

TIAMAT

a Tool for Interactive Analysis of Microdata Anonymization Techniques

Chenyun Dai¹, Gabriel Ghinita¹, Elisa Bertino¹, Ji-Won Byun², Ninghui Li¹

¹Purdue University

²Oracle Corp

A. Background

- Microdata publishing may compromise privacy
 - Data contain quasi-identifying (QID) attributes
- k*-anonymity: privacy paradigm that prevents re-identification
 - Indistinguishable QID among groups of *k* records
 - Values are generalized to ranges
 - Value-generalization hierarchies (VGH)

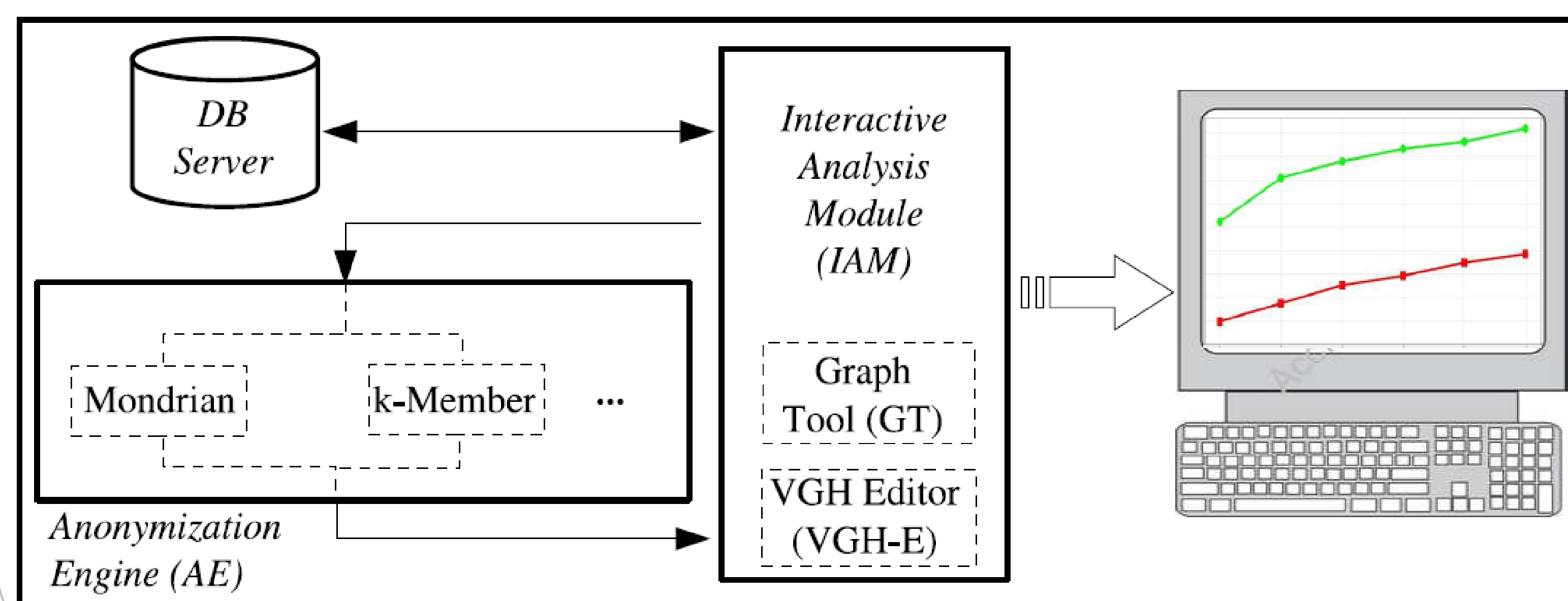
Age	ZipCode	Disease
42	52000	Ulcer
47	43000	Pneumonia
51	32000	Flu
55	27000	Gastritis
62	41000	Dyspepsia
67	55000	Dyspepsia

