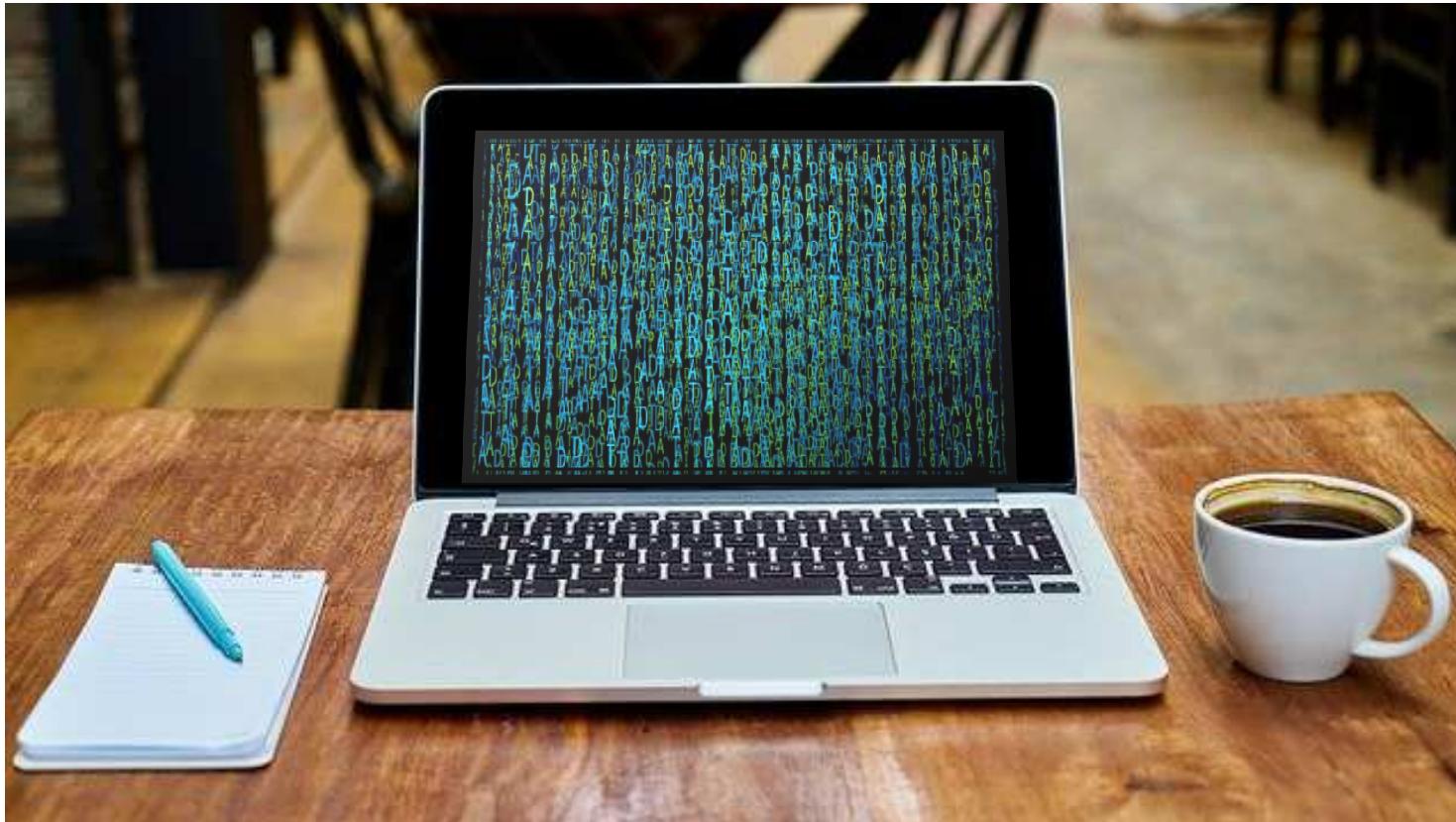


Wk 1. OT & Intro Data Science



a/b Instructor

- ▶ **Name:** Sam Han (3.1? => Sang Jin Han)
- ▶ **Education:**
 - ▶ BS. in Computer Science
 - ▶ MS. in Bio-Medical Informatics (Informatics + Bio + Medicine)
 - ▶ Ph.D in Bio-Medical Engineering (Vision Research)
- ▶ **Work Experience:**
 - ▶ Web DB Programmer (Korea)
 - ▶ Data Manager (NJ Med. School)
 - ▶ TA,RA (NJIT)
 - ▶ Associate Professor (MIU, Mongolia)
 - ▶ Software Developer/Entrepreneur - School ERP
 - ▶ Visiting Scholar (BITL), Lecturer @ Dankook Univ.



