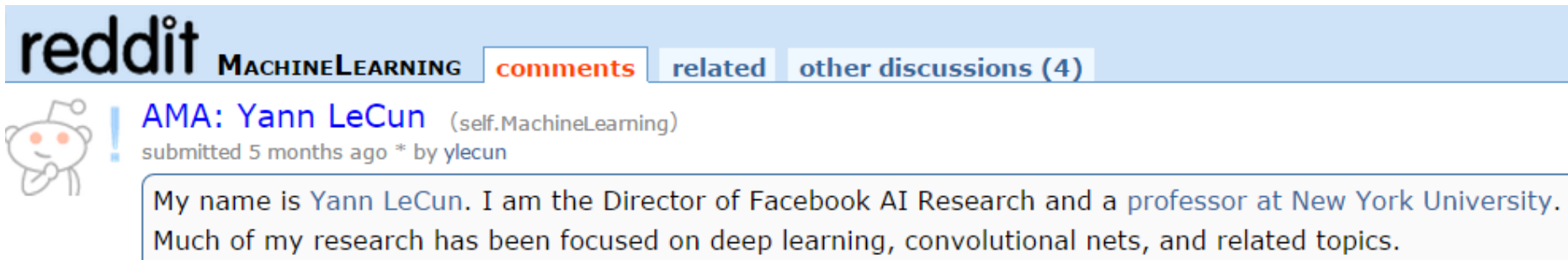


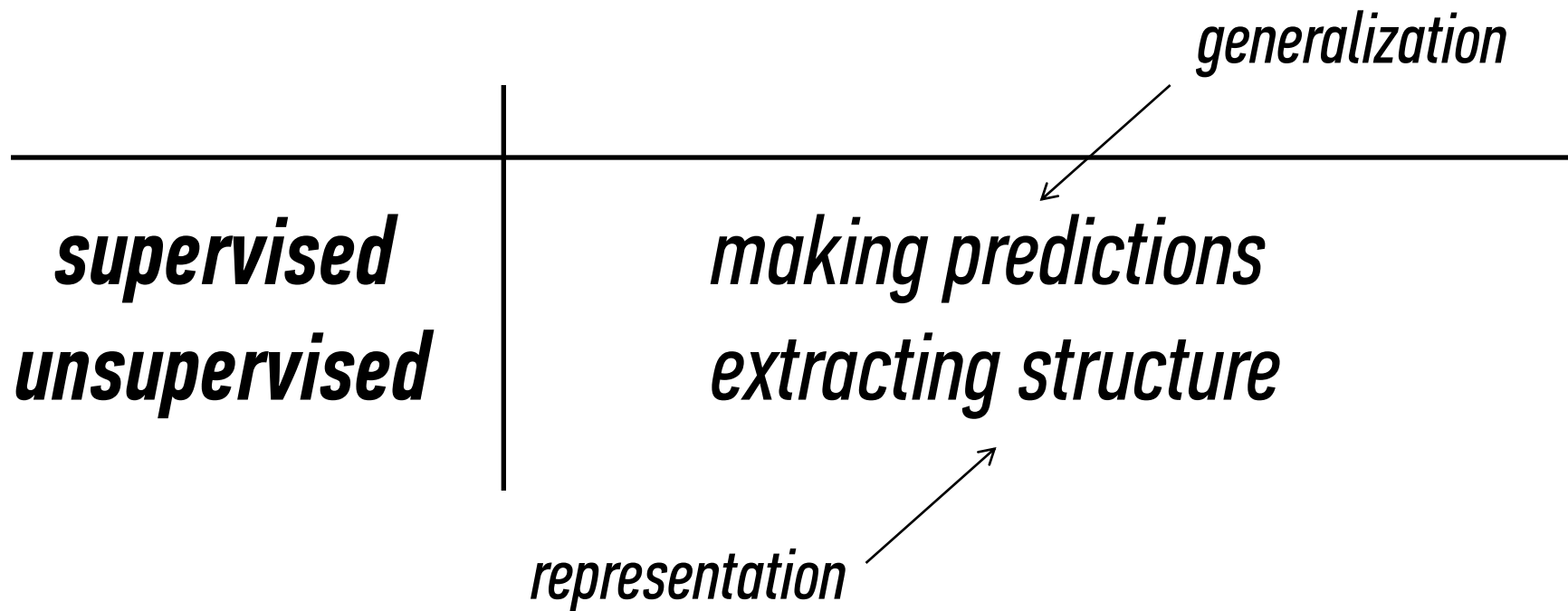
"A field of study that gives computers the ability to learn without being explicitly programmed." (1959)



