

Data Literacy

Prof. Dr. André Filipe M. Batista



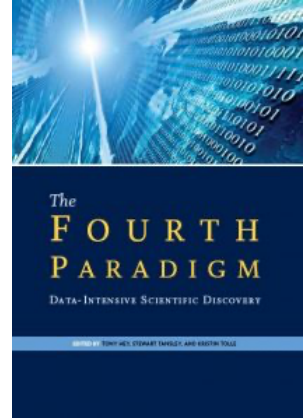
Motivação

Dados não são dados. Ou seja, não caem do céu, prontos e perfeitos para quem os consome. São o resultado, por vezes, de longos processos de construção que envolvem várias decisões metodológicas. Além disso, o consumo dos dados não é imediato. Quem lê ou analisa os dados precisa também estar capacitado para compreendê-los. Mesmo pesquisadores experimentados podem cair nas diversas armadilhas que uma nova e desconhecida base de dados geralmente apresenta.

--- Guia brasileiro de análise de dados: armadilhas & soluções / Editores Claudio D. Shikida, Leonardo Monasterio, Pedro Fernando Nery. Brasília: Enap, 2021.

Motivação (2)

- e-Science: quarto paradigma
 - Unificação de *teoria, experimentação e simulação*
 - Ciência orientada a um grande volume de dados
 - Dados capturados por instrumentos ou gerados por simulação
 - Dados processados por software
 - Informação/Conhecimentos armazenados em computadores – larga escala
 - Pesquisadores analisam os dados com auxílio de software e bases de dados
- A ciência é produtora e consumidora de dados!



Motivação (3)

- Citizen Science
 - Ciência Cidadã
 - Pessoas comuns que voluntariamente participam de atividades de coleta e compartilhamento de dados



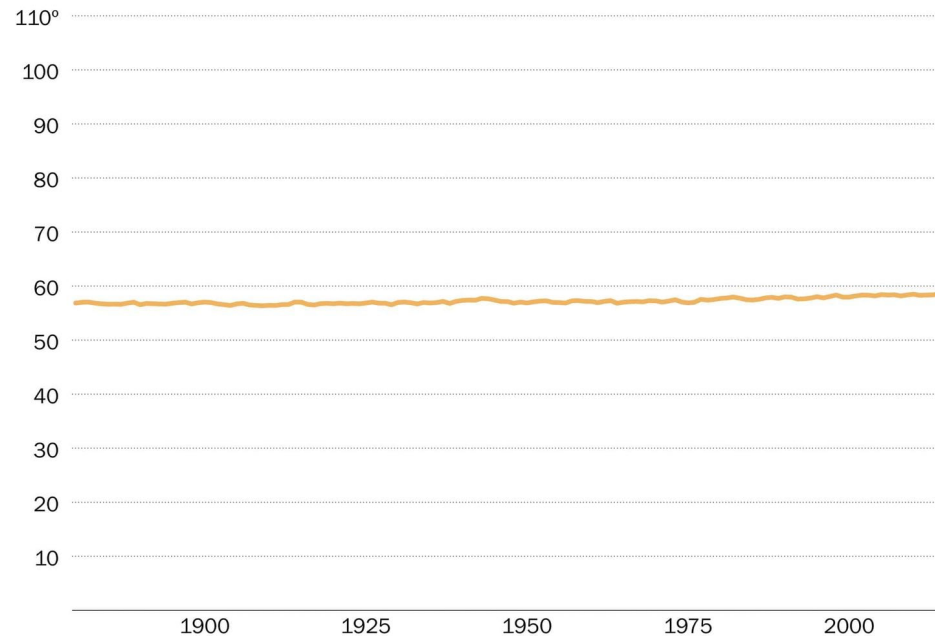
Motivação (4)

