

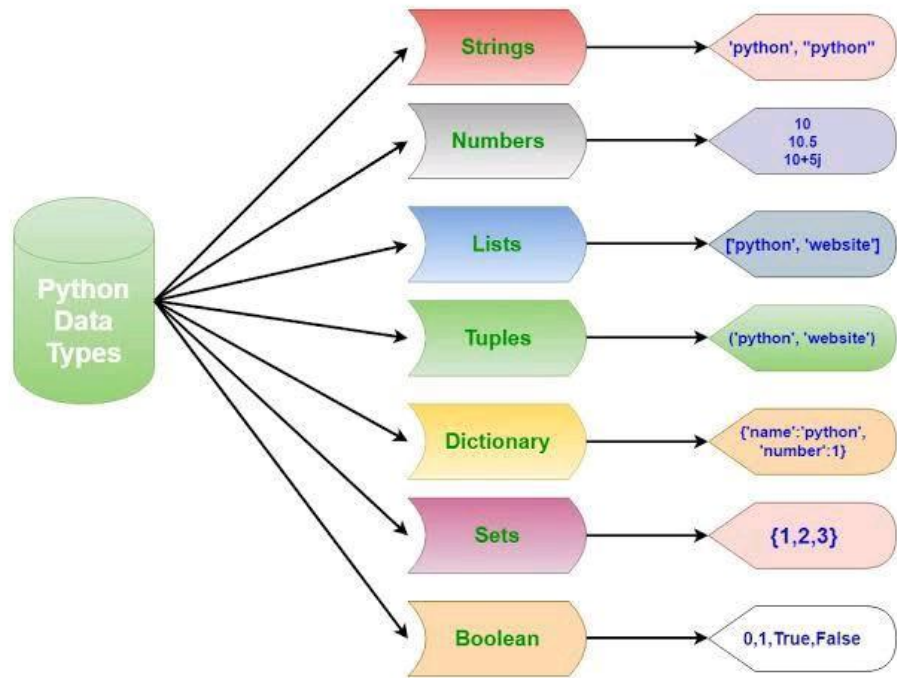


# Classifying Data

Stephen Downes

December 4, 2021

# Data Types



- **Personal Data**

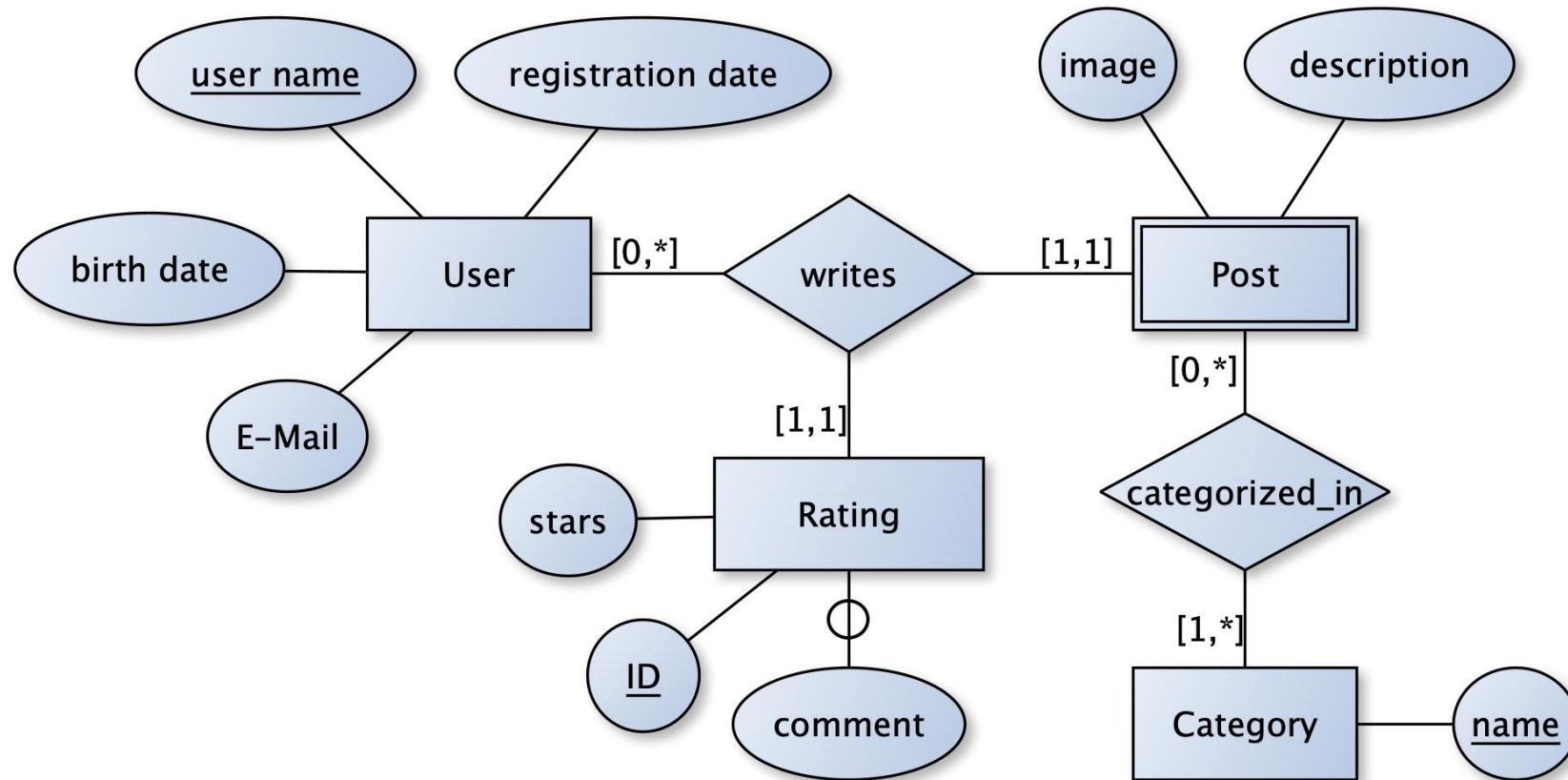
- Pentland in 2007 suggested to the World Economic Forum<sup>16</sup> that individuals might have control over their personal data, so that “a person's data would be equivalent to their ‘money.”

