

TMF Workshop: Anonymization tools and their practical relevance (for biomedical research)



ARX

A Comprehensive Tool for Anonymizing Biomedical Data

Fabian Prasser, Florian Kohlmayer, Klaus A. Kuhn

Lehrstuhl für Medizinische Informatik
Institut für Medizinische Statistik und Epidemiologie
Klinikum rechts der Isar der TU München

ARX: Towards useful data anonymization

- **Usability has many dimensions**
 - Ability to balance data utility with privacy requirements
- **Need to support a broad spectrum of methods**
 - Privacy models
 - Transformation models
 - Methods for analyzing data utility
 - Methods for analyzing risks
- **Further “non-functional” requirements**
 - Integrated and harmonized: ARX is not a “tool box”
 - Compatibility (syntactic and semantic)
 - Performance and scalability
 - Intuitive visualization and parameterization
 - Provide methods to end-users as well as programmers

ARX: Highlights

- **Compatibility**
 - **Built-in data import facilities**
 - Relational databases (MS SQL, DB2, SQLite, MySQL)
 - MS Excel
 - CSV (all common formats, auto-detection)
 - **Support for different data types and scales of measure**
 - Strings (with nominal and ordinal scale)
 - Dates (interval scale)
 - Numbers (ratio scale)
 - Automatic detection of data types and formats
 - **Methods for handling and cleaning low-quality data**
 - Handles missing and invalid values correctly
 - In privacy models, transformation methods, visualizations
 - Manual removal of tuples, query interface, find & replace

ARX: Highlights

- **Flexible transformation methods**
 - **Global recoding**
 - Full-domain generalization
 - Top- & bottom coding
 - **Local recoding**
 - Tuple suppression
 - **Fully integrated and parameterizable**
 - Importance of attributes, suitability of methods
- **Functional representations of transformation rules**
 - Especially functional representations of hierarchies
 - Support for categorical and continuous variables (categorization)
- **Multiple methods for measuring data utility**
 - Parametrizable, e.g., with different aggregate functions
 - Use functional representations of transformation rules

ARX: Highlights

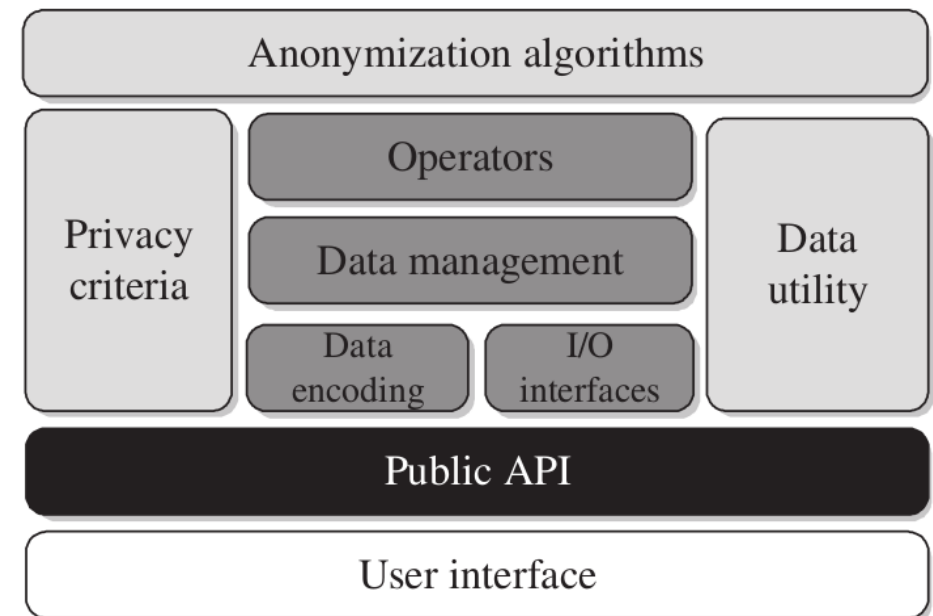
- **Multiple apriori privacy models**
 - k-Anonymity
 - ℓ -Diversity (distinct, entropy, recursive-(c, ℓ))
 - t-Closeness (equal and hierarchical ground distance)
 - δ -Presence
- **Multiple methods for risk-based anonymization**
 - Sample characteristics
 - Average cell size
 - Sample uniqueness
 - Super-population models
 - Decision rule by Dankar et al.
 - Based on models by Pitman, Zayatz and the SNB model
- **Support for arbitrary combinations of these models**
 - Optimal solution within our coding model

ARX: Highlights

- **Scalability:** ARX can handle large datasets (several million data entries) on commodity hardware
- **Efficient in-memory data management engine**
 - Works with compressed data representations
 - Tight coupling between transformation operators and the „database kernel”
 - Provides a space-time trade-off
- **Optimized search strategy:** Based on multiple pruning strategies
- **Efficient implementations of further complex tasks**
 - Evaluations of privacy criteria (e.g. t-closeness, δ -presence)
 - Methods for solving non-linear equation systems
 - Background jobs in the user interface

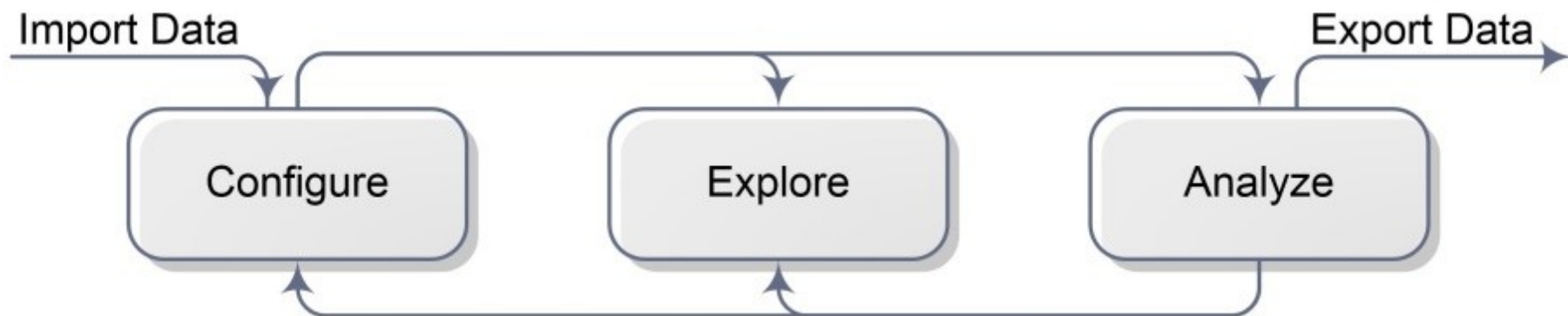
ARX: Highlights

- **Comprehensive graphical interface**
 - Scalability comparable to the ARX API
 - Supports all methods provided by the ARX API
 - Cross-platform (Windows, Linux/GTK, OSX) with native interfaces
 - Available as binary distributions with installers
- **Independent API**
 - User interface sits on top of the API
 - Java library
 - All methods provided by ARX are first-class citizens in both worlds



ARX: Anonymization workflow

- Iterative process to successively refine transformations
- Supported by the scalability of our framework
- Three (repeating) steps, mapped to four perspectives



- Define transformation model
- Define privacy model
- Define coding model

- Filter and analyze the solution space
- Organize transformations

- Compare and analyze input and output
- Regarding risks and utility

ARX: Wizard for data import

Import data

CSV

Please provide the information requested below

Location:

Delimiter:

Quote:

Escape:

Linebreak:

☒ First row contains column names

sex	age	race	marital-status	education	native-country
Male	39	White	Never-married	Bachelors	United-States
Male	50	White	Married-civ-s...	Bachelors	United-States
Male	38	White	Divorced	HS-grad	United-States
Male	53	Black	Married-civ-s...	11th	United-States
Female	28	Black	Married-civ-s...	Bachelors	Cuba
Female	37	White	Married-civ-s...	Masters	United-States
Female	40	Black	Married-civ-s...	9th	Jamaica

Import data

JDBC

Please provide the information requested below

Type:

Server:

Port:

Username:

Password:

Database:

Import data

Excel

Please provide the information requested below

Location:

Sheet:

☒ First row contains column names

age	gender	zipcode
34	male	81667
45	female	81675
66	male	81925
70	female	81931
34	female	81931
70	male	81931
45	male	81931

