

ARX

Data Anonymization Tool para anonimização de dados pessoais

Elizabeth Monteiro



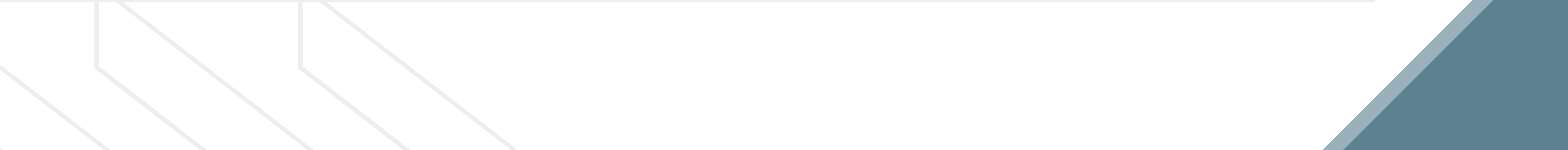


Anonimização

Uma técnica que remove ou modifica informações que possam identificar uma pessoa.

Técnica resulta em dados anonimizados, que não podem ser associados a nenhuma pessoa específica. O dado perde definitivamente a possibilidade de identificar uma pessoa.

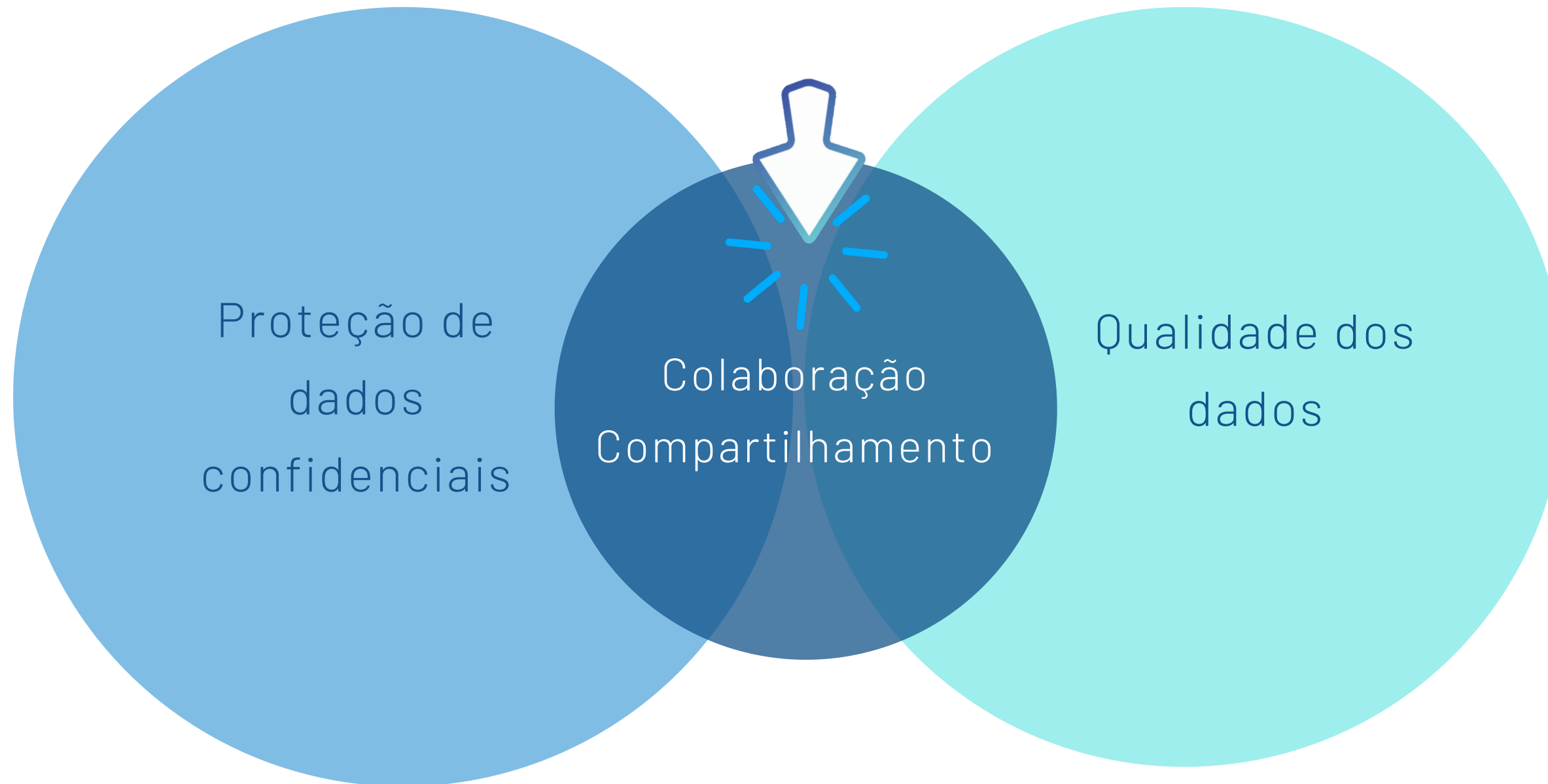
Dado é relativo a uma pessoa, mas passou por etapas que garantem a desvinculação dele a tal pessoa.



