# Discrimination Rate: An Attribute-Centric Metric to Measure Privacy

WODIAC
PETS 2017

Louis Philippe SONDECK

07-17-2017


# Summary

- Existing Metrics and Limitations
- The Discrimination Rate Metric (DR)
- An Attack Driven Privacy Assessment
- Results on Real Data

# Existing Metrics and Limitations

# Existing Metrics

- There exists **a *large amount of metrics*** for privacy measurements

- **Some of the most popular**:
  - ***k-anonymity-like metrics*** (k-anonymity, l-diversity, t-closeness…)
  - ***Distortion Rate metrics*** (Mutual Information, KL-divergence, Mean Squared Error…)
  - ***Differential Privacy metrics*** (based on the epsilon parameter)

***k-anonymity (Samarati, 2001); l-diversity (Machanavajjhala et al., 2007); t-closeness (Li et al., 2007); Distortion Rate (Rebollo-Monedero et al., 2010); Differential Privacy (C. Dwork, 2008)***

# Limitations

- **Common limitations:**
  - No measurement with respect to attacks which seems to be the most pragmatic approach
  - Average measurements, leading to the *worst case problem*

- **Specific limitations**
  - It is difficult to relate the measurements to the identification capacity *(Differential Privacy)*
  - Do not provide measurement over more than 2 variables *(k-anonymity-like metrics)*

# The Discrimination Rate Metric

# The Discrimination Rate Metric

- **Computes** the identification capability of attributes from their capability to refine an anonymity set

- The results are **scaled** between 0 and 1

- There are **3 versions**:
  - Simple DR **(SDR)**: the capability of **1 attribute**
  - Combined DR **(CDR)**: the capability of **N attributes**
  - Semantic DR **(SeDR):** enables measurements according **to subsets of the anonymity set**

# Definition: Simple Discrimination Rate

**Definition 3.** *(Simple Discrimination Rate)*
*Let* $X$ *and* $Y$ *be two d.r.v. The* ***Simple Discrimination Rate*** *of the key attribute* $Y$ *relatively to sensitive attribute* $X$*, is the capacity of the key attribute* $Y$ *to refine the set of outcomes of the sensitive attribute* $X$ *and is computed as follows:*

$$DR_X(Y) = \frac{H(X) - H(X|Y)}{H(X)} = 1 - \frac{H(X|Y)}{H(X)} \qquad (1)$$

- **Input :**
  - Y : key attribute
  - X : a sensitive attribute
- **Output :** SDR of Y over the set of outcomes of attribute X
- Capacity of SDR to measure down to the attribute' values

# SDR Computation

Example Table

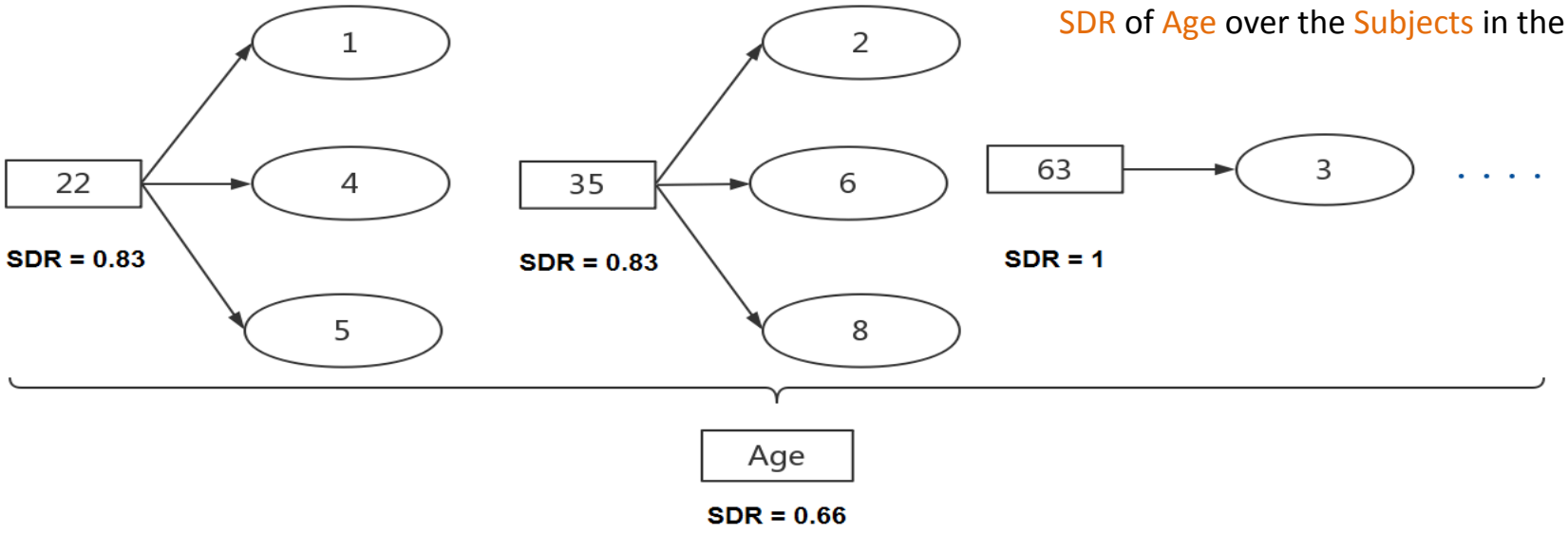
| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

