data: 5315 (69.62%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   MA_full_missing_all_numerical   s #[2]	2367	44.53

### Feature Set: custom-features-5 | Target Dataset: ma2019:

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']

Feature Space (possible combinations): 72

Number of Unique Records in Target Data: 5 (0.07%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   MA_full_missing_all_binary   s #[1]	0	0.0

### Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']

Feature Space (possible combinations): 1,587,600

Number of Unique Records in Target Data: 919 (12.04%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   Feat_set1_MA_1   s #[3]	242	26.33
YData   YData Fabric Synthesizers   Feat_set1_MA_2   s #[3]	31	3.37
YData   YData Fabric Synthesizers   Feat_set1_MA_6   s #[3]	71	7.73
YData   YData Fabric Synthesizers   Feat_set1_MA_8   s #[3]	23	2.5
YData   YData Fabric Synthesizers   Feat_set1_MA_9   s #[3]	34	3.7

**Feature Set: all-features I 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): 13,330,105,588,800,000,000,000,000

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

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData I YData Fabric Synthesizers I National_sex2_drem2.0 I s #[6]	453	1.67
YData I YData Fabric Synthesizers I National_sex1_drem1.0 I s #[6]	2	0.01
YData I YData Fabric Synthesizers I National_sex2_drem1.0 I s #[6]	2	0.01
YData I YData Fabric Synthesizers I National_sex1_drem2.0 I s #[6]	480	1.77

**Feature Set: custom-features-3 I Target Dataset: national2019:**

Features: ['AGEP', 'PINCP', 'POVPIP']  
Feature Space (possible combinations): 1,000,000

Number of Unique Records in Target Data: 18623 (68.33%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData I YData Fabric Synthesizers I NA_full_missing_all_numerical I s #[2]	7994	42.93

**Feature Set: custom-features-5 I Target Dataset: national2019:**

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']  
Feature Space (possible combinations): 72

Number of Unique Records in Target Data: 3 (0.01%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData I YData Fabric Synthesizers I NA_full_missing_all_binary I s #[1]	0	0.0

**Feature Set: demographic-focused-except-PINCP\_DECILE-EDU I Target Dataset: national2019:**

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600

Number of Unique Records in Target Data: 2847 (10.45%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData I YData Fabric Synthesizers I Feat_set1_National_8 I s #[4]	135	4.74
YData I YData Fabric Synthesizers I Feat_set1_National_9 I s #[4]	97	3.41
YData I YData Fabric Synthesizers I Feat_set1_National_5 I s #[4]	18	0.63
YData I YData Fabric Synthesizers I Feat_set1_National_6 I s #[4]	133	4.67
YData I YData Fabric Synthesizers I Feat_set1_National_7 I s #[4]	8	0.28
YData I YData Fabric Synthesizers I Feat_set1_National_1 I s #[4]	407	14.3
YData I YData Fabric Synthesizers I Feat_set1_National_3 I s #[4]	116	4.07
YData I YData Fabric Synthesizers I Feat_set1_National_2 I s #[4]	338	11.87



**Feature Set: custom-features-3 | Target Dataset: tx2019:**

Features: ['AGEP', 'PINCP', 'POVPIP']  
Feature Space (possible combinations): 1,000,000

Number of Unique Records in Target Data: 7515 (81.02%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   TX_full_missing_all_numerical   s #[2]	2885	38.39

**Feature Set: custom-features-5 | Target Dataset: tx2019:**

Features: ['DEAR', 'DEYE', 'DPHY', 'OWN\_RENT', 'SEX']  
Feature Space (possible combinations): 72

Number of Unique Records in Target Data: 5 (0.05%)

Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   TX_full_missing_all_binary   s #[1]	3	60.0

**Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: tx2019:**

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600

Number of Unique Records in Target Data: 1317 (14.2%)