2
Data Science

- ▶ **Interests**
 - ▶ Informatics, ML
 - ▶ Regional Development
 - ▶ Internet Addiction

- ▶ **Hobbies**
 - ▶ Bicycle
 - ▶ Guitar
 - ▶ JS Game Programming
 - ▶ Memorizing Proverbs

Syllabus (1)

Week	Date	Topic	Chapter	Activity
1	2-Sep	OT, Intro to Data Science	Ch1-1, Ch 4-1	
2	9-Sep	Using R studio, Big Data	Ch1-2,3, Ch 2	Install R & R-Studio
3	16-Sep	Variable, Vector, and Function, Open Data	Ch3	HW 1
4	23-Sep	Control / Loop statements and User defined function	Ch7	HW 2
5	30-Sep	Manipulations of vector	Ch4-2,3	HW 3
6	7-Oct	Matrix and Data Frame	Ch5	Quiz 1
7	14-Oct	Input/Output of data	Ch6	
8	21-Oct	Midterm Exam		Exam



Syllabus (2)

Data Science Syllabus (Updated) - by Prof. Sang Jin Han 2021 Fall				
Week	Date	Topic	Chapter	Activity
9	28-Oct	Basic Graph I, DS Ethics	Ch8	
10	4-Nov	Basic Graph II	Ch9	HW 4
11	11-Nov	Data Exploration	Ch10	HW 5
12	18-Nov	Data Preprocessing	Ch11	Quiz 2
13	25-Nov	Advanced Graph	Ch12	
14	2-Dec	Case Study	Ch13	HW 6
15	9-Dec	Final Exam		Exam



Grading Policy

Methods of Grading

sequence	Description	Percentage
1	Mid-tem Exam	25%
2	Final-exam	30%
3	Pop Quizzes	10%
4	Assignments	25%
5	Reports	0%
6	Presentations & Discussions	0%
7	Attendance	10%
8		0%
9	Others	0%
All		100%

5
Data Science

Lectures

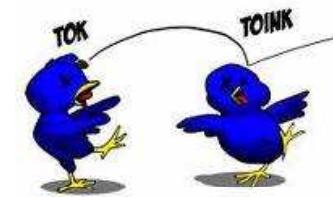
- ▶ Start with zoom
- ▶ Do f2f whenever possible
- ▶ Exams will be done by f2f

6
Data Science

Grade or Learning?

► GRADE-ORIENTED
APPROACH

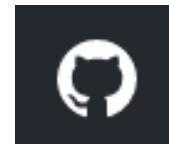
VS.



► LEARNING-ORIENTED
APPROACH

Secret Recipe for “Catching 2 Birds”

- ▶ Do Uni-Tasking
- ▶ Active Learning
- ▶ Ask Questions
- ▶ Make a note d/r class
- ▶ Do Journaling/Blogging
(eg, GitHub.io, Notion)



Secret Recipe for “Catching 2 Birds”