# SDR Computation

Example Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

SDR of Age over the Subjects in the table



# SDR Computation

Example Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

• Attribute Age can take 5 values:
- 22 -> 3 subjects, 35 -> 3 subjects, 63 -> 1 subject, 45 -> 1 subject, 40 -> 1 subject

• The corresponding conditional entropies:
-H(X|Y = 22) = H(X|Y = 35) = -log(1/3)
-H(X|Y = 63) = H(X|Y = 45) = H(X|Y = 40) = 0

# SDR Computation

$$SDR_X(Y) = 1 - \frac{H(X|Y)}{H(X)}$$

$$= 1 - \frac{-1/3\log_2(1/3) - 1/3\log_2(1/3)}{-\sum_{s=1}^{9} 1/9\log_2(1/9)}$$

$$= 1 - \frac{1/3\log_2(3) + 1/3\log_2(3)}{\log_2(9)}$$

$$= 0.66$$

Example Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

- Attribute Age can take 5 values:

- 22 -> 3 subjects, 35 -> 3 subjects, 63 -> 1 subject, 45 -> 1 subject, 40 -> 1 subject

- The corresponding conditional entropies:

-H(X|Y = 22) = H(X|Y = 35) = -log(1/3)

-H(X|Y = 63) = H(X|Y = 45) = H(X|Y = 40) = 0

# Definition: Combined DR

**Definition 4.** *(Combined Discrimination Rate)*
*Let* $\boldsymbol{X}, Y_1, \ldots, Y_n$ *be d.r.v. The* ***Combined Discrimination Rate*** *of key attributes* $Y_1, Y_2, \ldots, Y_n$ *relatively to the sensitive attribute* $\boldsymbol{X}$*, is the capacity of the combination of key attributes* $Y_1, \ldots, Y_n$ *to refine the set of outcomes of the sensitive attribute* $\boldsymbol{X}$ *and is computed as follows:*

$$DR_X(Y_1, Y_2, \ldots, Y_n) = 1 - \frac{H(X|Y_1, Y_2, \ldots, Y_n)}{H(X)} \qquad (2)$$