- Data Misleading

O gráfico publicado

Average global temperature by year, 50x scale

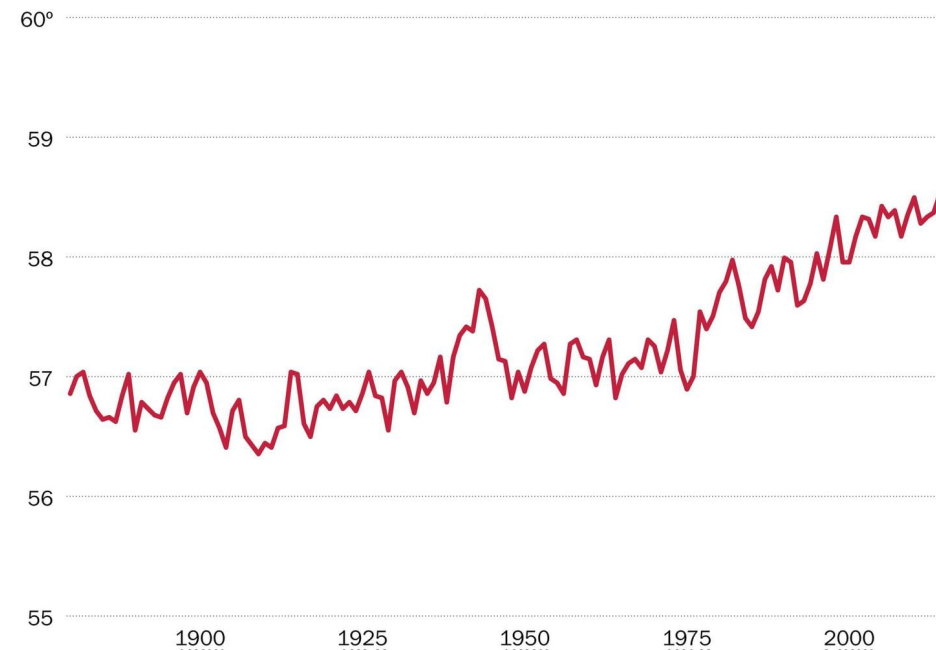
Data from NASA/GISS.



Os mesmos dados, gráfico diferente

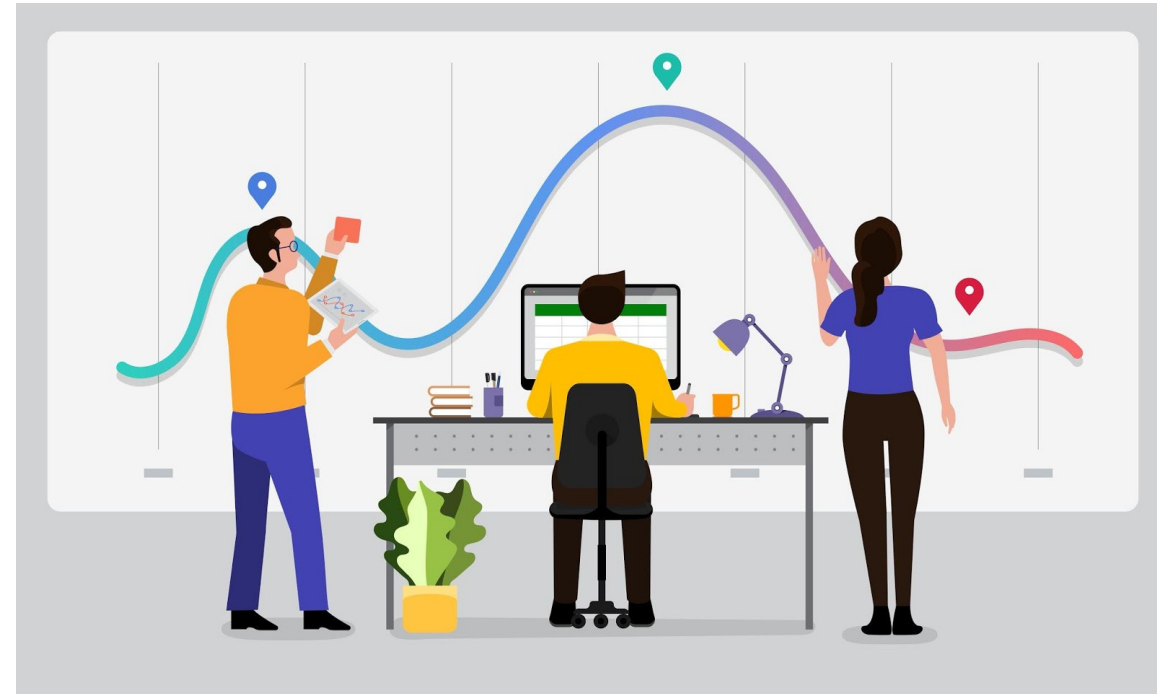
Average global temperature by year

Data from NASA/GISS.