Dados que contextualizam informações vinculadas a uma pessoa

Identificadores	Semi identificadores	Sensíveis
Caracterizam-se por identificar unicamente os indivíduos no conjunto de dados	Contém valores que, quando correlacionados e/ou combinados com dados externos, podem proporcionar a identificação do indivíduo e, desta forma, vincular o indivíduo a seus dados confidenciais	Representam os dados confidenciais, pois quando expostos podem causar algum tipo de dano ao indivíduo
nome, número de matrícula, número da seguridade social, CPF, RG	CEP, endereço, localização, cargo, função, sexo, data de nascimento	dados médicos, genéticos, doença, religião, crença, salário, opinião política, orientação sexual, vida sexual

Pesquisas



A divulgação de dados confidenciais (pessoais/sensíveis) pode acarretar danos aos indivíduos, principalmente quando há diferentes fontes de dados disponíveis para vinculação



ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#) [Overview](#) [Anonymization tool](#) [Development](#) [Publications](#) [Downloads](#)

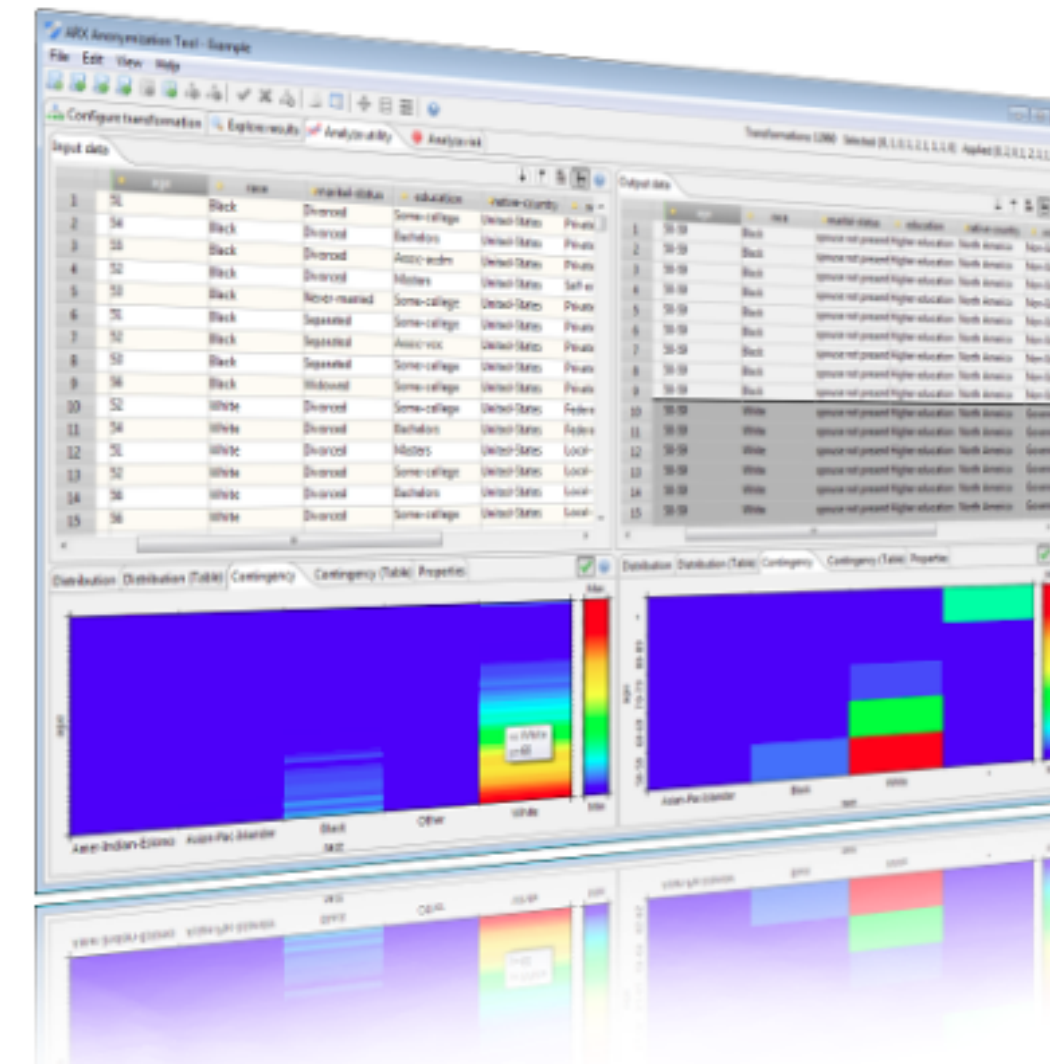
The current version 3.9.1 of ARX was released on 27/11/2022.

ARX Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports a wide variety of (1) privacy and risk models, (2) methods for transforming data and (3) methods for analyzing the usefulness of output data.

The software has been used in a variety of contexts, including commercial big data analytics platforms, research projects, clinical trial data sharing and for training purposes.

ARX is able to handle large datasets on commodity hardware and it features an intuitive cross-platform graphical user interface. You can find further information [here](#), or directly proceed to our [downloads](#) section.



<https://arx.deidentifier.org/>



ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[Development](#)[Publications](#)[Downloads](#)

The current version 3.9.1 of ARX was released on 27/11/2022.

[Home](#) » [Overview](#)

Overview

Overview of supported anonymization methods

ARX is an open source tool for transforming structured (i.e. tabular) personal data using selected methods from the broad areas of data anonymization and statistical disclosure control. It supports transforming datasets in ways that make sure that they adhere to user-specified privacy models and risk thresholds that mitigate attacks that may lead to privacy breaches. ARX can be used to remove direct identifiers (e.g. names) from datasets and to enforce further constraints on indirect identifiers. Indirect identifiers (or quasi-identifiers, or keys) are attributes that do not directly identify an individual but may together with other indirect identifiers form an identifier that can be used for linkage attacks. It is typically assumed that information about indirect identifiers is available to the attacker (in some form of background knowledge) and that they cannot simply be removed from the dataset (e.g. because they are required later for analyses). ARX also supports methods for protecting sensitive attributes from disclosure and semantic privacy models, which require fewer assumptions to be made about the goals and the background knowledge of attackers.

ARX supports (almost) arbitrary combinations the following privacy models:

- Syntactic privacy models, such as k-anonymity, l-diversity, t-closeness, δ -disclosure privacy, β -likeness and δ -presence.
- Statistical privacy models, such as k-map, thresholds on average risk and methods based on super-population models.
- Semantic privacy models, such as (ϵ, δ) -differential privacy and a game-theoretic de-identification approach.

