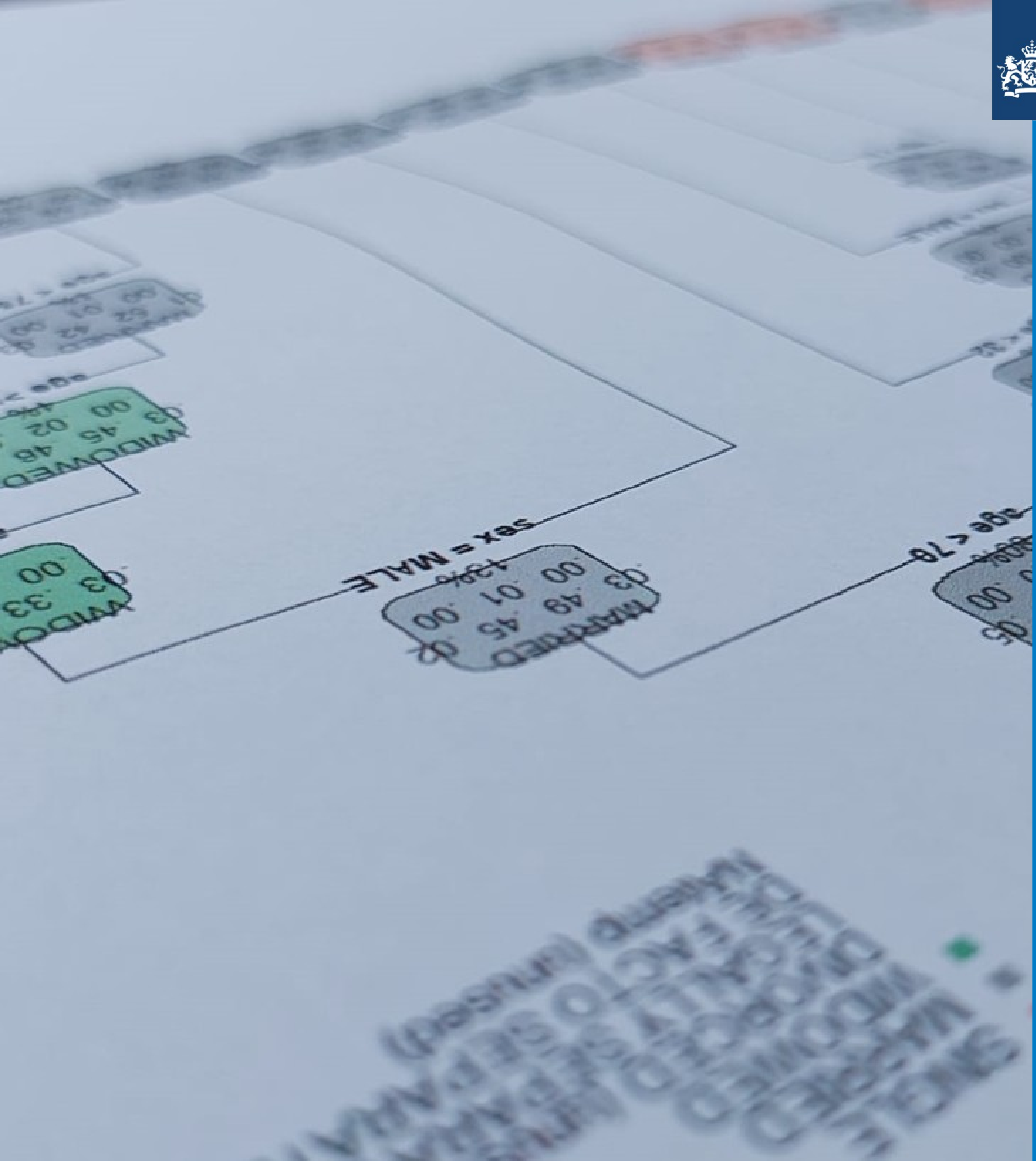
Practical Synthetic Data Generation

Lotte Pater, Ministry of Education / University of Groningen

25/03/2025

lotte.pater@duo.nl



Practical synthetic data generation

This talk has two points:

1. CART is a good method to generate synthetic data tables
2. Using synthetic data in practice is super multidisciplinary (and interesting!)

I’ll spend the next 30-40 minutes motivating these points

Why synthetic data?

1. Idea:
   - Privacy: individual level
   - Research: structural level
2. Synthetic data:
   - Fake on the individual level
     - GDPR doesn’t apply
   - Same(~ish) conclusions on the structural level

Synthetic data

Synthetic vs anonimized data

Synthesis process: from original to synthetic data

Synthesis method 1: GANs

InPrivate | make synthetic data - Google Zo

https://www.google.com/search?q=make+synthetic+data&sca_esv=570934989&ei=wn8eZcLqEcz9kwWIlbrIDQ&ved=0ahUKEwjC08mHyt6BAxXM_qQKHa

Google

make synthetic data

Video's | Afbeeldingen | Nieuws | Boeken | Maps | Vluchten | Financieel

Ongeveer 417.000.000 resultaten (0,20 seconden)

Gesponsord

Gretel.ai
https://www.gretel.ai

Synthetic Data Generation - Create Synthetic Training Data

Train models & produce better results at a fraction of the cost with smarter, safer **data**. Collaborate With Team. Run In The Cloud. Scale Workloads.

Synthetic Tabular Data

Request Early Access to Tabular LLM Generate Data From Scratch

Videos, podcasts

Read About Gretel In The News, Listen To Podcasts, Or Watch

Gesponsord

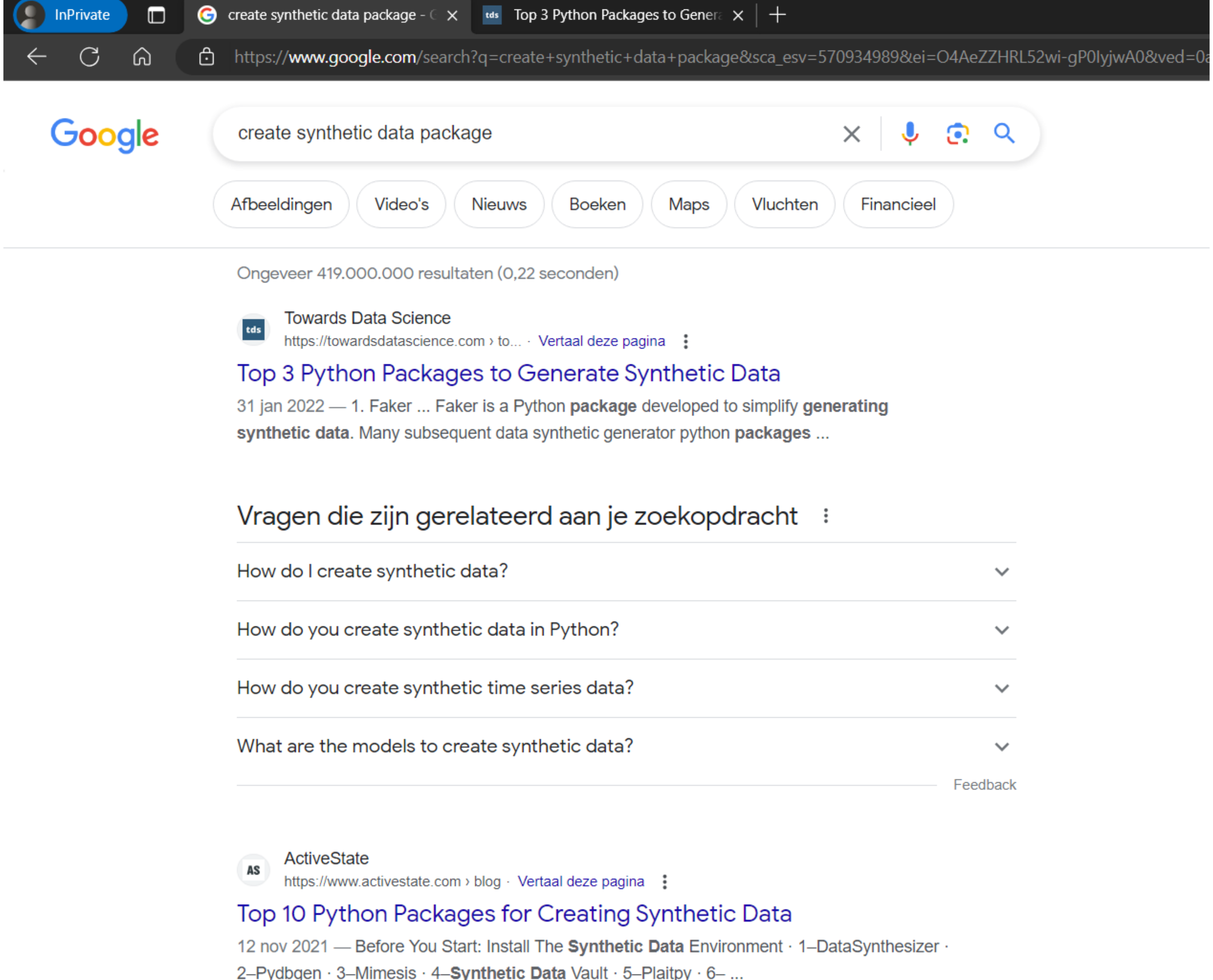
K2View
https://www.k2view.com › generate › synthetic-data