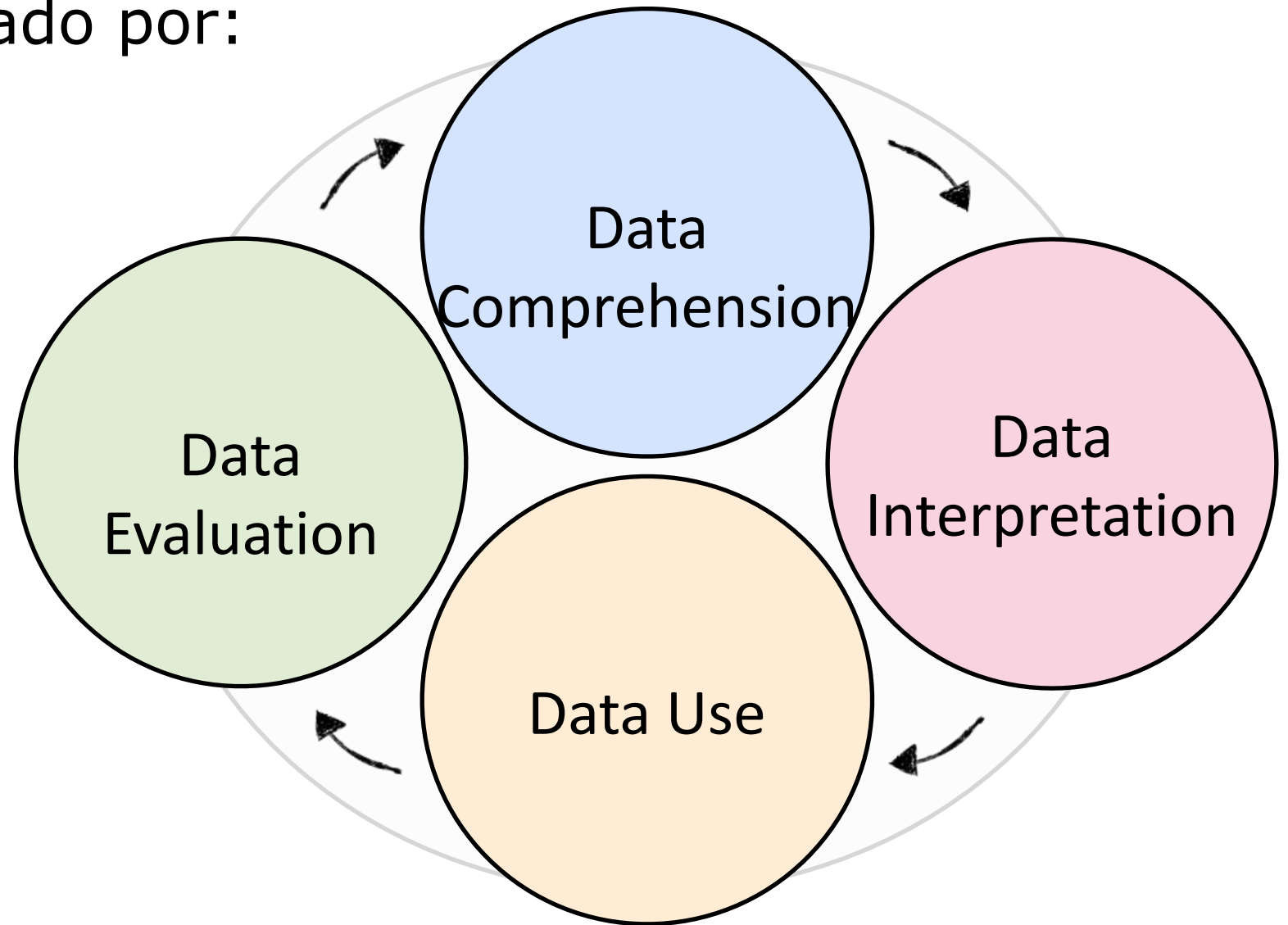
Data Literacy

- Proficiência em Dados
 - Capacidade de consumir dados para a geração de conhecimento, produzi-los de forma coerente e criticamente analisa-los