< Back Next > Finish Cancel

Import data

Columns

Please check and/or modify the detected columns

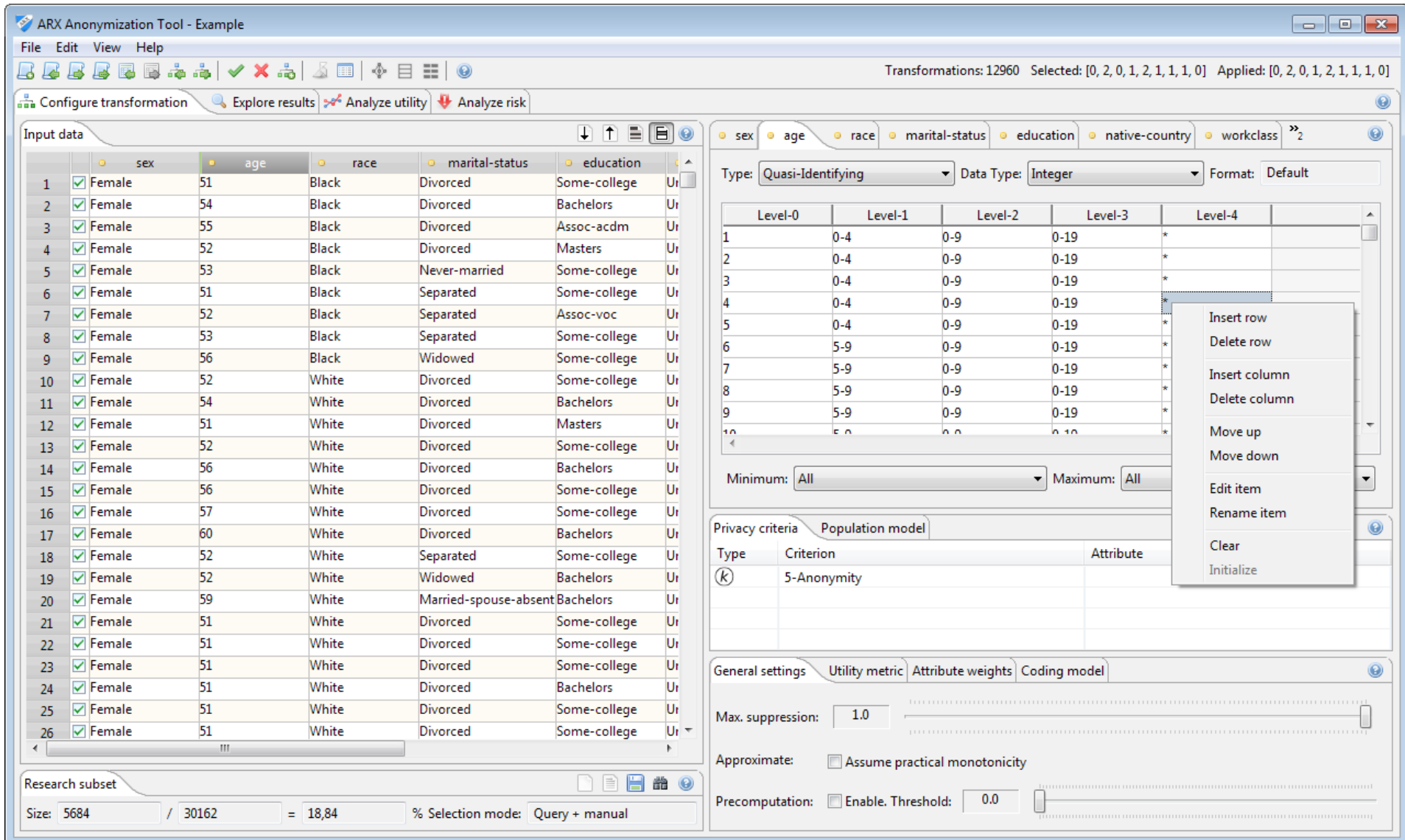
DA	Name	Datatype	Format
<input checked="" type="checkbox"/>	sex	String	
<input checked="" type="checkbox"/>	age	Integer 100%	
<input checked="" type="checkbox"/>	race	Integer 100%	
<input checked="" type="checkbox"/>	marital-status	Decimal (#,##0) 100%	
<input checked="" type="checkbox"/>	education	Decimal (#,##0,###) 100%	
<input checked="" type="checkbox"/>	native-country	String 100%	
<input checked="" type="checkbox"/>	workclass	String	
<input checked="" type="checkbox"/>	occupation	String	
<input checked="" type="checkbox"/>	salary-class	String	

Move up Move down

☒ Perform data cleansing

< Back Next > Finish Cancel

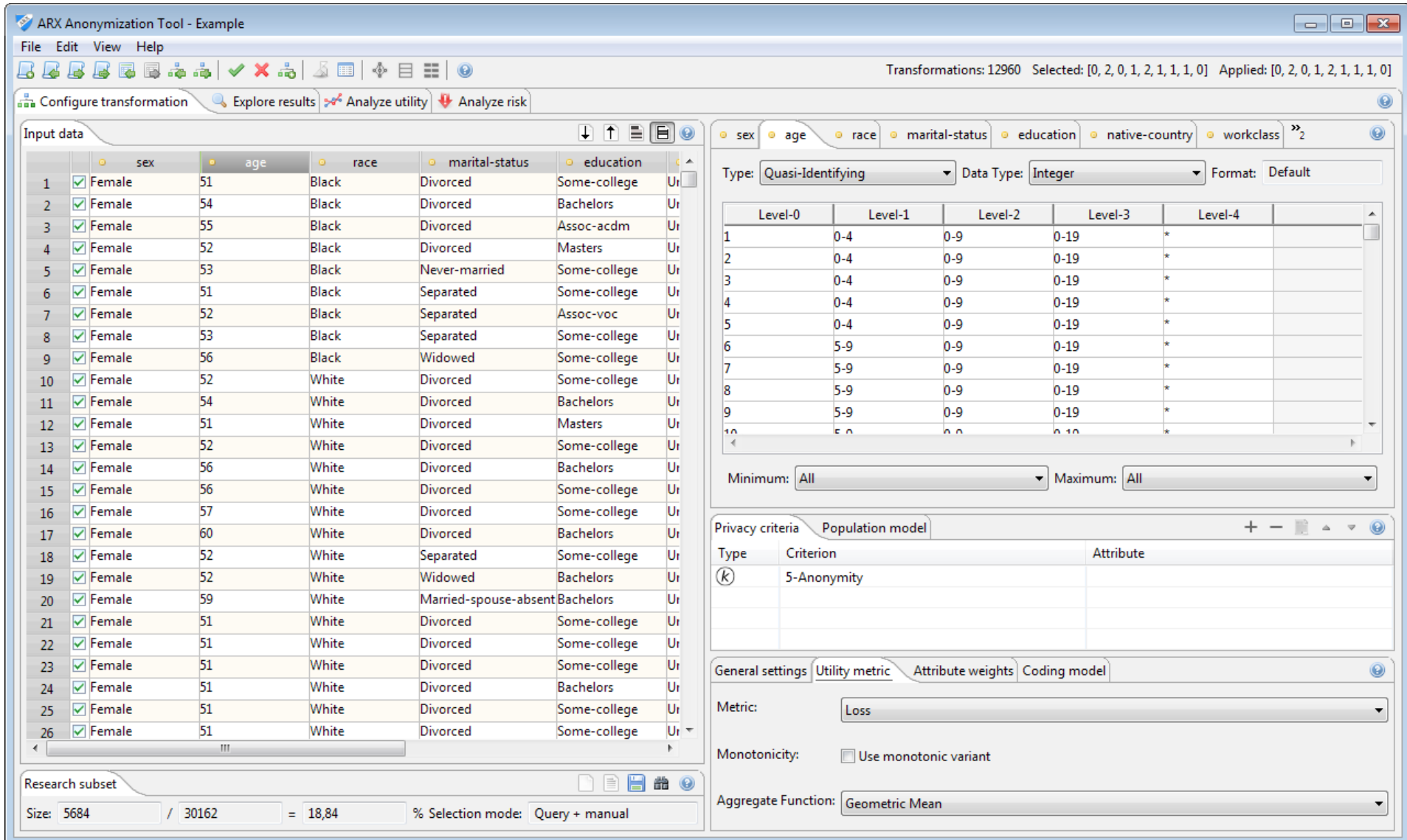
ARX: Configuration (1)



The screenshot displays the ARX Anonymization Tool interface. The main window is titled "ARX Anonymization Tool - Example". The top menu bar includes File, Edit, View, and Help. The toolbar contains various icons for file operations and analysis. The main workspace is divided into several panes:

- Input data:** A table showing a dataset with columns: sex, age, race, marital-status, education, and a selection column. The table contains 26 rows of data.
- Transformations:** A list of transformations with a status bar showing "Transformations: 12960 Selected: [0, 2, 0, 1, 2, 1, 1, 0] Applied: [0, 2, 0, 1, 2, 1, 1, 0]".
- Configuration pane:**
 - Type:** Quasi-Identifying
 - Data Type:** Integer
 - Format:** Default
 - Level-0 to Level-4:** A table showing the configuration for each level. A context menu is open over the Level-4 column, showing options: Insert row, Delete row, Insert column, Delete column, Move up, Move down, Edit item, Rename item, Clear, and Initialize.
 - Privacy criteria:** A table with columns Type, Criterion, and Attribute. It shows a criterion of 5-Anonymity.
 - General settings:**
 - Max. suppression:** 1.0
 - Approximate:** ☐ Assume practical monotonicity
 - Precomputation:** ☐ Enable. Threshold: 0.0
- Research subset:** A section at the bottom showing the size of the subset: Size: 5684 / 30162 = 18,84 % Selection mode: Query + manual

ARX: Configuration (2)



The screenshot shows the ARX Anonymization Tool - Example interface. The main window displays a list of input data with columns: sex, age, race, marital-status, education, and a utility score. The data is sorted by utility score in descending order. The right panel shows the configuration for a transformation on the 'age' attribute.

Input data:

	sex	age	race	marital-status	education	Utility
1	Female	51	Black	Divorced	Some-college	Ur
2	Female	54	Black	Divorced	Bachelors	Ur
3	Female	55	Black	Divorced	Assoc-acdm	Ur
4	Female	52	Black	Divorced	Masters	Ur
5	Female	53	Black	Never-married	Some-college	Ur
6	Female	51	Black	Separated	Some-college	Ur
7	Female	52	Black	Separated	Assoc-voc	Ur
8	Female	53	Black	Separated	Some-college	Ur
9	Female	56	Black	Widowed	Some-college	Ur
10	Female	52	White	Divorced	Some-college	Ur
11	Female	54	White	Divorced	Bachelors	Ur
12	Female	51	White	Divorced	Masters	Ur
13	Female	52	White	Divorced	Some-college	Ur
14	Female	56	White	Divorced	Bachelors	Ur
15	Female	56	White	Divorced	Some-college	Ur
16	Female	57	White	Divorced	Some-college	Ur
17	Female	60	White	Divorced	Bachelors	Ur
18	Female	52	White	Separated	Some-college	Ur
19	Female	52	White	Widowed	Bachelors	Ur
20	Female	59	White	Married-spouse-absent	Bachelors	Ur
21	Female	51	White	Divorced	Some-college	Ur
22	Female	51	White	Divorced	Some-college	Ur
23	Female	51	White	Divorced	Some-college	Ur
24	Female	51	White	Divorced	Bachelors	Ur
25	Female	51	White	Divorced	Some-college	Ur
26	Female	51	White	Divorced	Some-college	Ur

Transformation Configuration:

- Attributes: sex, age, race, marital-status, education, native-country, workclass
- Type: Quasi-Identifying
- Data Type: Integer
- Format: Default

Level-based Configuration:

Level-0	Level-1	Level-2	Level-3	Level-4
1	0-4	0-9	0-19	*
2	0-4	0-9	0-19	*
3	0-4	0-9	0-19	*
4	0-4	0-9	0-19	*
5	0-4	0-9	0-19	*
6	5-9	0-9	0-19	*
7	5-9	0-9	0-19	*
8	5-9	0-9	0-19	*
9	5-9	0-9	0-19	*
10	5-9	0-9	0-19	*

Privacy criteria:

Type	Criterion	Attribute
(k)	5-Anonymity	

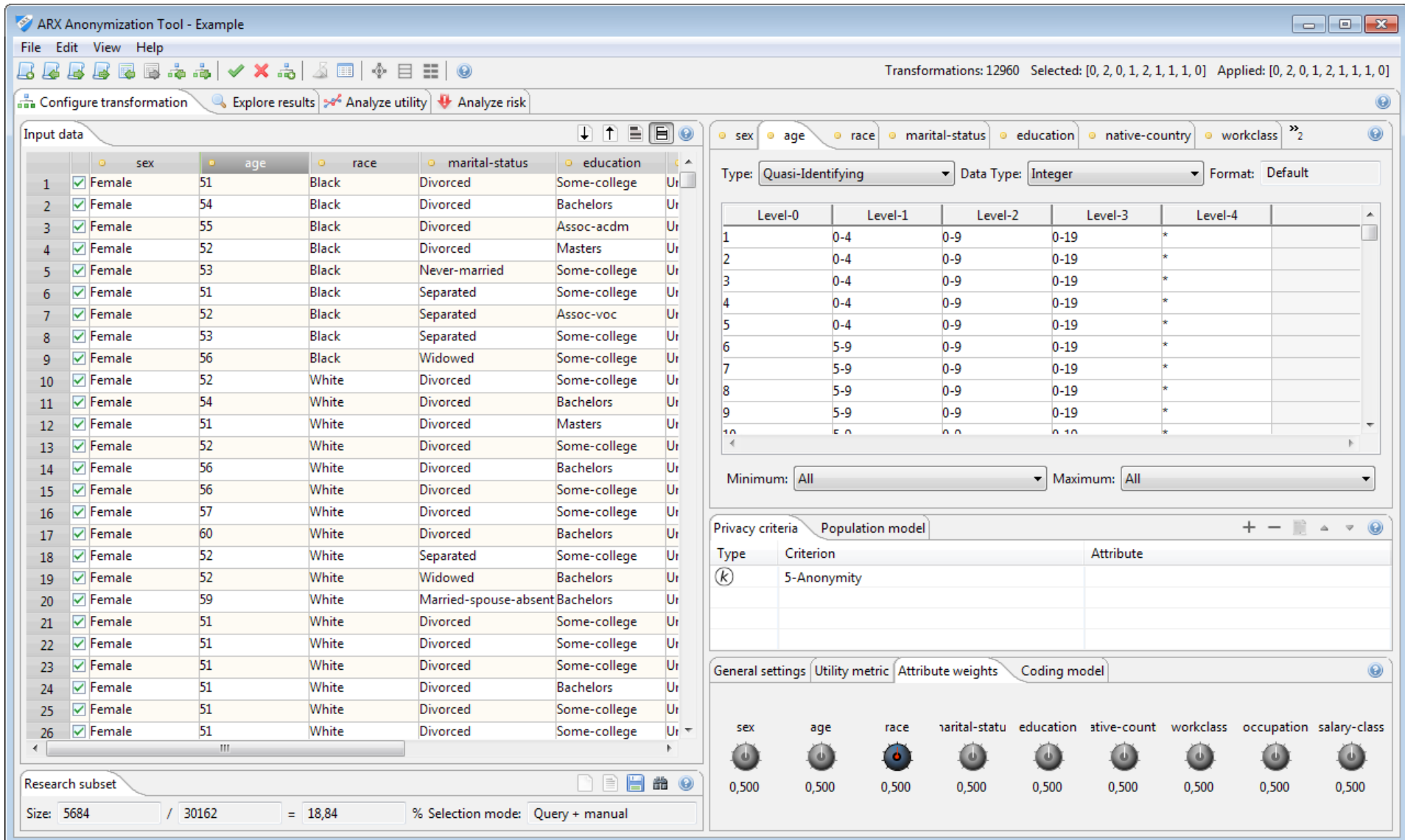
General settings:

- Metric: Loss
- Monotonicity: ☐ Use monotonic variant
- Aggregate Function: Geometric Mean

Research subset:

Size: 5684 / 30162 = 18,84 % Selection mode: Query + manual

ARX: Configuration (3)



ARX Anonymization Tool - Example

File Edit View Help

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

Configure transformation Explore results Analyze utility Analyze risk

Input data

	sex	age	race	marital-status	education	
1	Female	51	Black	Divorced	Some-college	Ur
2	Female	54	Black	Divorced	Bachelors	Ur
3	Female	55	Black	Divorced	Assoc-acdm	Ur
4	Female	52	Black	Divorced	Masters	Ur
5	Female	53	Black	Never-married	Some-college	Ur
6	Female	51	Black	Separated	Some-college	Ur
7	Female	52	Black	Separated	Assoc-voc	Ur
8	Female	53	Black	Separated	Some-college	Ur
9	Female	56	Black	Widowed	Some-college	Ur
10	Female	52	White	Divorced	Some-college	Ur
11	Female	54	White	Divorced	Bachelors	Ur
12	Female	51	White	Divorced	Masters	Ur
13	Female	52	White	Divorced	Some-college	Ur
14	Female	56	White	Divorced	Bachelors	Ur
15	Female	56	White	Divorced	Some-college	Ur
16	Female	57	White	Divorced	Some-college	Ur
17	Female	60	White	Divorced	Bachelors	Ur
18	Female	52	White	Separated	Some-college	Ur
19	Female	52	White	Widowed	Bachelors	Ur
20	Female	59	White	Married-spouse-absent	Bachelors	Ur
21	Female	51	White	Divorced	Some-college	Ur
22	Female	51	White	Divorced	Some-college	Ur
23	Female	51	White	Divorced	Some-college	Ur
24	Female	51	White	Divorced	Bachelors	Ur
25	Female	51	White	Divorced	Some-college	Ur
26	Female	51	White	Divorced	Some-college	Ur

Research subset

Size: 5684 / 30162 = 18,84 % Selection mode: Query + manual

sex age race marital-status education native-country workclass

Type: Quasi-Identifying Data Type: Integer Format: Default

Level-0	Level-1	Level-2	Level-3	Level-4
1	0-4	0-9	0-19	*
2	0-4	0-9	0-19	*
3	0-4	0-9	0-19	*
4	0-4	0-9	0-19	*
5	0-4	0-9	0-19	*
6	5-9	0-9	0-19	*
7	5-9	0-9	0-19	*
8	5-9	0-9	0-19	*
9	5-9	0-9	0-19	*
10	5-9	0-9	0-19	*

Minimum: All Maximum: All

Privacy criteria Population model

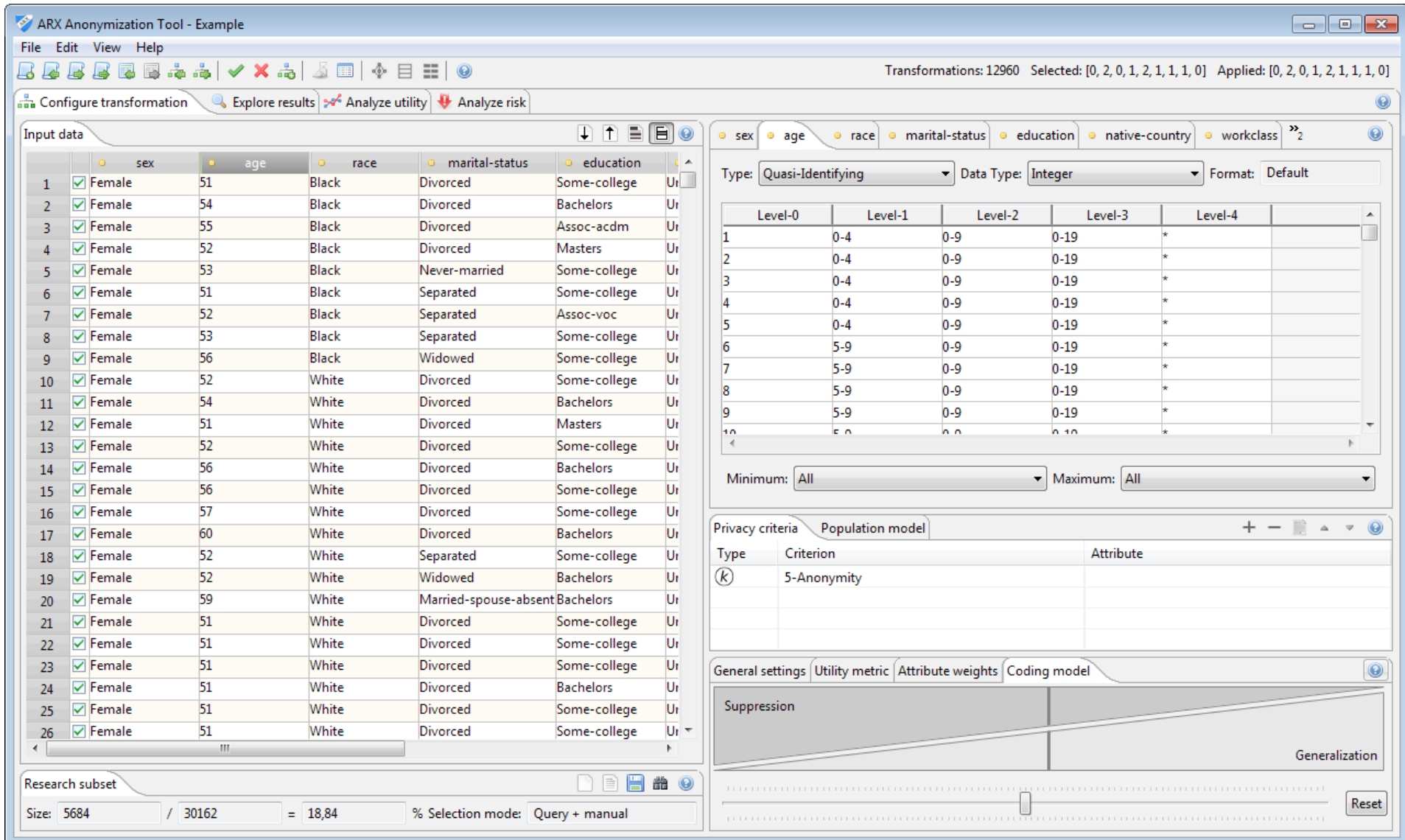
Type	Criterion	Attribute
(k)	5-Anonymity	

General settings Utility metric Attribute weights Coding model

sex age race marital-status education native-count workclass occupation salary-class

0,500 0,500 0,500 0,500 0,500 0,500 0,500 0,500 0,500

ARX: Configuration (4)