ARX supports (almost) arbitrary combinations of the following data transformation models:

- Global and local transformation schemes: ARX can apply the same transformation scheme to all records in a dataset or apply different transformation schemes to different subsets of records.
- Random sampling: Privacy risks can be reduced by drawing a random sample from the input dataset.
- Generalization: Records can be made less unique by generalizing attribute values based on user-specified hierarchies.
- Record, attribute and cell suppression: Privacy risks can be lowered by removing individual attribute values or complete records.



ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

Home

Overview

Anonymization tool

Development

Publications

Downloads

Privacy models

Data quality models

Transformation models

Related software

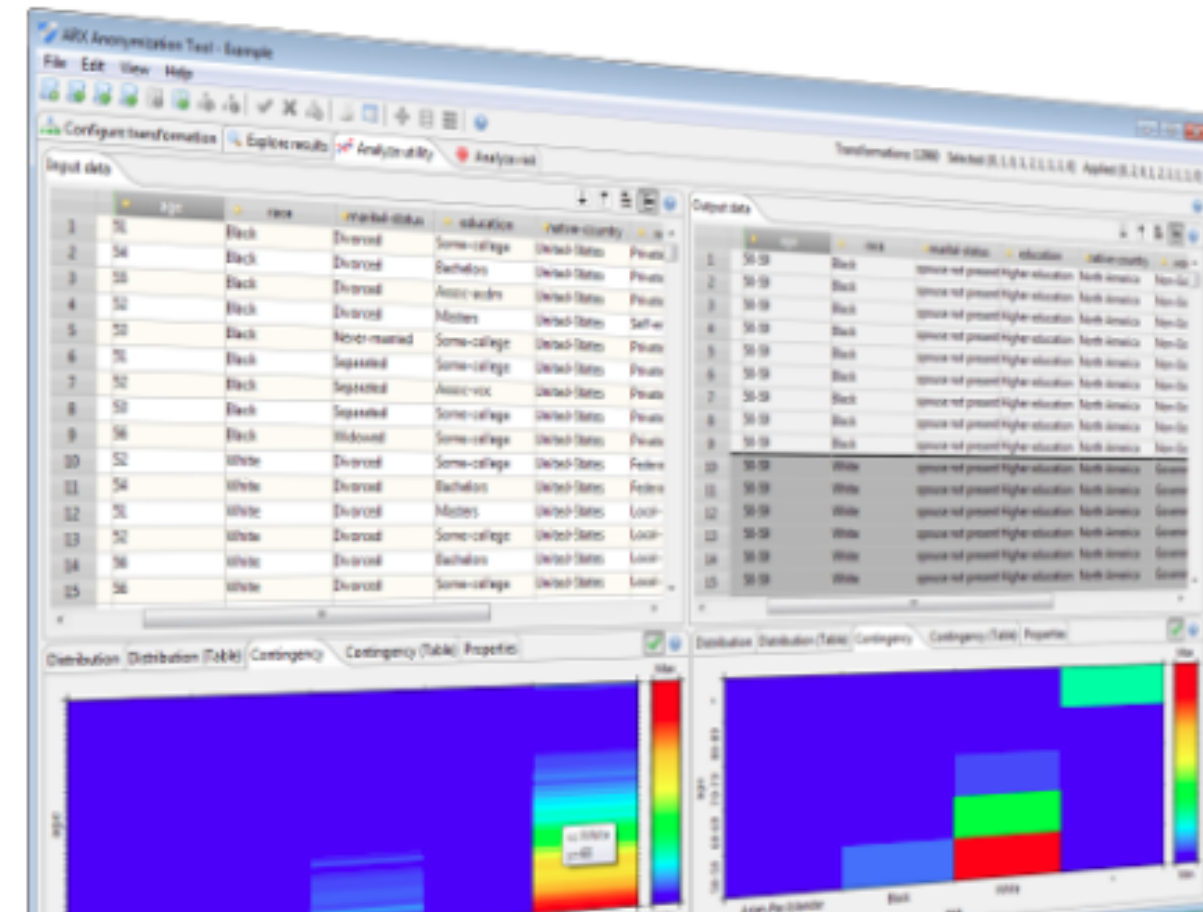
You can find more information here.

ARX

Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports a wide variety of (1) privacy and risk models, (2) methods for transforming data and (3) methods for analyzing the usefulness of output data.

The software has been used in a variety of contexts, including commercial big data analytics platforms, research projects, clinical trial data sharing and for training purposes.





ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[Development](#)[Publications](#)[Downloads](#)

Please report bugs

on GitHub page.