O processo de proficiência

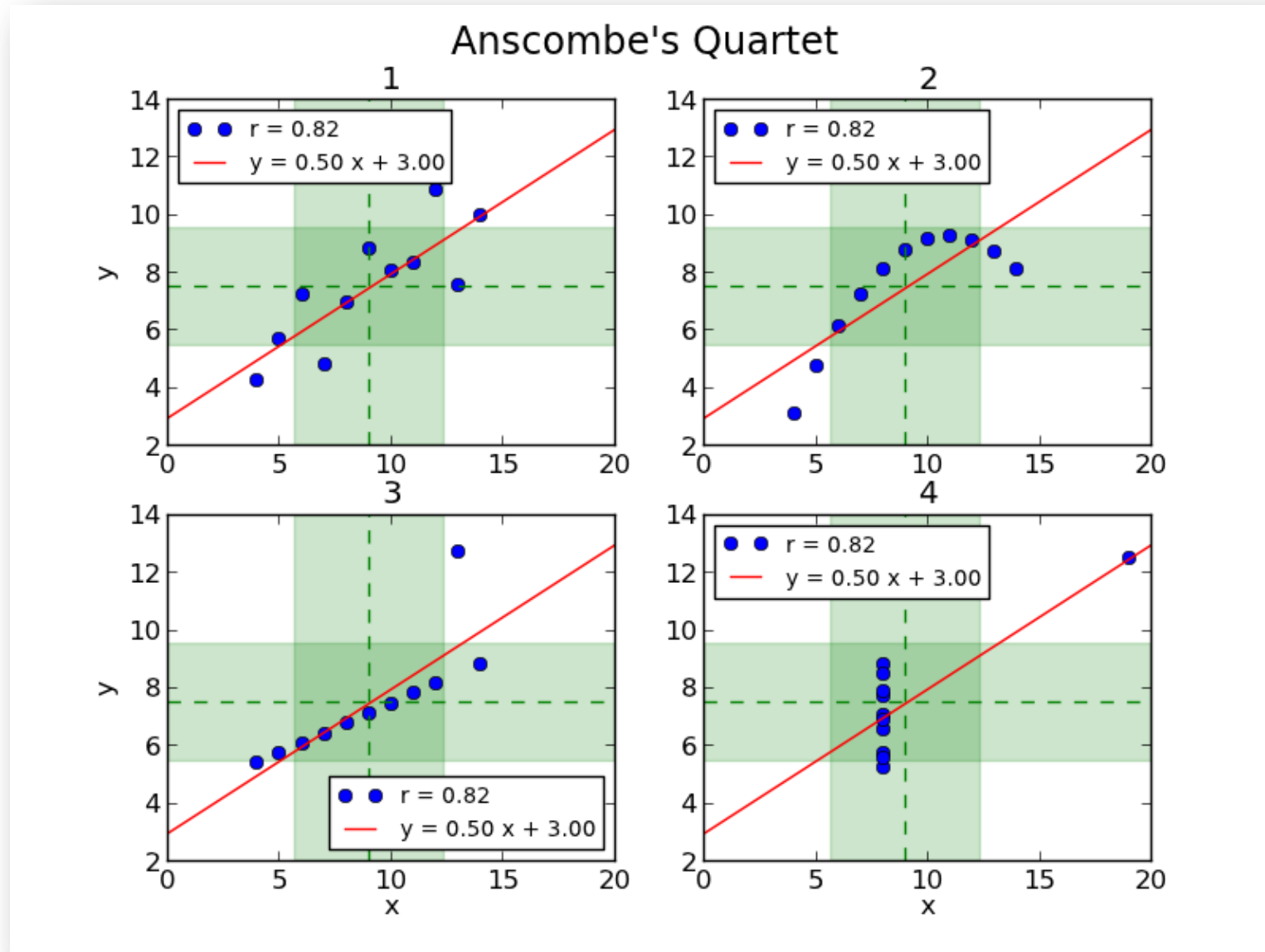
- Neste curso representado por:
 - Ler
 - Entender
 - Apresentar
 - Avaliar/Gestão de Projetos