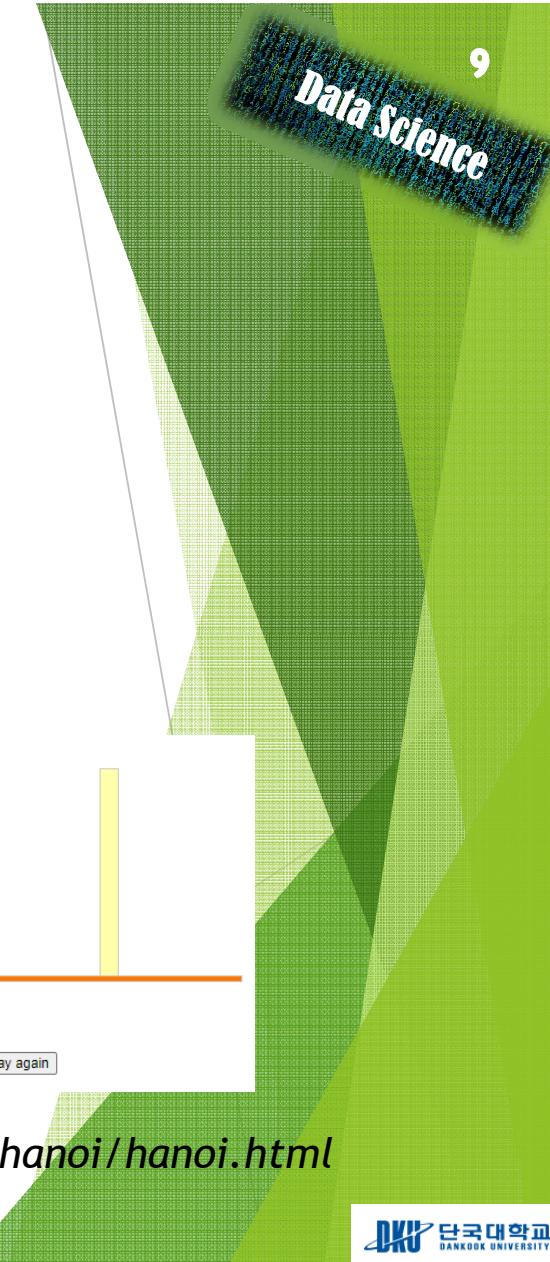
▶ Terminologies

▶ Make Games and Apps

“Use or Loose it”

The screenshot shows a web browser window with the URL wise4edu.com/init/. The page title is "Welcome to Magic Initials". A red box highlights the "Tutorial Videos" button. Below it is a language dropdown set to "EN". The main content area displays the Tower of Hanoi game with three yellow rods. The left rod has three gray disks stacked from bottom to top. The middle and right rods are empty. At the bottom, there are input fields for "Number of disks" (set to 3) and "# move" (set to 0), along with "Solve" and "Play again" buttons.

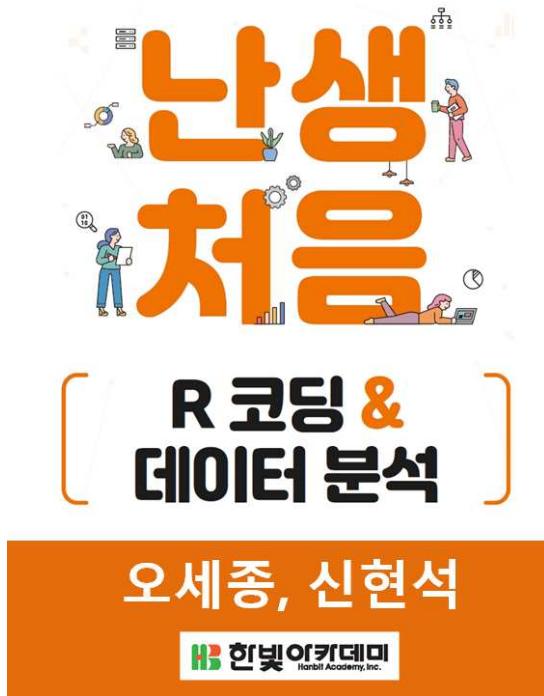
<https://wise4edu.com/jsgame/hanoi/hanoi.html>



Can you explain DS KEY Terminologies?

- ▶ Discrete vs. Continuous
- ▶ Accurate vs. Precise
- ▶ FP, FN
- ▶ H_0 vs. H_A
- ▶ Top down vs. Bottom up
- ▶ Trade-off

Textbook & Materials



+ Wiki + Youtube + Others

What is Data Science?

