

Data 311-Lecture 1: Properties of Data

Data Science Landscape

Properties of Data

Unstructured data

raw video stream

text of webpage

Structured data

list

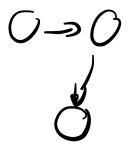
table

dictionary (map)

graph

tree

time series



Numerical vs Categorical

Temperature

Sky conditions

Wind direction (N, NE, ...)

Big vs Small

Data Types

integers

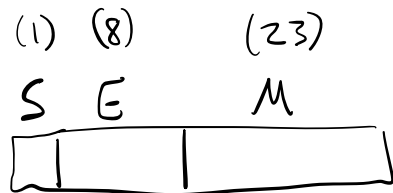
- signed vs unsigned

32, 64 (8)

↙ ↓
-2b to 2b -9q to +9q

floating-point

- float{32, 64}



$$1.M \times 2^{(E-127)}$$

$$0.1 + 0.2 \neq 0.3$$

str

object (pandas - wrapper around whatever)