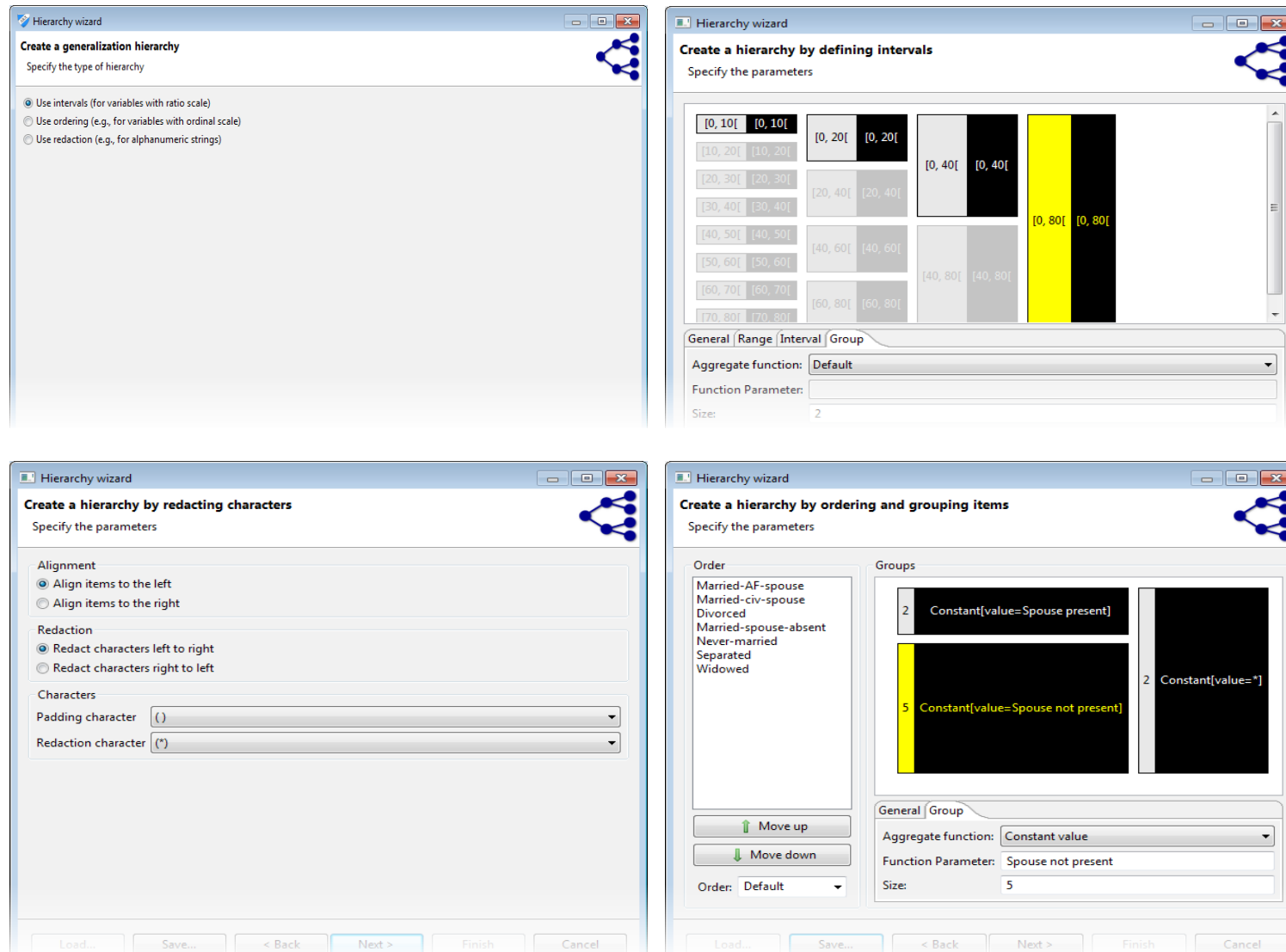


The screenshot displays the ARX Anonymization Tool interface. The main window is titled "ARX Anonymization Tool - Example". The top menu bar includes File, Edit, View, and Help. The toolbar contains various icons for file operations and analysis. The main workspace is divided into several panels:

- Input data:** A table showing 26 rows of data with columns: sex, age, race, marital-status, education, and a selection column. The data is filtered to show only females.
- Transformations:** A list of transformations applied to the data. The selected transformation is "Quasi-Identifying" with a data type of "Integer" and a format of "Default".
- Privacy criteria:** A table showing the privacy criteria for the transformation. The criteria are "5-Anonymity" and "Attribute".
- General settings:** A section for configuring the transformation, including a "Suppression" slider and a "Generalization" slider.
- Research subset:** A section showing the size of the research subset (5684) and the selection mode (Query + manual).

The bottom status bar shows the transformation count: 12960. The selected transformations are [0, 2, 0, 1, 2, 1, 1, 0] and the applied transformations are [0, 2, 0, 1, 2, 1, 1, 0].

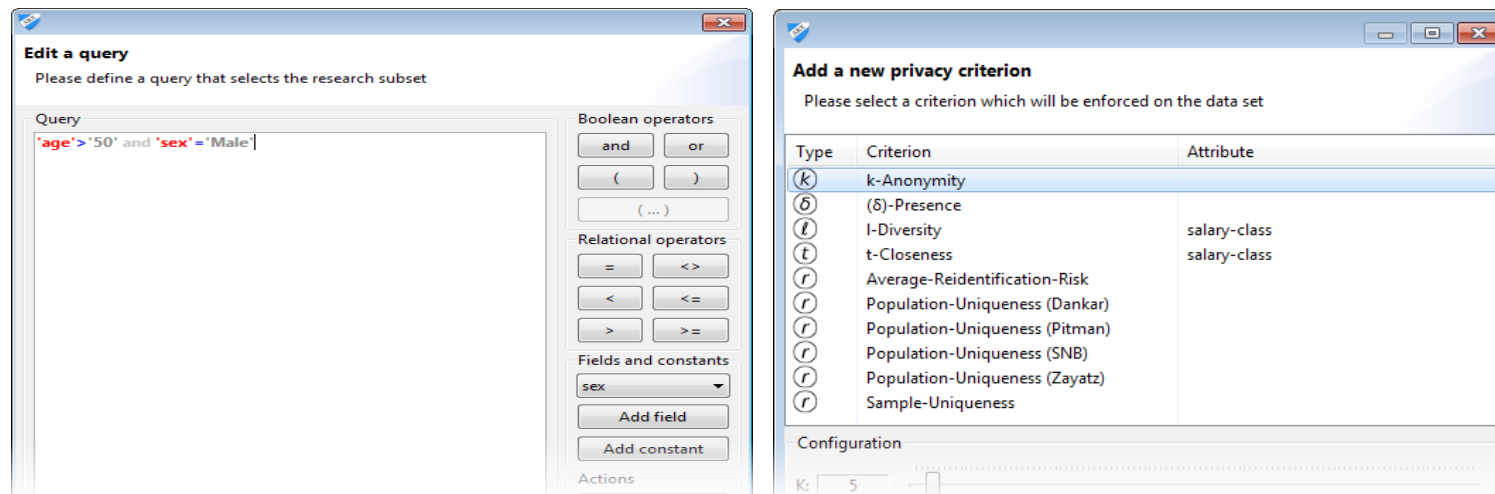
ARX: Wizard for transformation rules



The ARX Hierarchy Wizard provides four methods to create a hierarchy:

- Create a generalization hierarchy:** Specify the type of hierarchy.
 - ☒ Use intervals (for variables with ratio scale)
 - ☐ Use ordering (e.g., for variables with ordinal scale)
 - ☐ Use redaction (e.g., for alphanumeric strings)
- Create a hierarchy by defining intervals:** Specify the parameters.
 - General tab: Aggregate function: Default, Function Parameter: , Size: 2
 - Interval tab: Shows a grid of intervals (e.g., [0, 10[, [10, 20[, ..., [70, 80[) with a yellow highlight on [0, 80[.
- Create a hierarchy by redacting characters:** Specify the parameters.
 - Alignment: ☒ Align items to the left, ☐ Align items to the right
 - Redaction: ☒ Redact characters left to right, ☐ Redact characters right to left
 - Characters: Padding character: (), Redaction character: (*)
- Create a hierarchy by ordering and grouping items:** Specify the parameters.
 - Order: Married-AF-spouse, Married-civ-spouse, Divorced, Married-spouse-absent, Never-married, Separated, Widowed
 - Groups: 2 Constant[value=Spouse present], 5 Constant[value=Spouse not present], 2 Constant[value=*]
 - General tab: Aggregate function: Constant value, Function Parameter: Spouse not present, Size: 5

ARX: Further dialogs



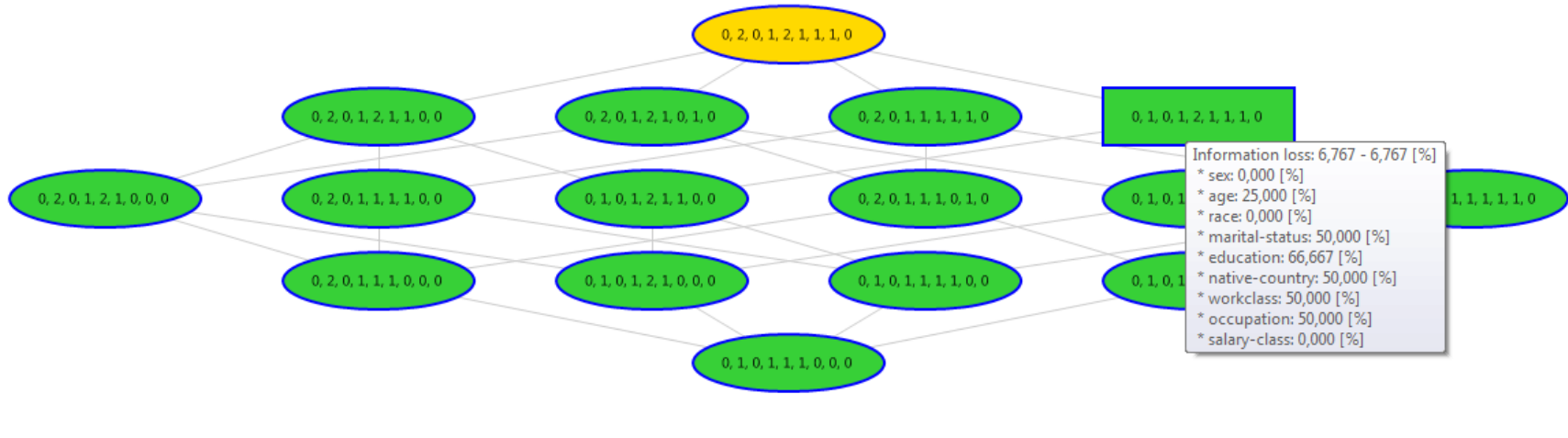
ARX: Exploration (1)

