Module Overview

- Scale of Measurements
- Data Structures
- Data Storage

Book Keeping

- Ongoing search for TA

- Still haven't done work reviews 🤦🏾‍♂️

- Apologies for the late work products

- Gentle reminder on your final project (Wk4/16)

Outcome

After this Module, you will;

- Understand the formal concepts and measures of describing data

- Understand how to technically represent and structure your data

- Get an overview of the evolving technologies for data storage

Scale of Measurements

- Nominal Scale

- Ordinal Scale

- Interval Scale

- Ranked Scale

> "Datafication is a technological trend turning many aspects of our life into data which is subsequently transferred into information realised as a new form of value"
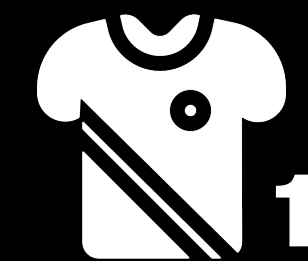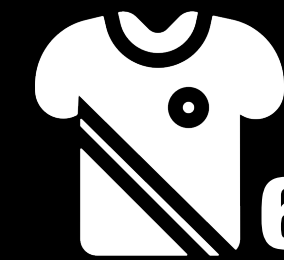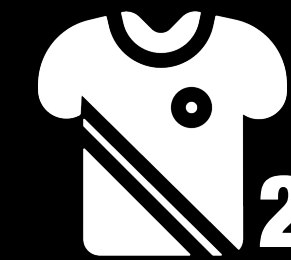
Nominal Scale

- This represents the assignment of numerals, words or letters as labels.

- Determination of equality

- Permutation group $x' = f(x)\, f(x)$ means any one-to-one substitution

- Number of cases, Mode & Contingency correlation
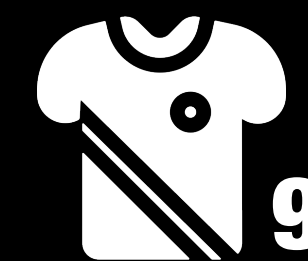
Liverpool Formation 1-4-3-3

Goal-keeper   1

Defenders   2   5   6   3

Midfielders   4   10   8

7   11

Attackers   9

Ordinal Scale

- This represents the assignment of numerals, words or letters as label based on the rule to preserve rank-ordering.

- Determination of greater or less

- Isotonic group $x' = f(x)\ f(x)$ means any monotonic increasing function

- Median & Percentile

Large
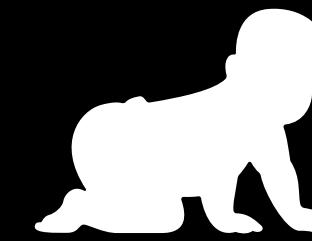
Regular

Small

Interval Scale

- This represents the quantitative representation based on the rule of relationships without the knowledge of a true zero point.

- Determination of equality of intervals or differences

- General linear group $x' = ax + b$

- Mean, Standard deviation, Rank-order correlation & Product-moment correlation

Toddler (1 yrs)

Boy (18 yrs)