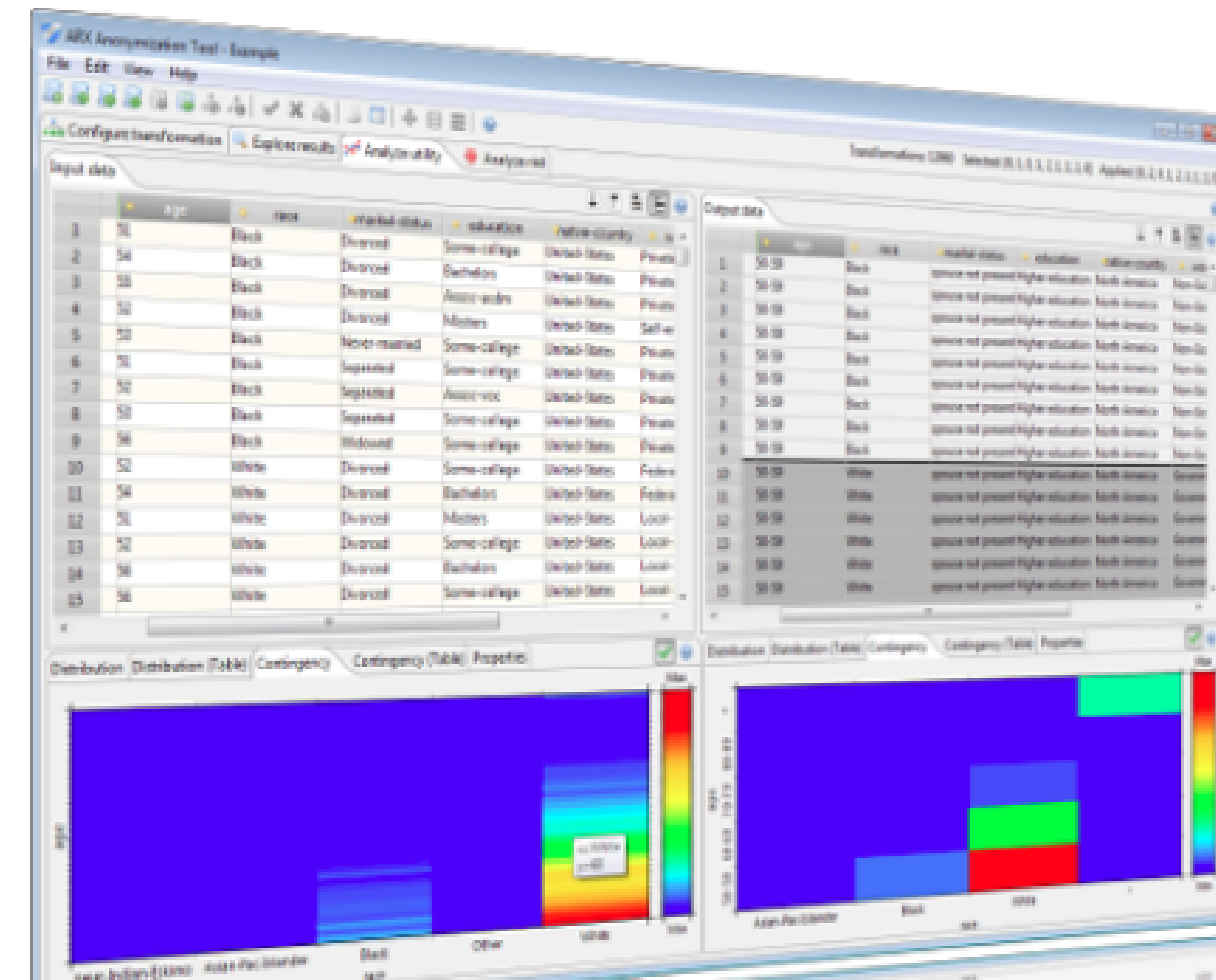
[Configuration](#)[Exploration](#)[Utility analysis](#)[Risk analysis](#)

ARX

Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports a wide variety of (1) privacy and risk models, (2) methods for transforming data and (3) methods for analyzing the usefulness of output data.

The software has been used in a variety of contexts, including commercial big data analytics platforms, research projects, clinical trial data sharing and for training purposes.





ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[Development](#)[Publications](#)[Downloads](#)

The current version 3.9.1 of ARX was released

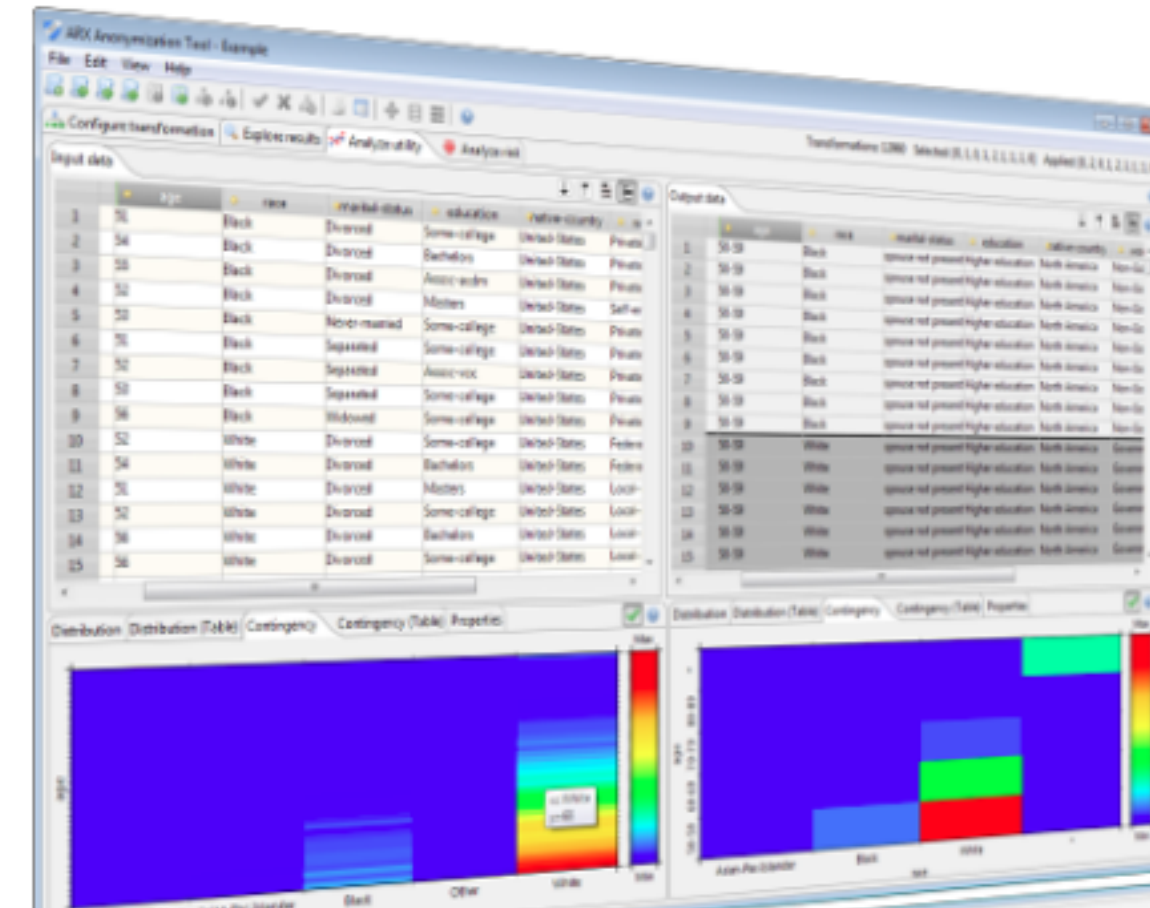
[API](#)[Dependencies](#)[Changelog](#)[Contributors](#)