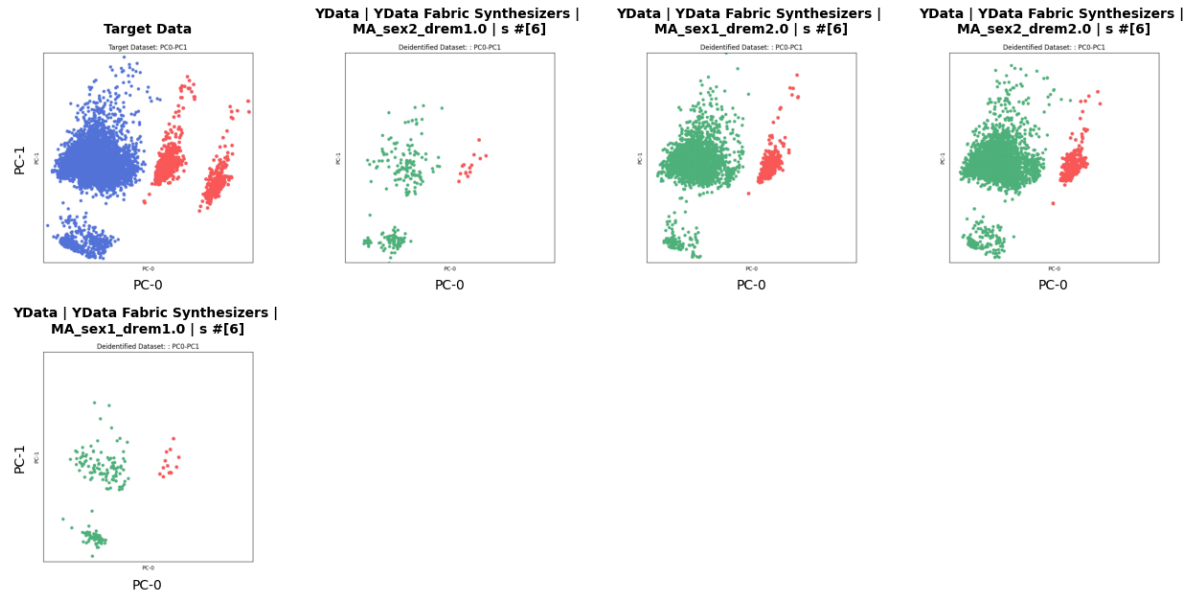
Variant	Records Matched In Target Data	Percent Records Matched In Target Data
YData   YData Fabric Synthesizers   Feat_set1_Texas_9   s #[5]	37	2.81
YData   YData Fabric Synthesizers   Feat_set1_Texas_8   s #[5]	56	4.25
YData   YData Fabric Synthesizers   Feat_set1_Texas_2   s #[5]	110	8.35
YData   YData Fabric Synthesizers   Feat_set1_Texas_3   s #[5]	6	0.46
YData   YData Fabric Synthesizers   Feat_set1_Texas_1   s #[5]	280	21.26
YData   YData Fabric Synthesizers   Feat_set1_Texas_6   s #[5]	26	1.97

## 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 [Principal 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 you'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) axes– 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.

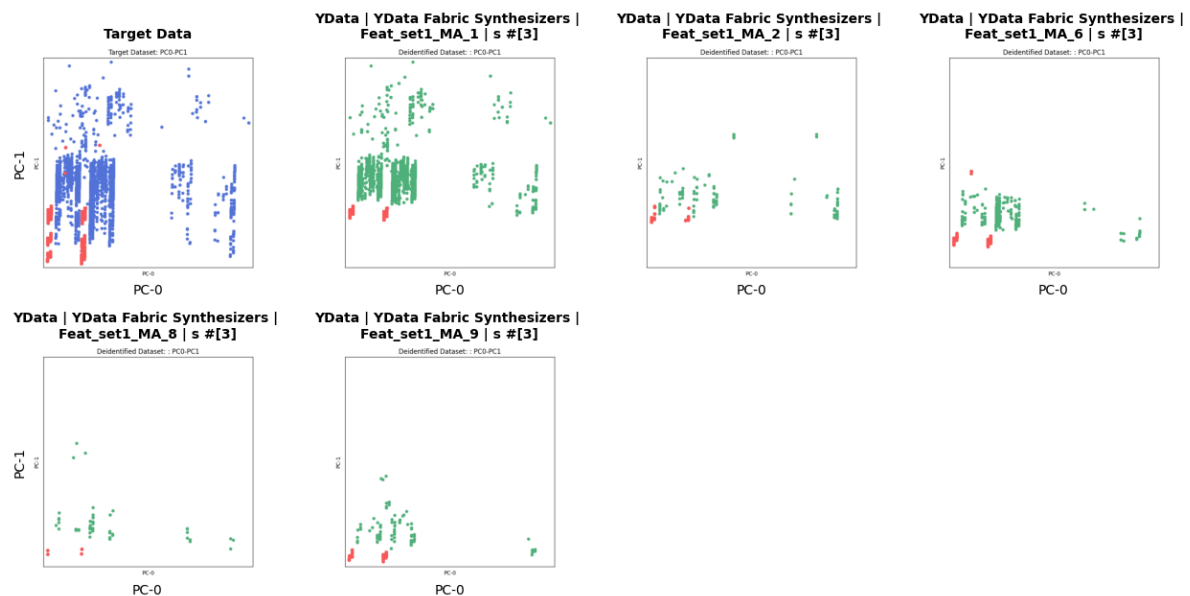
### Feature Set: all-features | Target Dataset: ma2019:

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): 13,330,105,588,800,000,000,000,000



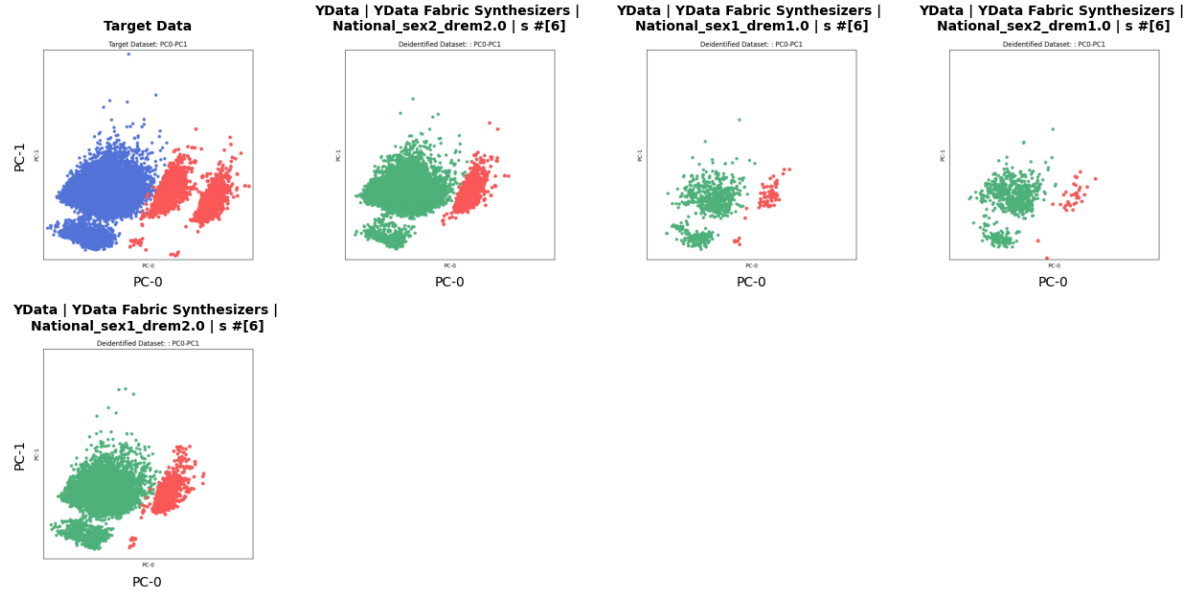
### Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: ma2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600