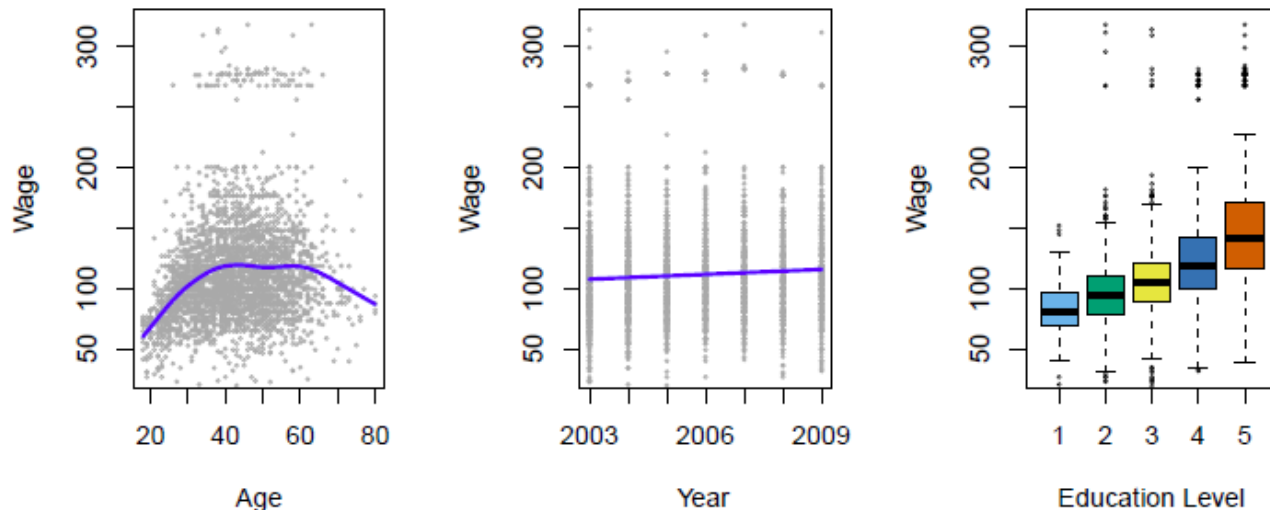
Arthur Samuel, AI pioneer  
Source: Stanford



Seriously, I don't like the phrase "Big Data". I prefer "**Data Science**", which is the **automatic (or semi-automatic) extraction of knowledge from data**. That is here to stay, it's not a fad. The amount of data generated by our digital world is growing exponentially with high rate (at the same rate our hard-drives and communication networks are increasing their capacity). But the amount of human brain power in the world is not increasing nearly as fast. This means that now or in the near future **most of the knowledge in the world will be extracted by machine and reside in machines**. It's inevitable. An entire industry is building itself around this, and a new academic discipline is emerging.



- Establish the relationship between salary and demographic variables in population survey data

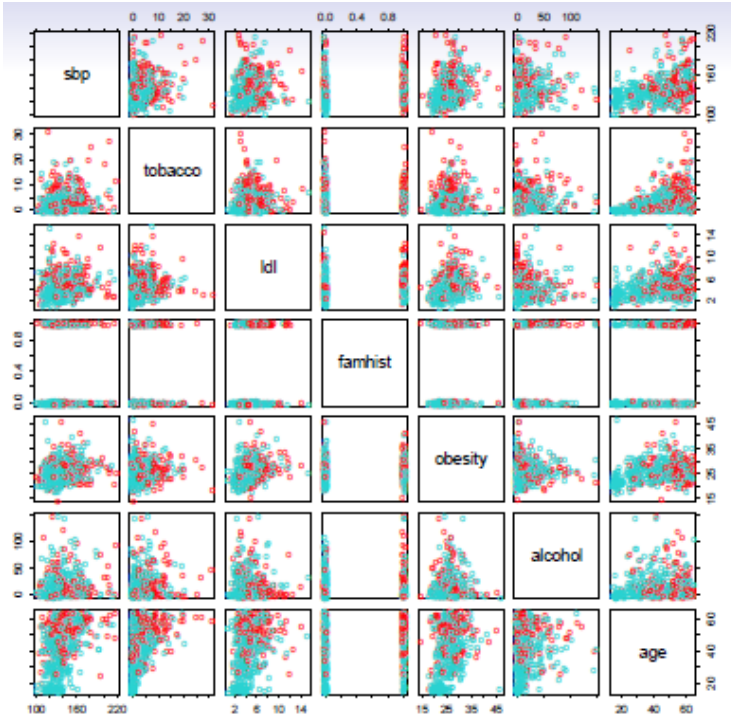


Income survey data for males from the central Atlantic region of the USA in 2009

## SUPERVISED LEARNING EXAMPLE: CLASSIFICATION

5

- Predict whether someone will have a heart attack on the basis of demographic, diet and clinical measurements



Case-control sample of men from South Africa  
Red = heart disease  
Blue = no heart disease

- Objectives of Supervised Learning:
  - Accurately predict unseen test cases
  - Understand which predictors affect the response, and how
  - Assess the quality of our predictions