Gartner Synthetic Data Guide - Synthetic Data Generation

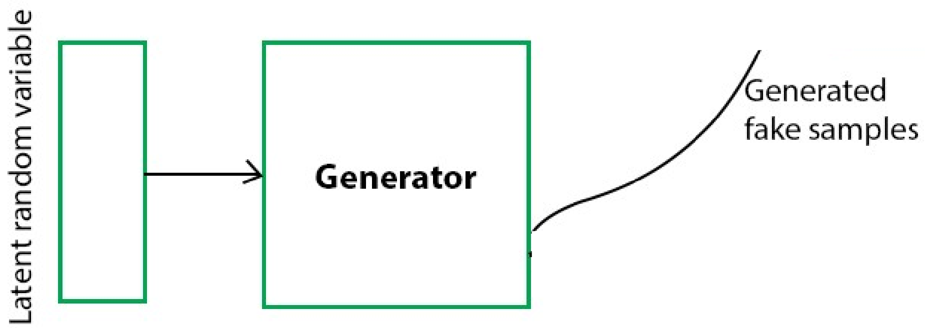
Read the Gartner© report: Pros and cons of **data** masking vs. **synthetic data** for TDM and ML.

Gesponsord

anyverse.ai
https://www.anyverse.ai › synthetic-data

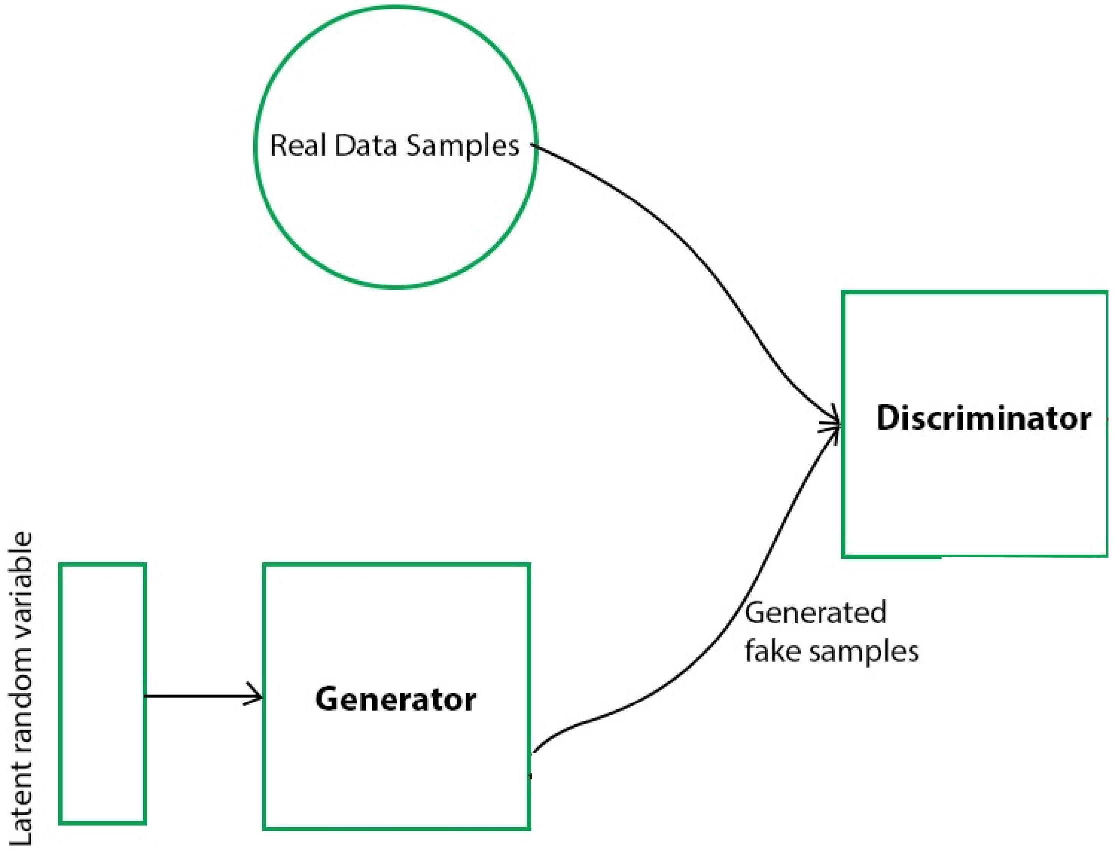

InPrivate
create synthetic data package - G
Top 3 Python Packages to Gener
https://www.google.com/search?q=create+synthetic+data+package&sca_esv=570934989&ei=O4AeZZHRL52wi-gP0lyjwA0&ved=0a
Google
create synthetic data package
Afbeeldingen
Video's
Nieuws
Boeken
Maps
Vluchten
Financieel
Ongeveer 419.000.000 resultaten (0,22 seconden)
Towards Data Science
https://towardsdatascience.com › to... · Vertaal deze pagina
Top 3 Python Packages to Generate Synthetic Data
31 jan 2022 — 1. Faker ... Faker is a Python package developed to simplify generating synthetic data. Many subsequent data synthetic generator python packages ...
Vragen die zijn gerelateerd aan je zoekopdracht
How do I create synthetic data?
How do you create synthetic data in Python?
How do you create synthetic time series data?
What are the models to create synthetic data?
Feedback
ActiveState
https://www.activestate.com › blog · Vertaal deze pagina
Top 10 Python Packages for Creating Synthetic Data
12 nov 2021 — Before You Start: Install The Synthetic Data Environment · 1–DataSynthesizer · 2–Pydbgen · 3–Mimesis · 4–Synthetic Data Vault · 5–Plaitpy · 6– ...