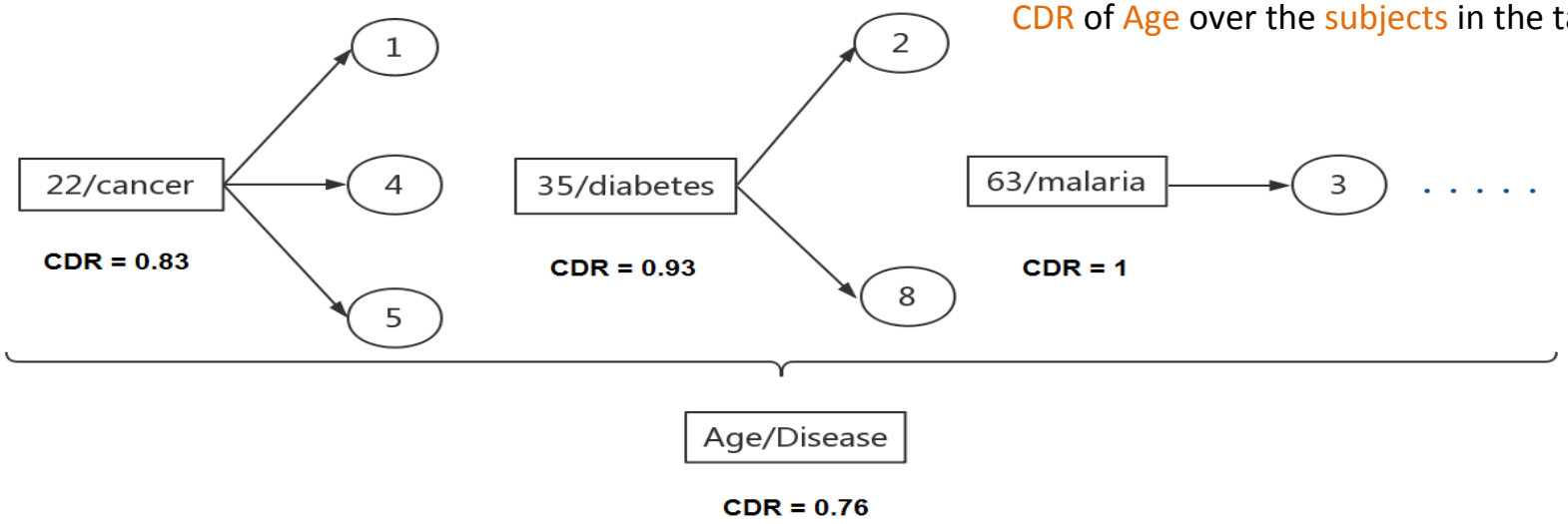
- **Input:**
  - Y1, …, Yn : N key attributes
  - X : a sensitive attribute
- **Output :** CDR of Y1, …, Yn over the set of outcomes of attribute X
- Capacity of CDR to measure down to the attribute values

# CDR Computation

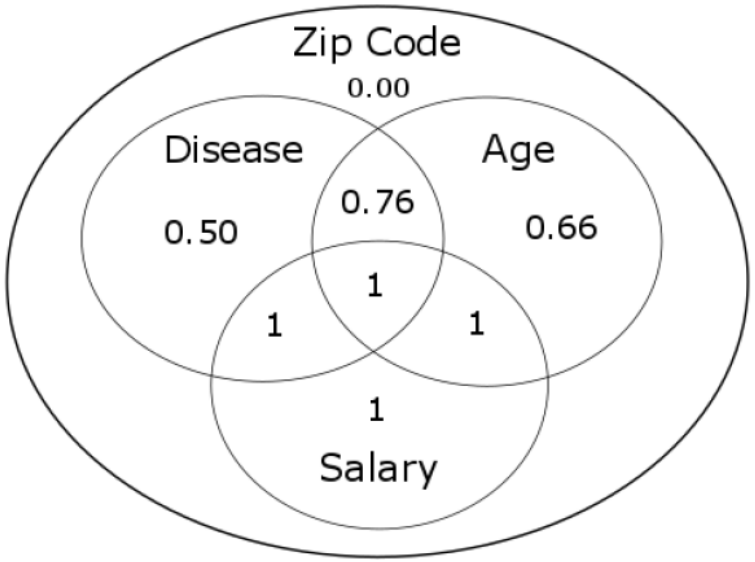
Example Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