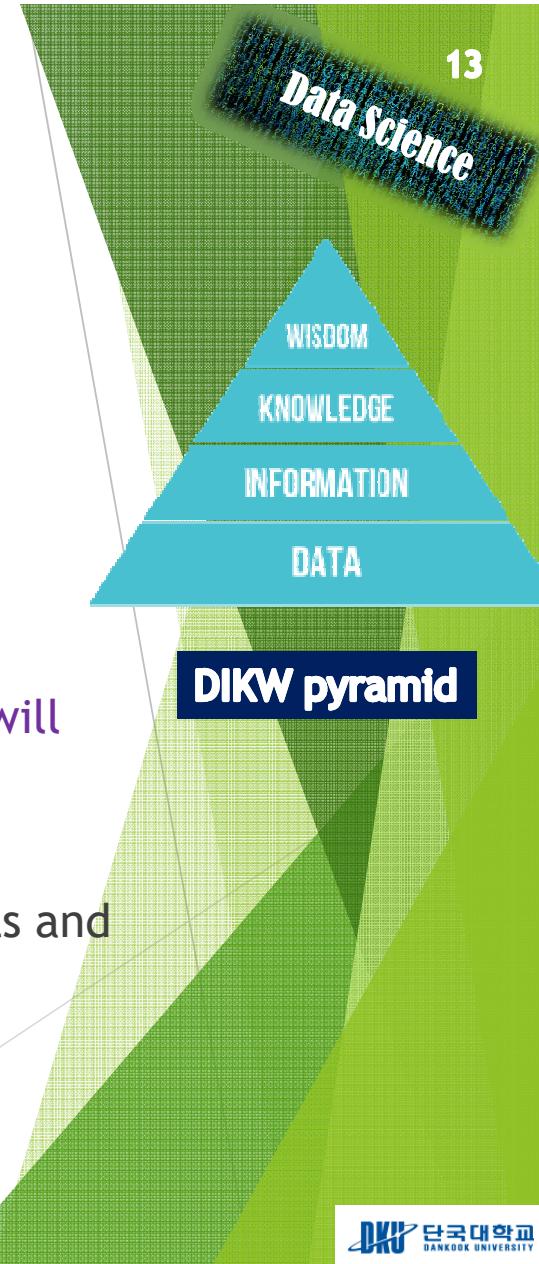
- ▶ IPO (?)
- ▶ Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to data mining, machine learning and **Big Data**. (Wiki)