Observe o conjunto de dados abaixo:

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

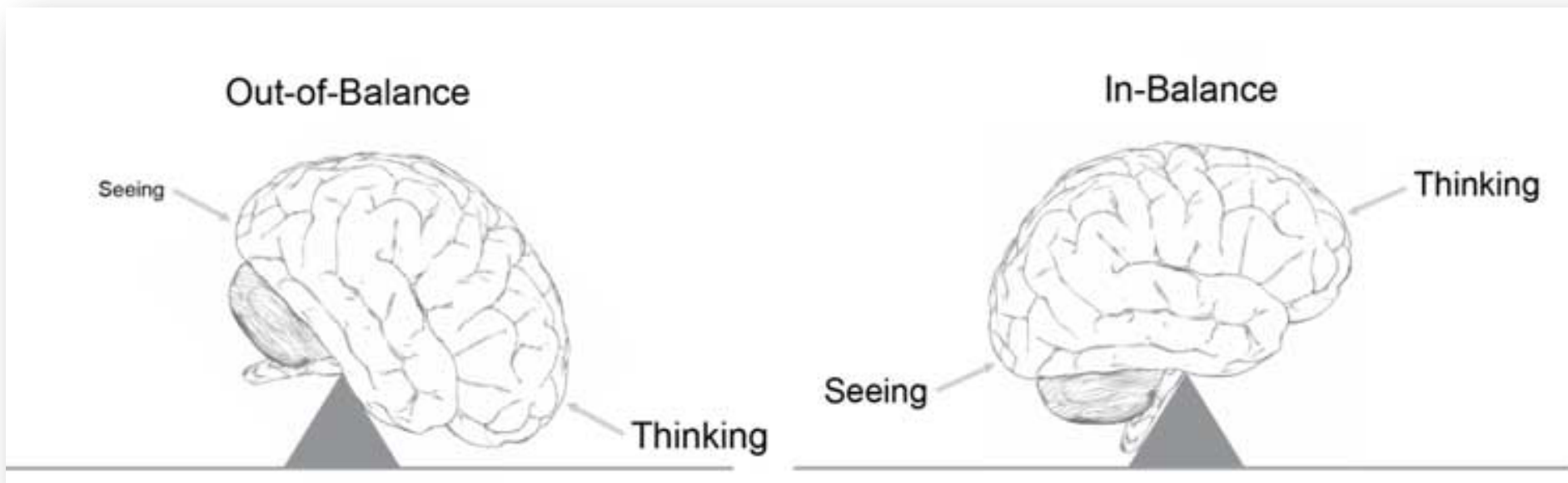
Agora visualize-os:



É preciso mais do que estatística...

- Métodos estatísticos isolados não apresentam uma quantidade de informação suficiente para seres humanos tomarem decisões → Precisamos de elementos de **Visualização**
- Algoritmos quando usados exclusivamente não conseguem codificar uma quantidade adequada de conhecimento humano sobre conceitos, fatos e contextos → Precisamos de elementos de **Interação**
- Elementos de visualização quando utilizados exclusivamente não podem gerenciar níveis de detalhes sobre os dados ou priorizar diferentes informações sobre os dados → Precisamos de mecanismos de **análise** e **interação**
- A interação direta com os dados, quando usada exclusivamente, não é viável em virtude do volume dos dados → Precisamos de mecanismos de **análise** e **visualização**