CDR of Age over the subjects in the table



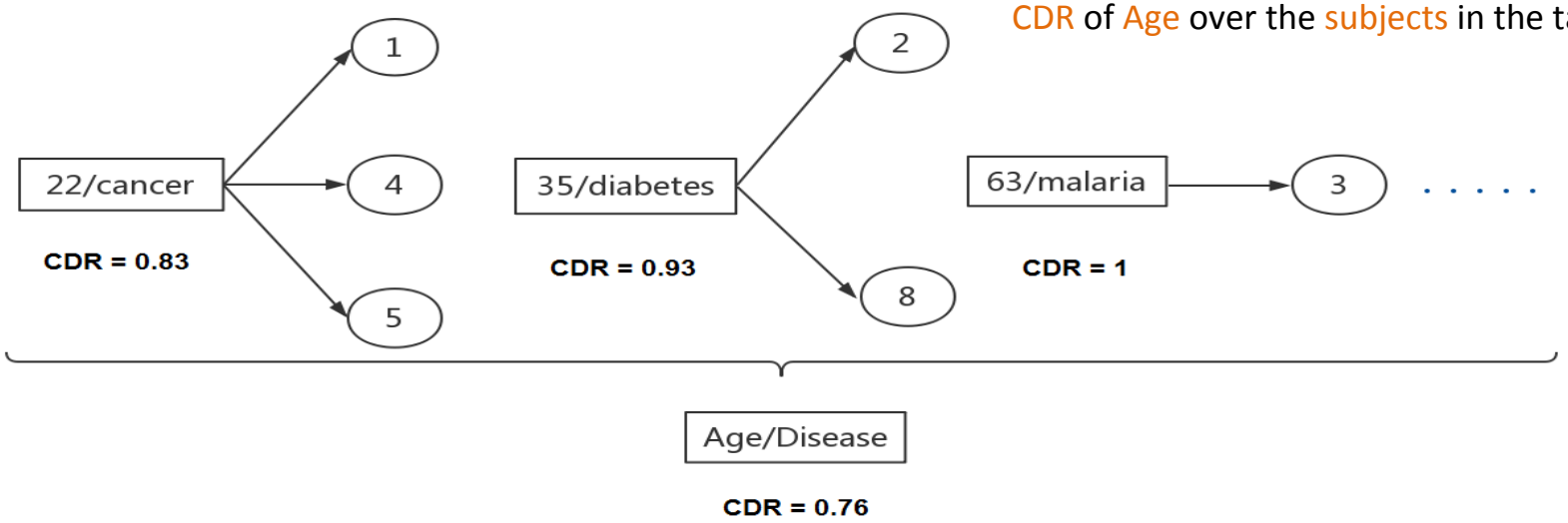
# CDR Computation




Example Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35000 | 22 | 4K | cancer |
| 2 | 35000 | 35 | 5K | diabetes |
| 3 | 35000 | 63 | 3K | malaria |
| 4 | 35000 | 22 | 13K | cancer |
| 5 | 35000 | 22 | 8K | cancer |
| 6 | 35000 | 35 | 15K | malaria |
| 7 | 35000 | 45 | 9K | malaria |
| 8 | 35000 | 35 | 7K | diabetes |
| 9 | 35000 | 40 | 11K | diabetes |

CDR of Age over the subjects in the table



# Comparison with existing metrics

| Metric | Granularity | Scope | Link with re-identification |
|---|---|---|---|
| Epsilon | - 2 attributes<br>- average | related to DP | weak |
| Mutual Information | - n attributes<br>- average | random variables in general | medium |
| K-anonymity | -1 attribute<br>- average | related to k-anonymity | medium |
| L-diversity | - 2 attributes<br>- average | related to k-anonymity | medium |
| T-closeness | - 2 attributes<br>- average | related to k-anonymity | medium |
| **DR** | - **n attributes**<br>- **fine** | **Random variables in general** | **High** |

# An Attack Driven Privacy Assessment

# k-anonymity model

- **Considers 3 types of attributes:**
  - Identifiers: attributes that can uniquely identify a subject (**e.g. security numbers**, **fingerprints**…)
  - Key attributes/ Quasi-identifiers: attributes that in combination can be used to identify a subject (**e.g. Age**, **Zip Code**, …)
  - Sensitive/Confidential Attributes: attributes that need to be protected (**e.g. health data**, **religion**, **salary**…)
- **k-anonymity** ensures that **each combination key attributes** is shared by **at least k subjects**

# Attacks Assessment

- DR enables assessment of **all the existing attacks** targeting the k-anonymity model**:**
  - Identity attack
  - homogeneity attack
  - background knowledge attack
  - skewness attack
  - semantic attack

- The **attacker's knowledge** is computed from the **identification capability of attributes** he owns

# Identity Attack (k-anonymity)

- Protects against: against Identity Attack
- Implements: generalization/suppression, aggregation…

(1) Originale Table

| | Age | Disease |
|---|---|---|
| 1 | 22 | lung cancer |
| 2 | 22 | lung cancer |
| 3 | 22 | lung cancer |
| 4 | 45 | stomach cancer |
| 5 | 63 | diabetes |
| 6 | 40 | aids |
| 7 | 35 | aids |
| 8 | 35 | flu |
| 9 | 32 | diabetes |

# Identity Attack (k-anonymity)

- Protects against: against Identity Attack
- Implements: generalization/suppression, aggregation…

(1) Originale Table

| | Age | Disease |
|---|---|---|
| 1 | 22 | lung cancer |
| 2 | 22 | lung cancer |
| 3 | 22 | lung cancer |
| 4 | 45 | stomach cancer |
| 5 | 63 | diabetes |
| 6 | 40 | aids |
| 7 | 35 | aids |
| 8 | 35 | flu |
| 9 | 32 | diabetes |