ARX

Data Anonymization Tool

ARX is a comprehensive open source software for anonymizing sensitive personal data. It supports a wide variety of (1) privacy and risk models, (2) methods for transforming data and (3) methods for analyzing the usefulness of output data.

The software has been used in a variety of contexts, including commercial big data analytics platforms, research projects, clinical trial data sharing and for training purposes.





ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[Development](#)[Publications](#)[Downloads](#)

Please report bugs and feature requests on our [GitHub page](#).

[Home](#) › Publications

Publications

The principles, methods, algorithms and data structures underlying the ARX anonymization tool have been published in peer-reviewed journals and conference proceedings:

- Carolin EM Koll, Sina M Hopff, Thierry Meurers, Chin Huang Lee, Mirjam Kohls, Christoph Stellbrink, Charlotte Thibeault, Lennart Reinke, Sarah Steinbrecher, Stefan Schreiber, Lazar Mitrov, Sandra Frank, Olga Miljukov, Johanna Erber, Johannes C Hellmuth, Jens-Peter Reese, Fridolin Steinbeis, Thomas Bahmer, Marina Hagen, Patrick Meybohm, Stefan Hansch, István Vadász, Lilian Krist, Steffi Jiru-Hillmann, Fabian Prasser, Jörg Janne Vehreschild & NAPKON Study Group.
Statistical biases due to anonymization evaluated in an open clinical dataset from COVID-19 patients.
Sci Data. 2022 Dec 21;9(1):776. ([DOI](#))
- Anna C Haber, Ulrich Sax, Fabian Prasser, NFDI4Health Consortium
Open Tools for Quantitative Anonymization of Tabular Phenotype Data: Literature Review
Brief Bioinformatics. bbac440 (2022). ([DOI](#))



ARX – Data Anonymization Tool

A comprehensive software for privacy-preserving microdata publishing

[Home](#)[Overview](#)[Anonymization tool](#)[Development](#)[Publications](#)[Downloads](#)

You can find a list of our scientific publications [here](#).

[Home](#) › [Downloads](#)

Downloads

The current version of ARX is 3.9.1 (last update: 27. November 2022). Distributions for various operating systems are available as downloads below.

To verify the integrity of a download you can use the checksums contained in this [file](#) which we have signed with [our PGP/GPG key](#) (signature [here](#)).

Note: To support our research, please cite one of our [papers](#) instead of referencing our website in scientific articles. Thanks.

License

Licensed under the Apache License, Version 2.0 (the “License”); you may not use ARX except in compliance with the License. You may obtain a copy of the License at <http://www.apache.org/licenses/LICENSE-2.0>.

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and



ARX – Ferramenta de Anonimização de Dados

1

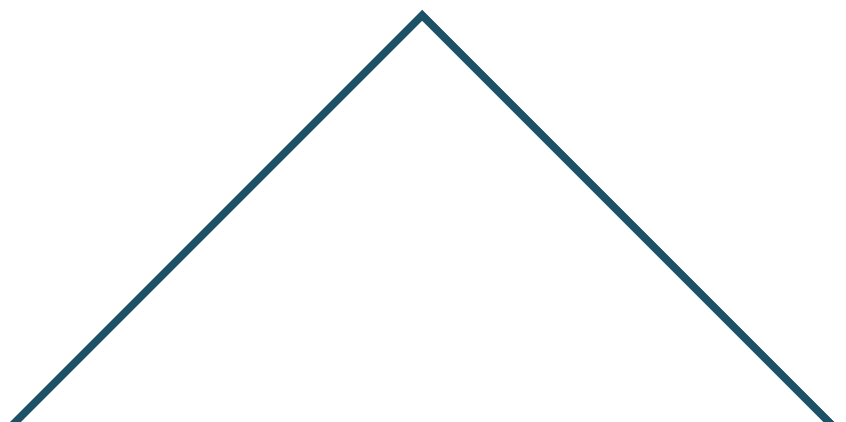
Software de código aberto para anonimizar dados pessoais.

2

Fornece suporte para os critérios de privacidade mais comuns e para combinações arbitrárias.

3

Oferece suporte à transformação de conjuntos de dados de forma a garantir que eles sigam os modelos de privacidade especificados pelo usuário.



ARX – Ferramenta de Anonimização de Dados

4 Os modelos contribuem para minimizar os riscos de ataques que podem levar a violações de privacidade.

5 Transforma dados pessoais estruturados (tabulares) usando métodos selecionados para a anonimização de dados e controle de divulgação estatística.

6 Disponibiliza informações sobre seu uso, configurações técnicas de anonimização, análise de utilidade e de risco.

O ARX pode ser usado para remover identificadores

DIRETOS

identificam diretamente os participantes (nomes, Documento de identidade, números de segurança social)

INDIRETOS

são atributos que não identificam diretamente um indivíduo, mas podem, junto com outros identificadores indiretos, formar um identificador que identifica o indivíduo.

Modelos de transformação de dados

ESQUEMAS DE TRANSFORMAÇÃO GLOBAL E LOCAL

Aplicação do mesmo esquema de transformação a todos os registros em um conjunto de dados ou aplicação de diferentes esquemas de transformação a diferentes subconjuntos de registros.

