



# **Synthetic Data – Even Better than The Real Thing**

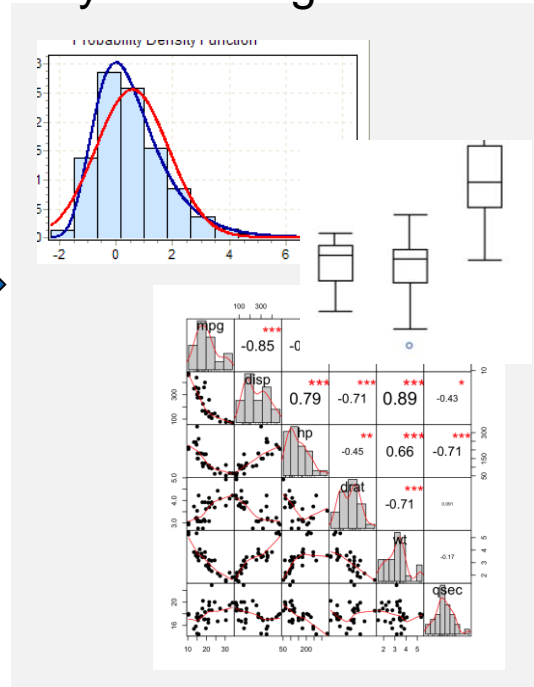
# What is Synthetic Data

- Synthetic data is a data that is generated from knowing the distributions & correlations within the original data but is generated so as it contains no actual observations from the original data

Original Data

X1	X2	X3	X4	X5	X6	X6	X7	X8	X9	X10

Synthesising Process



Synthetic Data

X1	X2	X3	X4	X5	X6	X6	X7	X8	X9	X10

# The Research

## The Synthetic data vault

Neha Patki  
CSAIL, MIT  
Cambridge, MA - 02139  
npatki@mit.edu

Roy Wedge  
LIDS, MIT  
Cambridge, MA - 02139  
rwedge@mit.edu

Kalyan Veeramachaneni  
LIDS, MIT  
Cambridge, MA - 02139

*Abstract*—The goal of this paper is to build a system that automatically creates synthetic data to enable data science endeavors. To achieve this, we present the Synthetic Data Vault (SDV), a system that builds generative models of relational databases. We are able to sample from the model and create synthetic data, hence the name SDV. When implementing the SDV, we also developed an algorithm that computes statistics at the intersection of related database tables. We then used a state-

Since  
sequen  
most  
strat  
data  
a p  
thes

Ref:  
<http://www.govtech.com/security/University-Researchers-Use-Fake-Data-for-Social-Good.html>

### University Researchers Use 'Fake' Data for Social Good

The third installment of MetroLab's Innovation of the Month series highlights how researchers are addressing the challenges of privacy, data storage and usability, and creating synthetic data sets to improve social outcomes.

BY BEN LEVINE / NOVEMBER 7, 2017



Ref:  
<http://www.govtech.com/security/University-Researchers-Use-Fake-Data-for-Social-Good.html>

# Our Problem

## New Generation Education



- **Generate data samples that replicated real world data without access to the data**

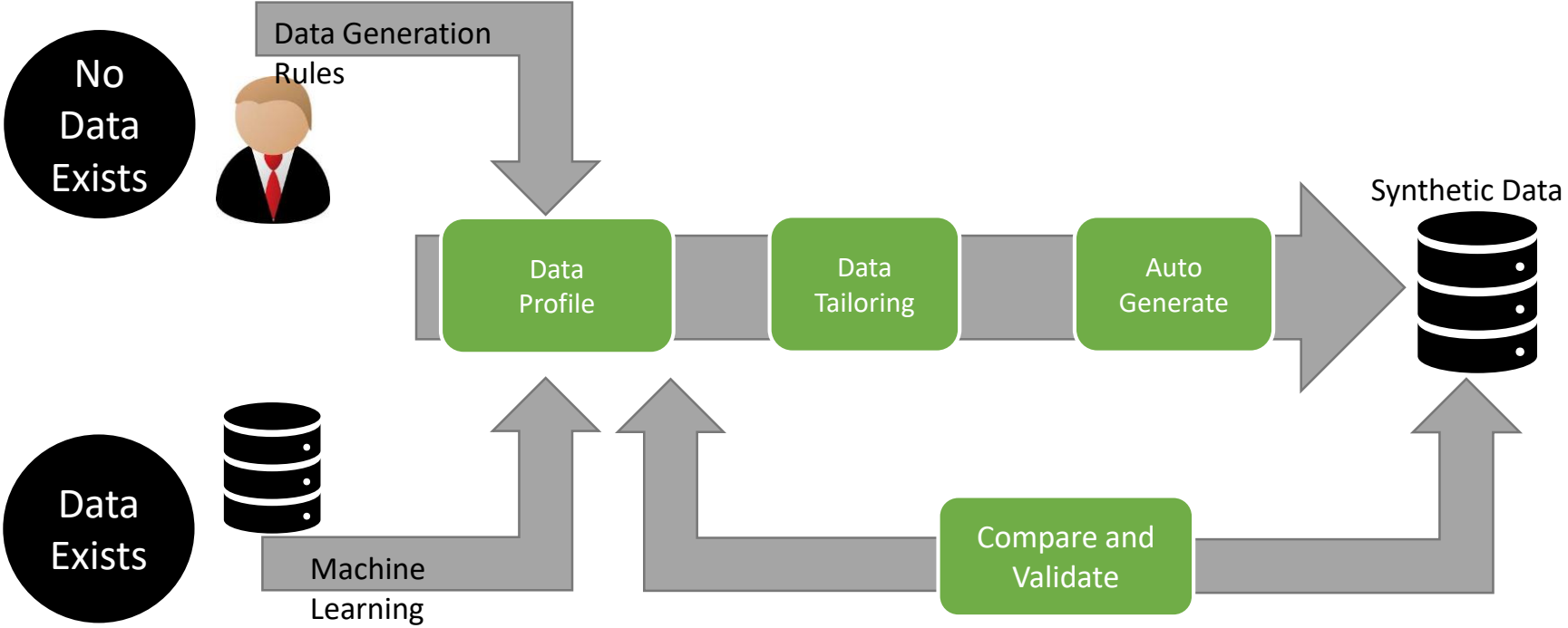
# Our Solution - Trevidata

Data on Demand

Easy to Use

Tailored by you

Small or Large



# Synthetic Data – Even Better than The Real Thing?

- We are currently looking for real world problems to test our software on?
- Do you have a data problem that synthetic data can help solve – Contact us, we would love to work with you –
- Contact details:
  - [Aoife@theanalyticsstore.com](mailto:Aoife@theanalyticsstore.com)
  - [www.trevidata.com](http://www.trevidata.com)

**Trevidata**