ARX Anonymization Tool - Example

File Edit View Help

Transformations: 12960 Selected: [0, 1, 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



Information loss: 6,767 - 6,767 [%]
 * sex: 0,000 [%]
 * age: 25,000 [%]
 * race: 0,000 [%]
 * marital-status: 50,000 [%]
 * education: 66,667 [%]
 * native-country: 50,000 [%]
 * workclass: 50,000 [%]
 * occupation: 50,000 [%]
 * salary-class: 0,000 [%]

Lattice List Tiles

Filter

	0	1	2	3	4
sex	✓	✗			
age	✗	✓	✓	✗	✗
race	✗	✗			
marital-status	✓	✓	✗		
education	✗	✓	✓	✗	

☒ Anonymous ☐ Non-anonymous ☐ Unknown

Information loss:

Clipboard

Node	Comment
[0, 2, 0, 1, 2, 1, 1, 1, 0]	Minimal information loss
[0, 1, 0, 1, 2, 1, 1, 1, 0]	Age is less generalized

Properties

Property	Value
Anonymous	ANONYMOUS
Min. info. loss	0.3315901556540435 [6,767%]
Max. info. loss	0.3315901556540435 [6,767%]
Successors	9
Predecessors	6
Transformation	[0, 1, 0, 1, 2, 1, 1, 1, 0]
Checked	true

ARX: Exploration (2)

ARX Anonymization Tool - Example

File Edit View Help

Transformations: 12960 Selected: [0, 1, 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

Transformation	Anonymity	Min. Info. Loss	Max. Info. Loss
[0, 2, 0, 1, 2, 1, 1, 1, 0]	ANONYMOUS	0.28307655989412295 [0,000%]	0.28307655989412295 [0,000%]
[0, 2, 0, 1, 2, 0, 1, 1, 0]	ANONYMOUS	0.28369947575043075 [0,087%]	0.28369947575043075 [0,087%]
[0, 2, 0, 1, 1, 1, 1, 1, 0]	ANONYMOUS	0.30521601292778766 [3,088%]	0.30521601292778766 [3,088%]
[0, 2, 0, 1, 1, 0, 1, 1, 0]	ANONYMOUS	0.3065634918757387 [3,276%]	0.3065634918757387 [3,276%]
[0, 2, 0, 0, 2, 1, 1, 1, 0]	ANONYMOUS	0.30757828082085226 [3,418%]	0.30757828082085226 [3,418%]
[0, 2, 0, 0, 2, 0, 1, 1, 0]	ANONYMOUS	0.31090204972648094 [3,881%]	0.31090204972648094 [3,881%]
[0, 2, 0, 1, 2, 1, 0, 1, 0]	ANONYMOUS	0.3311292560270289 [6,703%]	0.3311292560270289 [6,703%]
[0, 1, 0, 1, 2, 0, 1, 1, 0]	ANONYMOUS	0.33154208969794285 [6,760%]	0.33154208969794285 [6,760%]
[0, 1, 0, 1, 2, 1, 1, 1, 0]	ANONYMOUS	0.3315901556540435 [6,767%]	0.3315901556540435 [6,767%]
[0, 2, 0, 1, 2, 0, 0, 1, 0]	ANONYMOUS	0.3316659705759111 [6,777%]	0.3316659705759111 [6,777%]
[0, 2, 0, 1, 2, 1, 1, 0, 0]	ANONYMOUS	0.3385456412213499 [7,737%]	0.3385456412213499 [7,737%]
[0, 2, 0, 1, 2, 0, 1, 0, 0]	ANONYMOUS	0.3413197289976073 [8,124%]	0.3413197289976073 [8,124%]
[0, 2, 0, 0, 1, 1, 1, 1, 0]	ANONYMOUS	0.3454513242671491 [8,700%]	0.3454513242671491 [8,700%]
[0, 2, 0, 0, 1, 0, 1, 1, 0]	ANONYMOUS	0.3479633267296329 [9,051%]	0.3479633267296329 [9,051%]
[0, 2, 0, 0, 2, 1, 0, 1, 0]	ANONYMOUS	0.36136279212496913 [10,920%]	0.36136279212496913 [10,920%]
[0, 2, 0, 0, 2, 0, 0, 1, 0]	ANONYMOUS	0.36147737003757774 [10,936%]	0.36147737003757774 [10,936%]
[0, 1, 0, 1, 1, 1, 1, 1, 0]	ANONYMOUS	0.3623878009327741 [11,063%]	0.3623878009327741 [11,063%]
[0, 1, 0, 1, 1, 0, 1, 1, 0]	ANONYMOUS	0.3646654445341022 [11,380%]	0.3646654445341022 [11,380%]

Information loss: 6,760 - 6,760 [%]
 * sex: 0,000 [%]
 * age: 25,000 [%]
 * race: 0,000 [%]
 * marital-status: 50,000 [%]
 * education: 66,667 [%]
 * native-country: 0,000 [%]
 * workclass: 50,000 [%]
 * occupation: 50,000 [%]
 * salary-class: 0,000 [%]

Filter

	0	1	2	3	4
sex	✓	✗			
age	✓	✓	✓	✗	✗
race	✓	✗			
marital-status	✓	✓	✗		
education	✗	✓	✓	✗	

☒ Anonymous ☐ Non-anonymous ☐ Unknown

Information loss:

Clipboard

Node	Comment
[0, 2, 0, 1, 2, 1, 1, 1, 0]	Minimal information loss
[0, 1, 0, 1, 2, 1, 1, 1, 0]	Age is less generalized

Properties

Property	Value
Anonymous	ANONYMOUS
Min. info. loss	0.3315901556540435 [6,767%]
Max. info. loss	0.3315901556540435 [6,767%]
Successors	9
Predecessors	6
Transformation	[0, 1, 0, 1, 2, 1, 1, 1, 0]
Checked	true

ARX: Exploration (3)

ARX Anonymization Tool - Example

File Edit View Help

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

Configure transformation Explore results Analyze utility Analyze risk

Information loss: 6,760 - 6,760 [%]
 * sex: 0,000 [%]
 * age: 25,000 [%]
 * race: 0,000 [%]
 * marital-status: 50,000 [%]
 * education: 66,667 [%]
 * native-country: 0,000 [%]
 * workclass: 50,000 [%]
 * occupation: 50,000 [%]
 * salary-class: 0,000 [%]

Filter

	0	1	2	3	4
sex	✓	✓			
age	✓	✓	✓	✗	✗
race	✓	✗			
marital-status	✓	✓	✗		
education	✓	✓	✓	✗	

☒ Anonymous ☐ Non-anonymous ☐ Unknown

Information loss:

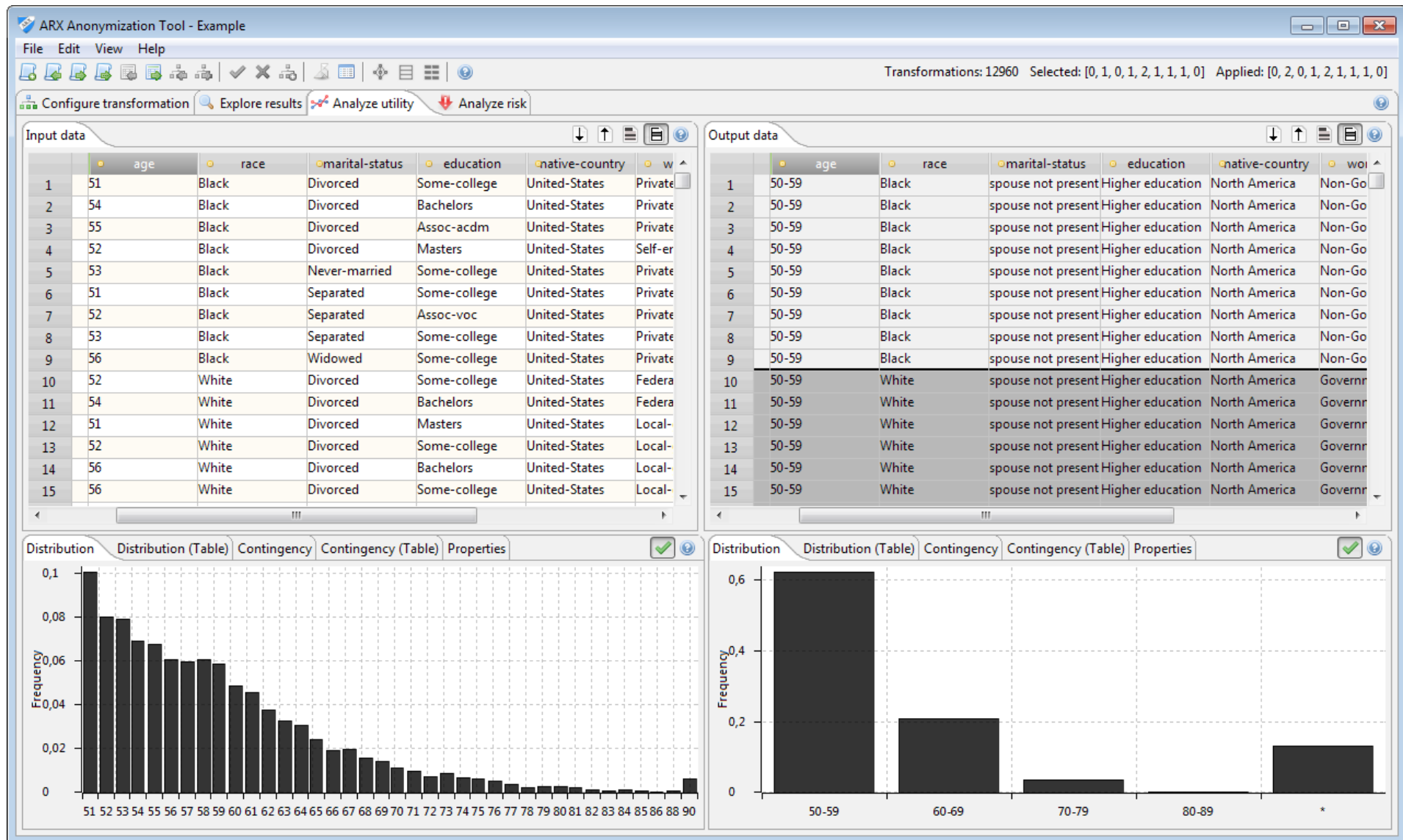
Clipboard

Node	Comment
[0, 2, 0, 1, 2, 1, 1, 0]	Minimal information loss
[0, 1, 0, 1, 2, 1, 1, 0]	Age is less generalized