AMOSTRAGEM ALEATÓRIA

Extração de uma amostra aleatória do conjunto de dados de entrada.

GENERALIZAÇÃO

Generalização dos valores de atributo com base nas hierarquias especificadas pelo usuário.

SUPRESSÃO DE REGISTROS, ATRIBUTOS E CÉLULAS

Remoção de valores de atributos individuais ou registros completos.

Modelos de transformação de dados

MICROAGREGAÇÃO

Clusters de valores de atributos numéricos podem ser combinados em um valor comum por funções de agregação especificadas pelo usuário.

CODIFICAÇÃO SUPERIOR E INFERIOR

Os valores que excedem um intervalo definido pelo usuário podem ser truncados.

CATEGORIZAÇÃO

As variáveis contínuas podem ser categorizadas automaticamente.

Input data

		sex	age	race	marital-status	education	native-country	
1	☒	Female	73	White	Married-civ-spo...	HS-grad	United-States	Self-e
2	☒	Female	76	Black	Widowed	10th	United-States	Priva
3	☒	Female	72	White	Married-civ-spo...	HS-grad	England	Priva
4	☒	Female	58	White	Married-civ-spo...	HS-grad	Greece	Priva
5	☒	Female	90	Black	Divorced	Some-college	United-States	Priva
6	☒	Female	60	White	Widowed	Some-college	England	Feder
7	☒	Female	59	White	Widowed	HS-grad	Germany	Feder
8	☒	Female	55	White	Divorced	HS-grad	England	Priva
9	☒	Female	84	White	Widowed	HS-grad	United-States	Local
10	☒	Female	81	White	Married-civ-spo...	10th	United-States	Self-e
11	☒	Male	52	White	Never-married	5th-6th	United-States	Priva
12	☒	Male	71	Black	Separated	Some-college	United-States	Self-e
13	☒	Male	74	White	Separated	Some-college	United-States	State
14	☒	Male	71	White	Widowed	HS-grad	United-States	Feder
15	☒	Male	71	White	Widowed	Masters	United-States	Local
16	☒	Male	51	Black	Married-civ-spo...	11th	United-States	Local
17	☒	Male	61	Asian-Pac-Island...	Married-civ-spo...	HS-grad	South	Self-e
18	☒	Male	60	White	Married-civ-spo...	Some-college	Columbia	Self-e
19	☒	Male	57	Asian-Pac-Island...	Married-civ-spo...	Bachelors	South	Priva
20	☒	Male	56	White	Married-civ-spo...	9th	Columbia	Self-e
21	☒	Male	55	Other	Married-civ-spo...	Bachelors	India	Priva
22	☒	Male	63	Black	Widowed	HS-grad	United-States	Feder

Data transformation

Attribute metadata

Type: Quasi-identifying

Transformation: Generalization

Minimum: All

Maximum: All

	Level-0	Level-1	Level-2	Level-3	Level-4	
1		[1, 5]	[1, 10]	[1, 20]	*	
2		[1, 5]	[1, 10]	[1, 20]	*	
3		[1, 5]	[1, 10]	[1, 20]	*	
4		[1, 5]	[1, 10]	[1, 20]	*	
5		[1, 5]	[1, 10]	[1, 20]	*	
6		[6, 10]	[1, 10]	[1, 20]	*	
7		[6, 10]	[1, 10]	[1, 20]	*	
8		[6, 10]	[1, 10]	[1, 20]	*	
9		[6, 10]	[1, 10]	[1, 20]	*	
10		[6, 10]	[1, 10]	[1, 20]	*	

Privacy models

Population

Costs and benefits

Type

(k)

Model

5-Anonymity

Attribute

Vermelho indica um atributo identificador
Amarelo indica um atributo quase identificador
Roxo indica um atributo sensível
Verde indica um atributo não sensível



Configure transformation Explore results Analyze utility Analyze risk

Input data						
	sex	age	race	marital-status	education	native-country
1	female	73	White	Married-civ-spo...	HS-grad	United-States
2	female	76	Black	Widowed	10th	United-States
3	female	72	White	Married-civ-spo...	HS-grad	England
4	female	58	White	Married-civ-spo...	HS-grad	Greece
5	female	90	Black	Divorced	Some-college	United-States
6	female	60	White	Widowed	Some-college	England
7	female	59	White	Widowed	HS-grad	Germany
8	female	55	White	Divorced	HS-grad	England
9	female	84	White	Widowed	HS-grad	United-States
10	female	81	White	Married-civ-spo...	10th	United-States
11	male	52	White	Never-married	5th-6th	United-States
12	male	71	Black	Separated	Some-college	United-States
13	male	74	White	Separated	Some-college	United-States
14	male	71	White	Widowed	HS-grad	United-States
15	male	71	White	Widowed	Masters	United-States
16	male	51	Black	Married-civ-spo...	11th	United-States
17	male	61	Asian-Pac-Islander	Married-civ-spo...	HS-grad	South
18	male	60	White	Married-civ-spo...	Some-college	Columbia
19	male	57	Asian-Pac-Islander	Married-civ-spo...	Bachelors	South
20	male	56	White	Married-civ-spo...	9th	Columbia
21	male	55	Other	Married-civ-spo...	Bachelors	India
22	male	63	Black	Widowed	HS-grad	United-States
23	male	53	Asian-Pac-Islander	Divorced	Bachelors	United-States
24	male	53	Other	Married-spouse...	12th	Dominican-Re
25	male	85	White	Married-civ-spo...	Bachelors	Poland
26	female	67	White	Never-married	Masters	United-States

Data transformation			Attribute metadata	
Type:	Quasi-identifying	Transformation:	Generalization	
Minimum:	All	Maximum:	All	
Level-0	Level-1			
Asian-Pac-Islander	*			
Amer-Indian-Eskimo	*			
Other	*			
Black	*			

Privacy models			Population	Costs and benefits
Type	Model	Attribute		
(k)	5-Anonymity			

General settings

Utility measure

Coding model

Attribute weights

?