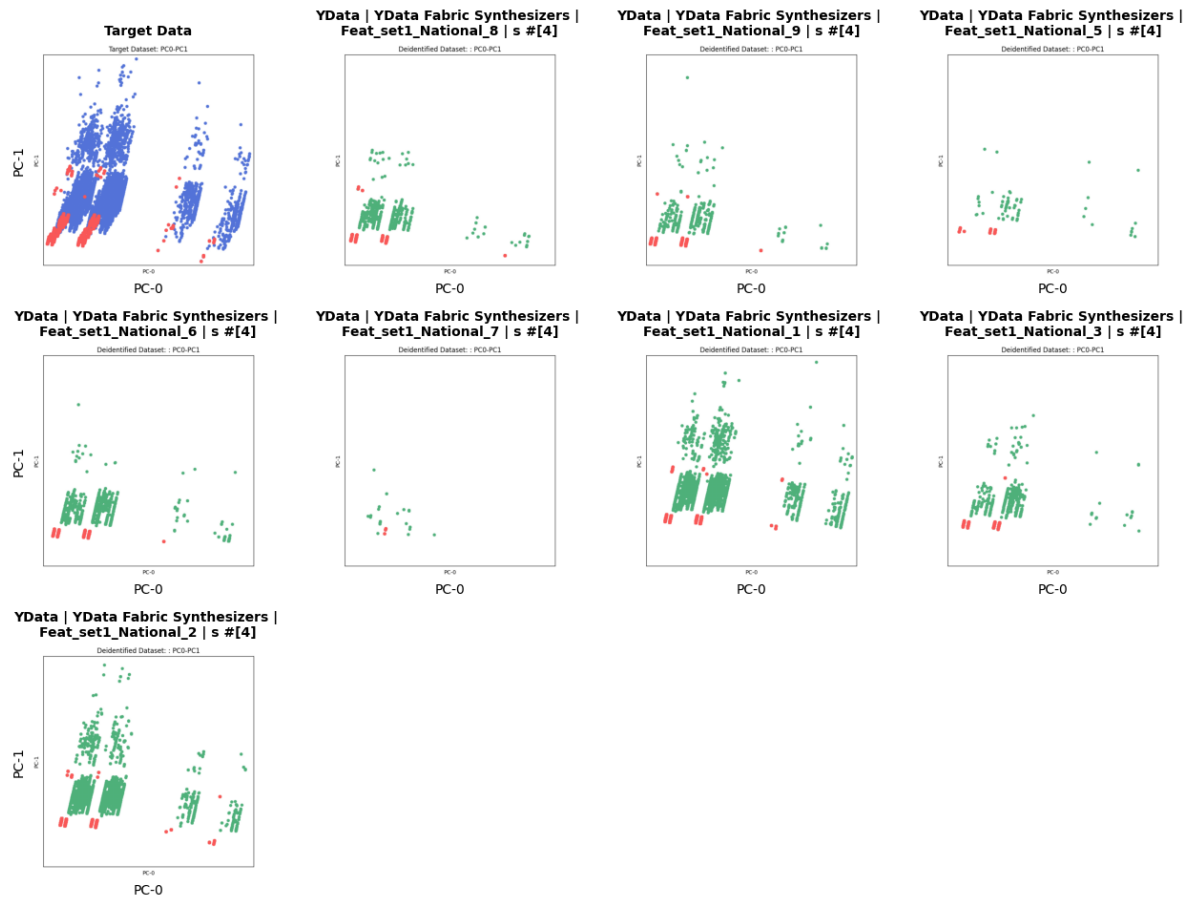
## Feature Set: all-features | Target Dataset: national2019:

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): 13,330,105,588,800,000,000,000,000



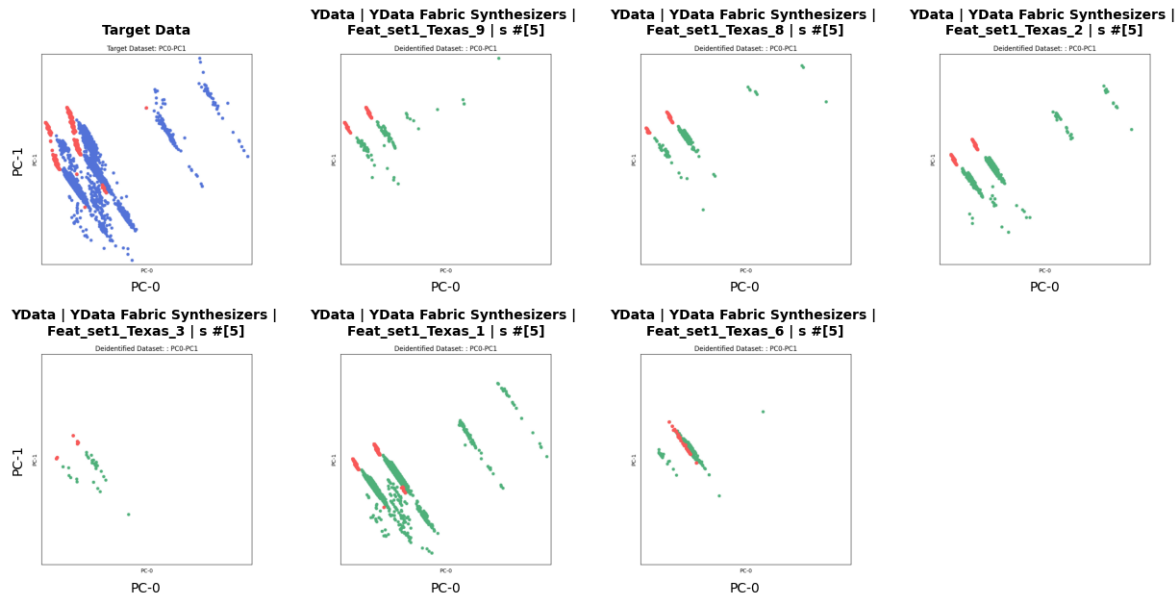
## Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: national2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
 Feature Space (possible combinations): 1,587,600



Feature Set: demographic-focused-except-PINCP\_DECILE-EDU | Target Dataset: tx2019:

Features: ['AGEP', 'DEYE', 'DVET', 'HOUSING\_TYPE', 'MSP', 'OWN\_RENT', 'RAC1P', 'SEX']  
Feature Space (possible combinations): 1,587,600



## 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:

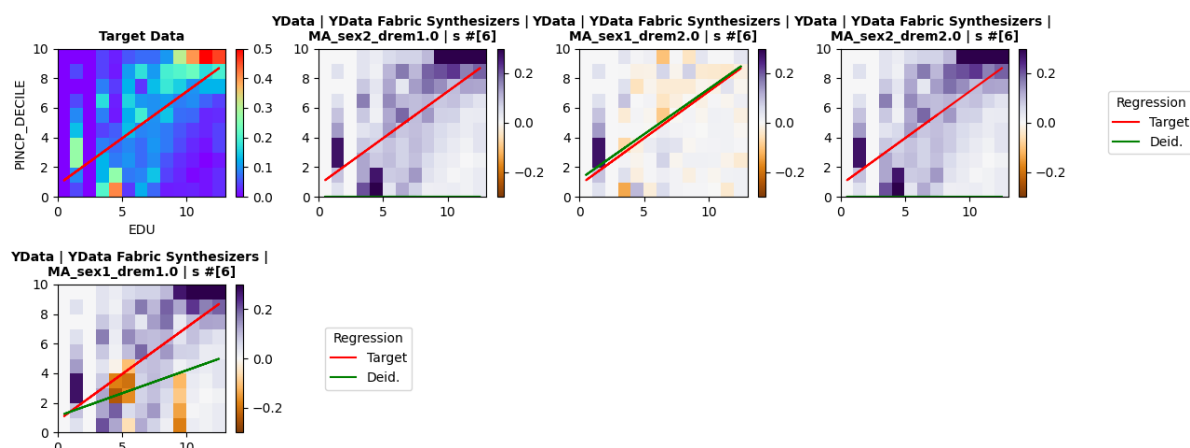
- EDU: The highest education level this individual has attained, ranging from 1 (elementary school) to 12 (PhD). See Appendix of this report for the full list of code values.
- PINCP\_DECILE: The individual's income decile relative to their PUMA. This helps us account for differences in cost of living 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.

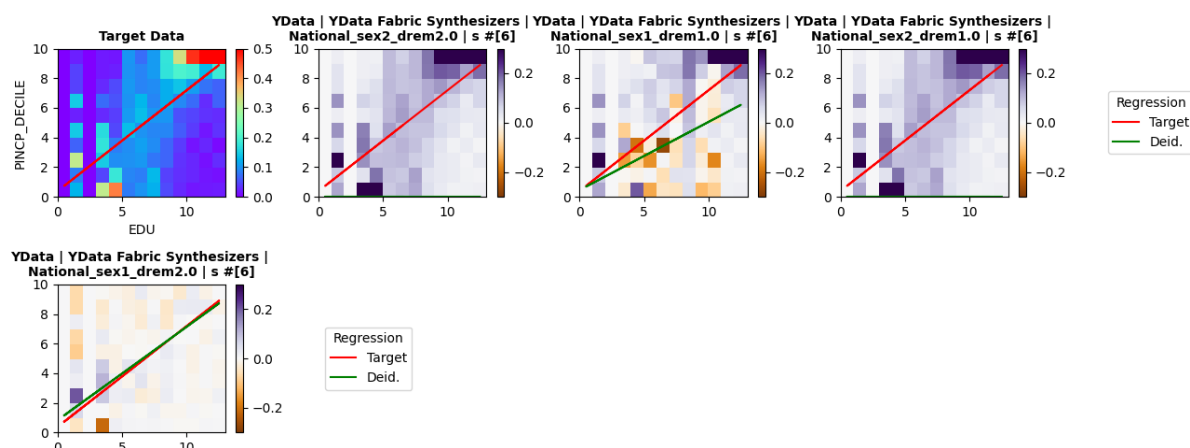
### Feature Set: all-features | Target Dataset: ma2019:

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): 13,330,105,588,800,000,000,000,000