- **Multimodal**

- not only logs of activities completed on a computer or mobile devices, but also by employing such technologies as biosensors, eye tracking, infrared imaging, or wearable cameras (Blikstein & Worsley, 2016)

[http://www3.weforum.org/docs/WEF\\_ITTC\\_PersonalDataNewAsset\\_Report\\_2011.pdf](http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf)

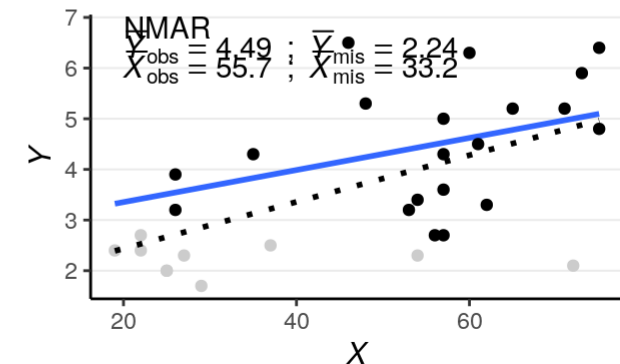
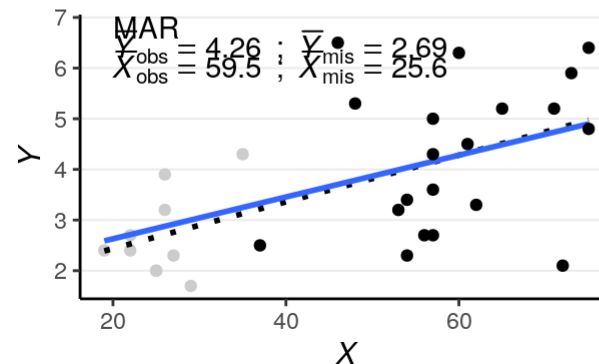
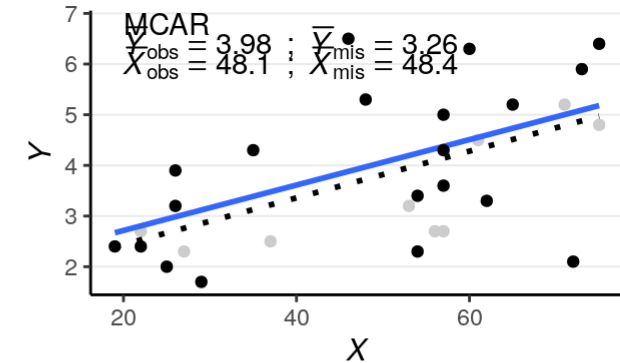
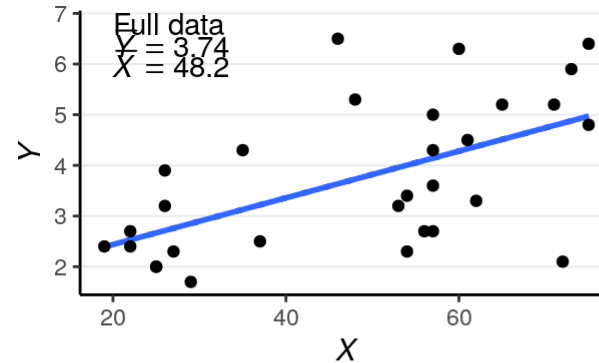
# Entities



<https://stackoverflow.com/questions/64726931/does-my-er-diagram-properly-represent-a-basic-website>

# Omitted Data

“Omitted variable bias occurs when an algorithm lacks sufficient input information to make a truly informed prediction about someone, and learns instead to rely on available but inadequate proxy variables.”  
(Eckersley, et.al., 2017)



[https://bookdown.org/marklhc/notes\\_bookdown/missing-data.html](https://bookdown.org/marklhc/notes_bookdown/missing-data.html)