GANs - Generative Adversarial Networks



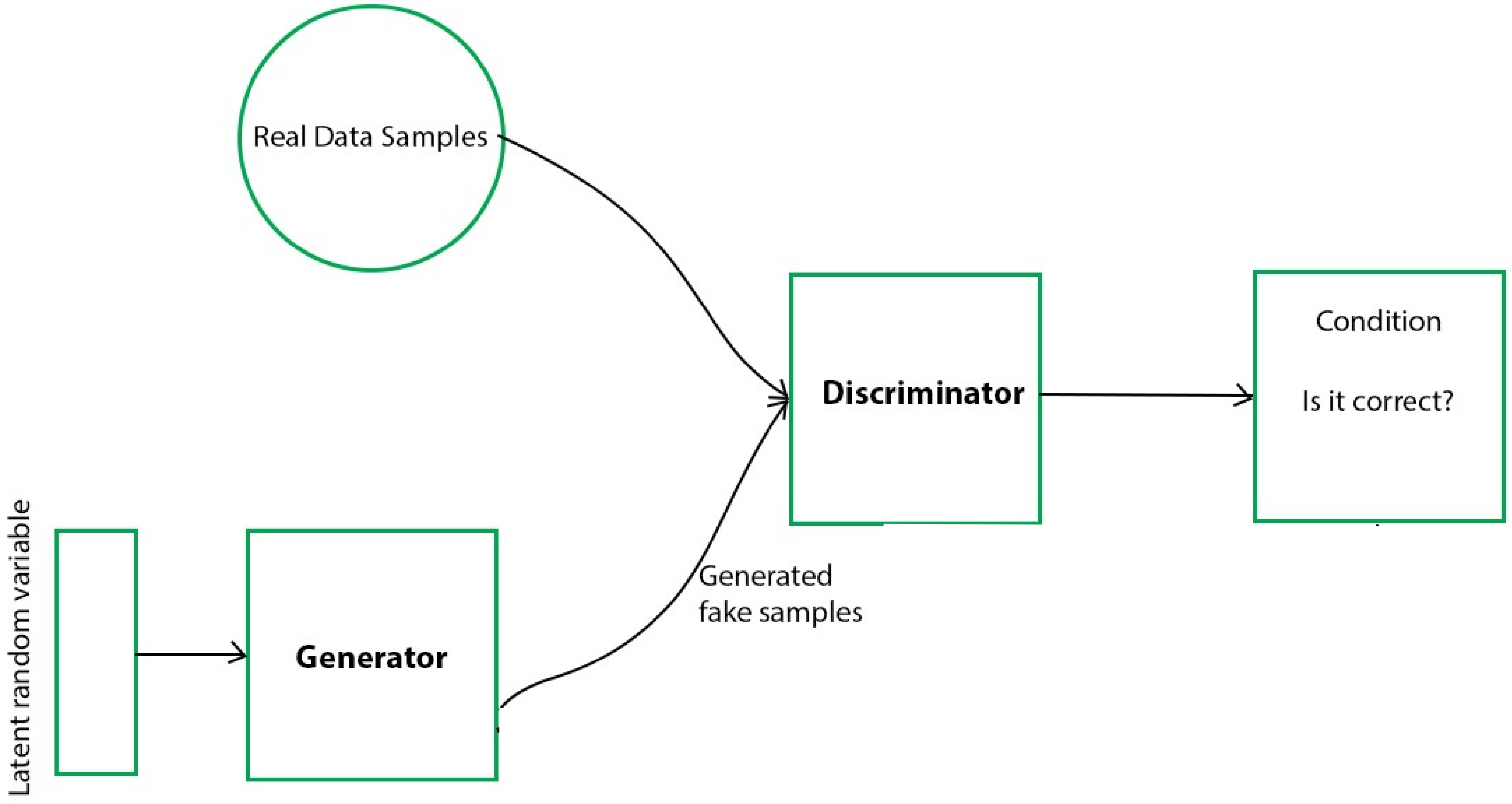
Sourc
ork-g

GANs - Generative Adversarial Networks



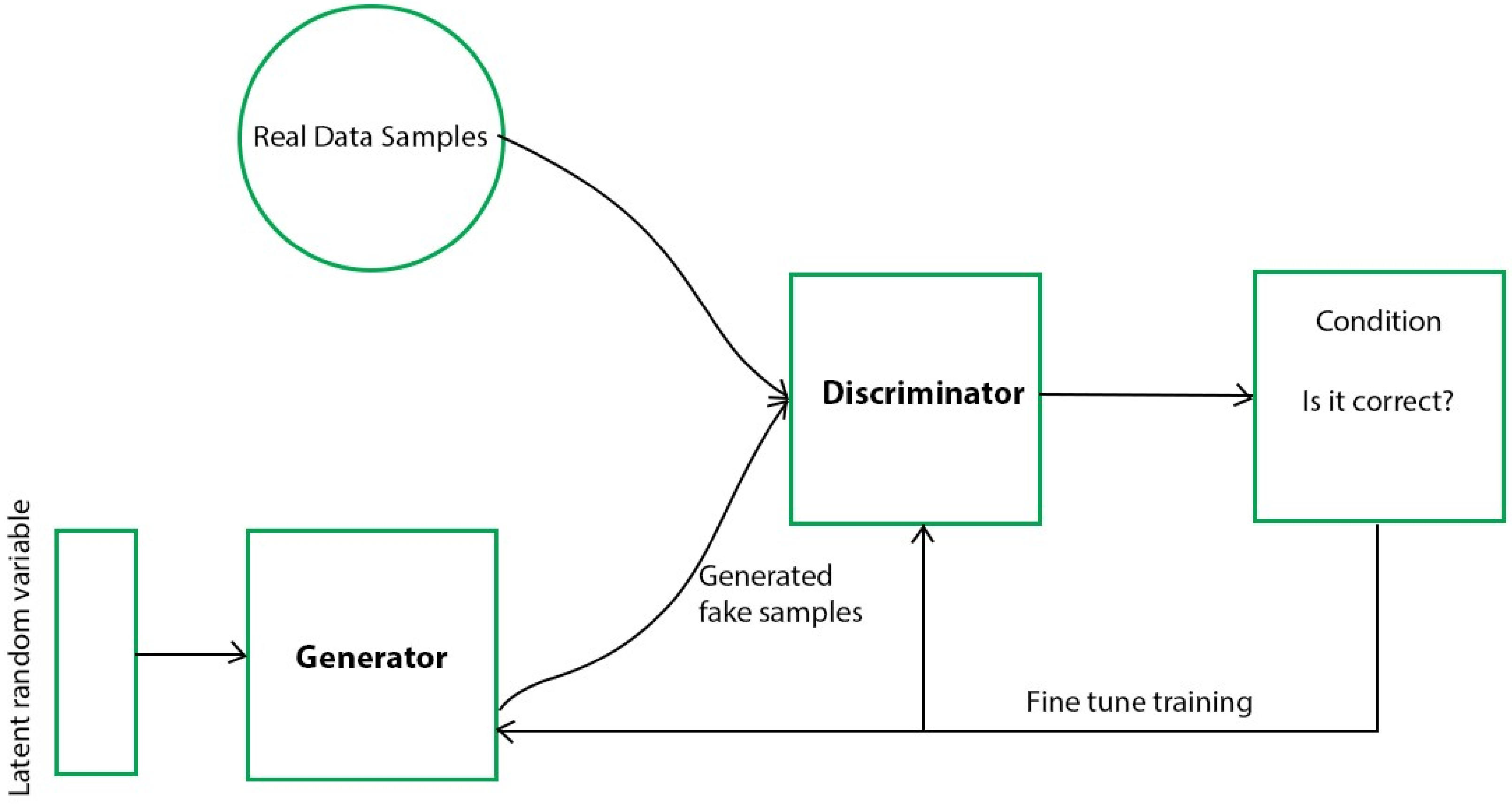
Sourc
ork-g

GANs - Generative Adversarial Networks



Sourc
ork-g

GANs - Generative Adversarial Networks



Sourc
ork-g

GANs work great for images…

…but not so much for tables.

The paper concludes that the CART model generates data with the highest utility for all the considered types of tabular datasets. In contrast, the Bayesian network model generates data with the lowest quality for all tabular datasets. Contrary to popular belief, the performance of GANs is unexceptional.

Source: 'Comparison of Tabular Synthetic Data Generation Techniques Using Propensity and Cluster Log Metric | Elsevier Enhanced Reader'. https://doi.org/10.1016/j.jjimei.2023.100177.

Why?

- GANs optimize for individuals that are similar to the synthetic dataset
- BUT: You want the distribution to be similar
- What happens?
  - Mode collapse
  - One-dimensional distributions that are very dissimilar

Synthesis method 2: CART

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Sex distribution

Generate **Sex**

Sex
MALE
MALE
FEMALE
FEMALE
FEMALE
FEMALE
MALE
FEMALE
MALE
FEMALE
MALE
MALE
MALE
FEMALE

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Age predicted from **Sex**

Sex
MALE
MALE
FEMALE
FEMALE
FEMALE
FEMALE
MALE
FEMALE
MALE
FEMALE
MALE
MALE
MALE
FEMALE

Generate **Age**

Age
81
54
32
98
50
37
28
62
78
29
59
41
18
73

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Education predicted from **Sex** and **Age**

Sex	Age
MALE	81
MALE	54
FEMALE	32
FEMALE	98
FEMALE	50
FEMALE	37
MALE	28
FEMALE	62
MALE	78
FEMALE	29
MALE	59
MALE	41
MALE	18
FEMALE	73

Generate **Education**

Education
PRIMARY/NO EDUCATION
VOCATIONAL/GRAMMAR
VOCATIONAL/GRAMMAR
PRIMARY/NO EDUCATION
PRIMARY/NO EDUCATION
VOCATIONAL/GRAMMAR
VOCATIONAL/GRAMMAR
PRIMARY/NO EDUCATION
PRIMARY/NO EDUCATION
SECONDARY
PRIMARY/NO EDUCATION
SECONDARY
SECONDARY
PRIMARY/NO EDUCATION