### 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): 13,330,105,588,800,000,000,000,000



## 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:

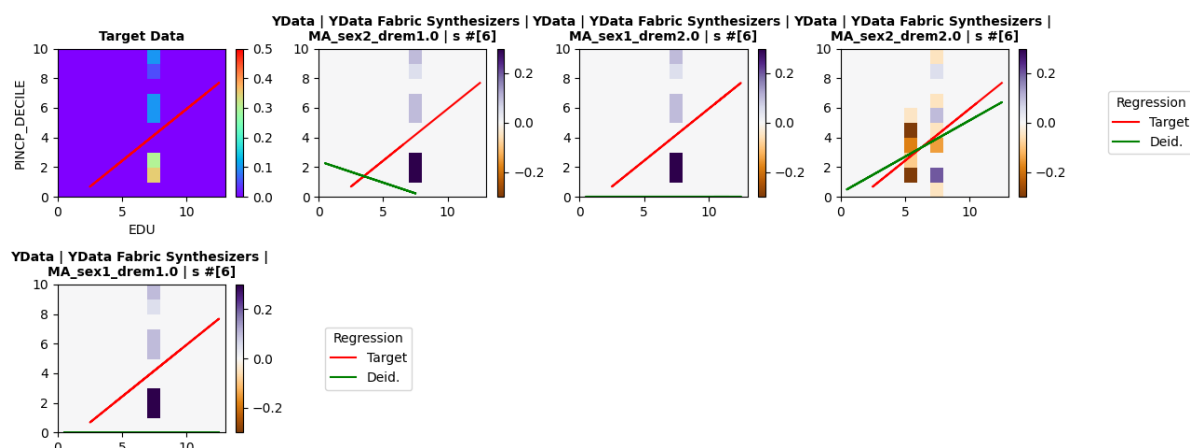
- EDU: The highest education level this individual has attained, ranging from 1 (elementary school) to 12 (PhD). See Appendix of this report for the full list of code values.
- PINCP\_DECILE: The individual's income decile relative to their PUMA. This helps us account for differences in cost of living 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.

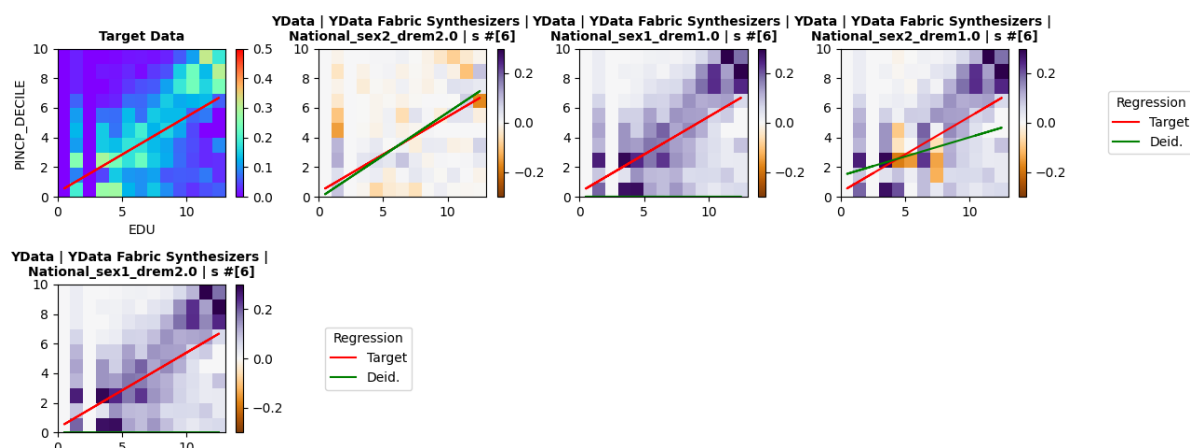
### Feature Set: all-features | Target Dataset: ma2019:

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): 13,330,105,588,800,000,000,000,000



### 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): 13,330,105,588,800,000,000,000,000





## Data Description

### Deidentified (Deid.) Datasets:

I YData Fabric Synthesizers I I MA\_sex2\_drem1.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_sex2_drem1.0
Submission Number	6
Feature Set Name	all-features
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVPI, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	1f309023d57a457e00264f0f2ad895075b9b9818
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_MA_sex2_drem1.0_MiriamSantos_6
Records	246
Features	24



YData Fabric Synthesizers | MA\_sex1\_drem2.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_sex1_drem2.0
Submission Number	6
Feature Set Name	all-features
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_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	b42e8a1338874ac21893b9616ad2126f45a96430
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_MA_sex1_drem2.0_MiriamSantos_6
Records	3216
Features	24

YData Fabric Synthesizers | MA\_sex2\_drem2.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_sex2_drem2.0
Submission Number	6
Feature Set Name	all-features
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_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	ef05266d70813f05b4cdf26a40c1afd6842a2237
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_MA_sex2_drem2.0_MiriamSantos_6
Records	3643
Features	24