# Data Quality

“The principle of ‘representative and high quality data,’ driven by what is colloquially referred to as the ‘garbage in, garbage out’ problem, is defined as the use of appropriate inputs to an AI system.” (Fjeld, et.al., 2020:49)



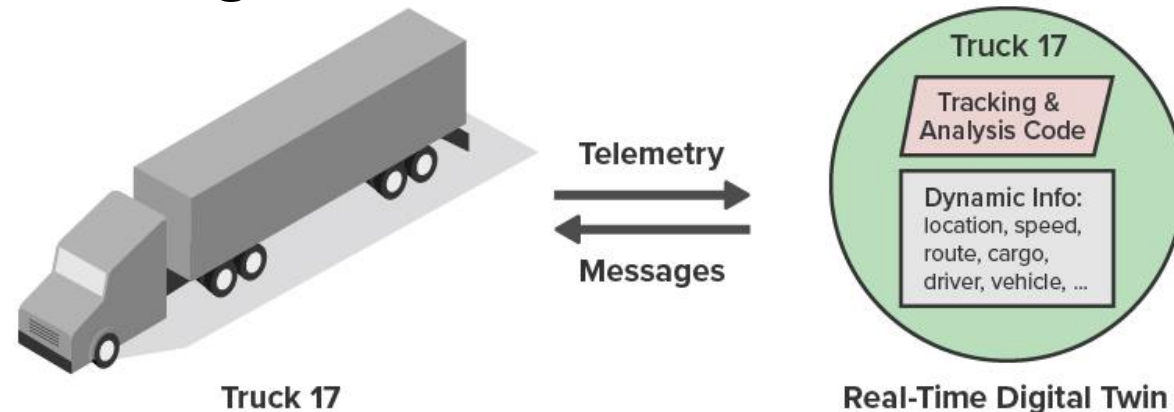
A Data Quality Mind Map

20th April 2019, Henrik Gabs Liliendahl

<https://liliendahl.com/2019/04/20/a-data-quality-mind-map/>

# Real-Time and Dynamic Data

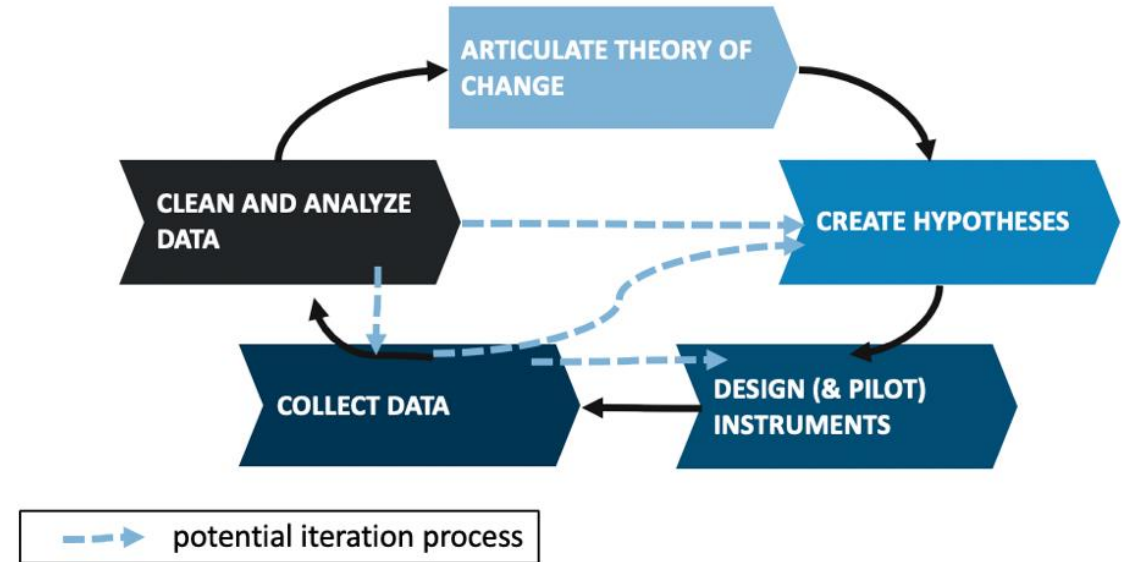
- Dynamic: data sets often evolve and accumulate, so they may permit discoveries in the future that they do not permit today.
- Real-time: online analysis and decision-making as data arrive post additional challenges



David J. Hand. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? Big Data, Vol. 6, No. 3. Published Online:17 Sep 2018 <https://doi.org/10.1089/big.2018.0083>  
<https://www.scaleoutsoftware.com/technology/why-use-real-time-digital-twins-for-streaming-analytics/>

# Data Assemblage

CDS asks us to consider the notion of the data assemblage, which “encompasses all of the technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation and deployment of data” Ackermann



Prinsloo 2019 <https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/6589/7307>

<https://medium.com/idinsight-blog/the-case-for-iteration-in-qualitative-research-design-e07ed1314756>