Motivation

Organizations need/want to publish their datasets without compromising users' privacy

DP: Weak utility (the "curse of dimensionality") [1]

k-Anonymity: Inadequate privacy

[1] Brickell & Shmatikov, The cost of privacy: destruction of data-mining utility in anonymized data publishing. In KDD 2008.

Raw Data → **Data Anonymizer** → Sanitized Data

Organizations need/want to publish their datasets without

compromising users' privacy

How about generating

synthetic dataset?