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Life satisfaction predicted from all other variables

Sex	Age	Education	Marital status	Income
MALE	81	PRIMARY/NO EDUCATION	MARRIED	2100
MALE	54	VOCATIONAL/GRAMMAR	MARRIED	1700
FEMALE	32	VOCATIONAL/GRAMMAR	DIVORCED	870
FEMALE	98	PRIMARY/NO EDUCATION	MARRIED	800
FEMALE	50	PRIMARY/NO EDUCATION	MARRIED	NA
FEMALE	37	VOCATIONAL/GRAMMAR	MARRIED	158
MALE	28	VOCATIONAL/GRAMMAR	NA	1500
FEMALE	62	PRIMARY/NO EDUCATION	MARRIED	830
MALE	78	PRIMARY/NO EDUCATION	MARRIED	NA
FEMALE	29	SECONDARY	MARRIED	580
MALE	59	PRIMARY/NO EDUCATION	MARRIED	1300
MALE	41	SECONDARY	UNMARRIED	1500
MALE	18	SECONDARY	UNMARRIED	-8
FEMALE	73	PRIMARY/NO EDUCATION	WIDOWED	1350

Generate **Life satisfaction**

Life satisfaction
PLEASED
PLEASED
MIXED
MOSTLY DISSATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
MIXED
PLEASED
MOSTLY SATISFIED

Original

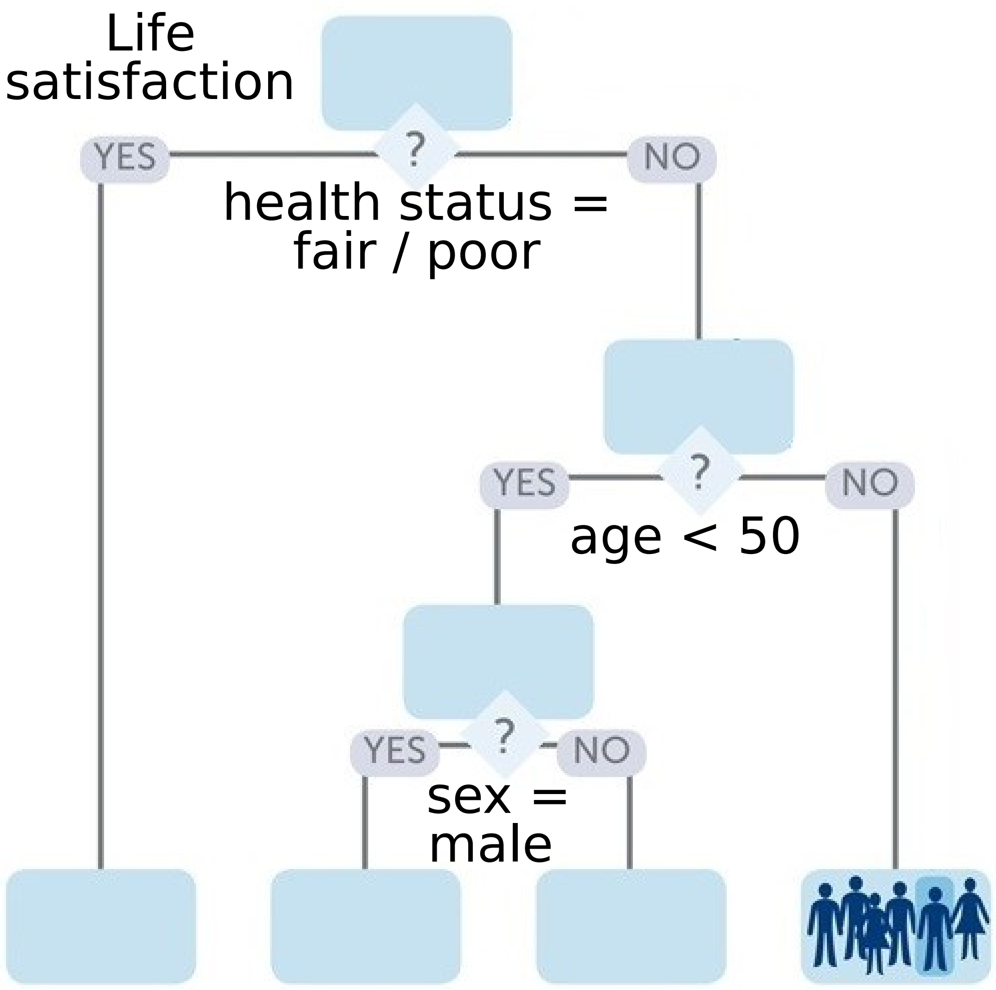
Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Joint distribution is approximated by a set of conditional distributions