(2) Generalization Table

| | Age | Age* |
|---|---|---|
| 1 | 22 | 2* |
| 2 | 22 | 2* |
| 3 | 22 | 2* |
| 4 | 45 | $\geq 40$ |
| 5 | 63 | $\geq 40$ |
| 6 | 40 | $\geq 40$ |
| 7 | 35 | 3* |
| 8 | 35 | 3* |
| 9 | 32 | 3* |

# Identity Attack (k-anonymity)

- Protects against: against Identity Attack
- Implements: generalization/suppression, aggregation…

(1) Originale Table

| | Age | Disease |
|---|---|---|
| 1 | 22 | lung cancer |
| 2 | 22 | lung cancer |
| 3 | 22 | lung cancer |
| 4 | 45 | stomach cancer |
| 5 | 63 | diabetes |
| 6 | 40 | aids |
| 7 | 35 | aids |
| 8 | 35 | flu |
| 9 | 32 | diabetes |

(2) Generalization Table

| | Age | Age* |
|---|---|---|
| 1 | 22 | 2* |
| 2 | 22 | 2* |
| 3 | 22 | 2* |
| 4 | 45 | $\geq 40$ |
| 5 | 63 | $\geq 40$ |
| 6 | 40 | $\geq 40$ |
| 7 | 35 | 3* |
| 8 | 35 | 3* |
| 9 | 32 | 3* |

(3) 3-anonymity Table

| | Age* | Disease |
|---|---|---|
| 1 | 2* | lung cancer |
| 2 | 2* | lung cancer |
| 3 | 2* | lung cancer |
| 4 | $\geq 40$ | stomach cancer |
| 5 | $\geq 40$ | diabetes |
| 6 | $\geq 40$ | flu |
| 7 | 3* | aids |
| 8 | 3* | aids |
| 9 | 3* | diabetes |

# Approaches for Identity attack Assessment

- 3 approaches for identity attack measurement
  - -Black box: the attacker only has the **anonymized table**
  - -Grey box: the attacker has the **anonymized table plus external data**
  - -White box: the attacker has the generalized table

# Approaches for Identity attack Assessment

- 3 approaches for identity attack measurement
  - -Black box: the attacker only has the **anonymized table**
  - -Grey box: the attacker has the **anonymized table plus external data**
  - -White box: the attacker has the generalized table

(2) Generalization Table

| | Age | Age* |
|---|---|---|
| 1 | 22 | 2* |
| 2 | 22 | 2* |
| 3 | 22 | 2* |
| 4 | 45 | $\geq 40$ |
| 5 | 63 | $\geq 40$ |
| 6 | 40 | $\geq 40$ |
| 7 | 35 | 3* |
| 8 | 35 | 3* |
| 9 | 32 | 3* |

# Identity Attack Assessment

• **Identity Attack:** capacity of an attacker to **refine** the set of the **original key attribute values (Age)** from the **generalized key attribute (Age*)**

Identity attack measurement

| X | Y | $DR_X(Y)$ |
|---|---|---|
| Age | 2* | 1 |
| Age | $\geq 40$ | 0.78 |
| Age | 3* | 0.87 |
| Age | Age* | 0.66 |

# Homogeneity & Background Knowledge (l-diversity)

- Protects against: homogeneity and background knowledge attacks

- Implements: diversification of sensitive values

(2) 3-anonymity Table

| | Age* | Disease |
|---|---|---|
| 1 | 2* | lung cancer |
| 2 | 2* | lung cancer |
| 3 | 2* | lung cancer |
| 4 | $\geq 40$ | stomach cancer |
| 5 | $\geq 40$ | diabetes |
| 6 | $\geq 40$ | flu |
| 7 | 3* | aids |
| 8 | 3* | aids |
| 9 | 3* | diabetes |

# Homogeneity & Background Knowledge (l-diversity)

- Protects against: homogeneity and background knowledge attacks

- Implements: diversification of sensitive values

(2) 3-anonymity Table

| | Age* | Disease |
|---|---|---|
| 1 | 2* | lung cancer |
| 2 | 2* | lung cancer |
| 3 | 2* | lung cancer |
| 4 | $\geq 40$ | stomach cancer |
| 5 | $\geq 40$ | diabetes |
| 6 | $\geq 40$ | flu |
| 7 | 3* | aids |
| 8 | 3* | aids |
| 9 | 3* | diabetes |