Age	ZipCode	Disease
42-47	43000-52000	Ulcer
42-47	43000-52000	Pneumonia
51-55	27000-32000	Flu
51-55	27000-32000	Gastritis
62-67	41000-55000	Dyspepsia
62-67	41000-55000	Dyspepsia

- Privacy-accuracy trade-off challenges:
 - Selecting a suitable *k*-anonymization algorithm
 - Choosing which QID attributes to release
 - Determining appropriate values of *k*

B. System Architecture

- IAM handles user input and visualization of results
 - Value Generalization Hierarchy Editor
 - Graph Tool (*JFreeChart*)
 - Import/Export data from/to DB Server
- AE performs QID generalization
 - Currently, supports *Mondrian* and *k-Member*
 - Can incorporate any generalization-based method



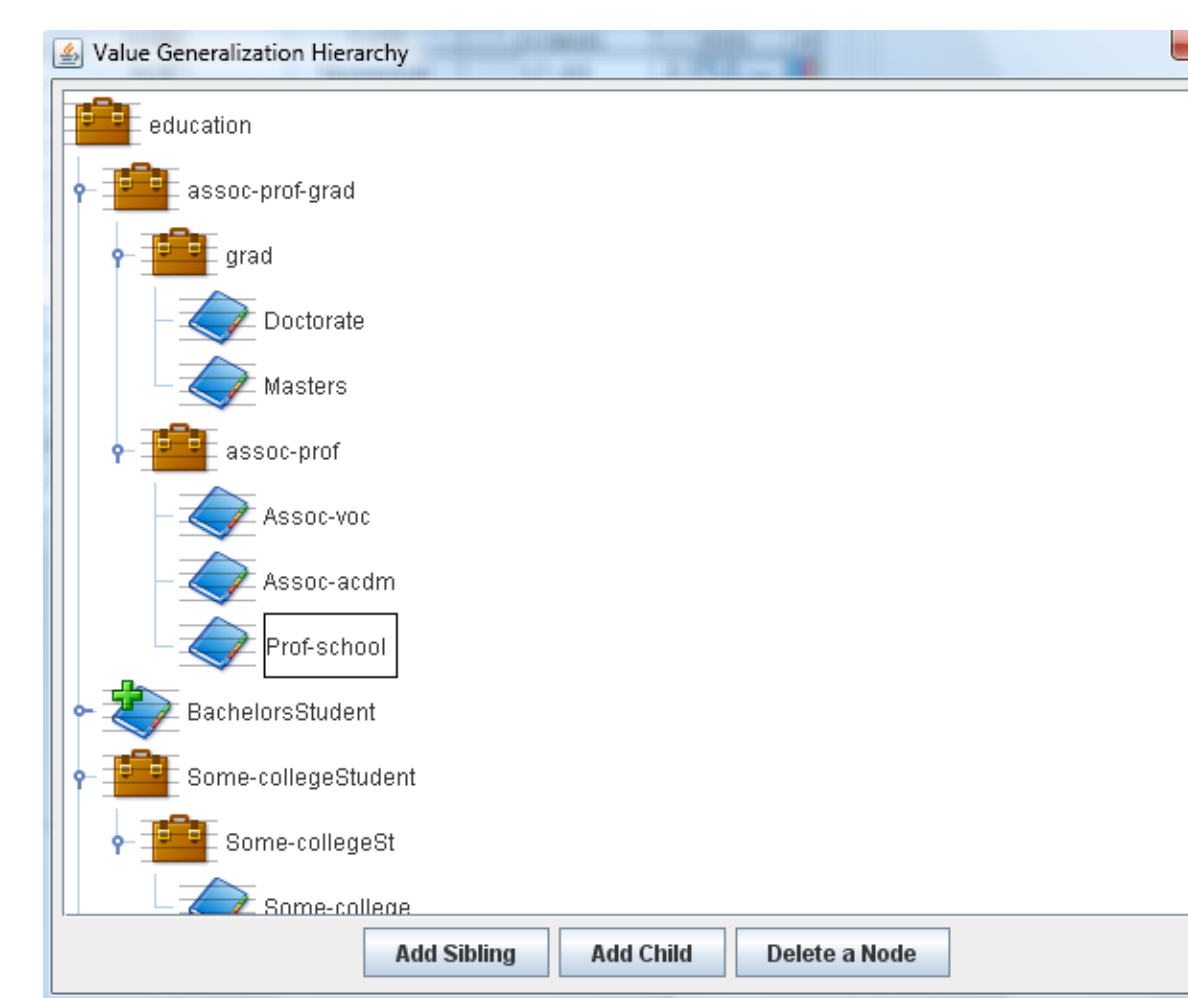
C. Dataset Preparation and Statistics

- Quasi-Identifier Selection

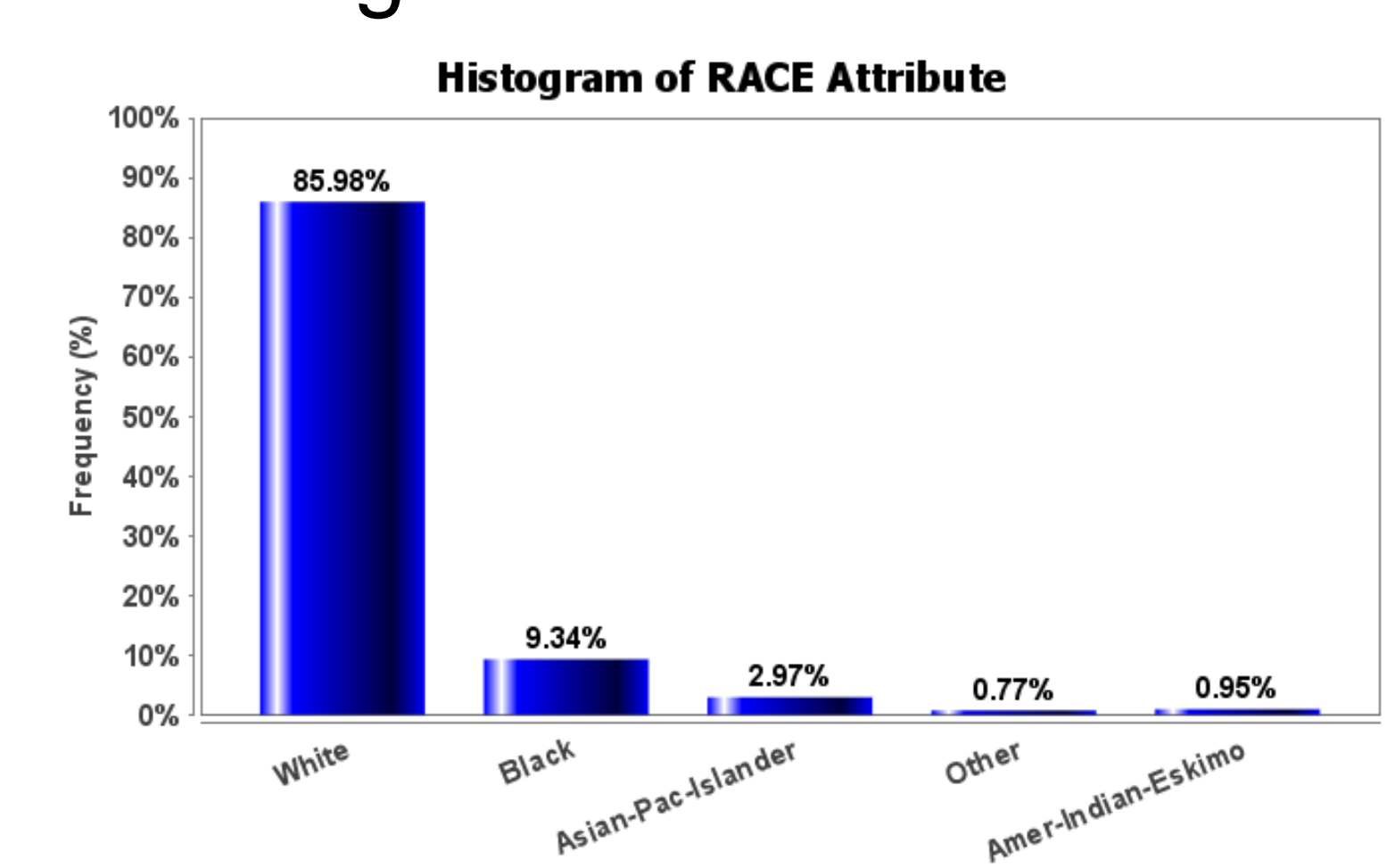
NAME	TYPE	DOMAIN	VGH
AGE	Numerical	17-90	0
WORKCLASS	Categorical	N/A	3
FNLWGT	Numerical	13769-1484705	0
EDUCATION	Categorical	N/A	4
EDUCATION_NUM	Numerical	1-16	0
MARITAL_STATUS	Categorical	N/A	3
OCCUPATION	Categorical	N/A	2
RELATIONSHIP	Categorical	N/A	0
RACE	Categorical	N/A	3
SEX	Categorical	N/A	2
CAPITAL_GAIN	Numerical	0-99999	0
CAPITAL_LOSS	Numerical	0-4356	0
HOURS_PER_WEEK	Numerical	1-99	0
NATIVE_COUNTRY	Categorical	N/A	3
INCOME	Categorical	N/A	0