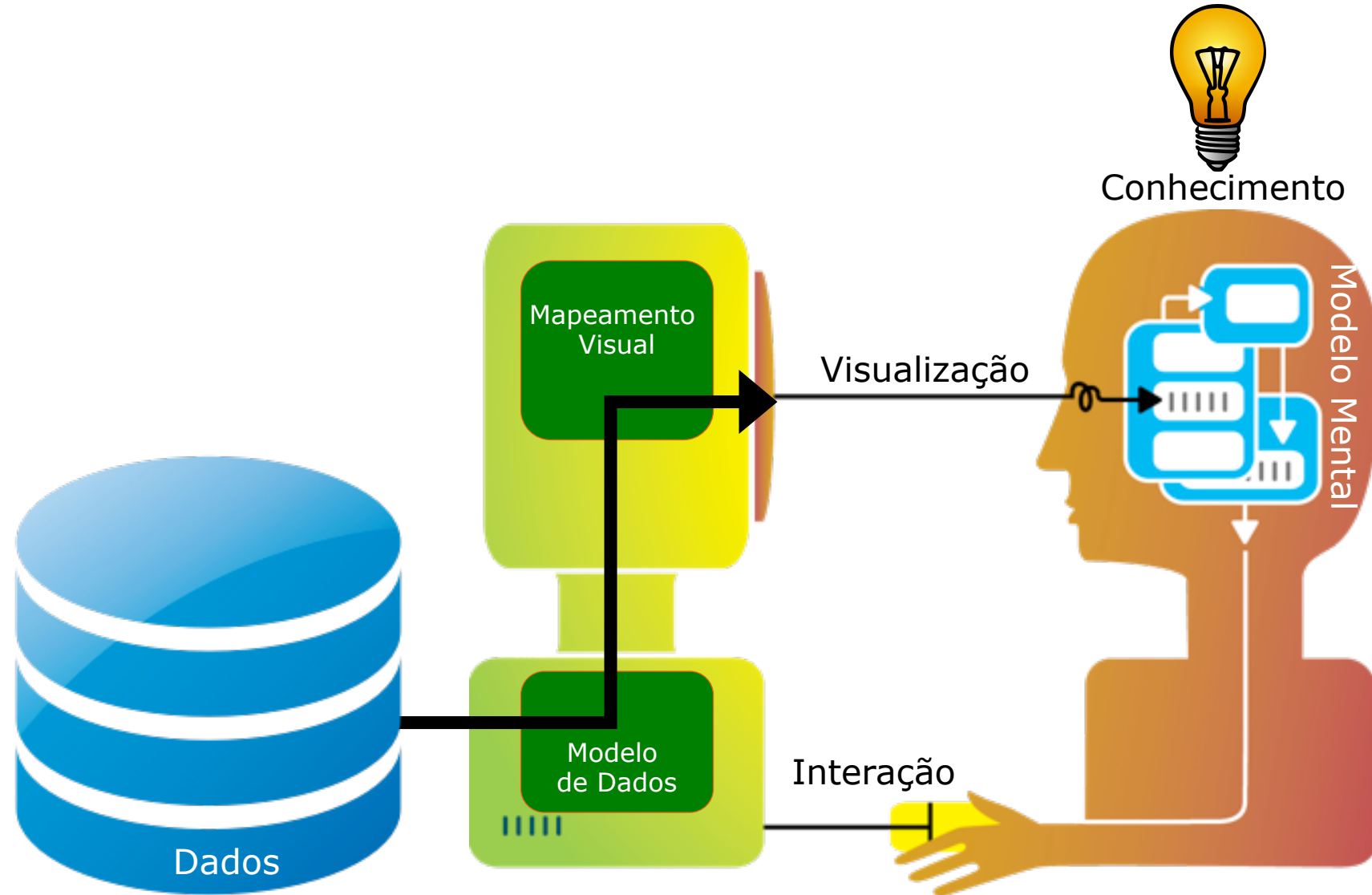
A importância da Visualização



“An estimated 50 percent of the brain's neurons are associated with vision. Visualisation <...> aims to put that neurological machinery to work.”

- B. McCormick, T. DeFanti, and M. Brown. Definition of Visualization. *ACM SIGGRAPH Computer Graphics*, 21(6), November 1987, p.3

Como desenvolvemos as habilidades de Data Literacy



Como trabalharemos estas habilidades nesse curso?

Ler

Quais dados de saúde estão disponíveis?

Como obtê-los?

O que são microdados?

Como ler microdados?

Compreender

Como se dá o processo de criação e transformação destes dados?

Há problemas com esses dados?

Como integrar e contextualizar tais dados com outras fontes/domínios do conhecimento?

Apresentar

Como apresento meus insights?

Qual é a maneira mais efetiva de comunicar os resultados?

Projetos

Como trabalho com equipes data-driven?

Como produzir actionable data?

Quais os desafios?

Aproveite o curso!

- Obrigado

- Contato:

- andrefmb@usp.br