# Homogeneity & Background Knowledge (l-diversity)

- Protects against: homogeneity and background knowledge attacks
- Implements: diversification of sensitive values

(1) Original Table

| | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 35567 | 22 | 4K | colon cancer |
| 2 | 35502 | 22 | 5K | stomach cancer |
| 3 | 35560 | 22 | 6K | lung cancer |
| 4 | 35817 | 45 | 7K | stomach cancer |
| 5 | 35810 | 63 | 12K | diabetes |
| 6 | 35812 | 40 | 9K | aids |
| 7 | 35502 | 35 | 8K | aids |
| 8 | 35568 | 35 | 10K | flu |
| 9 | 35505 | 32 | 11K | lung cancer |

(2) 3-anonymity Table

| | Age* | Disease |
|---|---|---|
| 1 | 2* | lung cancer |
| 2 | 2* | lung cancer |
| 3 | 2* | lung cancer |
| 4 | $\geq 40$ | stomach cancer |
| 5 | $\geq 40$ | diabetes |
| 6 | $\geq 40$ | flu |
| 7 | 3* | aids |
| 8 | 3* | aids |
| 9 | 3* | diabetes |

(3) 3-diversity Table

| | ZIP Code* | Age* | Salary | Disease |
|---|---|---|---|---|
| 1 | 355** | 2* | 4K | colon cancer |
| 2 | 355** | 2* | 5K | stomach cancer |
| 3 | 355** | 2* | 6K | lung cancer |
| 4 | 3581* | $\geq 40$ | 7K | stomach cancer |
| 5 | 3581* | $\geq 40$ | 12K | diabetes |
| 6 | 3581* | $\geq 40$ | 9K | aids |
| 7 | 355** | 3* | 8K | aids |
| 8 | 355** | 3* | 10K | flu |
| 9 | 355** | 3* | 11K | lung cancer |

# Homogeneity Attack Assessment

• **Homogeneity attack:** capacity of an attacker to **refine the set of the sensitive values (Disease)** from the **key attribute (Age*)**

| X | Y | $DR_X(Y)$ |
|---|---|---|
| k-anonymity Table | | |
| Disease | 2* | 1 |
| Disease | $\geq 40$ | 0.70 |
| Disease | 3* | 0.83 |
| Disease | Age* | 0.52 |
| l-diverse Table | | |
| Disease | 2* | 0.78 |
| Disease | $\geq 40$ | 0.78 |
| Disease | 3* | 0.78 |
| Disease | Age* | 0.36 |

# Background Knowledge Attack Assessment

• **Background knowledge:** computed from the **rest of information needed after** applying an **Homogeneity attack (1 - DR)**

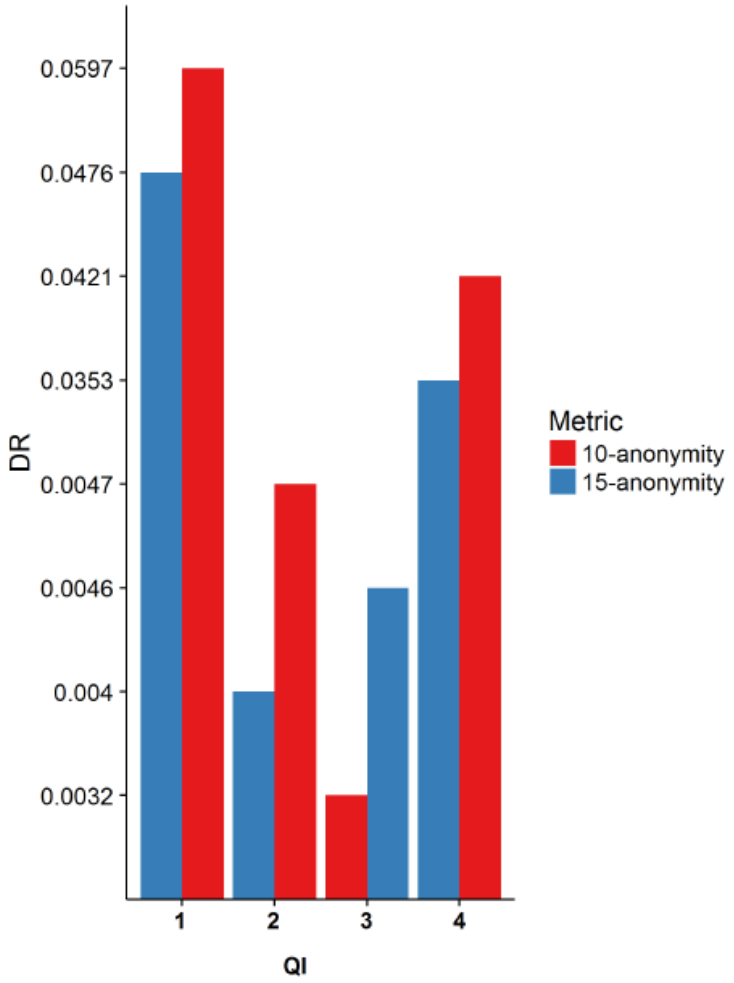
| X | Y | 1 - $DR_X(Y)$ |
|---|---|---|
| k-anonymity Table | | |
| Disease | 2* | 0 |
| Disease | $\geq 40$ | 0.30 |
| Disease | 3* | 0.17 |
| Disease | Age* | 0.48 |
| l-diverse Table | | |
| Disease | 2* | 0.22 |
| Disease | $\geq 40$ | 0.22 |
| Disease | 3* | 0.22 |
| Disease | Age* | 0.64 |