NAME	TYPE	VGH Levels
AGE	Numerical	0
OCCUPATION	Categorical	2
RACE	Categorical	3

- VGH Editor Tool

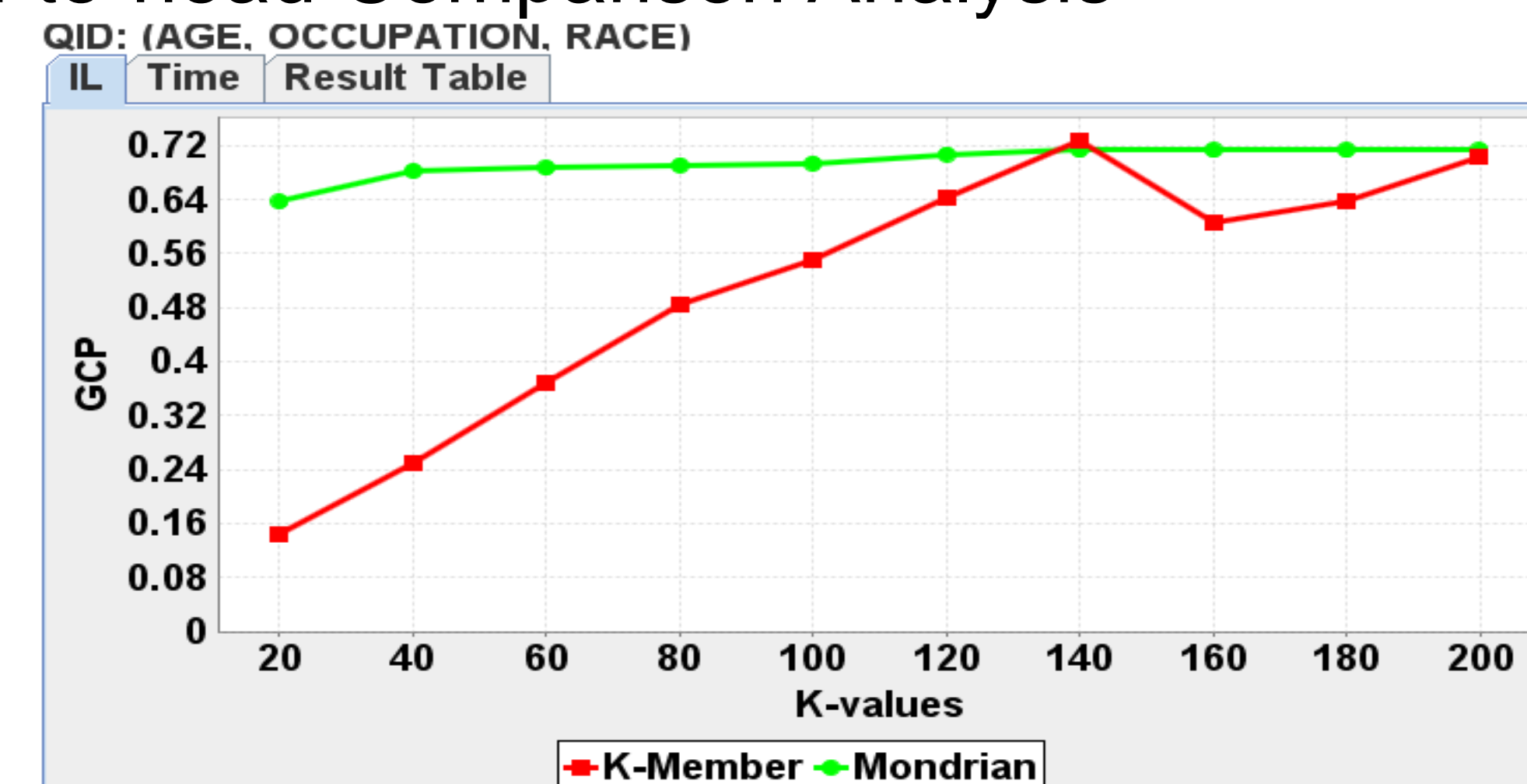


- Histogram Tool



D. Analysis of *k*-anonymization

- Head-to-head Comparison Analysis



- QID Change-Impact Analysis



- Inspecting the original and anonymized data

AGE	WORKCLASS	FNLWGT	EDUCATION	MARITAL_STATUS	OCCUPATION	RELATIONSHIP
49.0	Private	81973.0	Some-college	Married-cw-spouse	Craft-repair	Husband
61.0	Self-emp-inc	66814.0	HS-grad	Married-cw-spouse	Craft-repair	Husband
27.0	Private	232792.0	Some-college	Never-married	Sales	Own-child
19.0	Private	316888.0	Some-college	Never-married	Other-service	Own-child
45.0	Private	196594.0	Assoc-voc	Never-married	Prof-specialty	Not-in-family
20.0	Private	105376.0	Some-college	Never-married	Tech-support	Other-relative
31.0	Private	105814.0	HS-grad	Never-married	Transport-moving	Unmarried
22.0	Private	175374.0	Some-college	Married-cw-spouse	Other-service	Husband
16.0	Private	1100293.0	HS-grad	Widowed	Other-service	Unmarried
84.0	Private	181232.0	11th	Married-cw-spouse	Craft-repair	Husband
