



Hyrje në shkencën e të dhënave

Pjesa 1 - Hyrje

Prof. Asoc. Dr. Ermir Rogova

Datum, data, and science



Data = plural of datum, but we will use 'data' for both singular and plural versions



Information = meaningful data



Science = systematic study of the structure and behavior of the physical and natural world through observation and experimentation (Oxford English Dictionary)

What is Data Science?



A field of study and practice that involves collection, storage, and processing of data in order to derive important insights into a problem or a phenomenon.

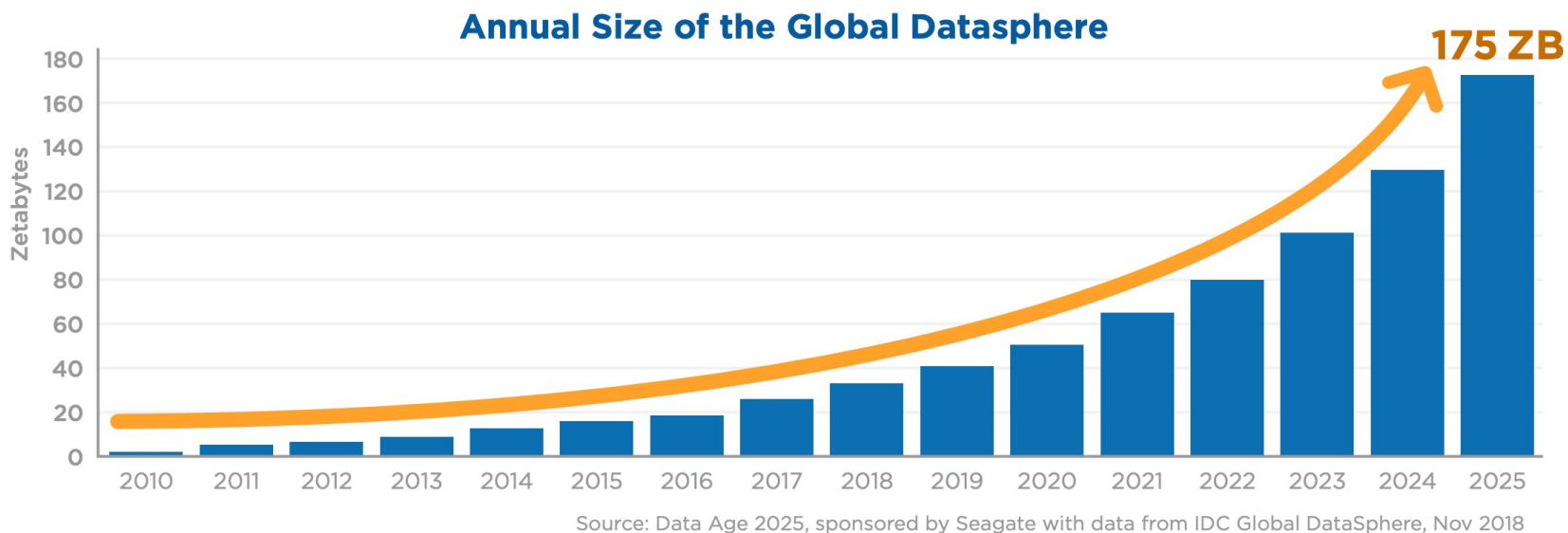


Data may be generated by **humans** (surveys, logs, etc.) or **machines** (weather data, road vision, etc.).



Data may be in different **formats** (text, audio, video, augmented or virtual reality, etc.).

Let's talk about data!



3V model of data



Velocity: The speed in which data is accumulated increased



Volume: The size and scope of the data increased



Variety: The massive array of data and types (structured and unstructured) increased

Data Science applications in different domains

- Finance
- Public policy
- Politics
- Healthcare
- Urban planning
- Education
- Libraries
- ...

Relation of Data Science with other fields



Statistics



Computer Science



Engineering



Business Analytics



(Computational) Social Science

What skills do we need for Data Science?