# Results on Real Data
# Adult Data Set ~30000 records

# Identity Attack Assessment (10-anon vs. 15-anon)

| | Attribute | Type | #Values |
|---|---|---|---|
| 1 | Marital Status | key attribute | 7 |
| 2 | Native Country | key attribute | 41 |
| 3 | Race | key attribute | 5 |
| 4 | Work Class | key attribute | 8 |

# Identity Attack Assessment (10-anon vs. 15-anon)

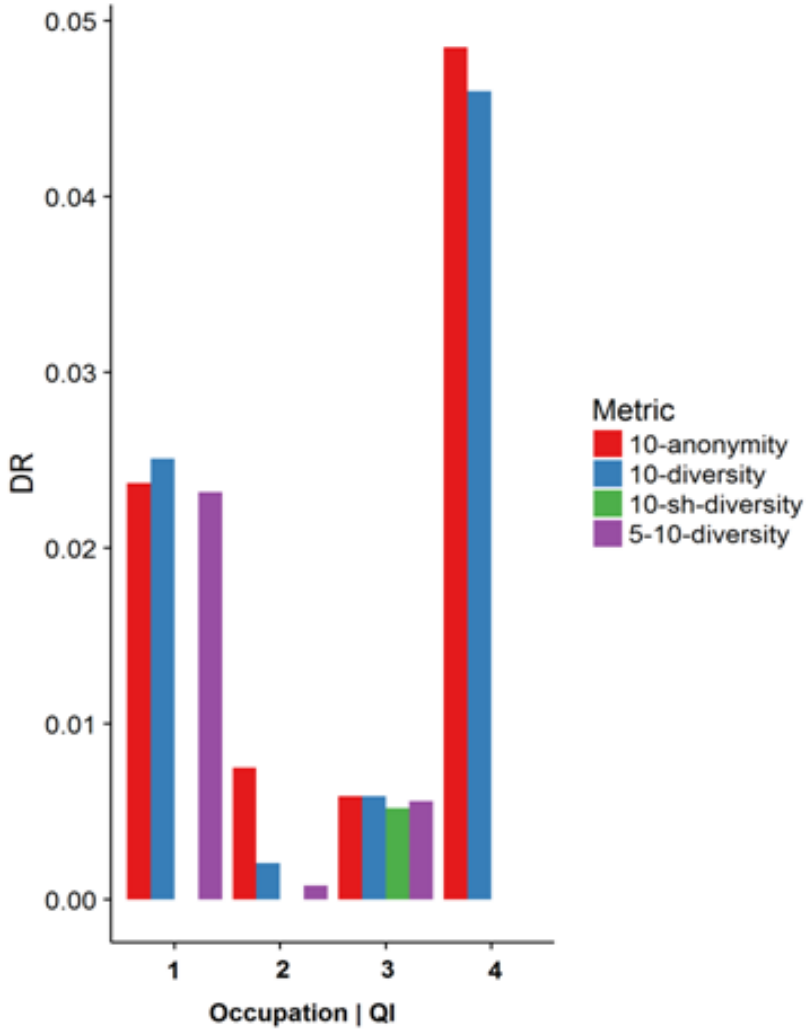



| | Attribute | Type | #Values |
|---|---|---|---|
| 1 | Marital Status | key attribute | 7 |
| 2 | Native Country | key attribute | 41 |
| 3 | Race | key attribute | 5 |
| 4 | Work Class | key attribute | 8 |