Properties

Property	Value
Anonymous	ANONYMOUS
Min. info. loss	0.3315901556540435 [6,767%]
Max. info. loss	0.3315901556540435 [6,767%]
Successors	9
Predecessors	6
Transformation	[0, 1, 0, 1, 2, 1, 1, 0]
Checked	true

ARX: Utility analysis (1)



ARX: Utility analysis (2)

ARX Anonymization Tool - Example

File Edit View Help

Transformations: 12960 Selected: [0, 1, 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

	age	race	marital-status	education	native-country	work-class
1	51	Black	Divorced	Some-college	United-States	Private
2	54	Black	Divorced	Bachelors	United-States	Private
3	55	Black	Divorced	Assoc-acdm	United-States	Private
4	52	Black	Divorced	Masters	United-States	Self-er
5	53	Black	Never-married	Some-college	United-States	Private
6	51	Black	Separated	Some-college	United-States	Private
7	52	Black	Separated	Assoc-voc	United-States	Private
8	53	Black	Separated	Some-college	United-States	Private
9	56	Black	Widowed	Some-college	United-States	Private
10	52	White	Divorced	Some-college	United-States	Federa
11	54	White	Divorced	Bachelors	United-States	Federa
12	51	White	Divorced	Masters	United-States	Local-
13	52	White	Divorced	Some-college	United-States	Local-
14	56	White	Divorced	Bachelors	United-States	Local-
15	56	White	Divorced	Some-college	United-States	Local-

Output data

	age	race	marital-status	education	native-country	work-class
1	50-59	Black	spouse not present	Higher education	North America	Non-Go
2	50-59	Black	spouse not present	Higher education	North America	Non-Go
3	50-59	Black	spouse not present	Higher education	North America	Non-Go
4	50-59	Black	spouse not present	Higher education	North America	Non-Go
5	50-59	Black	spouse not present	Higher education	North America	Non-Go
6	50-59	Black	spouse not present	Higher education	North America	Non-Go
7	50-59	Black	spouse not present	Higher education	North America	Non-Go
8	50-59	Black	spouse not present	Higher education	North America	Non-Go
9	50-59	Black	spouse not present	Higher education	North America	Non-Go
10	50-59	White	spouse not present	Higher education	North America	Governr
11	50-59	White	spouse not present	Higher education	North America	Governr
12	50-59	White	spouse not present	Higher education	North America	Governr
13	50-59	White	spouse not present	Higher education	North America	Governr
14	50-59	White	spouse not present	Higher education	North America	Governr
15	50-59	White	spouse not present	Higher education	North America	Governr

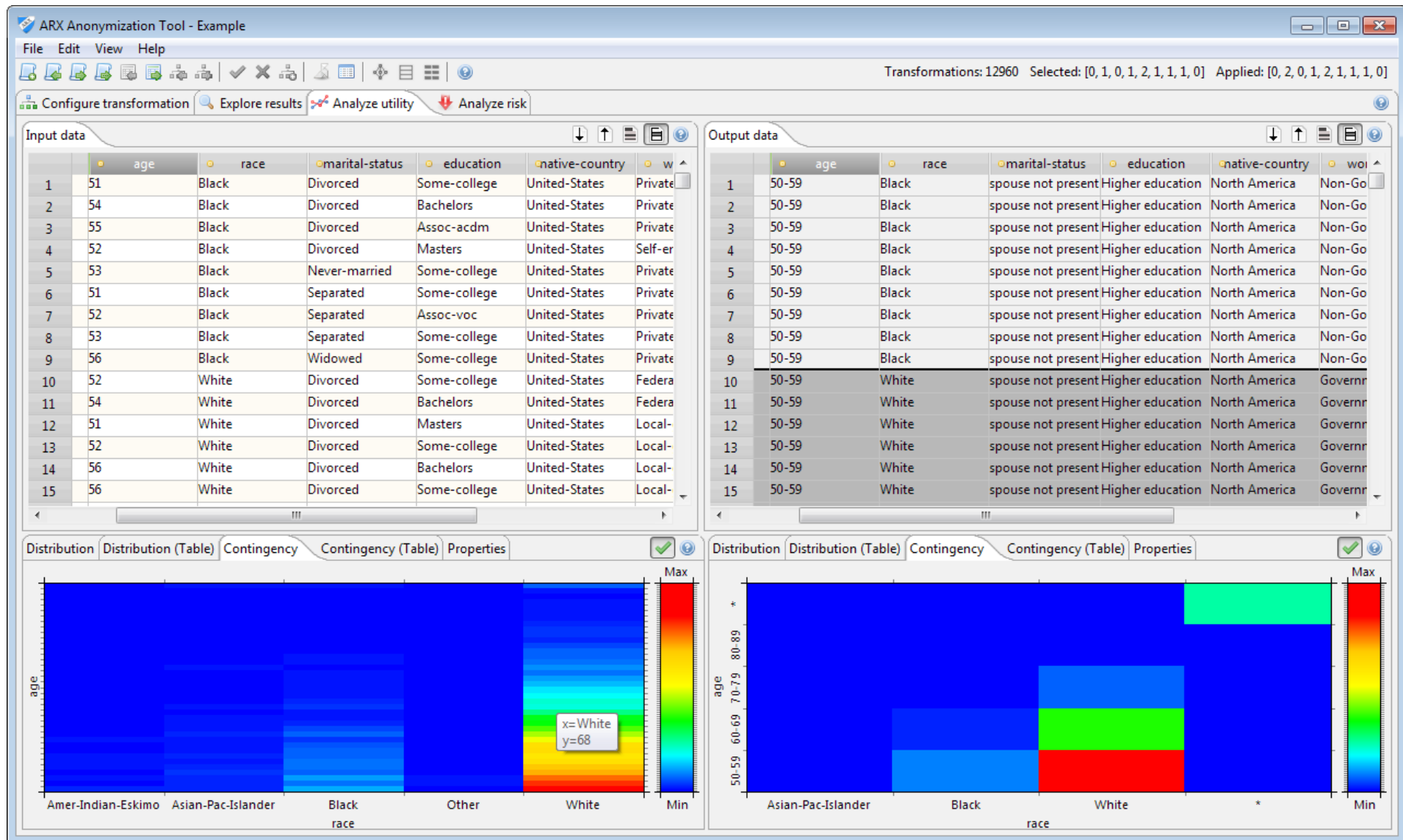
Distribution

Value	Frequency
51	10,01056%
52	7,98733%
53	7,86418%
54	6,89655%
55	6,75581%
56	6,03448%
57	5,92892%
58	6,05208%
59	5,84096%
60	4,85574%
61	4,55555%

Distribution

Value	Frequency
50-59	62,12175%
60-69	20,93596%
70-79	3,62421%
80-89	0,28149%
*	13,03659%

ARX: Utility analysis (3)



ARX: Utility analysis (4)

ARX Anonymization Tool - Example

File Edit View Help

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

Configure transformation Explore results Analyze utility Analyze risk

Input data

	age	race	marital-status	education	native-country	work-class
1	51	Black	Divorced	Some-college	United-States	Private
2	54	Black	Divorced	Bachelors	United-States	Private
3	55	Black	Divorced	Assoc-acdm	United-States	Private
4	52	Black	Divorced	Masters	United-States	Self-er
5	53	Black	Never-married	Some-college	United-States	Private
6	51	Black	Separated	Some-college	United-States	Private
7	52	Black	Separated	Assoc-voc	United-States	Private
8	53	Black	Separated	Some-college	United-States	Private
9	56	Black	Widowed	Some-college	United-States	Private
10	52	White	Divorced	Some-college	United-States	Federa
11	54	White	Divorced	Bachelors	United-States	Federa
12	51	White	Divorced	Masters	United-States	Local-
13	52	White	Divorced	Some-college	United-States	Local-
14	56	White	Divorced	Bachelors	United-States	Local-
15	56	White	Divorced	Some-college	United-States	Local-

Output data

	age	race	marital-status	education	native-country	work-class
1	50-59	Black	spouse not present	Higher education	North America	Non-Go
2	50-59	Black	spouse not present	Higher education	North America	Non-Go
3	50-59	Black	spouse not present	Higher education	North America	Non-Go
4	50-59	Black	spouse not present	Higher education	North America	Non-Go
5	50-59	Black	spouse not present	Higher education	North America	Non-Go
6	50-59	Black	spouse not present	Higher education	North America	Non-Go
7	50-59	Black	spouse not present	Higher education	North America	Non-Go
8	50-59	Black	spouse not present	Higher education	North America	Non-Go
9	50-59	Black	spouse not present	Higher education	North America	Non-Go
10	50-59	White	spouse not present	Higher education	North America	Governr
11	50-59	White	spouse not present	Higher education	North America	Governr
12	50-59	White	spouse not present	Higher education	North America	Governr
13	50-59	White	spouse not present	Higher education	North America	Governr
14	50-59	White	spouse not present	Higher education	North America	Governr
15	50-59	White	spouse not present	Higher education	North America	Governr

Distribution

	Amer-Indian-Eskimo	Asian-Pac-Islander	Black	Other	White
51	0,17593%	0,14075%	0,96763%	0,01759%	8,70866%
52	0,03519%	0,19353%	0,61576%	0,08797%	7,05489%
53	0,08797%	0,21112%	0,91485%	0,05278%	6,59747%
54	0,00000%	0,22871%	0,56298%	0,00000%	6,10486%
55	0,05278%	0,19353%	0,58058%	0,01759%	5,91133%
56	0,05278%	0,12315%	0,58058%	0,01759%	5,26038%
57	0,05278%	0,19353%	0,56298%	0,03519%	5,08445%
58	0,03519%	0,10556%	0,52780%	0,03519%	5,34835%
59	0,03519%	0,12315%	0,40464%	0,00000%	5,27797%
60	0,08797%	0,12315%	0,29909%	0,03519%	4,31034%