Man (35 yrs)
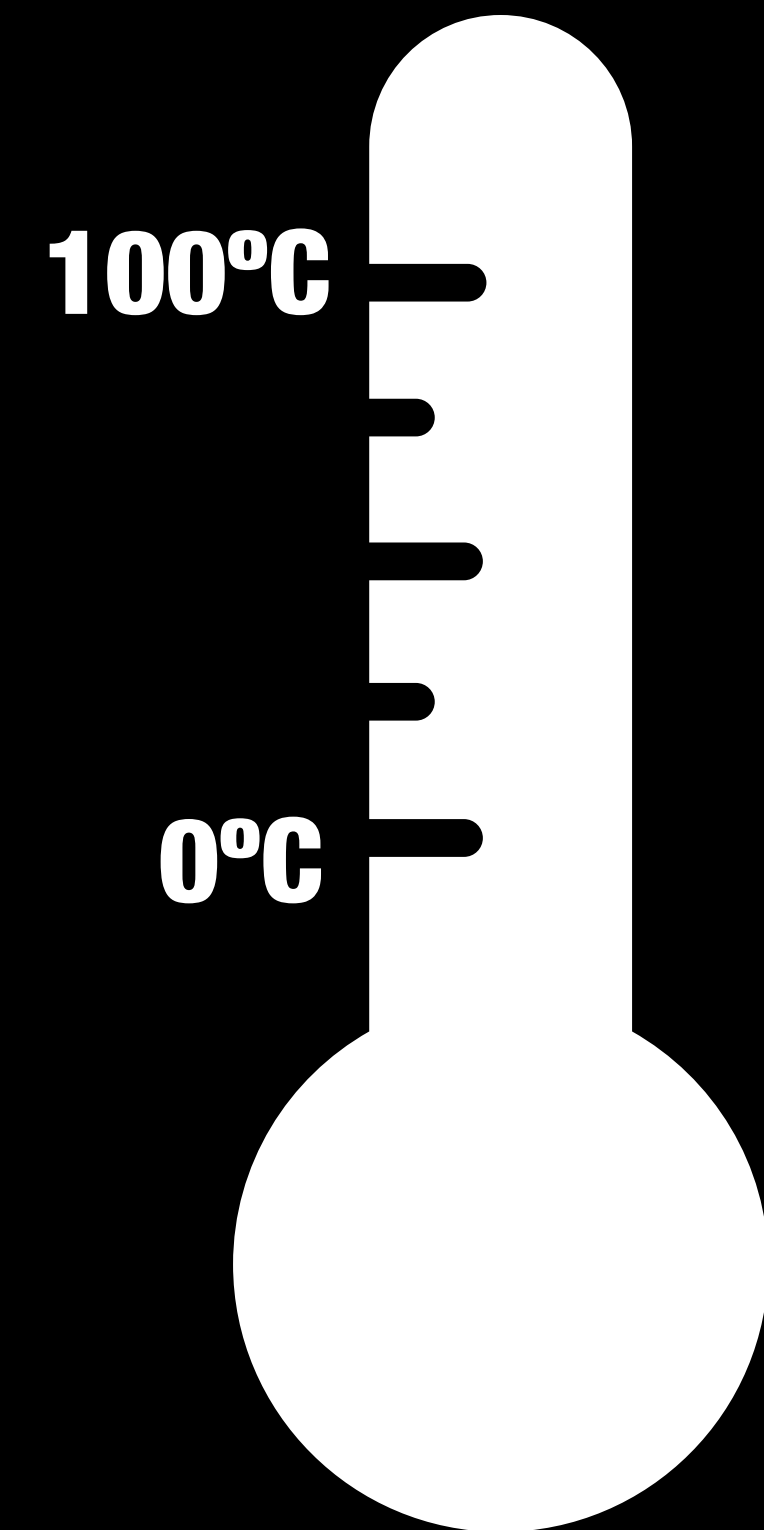
Old man (85 yrs)

Ratio Scale

- This represent the quantitive representation based on the rule of relationships with the knowledge that an absolute zero is always implied.

- Determination of equality of ratios

- Similarity group $x' = ax$

- Coefficient of variation

100°C

0°C

Noteworthy Quote

"

Any particular scale used for arbitrary or explicit measurement by mortal is not perfectly free of their taint.

S. S. Stevens

"

Data Structure

Primitive

- Boolean
- Character
- Integer
- Floating-point

Abstract

- Array
- List
- Dictionary

Boolean

" This is a data type that has one of two possible values intended to represent the two truth values of logic based Boolean algebra. "

$[true, false]$

$[1,0]$

$[yes, no]$

Character

> This is a data type that corresponds to symbol such as alphabets, or syllabary in the written form of natural language.