# Homogeneity Attack Assessment (k-anonymity vs. l-diversity)

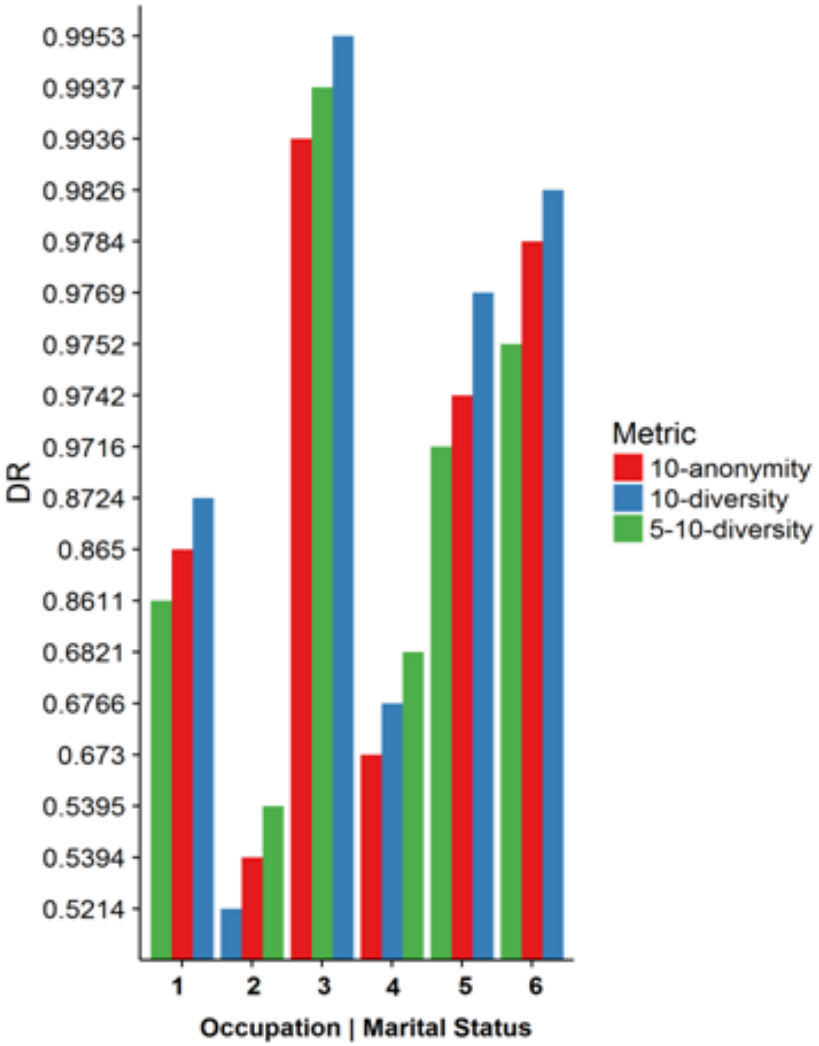
| | Attribute | Type | #Values |
|---|---|---|---|
| 1 | Marital Status | key attribute | 7 |
| 2 | Native Country | key attribute | 41 |
| 3 | Race | key attribute | 5 |
| 4 | Work Class | key attribute | 8 |
| 5 | Occupation | Sensitive | 14 |

# Homogeneity Attack Assessment (k-anonymity vs. l-diversity)




| | Attribute | Type | #Values |
|---|---|---|---|
| 1 | Marital Status | key attribute | 7 |
| 2 | Native Country | key attribute | 41 |
| 3 | Race | key attribute | 5 |
| 4 | Work Class | key attribute | 8 |
| 5 | Occupation | Sensitive | 14 |

# Homogeneity Attack Assessment (Marital Status' values)




| | Marital Status |
|---|---|
| 1 | Divorced |
| 2 | Married-civ-spouse |
| 3 | Married spouse absent |
| 4 | Never Married |
| 5 | Seperated |
| 6 | Widowed |

# Conclusion

- Anonymization refers to the trade-off between privacy and utility
- Metrics are therefore at the center of anonymization
- Discrimination Rate provides practical, flexible and accurate measurements for privacy assessment
- Discrimination Rate enables tackling assessment from the attacker's perspective which is the most pragmatic approach

# Q & A time