Contingency

	Amer-Indian-Eskimo	Asian-Pac-Islander	Black	Other	White
51	0,17593%	0,14075%	0,96763%	0,01759%	8,70866%
52	0,03519%	0,19353%	0,61576%	0,08797%	7,05489%
53	0,08797%	0,21112%	0,91485%	0,05278%	6,59747%
54	0,00000%	0,22871%	0,56298%	0,00000%	6,10486%
55	0,05278%	0,19353%	0,58058%	0,01759%	5,91133%
56	0,05278%	0,12315%	0,58058%	0,01759%	5,26038%
57	0,05278%	0,19353%	0,56298%	0,03519%	5,08445%
58	0,03519%	0,10556%	0,52780%	0,03519%	5,34835%
59	0,03519%	0,12315%	0,40464%	0,00000%	5,27797%
60	0,08797%	0,12315%	0,29909%	0,03519%	4,31034%

Properties

	Amer-Indian-Eskimo	Asian-Pac-Islander	Black	Other	White
51	0,17593%	0,14075%	0,96763%	0,01759%	8,70866%
52	0,03519%	0,19353%	0,61576%	0,08797%	7,05489%
53	0,08797%	0,21112%	0,91485%	0,05278%	6,59747%
54	0,00000%	0,22871%	0,56298%	0,00000%	6,10486%
55	0,05278%	0,19353%	0,58058%	0,01759%	5,91133%
56	0,05278%	0,12315%	0,58058%	0,01759%	5,26038%
57	0,05278%	0,19353%	0,56298%	0,03519%	5,08445%
58	0,03519%	0,10556%	0,52780%	0,03519%	5,34835%
59	0,03519%	0,12315%	0,40464%	0,00000%	5,27797%
60	0,08797%	0,12315%	0,29909%	0,03519%	4,31034%

Distribution

	Asian-Pac-Islander	Black	White	*
50-59	0,22871%	4,73258%	57,16045%	0,00000%
60-69	0,00000%	1,00281%	19,93315%	0,00000%
70-79	0,00000%	0,08797%	3,53624%	0,00000%
80-89	0,00000%	0,00000%	0,28149%	0,00000%
*	0,00000%	0,00000%	0,00000%	13,03659%

ARX: Utility analysis (5)

ARX Anonymization Tool - Example

File Edit View Help

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

Configure transformation Explore results Analyze utility Analyze risk

Input data

	age	race	marital-status	education	native-country	work-class
1	51	Black	Divorced	Some-college	United-States	Private
2	54	Black	Divorced	Bachelors	United-States	Private
3	55	Black	Divorced	Assoc-acdm	United-States	Private
4	52	Black	Divorced	Masters	United-States	Self-er
5	53	Black	Never-married	Some-college	United-States	Private
6	51	Black	Separated	Some-college	United-States	Private
7	52	Black	Separated	Assoc-voc	United-States	Private
8	53	Black	Separated	Some-college	United-States	Private
9	56	Black	Widowed	Some-college	United-States	Private
10	52	White	Divorced	Some-college	United-States	Federa
11	54	White	Divorced	Bachelors	United-States	Federa
12	51	White	Divorced	Masters	United-States	Local-
13	52	White	Divorced	Some-college	United-States	Local-
14	56	White	Divorced	Bachelors	United-States	Local-
15	56	White	Divorced	Some-college	United-States	Local-

Output data

	age	race	marital-status	education	native-country	work-class
1	50-59	Black	spouse not present	Higher education	North America	Non-Go
2	50-59	Black	spouse not present	Higher education	North America	Non-Go
3	50-59	Black	spouse not present	Higher education	North America	Non-Go
4	50-59	Black	spouse not present	Higher education	North America	Non-Go
5	50-59	Black	spouse not present	Higher education	North America	Non-Go
6	50-59	Black	spouse not present	Higher education	North America	Non-Go
7	50-59	Black	spouse not present	Higher education	North America	Non-Go
8	50-59	Black	spouse not present	Higher education	North America	Non-Go
9	50-59	Black	spouse not present	Higher education	North America	Non-Go
10	50-59	White	spouse not present	Higher education	North America	Governr
11	50-59	White	spouse not present	Higher education	North America	Governr
12	50-59	White	spouse not present	Higher education	North America	Governr
13	50-59	White	spouse not present	Higher education	North America	Governr
14	50-59	White	spouse not present	Higher education	North America	Governr
15	50-59	White	spouse not present	Higher education	North America	Governr

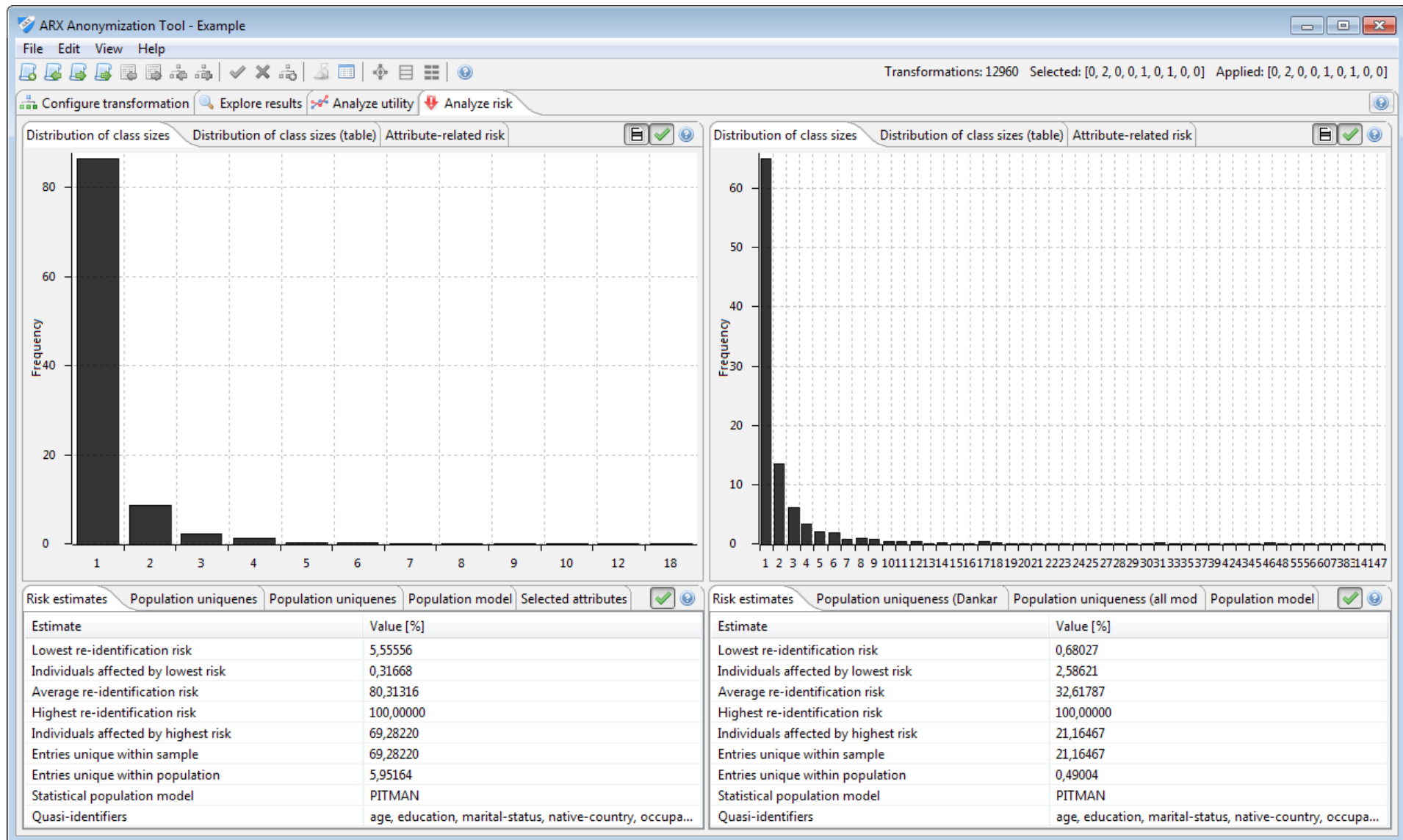
Distribution Distribution (Table) Contingency Contingency (Table) Properties

Property	Value	Datatype	Format	Height	Min	Max	V
Tuples	30162						
Max. Outliers	100.0 [%]						
Metric	Loss (0.5/1.0/1.0)						
Attributes	9						
Identifying	0						
Quasi-Identifying	9						
QI-0	sex	String		2	0	1	0
QI-1	age	Integer		5	0	4	0
QI-2	race	String		2	0	1	0
QI-3	marital-status	String		3	0	2	0
QI-4	education	String		4	0	3	0