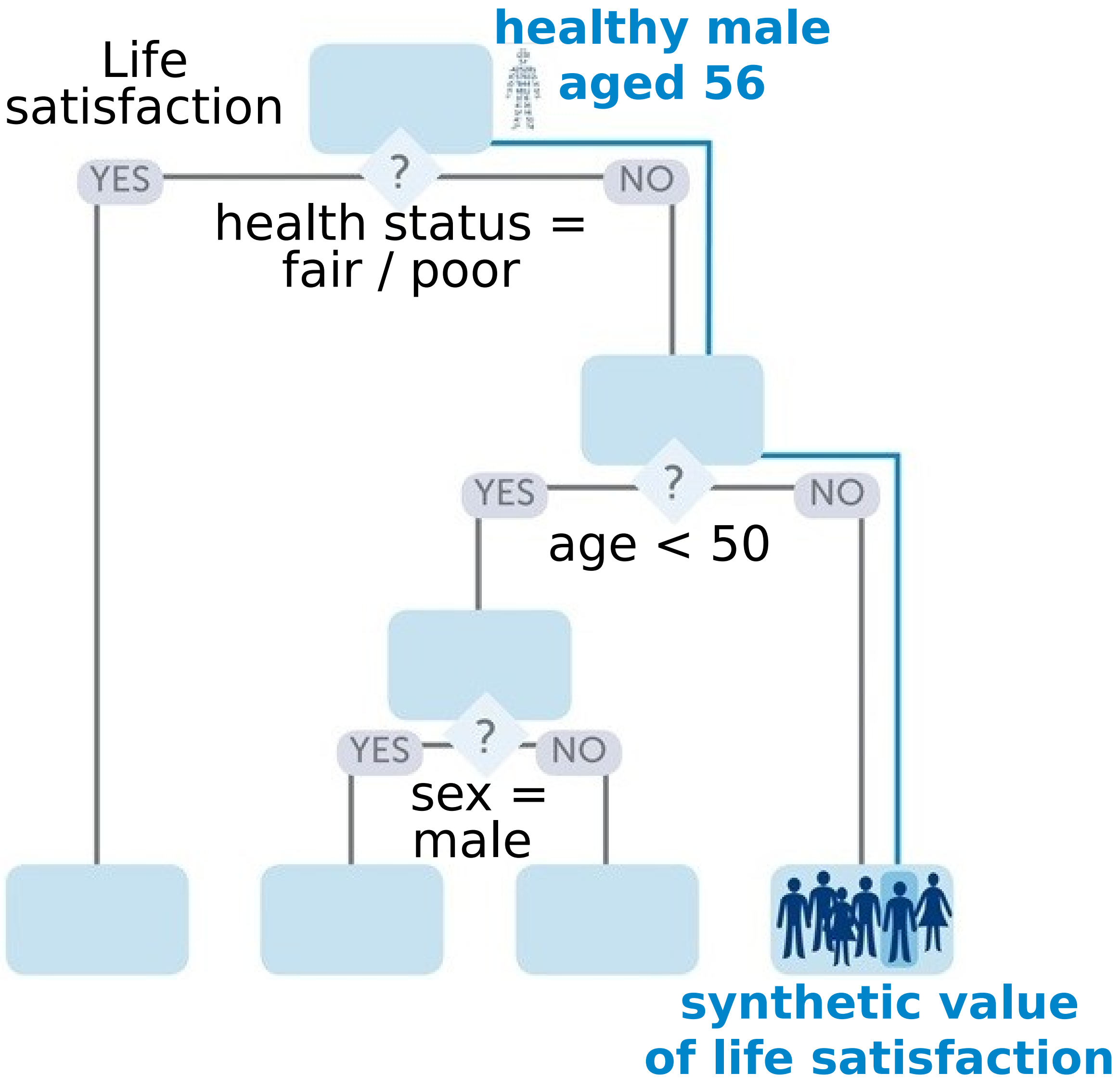
Synthetic

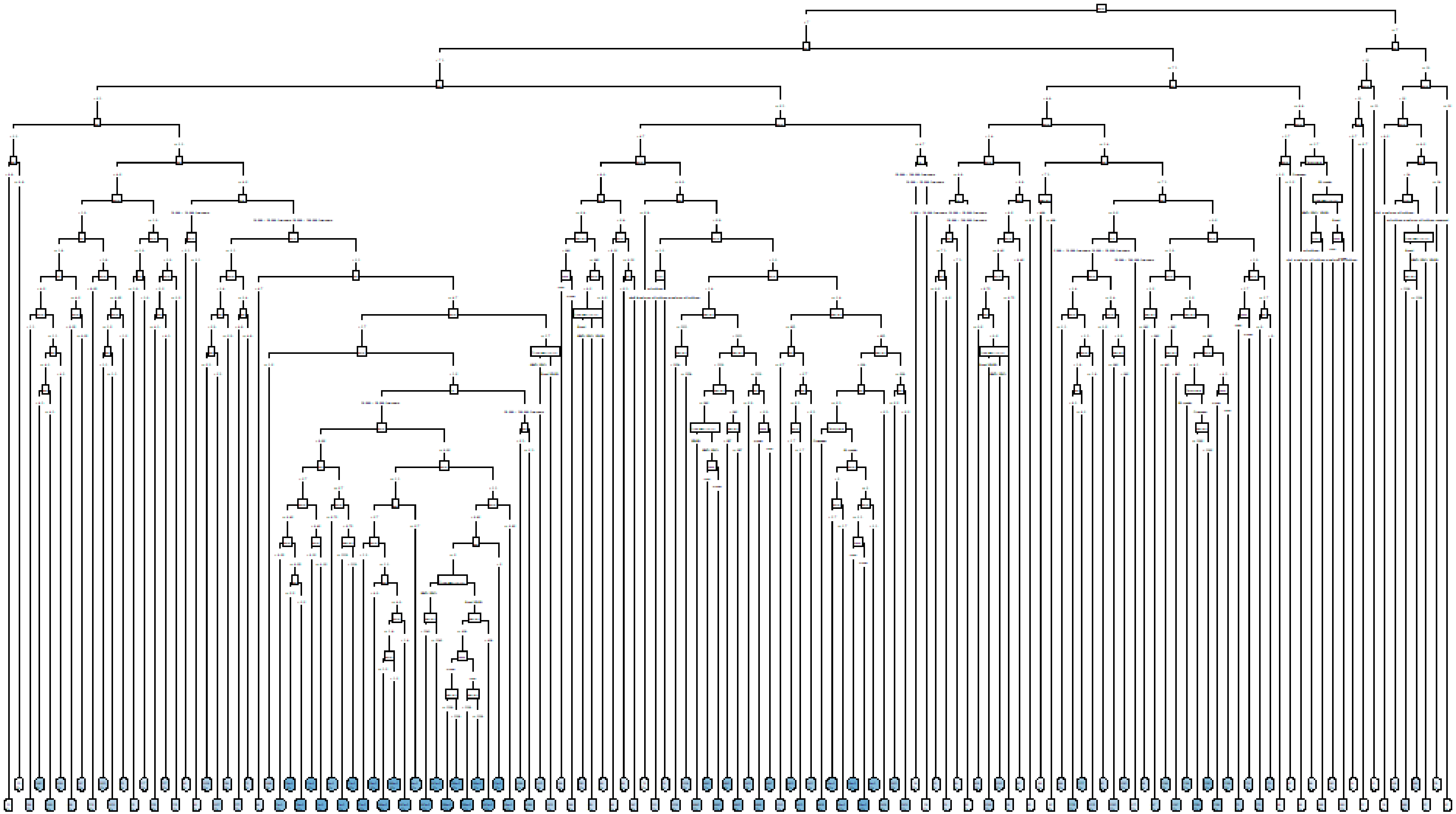
Sex	Age	Education	Marital status	Income	Life satisfaction
MALE	81	PRIMARY/NO EDUCATION	MARRIED	2100	PLEASED
MALE	54	VOCATIONAL/GRAMMAR	MARRIED	1700	PLEASED
FEMALE	32	VOCATIONAL/GRAMMAR	DIVORCED	870	MIXED
FEMALE	98	PRIMARY/NO EDUCATION	MARRIED	800	MOSTLY DISSATISFIED
FEMALE	50	PRIMARY/NO EDUCATION	MARRIED	NA	MOSTLY SATISFIED
FEMALE	37	VOCATIONAL/GRAMMAR	MARRIED	158	PLEASED
MALE	28	VOCATIONAL/GRAMMAR	NA	1500	MOSTLY SATISFIED
FEMALE	62	PRIMARY/NO EDUCATION	MARRIED	830	MOSTLY SATISFIED
MALE	78	PRIMARY/NO EDUCATION	MARRIED	NA	PLEASED
FEMALE	29	SECONDARY	MARRIED	580	MOSTLY SATISFIED
MALE	59	PRIMARY/NO EDUCATION	MARRIED	1300	MOSTLY SATISFIED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
MALE	18	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	73	PRIMARY/NO EDUCATION	WIDOWED	1350	MOSTLY SATISFIED

Classification and regression trees (CART)



Classification and regression trees (CART)





CART is the best method, but…

PROS

- One-dimensional distributions pretty much always as desired
- Two-dimensional distributions usually as well
  - Although it sometimes takes a bunch of work
- Probabilistic character works well for tables
- GDPR compliant by design

CONS

- Does badly for variables with many categories
  - Partly runs in exponential time
- Only R implementation: *synthpop*
- You need to know your data for the best result

Synthesis method 3:

Preserving correlations: A statistical method for generating synthetic data

Nicklas Jävergård, Rainey Lyons, Adrian Muntean[1] and Jonas Forsman[2]

[1]Department of Mathematics and Computer Science, Karlstad University, Sweden
[2]CGI, Data Advantage, Karlstad, Sweden

Abstract

We propose a method to generate statistically representative synthetic data. The main goal is to be able to maintain in the synthetic dataset the correlations of the features present in the original one, while offering a comfortable privacy level that can be eventually tailored on specific customer demands.

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Life satisfaction sampled from observed data

Life satisfaction
PLEASED
PLEASED
MIXED
MOSTLY DISSATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
MIXED
PLEASED
MOSTLY SATISFIED

Original

Sex	Age	Education	Marital status	Income	Life satisfaction
FEMALE	57	VOCATIONAL/GRAMMAR	MARRIED	800	PLEASED
MALE	41	SECONDARY	UNMARRIED	1500	MIXED
FEMALE	18	VOCATIONAL/GRAMMAR	UNMARRIED	NA	PLEASED
FEMALE	78	PRIMARY/NO EDUCATION	WIDOWED	900	MIXED
FEMALE	54	VOCATIONAL/GRAMMAR	MARRIED	1500	MOSTLY SATISFIED
MALE	20	SECONDARY	UNMARRIED	-8	PLEASED
FEMALE	39	SECONDARY	MARRIED	2000	MOSTLY SATISFIED
MALE	39	SECONDARY	MARRIED	1197	MIXED
FEMALE	38	VOCATIONAL/GRAMMAR	MARRIED	NA	MOSTLY DISSATISFIED
FEMALE	73	VOCATIONAL/GRAMMAR	WIDOWED	1700	PLEASED
FEMALE	54	SECONDARY	WIDOWED	2000	MOSTLY SATISFIED
MALE	30	VOCATIONAL/GRAMMAR	UNMARRIED	900	MOSTLY SATISFIED
MALE	68	SECONDARY	MARRIED	-8	DELIGHTED
MALE	61	PRIMARY/NO EDUCATION	MARRIED	-8	MIXED

Life satisfaction sampled from observed data

Life satisfaction
PLEASED
PLEASED
MIXED
MOSTLY DISSATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
PLEASED
MOSTLY SATISFIED
MOSTLY SATISFIED
MIXED
PLEASED
MOSTLY SATISFIED