Suppression limit:

100%

Approximate:

☒

 Assume practical monotonicity



Attribute: education Transformations: 12960 Selected: [0, 2, 0, 1, 2, 1, 1, 1, 0] Applied: [0, 2, 0, 1, 2, 1, 1, 1, 0]

Configure transformation Explore results Analyze utility Analyze risk

Input data

		sex	age	race	marital-status	education	native-country
1	☒	Female	60	White	Divorced	bachelors	United-St
2	☒	Female	59	White	Married-spouse...	Bachelors	United-St
3	☒	Female	57	White	Divorced	Some-college	United-St
4	☒	Female	56	White	Divorced	Bachelors	United-St
5	☒	Female	56	White	Divorced	Some-college	United-St
6	☒	Female	54	White	Divorced	Bachelors	United-St
7	☒	Female	52	White	Divorced	Some-college	United-St
8	☒	Female	52	White	Divorced	Some-college	United-St
9	☒	Female	52	White	Widowed	Bachelors	United-St
10	☒	Female	52	White	Separated	Some-college	United-St
11	☒	Female	51	White	Divorced	Masters	United-St
12	☒	Male	59	White	Married-civ-spo...	Some-college	United-St
13	☒	Male	58	White	Married-civ-spo...	Bachelors	United-St
14	☒	Male	58	White	Married-civ-spo...	Bachelors	United-St
15	☒	Male	57	White	Married-civ-spo...	Bachelors	United-St
16	☒	Male	56	White	Married-civ-spo...	Bachelors	United-St
17	☒	Male	56	White	Married-civ-spo...	Bachelors	United-St
18	☒	Male	55	White	Married-civ-spo...	Masters	United-St
19	☒	Male	54	White	Married-civ-spo...	Bachelors	United-St
20	☒	Male	54	White	Married-civ-spo...	Bachelors	United-St
21	☒	Male	52	White	Married-civ-spo...	Assoc-voc	United-St
22	☒	Male	52	White	Married-civ-spo...	Masters	United-St
23	☒	Male	51	White	Married-civ-spo...	Some-college	United-St
24	☒	Male	51	White	Married-civ-spo...	Bachelors	United-St
25	☒	Male	58	Black	Married-civ-spo...	Bachelors	United-St

Data transformation

Attribute metadata

Type: Quasi-identifying Transformation: Generalization

Minimum: All Maximum: All

Level-0	Level-1	Level-2	Level-3	
Some-college	Undergraduate	Higher education	*	
11th	High School	Secondary educ...	*	
HS-grad	High School	Secondary educ...	*	
Prof-school	Professional Education	Higher education	*	
Assoc-acdm	Professional Education	Higher education	*	
Assoc-voc	Professional Education	Higher education	*	
9th	High School	Secondary educ...	*	
7th-8th	High School	Secondary educ...	*	
12th	High School	Secondary educ...	*	
Masters	Graduate	Higher education	*	

Privacy models

Population

Costs and benefits

Type	Model	Attribute
(k)	5-Anonymity	

General settings

Utility measure

Coding model

Attribute weights

Measure: Loss

File Edit View Help



Attribute: education Transformations: 90720 Selected: [0, 1, 0, 0, 0, 0, 0, 0, 0] Applied: [0, 1, 0, 0, 0, 0, 0, 0, 0]

Configure transformation Explore results Analyze utility Analyze risk

Input data

	sex	age	race	marital-status	education	native-country	
1	female	73	White	Married-civ-spo...	HS-grad	United-States	Self-e
2	female	76	Black	Widowed	10th	United-States	Privat
3	female	72	White	Married-civ-spo...	HS-grad	England	Privat
4	female	58	White	Married-civ-spo...	HS-grad	Greece	Privat
5	female	90	Black	Divorced	Some-college	United-States	Privat
6	female	60	White	Widowed	Some-college	England	Feder
7	female	59	White	Widowed	HS-grad	Germany	Feder
8	female	55	White	Divorced	HS-grad	England	Privat
9	female	84	White	Widowed	HS-grad	United-States	Local
10	female	81	White	Married-civ-spo...	10th	United-States	Self-e
11	male	52	White	Never-married	5th-6th	United-States	Privat
12	male	71	Black	Separated	Some-college	United-States	Self-e
13	male	74	White	Separated	Some-college	United-States	State-
14	male	71	White	Widowed	HS-grad	United-States	Feder
15	male	71	White	Widowed	Masters	United-States	Local
16	male	51	Black	Married-civ-spo...	11th	United-States	Local
17	male	61	Asian-Pac-Island...	Married-civ-spo...	HS-grad	South	Self-e
18	male	60	White	Married-civ-spo...	Some-college	Columbia	Self-e
19	male	57	Asian-Pac-Island...	Married-civ-spo...	Bachelors	South	Privat
20	male	56	White	Married-civ-spo...	9th	Columbia	Self-e
21	male	55	Other	Married-civ-spo...	Bachelors	India	Privat
22	male	63	Black	Widowed	HS-grad	United-States	Feder
23	male	53	Asian-Pac-Island...	Divorced	Bachelors	United-States	Privat
24	male	53	Other	Married-spouse...	12th	Dominican-Rep...	Privat
25	male	85	White	Married-civ-spo...	Bachelors	Poland	Privat

Data transformation

Attribute metadata

Type:

Sensitive

Transformation:

Generalization

Minimum:

All

Maximum:

All

Level-0	Level-1	Level-2	Level-3	Level-4	Level-5
HS-grad	HS-grad	HS-grad	HS-grad	HS-grad	HS-grad
Some-college	Some-college	Some-college	Some-college	Some-college	Some-college
Bachelors	Bachelors	Bachelors	Bachelors	Bachelors	Bachelors
Masters	Masters	Masters	Masters	Masters	Masters
Assoc-voc	Assoc-voc	Assoc-voc	Assoc-voc	Assoc-voc	Assoc-voc
11th	11th	11th	11th	11th	11th
Assoc-acdm	Assoc-acdm	Assoc-acdm	Assoc-acdm	Assoc-acdm	Assoc-acdm
10th	10th	10th	10th	10th	10th
7th-8th	7th-8th	7th-8th	7th-8th	7th-8th	*
Prof-school	Prof-school	Prof-school	Prof-school	Prof-school	*

Privacy models

Population

Costs and benefits

Type

Model

Attribute

(k)

5-Anonymity

General settings

Utility measure

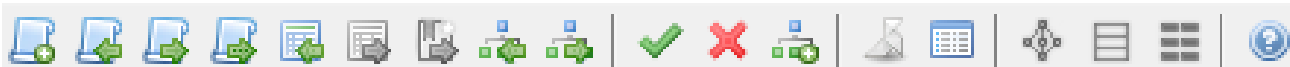
Coding model

Attribute weights

Suppression limit:

100%

File Edit View Help



Attribute: native-country Transformations: 12960 Selected: [0, 2, 0, 1, 2, 1, 1, 1, 0] Applied: [0, 2, 0, 1, 2, 1, 1, 1, 0]

Configure transformation Explore results Analyze utility Analyze risk

Input data

	sex	age	race	marital-status	education	native-country
1	e	60	White	Divorced	Bachelors	United-States
2	e	59	White	Married-spouse...	Bachelors	United-States
3	e	57	White	Divorced	Some-college	United-States
4	e	56	White	Divorced	Bachelors	United-States
5	e	56	White	Divorced	Some-college	United-States
6	e	54	White	Divorced	Bachelors	United-States
7	e	52	White	Divorced	Some-college	United-States
8	e	52	White	Divorced	Some-college	United-States
9	e	52	White	Widowed	Bachelors	United-States
10	e	52	White	Separated	Some-college	United-States
11	e	51	White	Divorced	Masters	United-States
12		59	White	Married-civ-spo...	Some-college	United-States
13		58	White	Married-civ-spo...	Bachelors	United-States
14		58	White	Married-civ-spo...	Bachelors	United-States
15		57	White	Married-civ-spo...	Bachelors	United-States
16		56	White	Married-civ-spo...	Bachelors	United-States
17		56	White	Married-civ-spo...	Bachelors	United-States
18		55	White	Married-civ-spo...	Masters	United-States
19		54	White	Married-civ-spo...	Bachelors	United-States
20		54	White	Married-civ-spo...	Bachelors	United-States
21		52	White	Married-civ-spo...	Assoc-voc	United-States
22		52	White	Married-civ-spo...	Masters	United-States
23		51	White	Married-civ-spo...	Some-college	United-States
24		51	White	Married-civ-spo...	Bachelors	United-States
25		58	Black	Married-civ-spo...	Bachelors	United-States

Data transformation Attribute metadata

Type: Quasi-identifying Transformation: Generalization

Minimum: All Maximum: All

Level-0	Level-1	Level-2	
Cambodia	Asia	*	
England	Europa	*	
Puerto-Rico	North America	*	
Canada	North America	*	
Germany	Europe	*	
Outlying-US(Guam-U...	North America	*	
India	Asia	*	
Japan	Asia	*	
Greece	Europe	*	
South	Africa	*	

Privacy models Population Costs and benefits

Type	Model	Attribute
(K)	5-Anonymity	

General settings Utility measure Coding model Attribute weights

Measure: Loss

Input data			
Classification performance			
Quality models			
Target variable	Baseline accuracy	Accuracy	
sex	70.90654%	86.19115%	
age	10.03162%	7.46662%	
race	88.15882%	89.75755%	
marital-status	62.29796%	70.67814%	
education	34.17077%	36.22628%	
native-country	92.04146%	92.04146%	
workclass	62.052%	64.07238%	
occupation	16.56711%	32.57203%	
salary-class	66.1279%	77.03795%	
Class	Sensitivity	Specificity	Brier score
Divorced	43.75%	88.58102%	11.84029%
Married-civ-spouse	93.20361%	64.0261%	14.17796%
Married-spouse-absent	0%	100%	1.19474%
Never-married	5.45455%	98.98248%	6.34535%
Separated	0%	99.96395%	2.40049%
Widowed	46.93548%	94.14432%	7.3861%
Minimum	46.93548%	64.0261%	1.19474%
Average	31.55727%	90.94964%	7.22416%
Maximum	93.20361%	100%	14.17796%

Output data					
Classification performance					
Quality models					
Target variable	Baseline accuracy	Accuracy	Original accuracy	Relative accuracy	Brier skill score
sex	70.90654%	83.16936%	86.19115%	80.22989%	-0.2219
age	68.27126%	68.27126%	68.53479%	0%	-0.05385
race	88.15882%	88.52776%	89.75755%	23.07692%	-0.14116
marital-status	62.29796%	82.37878%	81.69361%	103.53261%	-0.00732
education	46.8201%	63.66831%	70.39705%	71.46051%	-0.1655
native-country	95.52003%	95.73085%	96.18763%	31.57895%	-0.10175
workclass	82.85313%	82.88826%	83.04638%	18.18182%	-0.09261
occupation	34.68025%	46.08222%	49.70134%	75.90643%	-0.05943
salary-class	66.1279%	73.27829%	77.03795%	65.53945%	-0.10715
Class	Sensitivity	Specificity	Brier score		
Spouse not present	65.61044%	92.52679%	14.04454%		
Spouse present	92.52679%	65.61044%	14.04454%		
Minimum	65.61044%	65.61044%	14.04454%		
Average	79.06861%	79.06861%	14.04454%		
Maximum	92.52679%	92.52679%	14.04454%		

OBRIGADA



ecsamonteiro@gmail.com