I YData Fabric Synthesizers I I MA\_sex1\_drem1.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_sex1_drem1.0
Submission Number	6
Feature Set Name	all-features
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	3862b906e306c2d5933f618cec2d9f05973b69ce
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_MA_sex1_drem1.0_MiriamSantos_6
Records	182
Features	24

I YData Fabric Synthesizers I I National\_sex2\_drem2.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	National_sex2_drem2.0
Submission Number	6
Feature Set Name	all-features
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, SEX, MSP, HISP, RAC1P, NOC, NPF, HOUSING_TYPE, OWN_RENT, DENSITY, INDP, INDP_CAT, EDU, PINCP, PINCP_DECILE, POVIP, DVET, DREM, DPHY, DEYE, DEAR, WGTP, PWGTP
Deid Data Id	833046261dd0a19fbc32b34f537bbb698da3137
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_National_sex2_drem2.0_MiriamSantos_6
Records	12623
Features	24

I YData Fabric Synthesizers I I National\_sex1\_drem1.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	National_sex1_drem1.0
Submission Number	6
Feature Set Name	all-features
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, 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	835305201d5059717bbaef71e4c22d46a34d6db4
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_National_sex1_drem1.0_MiriamSantos_6
Records	769
Features	24

I YData Fabric Synthesizers I I National\_sex2\_drem1.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	National_sex2_drem1.0
Submission Number	6
Feature Set Name	all-features
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, 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	d3171bf9f0b7e90f81b7433ff9f1cdad6b4bb161
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_National_sex2_drem1.0_MiriamSantos_6
Records	752
Features	24

I YData Fabric Synthesizers I | National\_sex1\_drem2.0:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:45:49
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	National_sex1_drem2.0
Submission Number	6
Feature Set Name	all-features
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	PUMA, AGE, 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	3eacd0e977faaac06188dc7eb1cf681a0867fdb9
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_National_sex1_drem2.0_MiriamSantos_6
Records	11756
Features	24

I YData Fabric Synthesizers I | MA\_full\_missing\_all\_numerical:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:42:44
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_full_missing_all_numerical
Submission Number	2
Feature Set Name	custom-features-3
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGE, PINCP, POVPIP
Deid Data Id	207be371f52cf013900cb2170484e7f3218d6099
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_numerical_MiriamSantos_2
Records	8000
Features	3

**I YData Fabric Synthesizers I I NA\_full\_missing\_all\_numerical:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:42:44
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	NA_full_missing_all_numerical
Submission Number	2
Feature Set Name	custom-features-3
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, PINCP, POVPIP
Deid Data Id	e87dade6905b8e4ce1d5997e4e62742b16da1ee4
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_numerical_MiriamSantos_2
Records	28000
Features	3

**I YData Fabric Synthesizers I I TX\_full\_missing\_all\_numerical:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:42:44
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	TX_full_missing_all_numerical
Submission Number	2
Feature Set Name	custom-features-3
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, PINCP, POVPIP
Deid Data Id	f58ff04d4663df2245c580cd7fb0c500f3aed827
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_numerical_MiriamSantos_2
Records	8000
Features	3

**I YData Fabric Synthesizers I I MA\_full\_missing\_all\_binary:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:33:36
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	MA_full_missing_all_binary
Submission Number	1
Feature Set Name	custom-features-5
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	SEX, OWN_RENT, DPHY, DEYE, DEAR
Deid Data Id	96f331b66cccf0dcabbd94791974fde62a26d7ff
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_binary_MiriamSantos_1
Records	8000
Features	5

**I YData Fabric Synthesizers I I NA\_full\_missing\_all\_binary:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:33:36
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	NA_full_missing_all_binary
Submission Number	1
Feature Set Name	custom-features-5
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	SEX, OWN_RENT, DPHY, DEYE, DEAR
Deid Data Id	5e27a2e2e2c2ec8ec8fecf73f86b3810656a5c6f
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_binary_MiriamSantos_1
Records	28000
Features	5

**I YData Fabric Synthesizers I | TX\_full\_missing\_all\_binary:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:33:36
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	TX_full_missing_all_binary
Submission Number	1
Feature Set Name	custom-features-5
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	SEX, OWN_RENT, DPHY, DEYE, DEAR
Deid Data Id	897d22d1538b72f309ac60bc647ee9f486d27bb0
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_full_missing_all_binary_MiriamSantos_1
Records	8000
Features	5