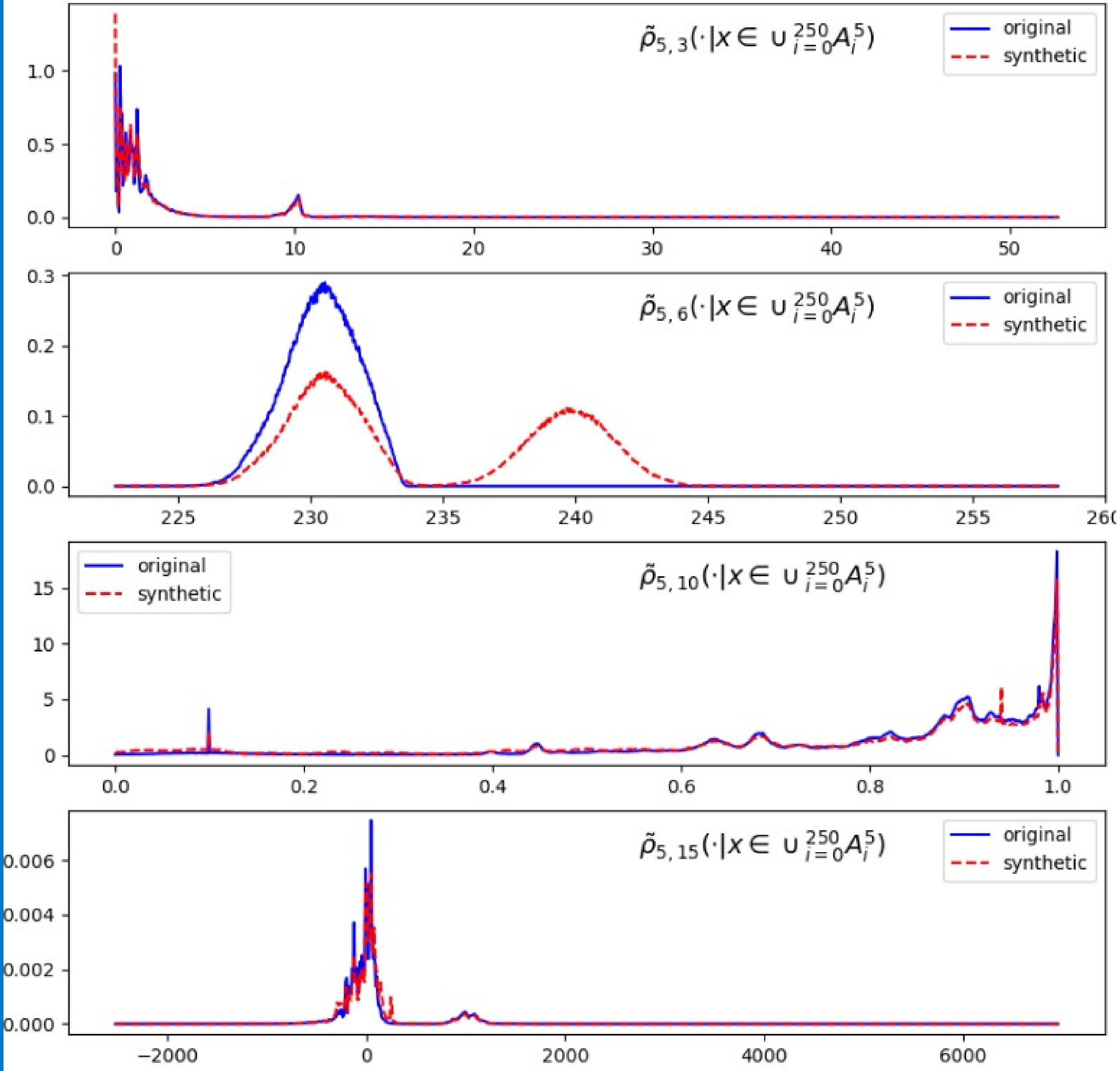
Generate **sex, age, etc**

Sex	Age	Education	Marital status	Income
MALE	81	PRIMARY/NO EDUCATION	MARRIED	2100
MALE	54	VOCATIONAL/GRAMMAR	MARRIED	1700
FEMALE	32	VOCATIONAL/GRAMMAR	DIVORCED	870
FEMALE	98	PRIMARY/NO EDUCATION	MARRIED	800
FEMALE	50	PRIMARY/NO EDUCATION	MARRIED	NA
FEMALE	37	VOCATIONAL/GRAMMAR	MARRIED	158
MALE	28	VOCATIONAL/GRAMMAR	NA	1500
FEMALE	62	PRIMARY/NO EDUCATION	MARRIED	830
MALE	78	PRIMARY/NO EDUCATION	MARRIED	NA
FEMALE	29	SECONDARY	MARRIED	580
MALE	59	PRIMARY/NO EDUCATION	MARRIED	1300
MALE	41	SECONDARY	UNMARRIED	1500
MALE	18	SECONDARY	UNMARRIED	-8
FEMALE	73	PRIMARY/NO EDUCATION	WIDOWED	1350

Joint distribution is approximated by a set of conditional distributions

Compared to CART

- Pro:
  - Less finicky to use
- Con:
  - Worse results for two dimensional distributions



Questions?

Integral privacy decisions

Question: **How do you judge the privacy impact of synthetic data?**

- Many privacy measures exist in the literature
- Our first approach: we picked one that seemed to worked well and set a treshhold
- Unsatisfactory
  - Hard to interpret
  - Behaved weirdly
  - Does not include context

Alternative: Integral privacy judgement

LEGAL

Memo

(in collaboration with legal professionals)

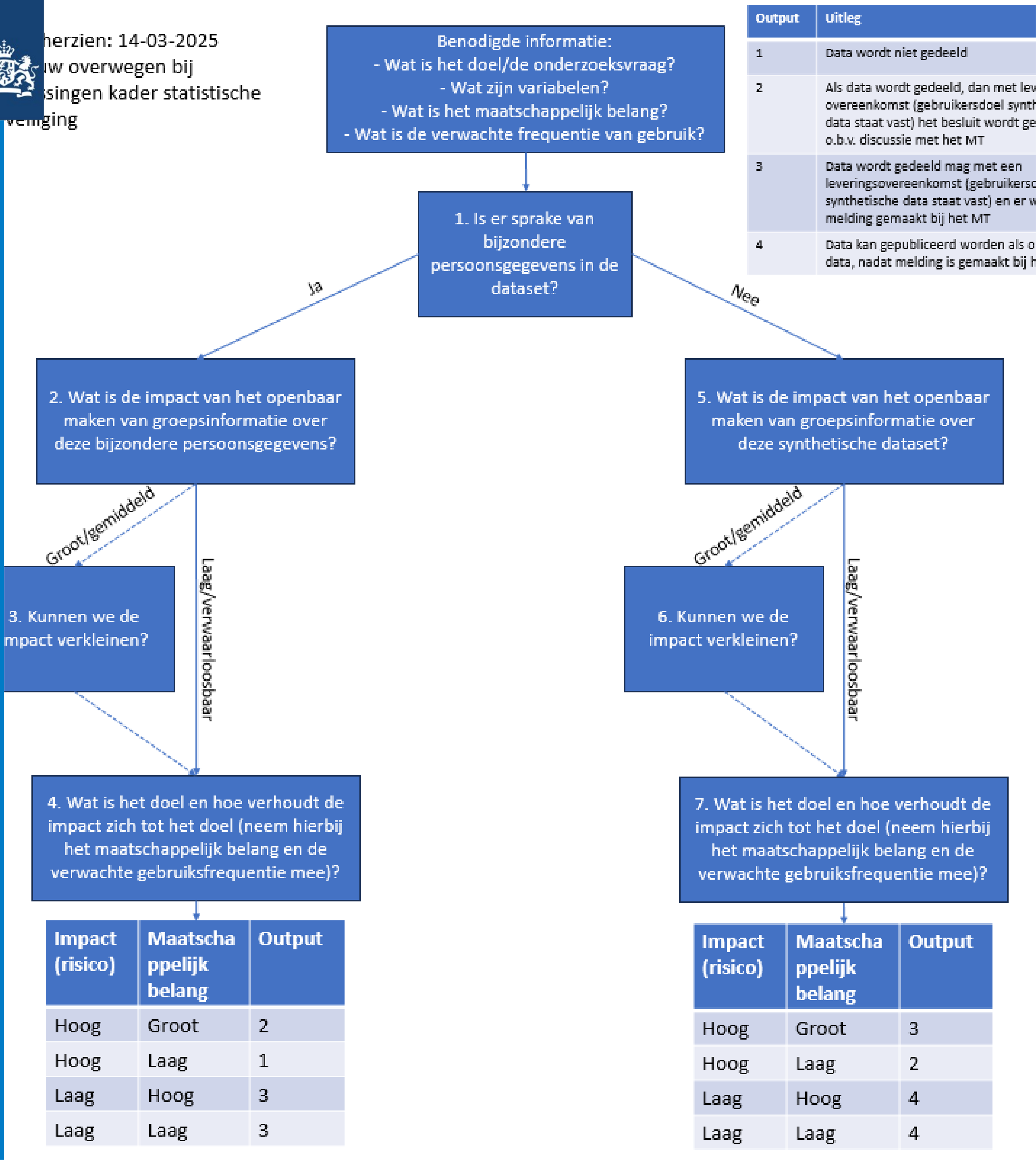
Conclusion: our synthetic data has more or less the same legal status as our aggregated data