$[hello, goodmorning, stay, bye]$

$[lagos, ogun, edo, kaduna, anambra]$

$[developer, designer, scientist, manager]$

Integer

"

This is a data type that represents some range of mathematical whole numbers that signed $(-,0,+)$ or unsigned $(0,+)$

"

$$[1,5,10,16,21,30,40,55,65,80]$$

$$[-5,-4,-3,-2,-1,1,2,3,4,5]$$

$$[150,156,152,159,153,155,156]$$

Floating-point

" This is a data type that depicts formulaic representation of real numbers as approximation to support the trade-off between range and precision. "

[12.4,22.6,30.5,79.6,87.5,99.9]

Array

> This is a collection of primitive data structure stored at contiguous memory location that represents the storage of multiple values of the same data type together.

$$[1,0,0,1,1,0,1,0]$$

$$[lagos, ogun, edo, kaduna, anambra]$$

$$[150,156,152,159,153,155,156]$$

List

> This is a collection of primitive data structure and array but not stored at contiguous memory location.

$$[1,0,0,[1,1],0,1,0]$$

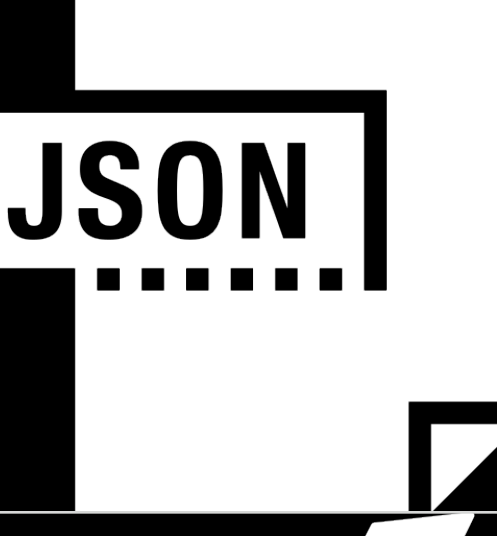$$[[lagos, ogun, edo], [kaduna, kastina, Borno], [anambra, abia, ebonyi]]$$

Dictionary

"
This is a general-purpose data structure for storing a group of objects. A dictionary has a set of keys and each key has a single associated value.
"

{'Detra' : 17, Nova' : 84, Charlie' : 22, Henry' : 75, Roxanne' : 92, Elsa' : 29}

Data Storage Technologies

	Flat Files (.CSV)	A CSV is a comma-separated values file, which allows data to be saved in a tabular format. CSVs look like a garden-variety spreadsheet but with a .csv extension.
	RDMS	A relational database is a collection of tables, each of which is assigned a unique name that uses Structured Query Language (SQL) statements to query and maintain.
	NoSQL Database (.JSON)	NoSQL databases are schema agnostic and provide the flexibility needed to store and manipulate large volumes of unstructured and semi-structured data.
	Data Warehouse	A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and usually residing at a single site.

Practice Lab

Understand the basic of data measurement, structure and storage

Use the following Instructions:

- Make a list of human and physical interaction that can measure.
- Create an array to collect some data values for each of the interactions identified above.
- Extend your work so far to create a dataframe (if your using R) or dictionary (if you are using Python)
- Save your table to a flat file stored in course directory (SGA07_DATASCI) as data.csv

Recap/Summary

At the end of this Module, you should understand;

- Introduce you to various scales of measurements

- Provide some examples of data structure

- Give an overview of data storage technologies

- Practical implementation of initial stage of data preprocessing

Suggested Material

- On the theory of Scales of Measurement by S.S. Stevens, 1974 as part of the Journal of Science

- https://en.wikipedia.org/wiki/Data_structure

- https://en.wikipedia.org/wiki/List_of_data_structures

- https://towardsdatascience.com/everything-a-data-scientist-should-know-about-data-management-6877788c6a42

- Data Mining Concepts and Techniques by Morgan Kaufmann 3rd Edition (Chapter 1: Introduction to Data Mining and Database Technologies)