**I YData Fabric Synthesizers I | Feat\_set1\_MA\_1:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:53:07
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_MA_1
Submission Number	3
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	d719f6ae8e15d1892ce2f033dfd92f8c24e78f81
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_MA_1_MiriamSantos_3
Records	6658
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_MA\_2:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:53:07
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_MA_2
Submission Number	3
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	0202462fa3a7b5e2fe05c5c97f4aa3f23711f5e0
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_MA_2_MiriamSantos_3
Records	180
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_MA\_6:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:53:07
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_MA_6
Submission Number	3
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	1e77f8db4fcc3332655bdb88feb67da43f5e9c64
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_MA_6_MiriamSantos_3
Records	570
Features	8



# YData Fabric Synthesizers I | Feat\_set1\_MA\_8:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:53:07
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_MA_8
Submission Number	3
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	7db8002785979f659ef1e35dff23820d58bbf941
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_MA_8_MiriamSantos_3
Records	68
Features	8

# YData Fabric Synthesizers I | Feat\_set1\_MA\_9:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 10:53:07
Target Dataset	ma2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_MA_9
Submission Number	3
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	424ec3d878c09ff5e42cad2f2c1c2fb8147bb875
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_MA_9_MiriamSantos_3
Records	152
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_8:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_8
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	1ea3bed009e0b7c94488e01f16b085eb045bfeb7
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_8_MiriamSantos_4
Records	904
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_9:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_9
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	ecf9ada0f3d2cf5ed8e87cb37be65e7b61c531da
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_9_MiriamSantos_4
Records	876
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_5:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_5
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	87651779cc4e58bad153f9a461a5ec27b14e5aa5
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_5_MiriamSantos_4
Records	97
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_6:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_6
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	93543158cec4dc16215dd37b754050ef9cb41a22
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_6_MiriamSantos_4
Records	2211
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_7:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_7
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	4e8a54b5d3d35177b4132b98a7feff855ce02c75
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_7_MiriamSantos_4
Records	31
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_1:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_1
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	6740117729aa9bdff12567c0483fa315a03fbc78
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_1_MiriamSantos_4
Records	15094
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_3:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_3
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	70f4feb3ec9afd6058168cafe13be1ba581494a9
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_3_MiriamSantos_4
Records	877
Features	8

**I YData Fabric Synthesizers I | Feat\_set1\_National\_2:**

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:26:54
Target Dataset	national2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_National_2
Submission Number	4
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	ydata-sdk
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	e500382619ad12b29d7e7f1d66b59dc7978d1211
Privacy Label Detail	These CRC submissions were made with High-Fidelity settings and use the original Deep NN architecture and optimal parameters to produce non-differentially private synthetic data. The YData Fabric Synthesizer tool also has embedded options for user selection of High-Privacy and Balanced Fidelity-Privacy; those options use a differential privacy mechanism to inject noise into the process of synthetic data generation. More information in the public documentation ( <a href="https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/">https://docs.sdk.ydata.ai/0.6/examples/synthesize_with_anonymization/</a> )
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_National_2_MiriamSantos_4
Records	7161
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_9:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_9
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	92003a45a813093bddbf143f0729107547aa42fc
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_9_MiriamSantos_5
Records	253
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_8:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_8
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	2ec9c2052c3e88a31951e22a9d16038e1040941c
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_8_MiriamSantos_5
Records	255
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_2:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_2
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	b10303c58fa609dd78cc789f829f470d3e4fea31
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_2_MiriamSantos_5
Records	627
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_3:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_3
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	892a804a70d565ffb359d69edecb222a9b65e71
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_3_MiriamSantos_5
Records	43
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_1:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_1
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	ce06806374d87c0d9c86bbaa270406c3064f316c
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_1_MiriamSantos_5
Records	7831
Features	8

I YData Fabric Synthesizers I | Feat\_set1\_Texas\_6:

Label Name	Label Value
Team	YData
Submission Timestamp	7/7/2023 14:33:04
Target Dataset	tx2019
Algorithm Name	YData Fabric Synthesizers
Variant Label	Feat_set1_Texas_6
Submission Number	5
Feature Set Name	demographic-focused-except-PINCP_DECILE-EDU
Library Name	YData Fabric Synthesizers
Algorithm Type	neural net
Privacy Category	non_dp
Features List	AGEP, SEX, MSP, RAC1P, HOUSING_TYPE, OWN_RENT, DVET, DEYE
Deid Data Id	193b61affe90354db0aad09b855d851257f84210
Privacy Label Detail	
Research Papers	

Property	Value
Filename	ydata_fabric_synthesizers_Feat_set1_Texas_6_MiriamSantos_5
Records	249
Features	8

## 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