ETHICAL-STATISTICAL

- A) Evaluation before synthetisizing: mostly ethical
- B) Evaluation after synthesizing: mostly statistical

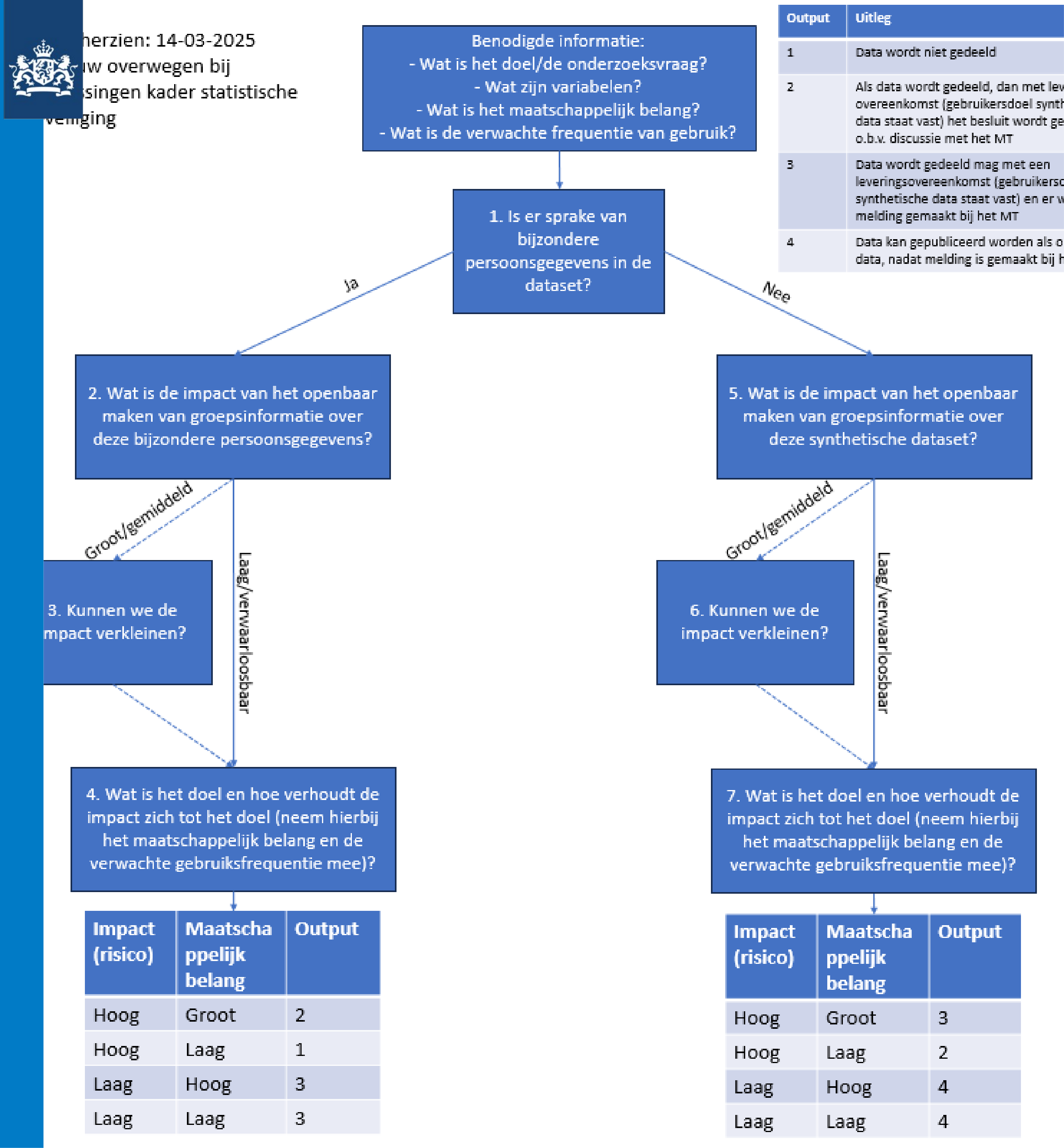
Evaluation before

- Mainly concerns group information
- Questions:
  - Does the dataset contain sensitive personal data (i.e. etnicity)
  - What's the possible impact of publication on people the original data?
  - What are the societal benefits and frequency of use?



Evaluation before: outcomes

- Four categories:
  - 1: Don’t share synthetic data
  - 2: Synthetic data
- Always a ‘comply or explain’, ethics can’t be fully captured with a flow chart ⏪



Evaluation after: metrics from literature.

- Atrribute disclose
- Chose metrics from literature
- From three categories:
  - Identity disclosure & singling out
  - Attribute disclosure
  - General similarity & outliers

DCR

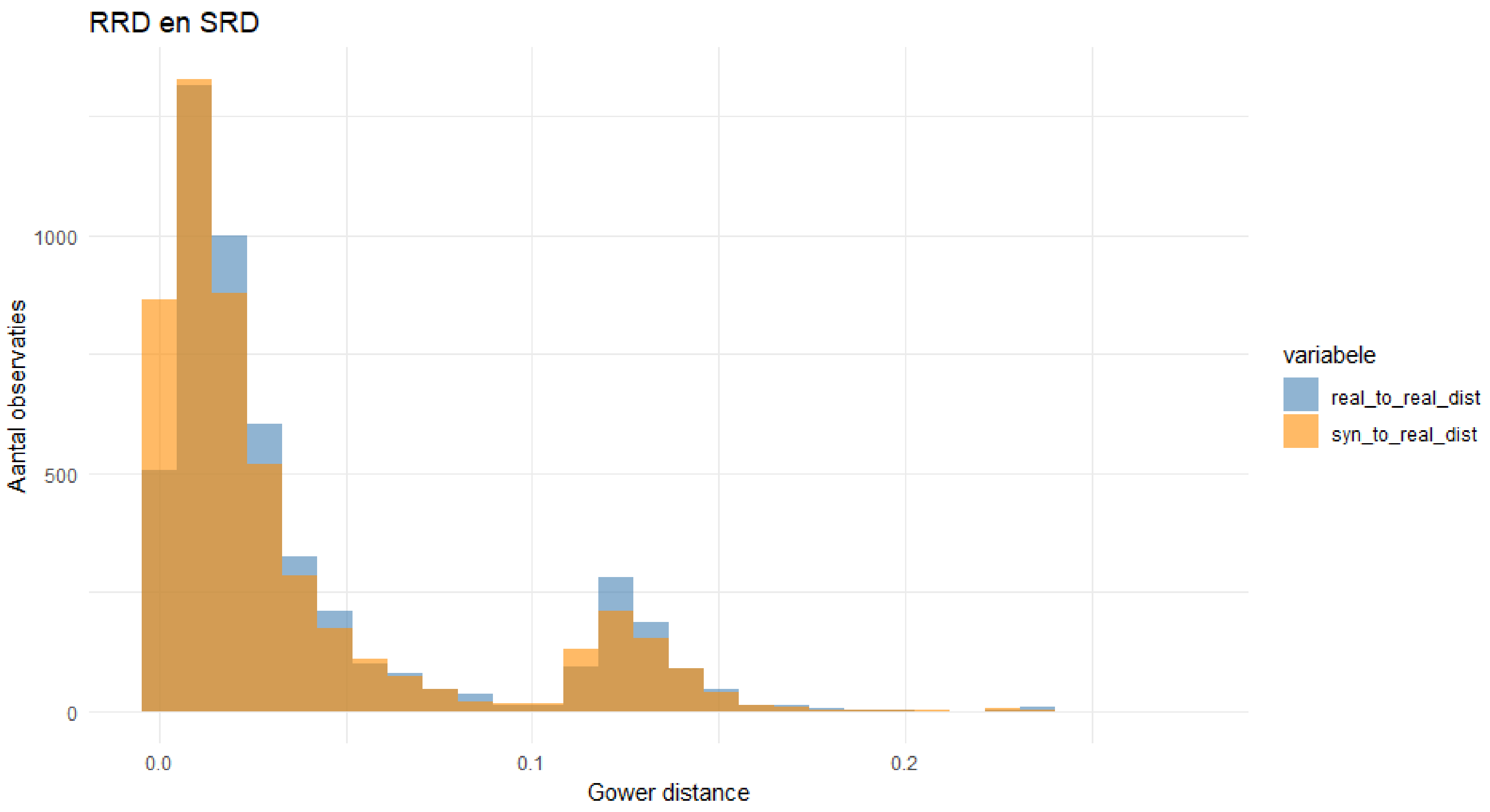
- Distance to Closest Record (DCR)
  - “General similarity & outliers”

