Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#111 Observations & Experiments

été 2020

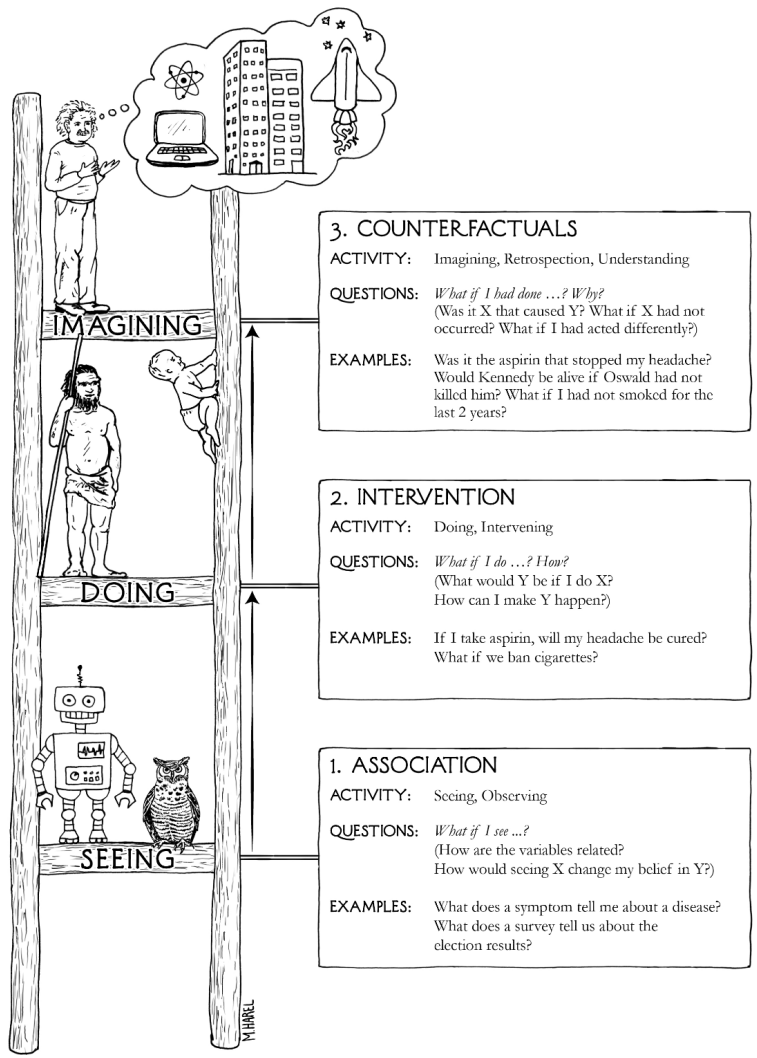
Observational vs Experimental Data

istockphoto

Causal Inference

"*Social scientists know that large amounts of data will not overcome the selection problems that make causal inference so difficult*", Grimmer (2015, We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together)

"*no causation without manipulation*", Holland (1986, Statistics and Causal Inference)



(source Pearl & Mackenzie (2018, The Book of Why))

Causal Inference

See Une courte histoire des expériences randomisées,
Polio hit the U.S. in 1916, caused hundreds of thousands of fatalities over 40 years.
Jonas Salk developed a vaccine, ready to test it in 1954



(images: https://csnbbs.com/)

Causal Inference & Polio

Incidence of disease vary from year to year (1954 vs. 1953)
used controlled experiment, treatment group vs. control group
initial design : select 2 million children (targeted school districts, high risk), vulnerable age (grades 1,2,3)
grade 2 gets the vaccine, grades 1 & 3 are the controls
(problem parental consent, correlated with higher income)
alternative design : randomized controlled double-blind
control group in the same population as treatment group
random allocation, children, parents and doctors should not know the group

Causal Inference & Polio

	grades 1,2,3	
	size	rate*
treatment	221,998	25
control	725,173	54
no consent	123,605	44

	randomized	
	size	rate*
treatment	200,745	28
control	201,229	71
no consent	338,778	46

* rate per 100,000

Source : Meier (1972)

Controlled experiment: investigators decide who is in the treatment group and who is in the control group.

Observational study: the subjects assign themselves to these two groups. The investigators just watch.

In some cases, randomized experiments are impossible

Observational vs Experimental

A Comparison of Observational Studies and Randomized, Controlled Trials (in 2000), and Randomised controlled trials the gold standard for effectiveness research (in 2018)