47.0	Local-gov	186009.0	Some-college	Divorced	Adm-clerical	Unmarried
34.0	Private	198163.0	HS-grad	Never-married	Adm-clerical	Not-in-family

AGE	WORKCLASS	FNLWGT	EDUCATION	MARITAL_STATUS	OCCUPATION	RELATIONSHIP
25.0-36.0	Private	211265.0	Some-college	other	Occupation	Other-relative
25.0-36.0	Private	189775.0	Some-college	other	Occupation	Own-child
25.0-36.0	Private	212563.0	Some-college	other	Occupation	Unmarried
25.0-36.0	Private	170282.0	Some-college	other	Occupation	Unmarried
28.0-47.0	Private	88419.0	education	Never-married	Exec-managerial	Not-in-family
28.0-47.0	Private	163003.0	education	Never-married	Exec-managerial	Not-in-family
28.0-47.0	Private	348491.0	education	Never-married	Exec-managerial	Not-in-family
28.0-47.0	Private	200734.0	education	Never-married	Exec-managerial	Unmarried
29.0-43.0	Private	419721.0	HS-grad	Never-married	Other-service	Unmarried
29.0-43.0	Private	206365.0	HS-grad	Never-married	Other-service	Not-in-family
29.0-43.0	Private	39581.0	HS-grad	Never-married	Other-service	Not-in-family
29.0-43.0	Private	170240.0	HS-grad	Never-married	Other-service	Own-child