- Idea:
  - Measure distance between individual rows
    - Gower distance
  - Real to Real Distance (RRD)
  - Synthetic to Real Distance (SRD)

DCR

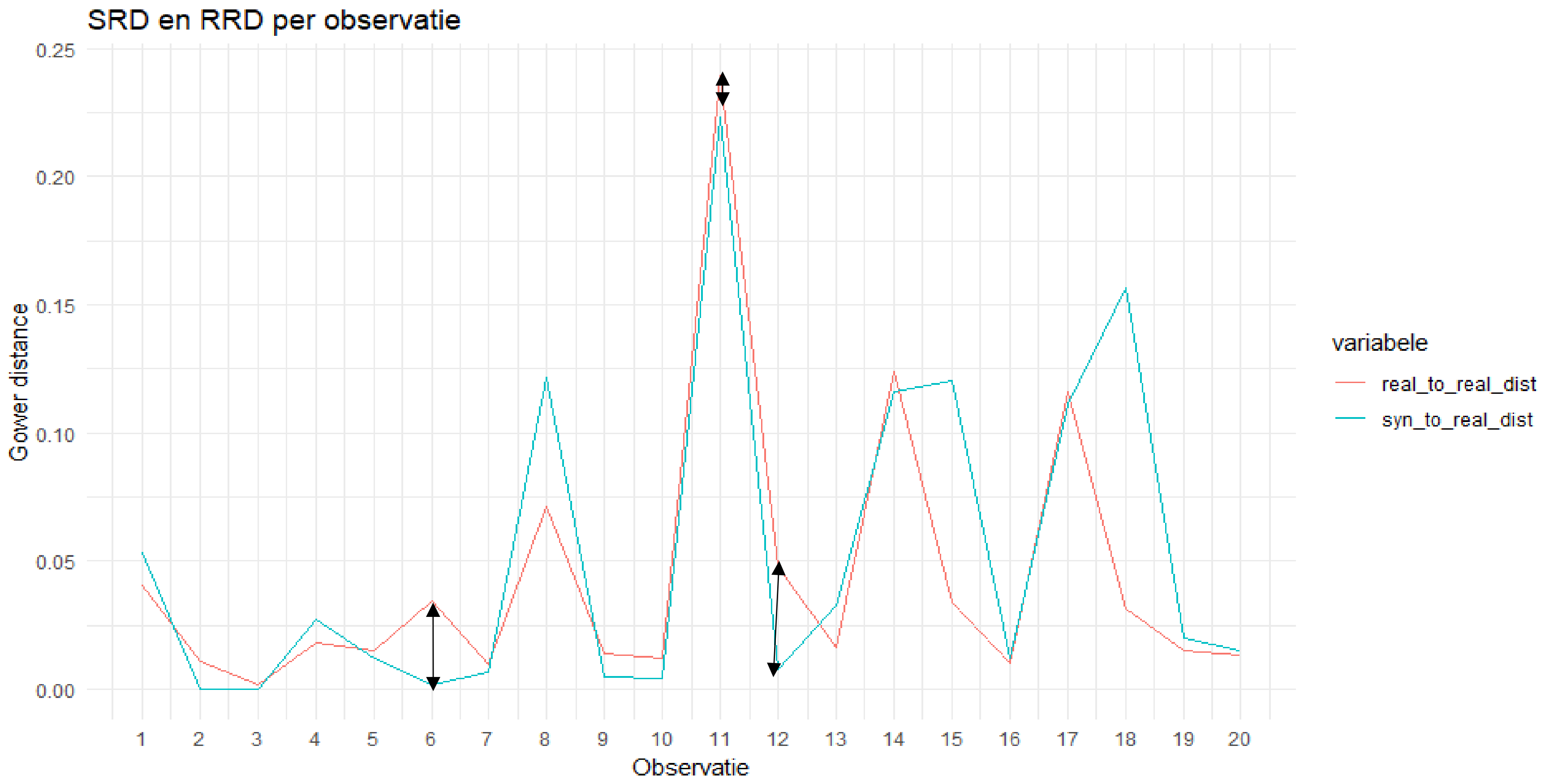
sex	age	region	placesize	depress	income	ls	marital	workab	syn_to_real_dist	real_to_real_dist
FEMALE	57	Lubuskie	URBAN 100,000-200,000	6	800	PLEASED	MARRIED	NO	0.0531994048	0.040197861
MALE	20	Podlaskie	RURAL AREAS	0	350	MOSTLY SATISFIED	SINGLE	NO	0.0000000000	0.010582011
FEMALE	18	Mazowieckie	URBAN 500,000 AND OVER	0	NA	PLEASED	SINGLE	NA	0.0000000000	0.001763668
FEMALE	78	Podlaskie	RURAL AREAS	16	900	MIXED	WIDOWED	NO	0.0274136741	0.017804223
FEMALE	54	Zachodnio-pomorskie	URBAN 100,000-200,000	4	1500	MOSTLY SATISFIED	MARRIED	NO	0.0122023810	0.015023463
MALE	20	Slaskie	URBAN 100,000-200,000	5	-8	PLEASED	SINGLE	NO	0.0017857143	0.034636268
FEMALE	39	Wielkopolskie	RURAL AREAS	2	2000	MOSTLY SATISFIED	MARRIED	NO	0.0066798942	0.009635100
MALE	39	Lubuskie	URBAN 100,000-200,000	4	1197	MIXED	MARRIED	NO	0.1217139446	0.071563829
FEMALE	43	Swietokrzyskie	RURAL AREAS	0	580	MOSTLY SATISFIED	MARRIED	NO	0.0046875000	0.013683487
FEMALE	63	Dolnoslaskie	URBAN BELOW 20,000	6	1400	PLEASED	MARRIED	NO	0.0041666667	0.012215527
FEMALE	38	Kujawsko-pomorskie	URBAN 100,000-200,000	0	1500	MOSTLY DISSATISFIED	MARRIED	YES	0.2231303991	0.239418444
FEMALE	73	Slaskie	URBAN 200,000-500,000	6	1700	PLEASED	WIDOWED	NO	0.0079848964	0.047823972
FEMALE	54	Slaskie	URBAN 200,000-500,000	4	2000	MOSTLY SATISFIED	WIDOWED	NO	0.0325389884	0.016534392
MALE	30	Zachodnio-pomorskie	URBAN 200,000-500,000	3	900	MOSTLY SATISFIED	SINGLE	NO	0.1159718752	0.123816105

DCR



RRD en SRD
1000
500
0
Aantal observaties
0.0
0.1
0.2
Gower distance
variabele
real_to_real_dist
syn_to_real_dist

DCR



SRD en RRD per observatie
Gower distance
0.25
0.20
0.15
0.10
0.05
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Observatie
variabele
real_to_real_dist
syn_to_real_dist

Questions?