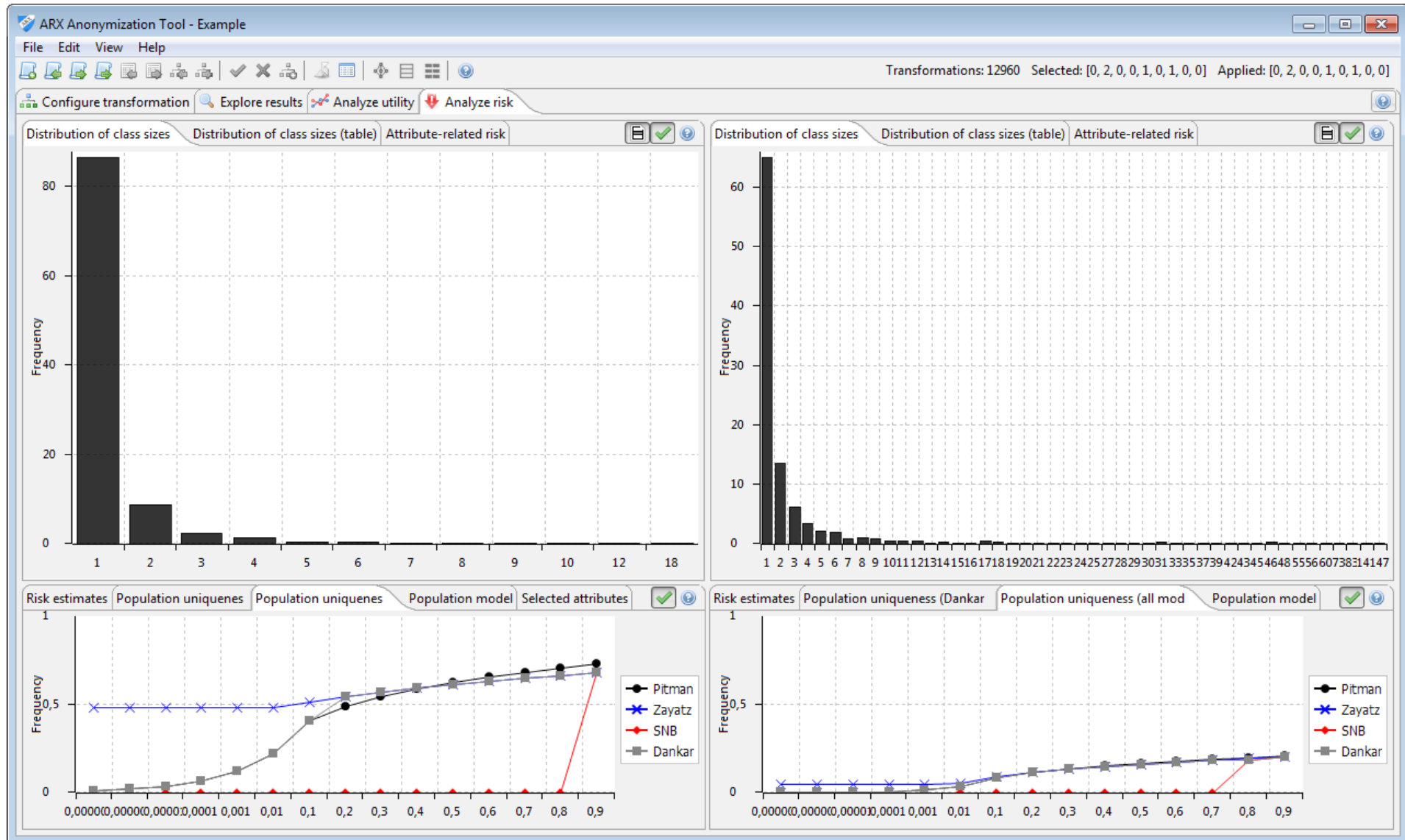
Distribution Distribution (Table) Contingency Contingency (Table) Properties

Property	Value
Outliers	741
Equivalence classes	654
Outlying classes	481
Min. class size	5 (5)
Max. class size	311 (741)
Avg. class size	28.57225433526...
Information loss	0.331590155654...
Successors	9
Predecessors	6
Transformation	[0, 1, 0, 1, 2, 1, 1, ...]
Anonymity	k-Anonymity
k	5

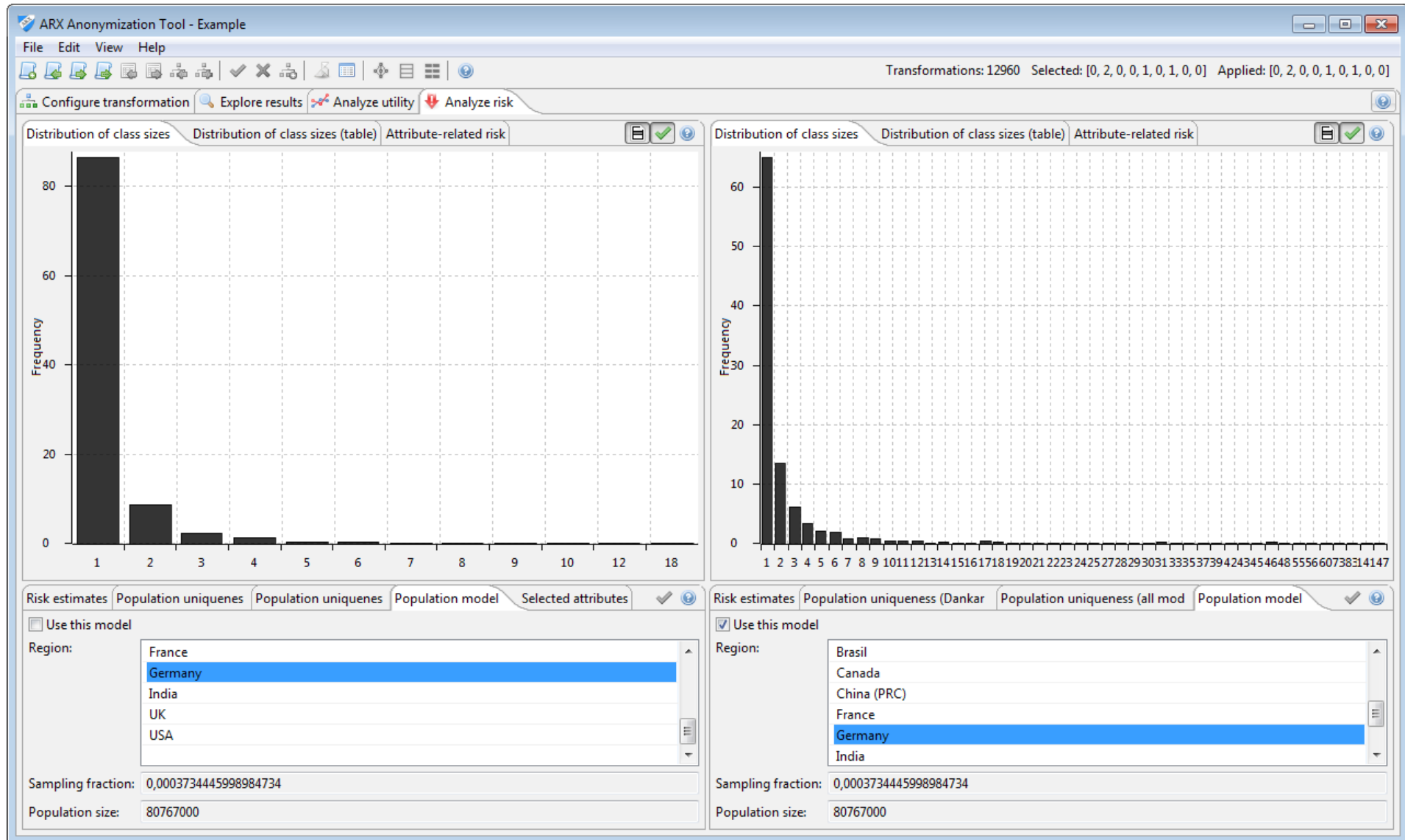
ARX: Risk analysis (1)



ARX: Risk analysis (2)



ARX: Risk analysis (3)



ARX: Risk analysis (4)

ARX Anonymization Tool - Example

File Edit View Help

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

Configure transformation Explore results Analyze utility Analyze risk

Distribution of class sizes Distribution of class sizes (table) Attribute-related risk

Quasi-identifier	Score [...]	Highest...	Average...
workclass	0,591	98,730	27,471
salary-class	0,733	97,432	28,014
sex	0,813	97,537	28,920
native-country	0,965	100,000	29,628
education	1,016	98,867	35,483
age	1,337	100,000	44,717
occupation	1,419	99,614	43,965
marital-status	1,433	99,262	40,100
race	2,364	99,625	52,870
occupation, salary-class	0,591	99,258	33,808
age, sex	0,871	100,000	39,209
native-country, race	0,887	100,000	28,497
marital-status, occupation	0,908	100,000	37,280
marital-status, native-country	0,996	100,000	31,573
marital-status, workclass	1,011	100,000	35,128
native-country, workclass	1,025	100,000	32,461
occupation, workclass	1,031	100,000	40,034
occupation, sex	1,209	100,000	40,953
age, education	1,306	100,000	50,255

Risk estimates Population uniqueness Population uniqueness (Dankar) Population uniqueness (all mod) Population model

☒ Use this model

Region:
 Brasil
 Canada
 China (PRC)
 France
 Germany
 India

Sampling fraction: 0,0003734445998984734

Population size: 80767000

Resulting quasi-identifiers: 511

ARX: Context-sensitive help

ARX Help Center

Please select a topic

1. Perspectives

1.1. Viewing and Manipulating Data

2. Defining the transformation

2.1. Defining attribute properties

2.2. Creating generalization hierarchies

2.3. Defining privacy criteria

2.4. Defining general properties

2.5. Defining a research subset

3. Exploring the solution space

3.1. Exploring the lattice

3.2. Filtering the lattice

3.3. Using the clipboard

3.4. Properties of transformations

4. Analyzing transformed datasets

4.1. Visualizations and properties

Back

Forward

I-Diversity

Selecting this tab allows enforcing I-diversity for the currently selected sensitive attribute. To specify the details of the criterion, a slider can be used to define the diversity parameter I . A dropdown list allows selecting the type of I-diversity, i.e., distinct I-diversity, entropy I-diversity or recursive-(c,I)-diversity. If recursive (c,I)-diversity is selected, an additional slider can be used to select an appropriate value for the parameter c :

k-Anonymity

d-Presence

I-Diversity

t-Closeness

L: 3

Variant: Recursive-(c,I)-Diversity

C: 4.0

t-Closeness

Selecting this tab allows enforcing t-closeness for the currently selected sensitive attribute. Again, a dropdown list allows selecting the function that should be used for computing the closeness of data distributions. If "Hierarchical Earth Mover's Distance" is selected, a generalization hierarchy must be specified for the respective sensitive attribute. A slider enables selecting an appropriate value for the closeness parameter t .

k-Anonymity

d-Presence

I-Diversity

t-Closeness

Measure: Hierarchical Earth Mover's Distance

T: 0.2

d-Presence

Analogously to k-anonymity, d-presence is enforced for all quasi-identifiers. Two sliders can be used for selecting appropriate values for the d-min and d-max parameters:

OK

ARX: Further developments

- **Current projects**

- Non-interactive Differential Privacy
- Support for high-dimensional data: heuristic algorithms
- Further analyses and visualizations: utility and risks
- Support for transactional attributes: (k, k^m) -anonymity
- Integrated data masking methods
- More flexible definition of quasi-identifiers
- More risk models

- **Planned projects**

- Auto-detection of HIPAA identifiers
- Implement more flexible privacy criteria

Thank you for your attention

- **ARX is open source software**
- **Contribute:** feature requests, code reviews, criticism, enhancements, questions
- **Repository:** <https://github.com/arx-deidentifier/arx>
- **Further information & download:** <http://arx.deidentifier.org>
- **Get in touch**
 - Fabian Prasser (prasser@in.tum.de)
 - Florian Kohlmayer (florian.kohlmayer@tum.de)
- **Any questions?**