Willing to experiment

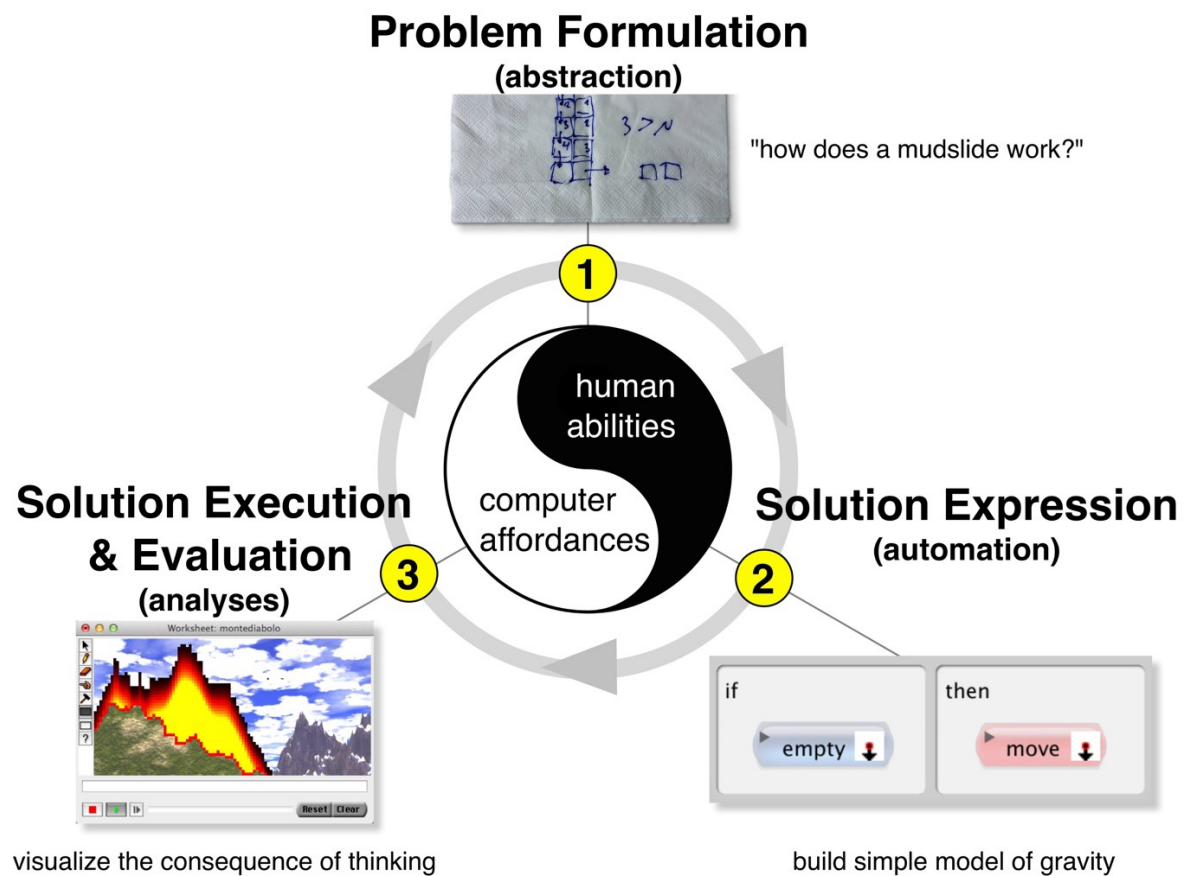


Proficiency in mathematical reasoning



Data literacy

Computational thinking



Source: WikiMedia: A. Repenning, A. Basawapatna, and N. Escherle, "Computational Thinking Tools"

Computational thinking examples

- Find the largest number:

7, 24, 62, 11, 4, 39, 42, 5, 97, 54

- Sort the numbers

Hands-on exercise

OBSERVATION	HEIGHT (INCHES)	WEIGHT (LBS)
1	58	115
2	59	117
3	60	120
4	61	123
5	62	126
6	63	129
7	64	132
8	65	135
9	66	139
10	67	142
11	68	146
12	69	150
13	70	154
14	71	159
15	72	164

1. How much increase can we expect in weight with an increase of one inch in height?
2. What would you expect the weight to be an American woman who is 57” tall?
3. What would you expect the weight of someone who is 73” tall?

Summary

- Data science is the field of study and practice that involves collection, storage, and processing of data in order to derive important insights into a problem or a phenomenon.
- It calls for computational thinking and understanding of statistics.
- Any tool that can represent and process data can be useful for doing data science, including spreadsheet programs (e.g., Excel), databases (e.g., MySQL), Python, R, and even the back of an envelop!

Temat e javës së dytë

- Data types
- Data collections
- Data pre-processing



Pyetje???