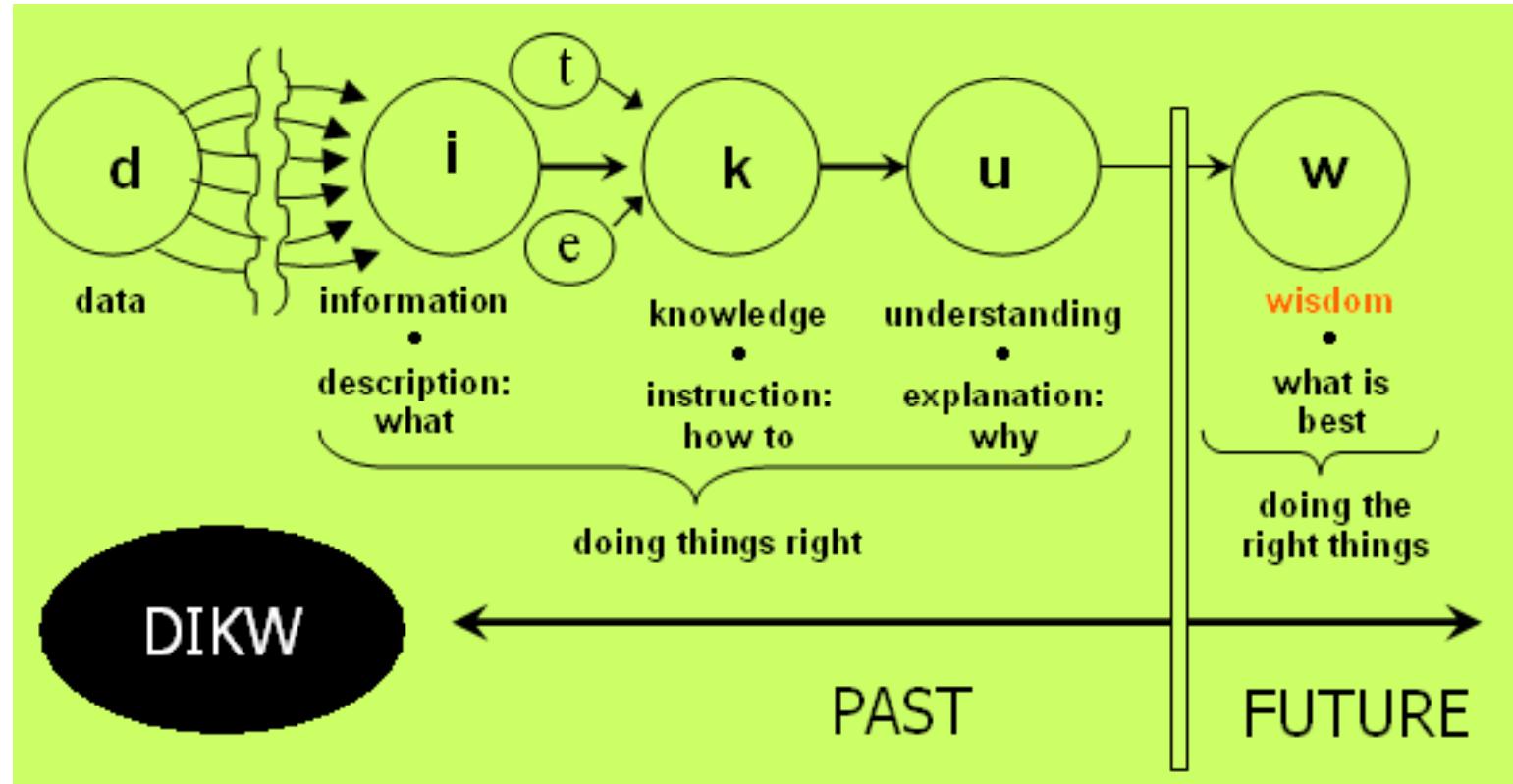
DIKW

1. **Data** : symbols, signs, signal (**Objective**)
e.g., 14:00, -15 °C, 17:00, -20 °C
2. **Information**: can answer “who”, “what”, “where”, “how many”, “when”, “organized or structured data” (**Objective**)
e.g., the temperature is -20 °C and it is going down.
3. **Knowledge**: “organized or structured information”, information + experience, inter-connected information (**Subjective**)
e.g., The temperature will become -30 °C tomorrow morning. There will be heavy smoke on the air.
4. **Wisdom**: “integrated knowledge--information made super-useful”, “knowing the right things to do”, “the ability to make sound judgments and decisions” (**Subjective**)
e.g., Tomorrow morning I need to wear Ugg boots and a mask.

(Wikipedia: DIKW)



Information Hierarchy

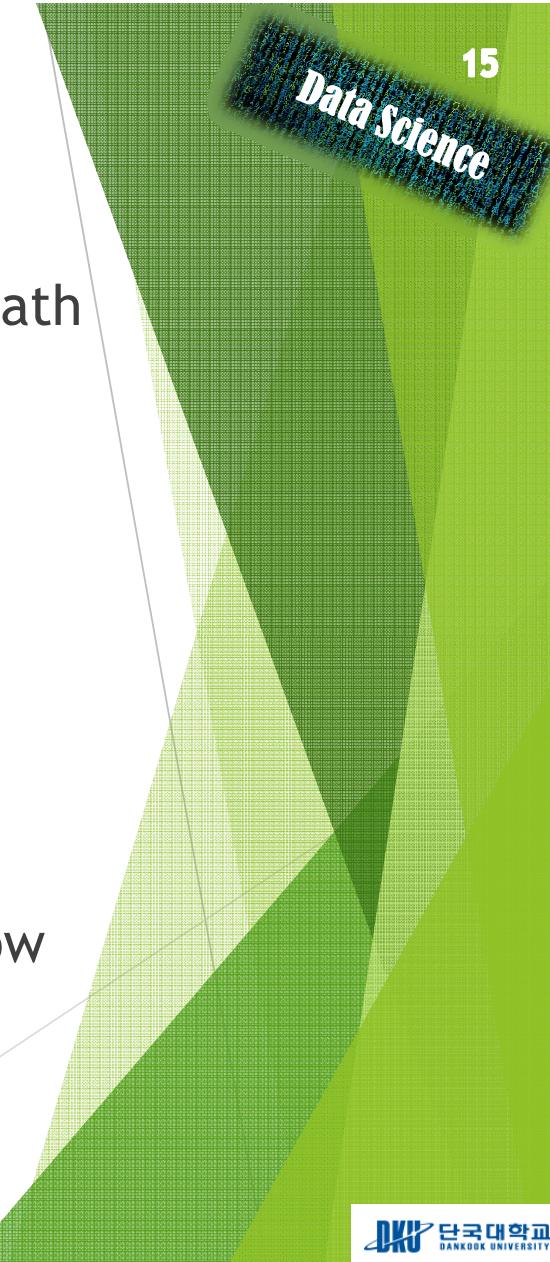


<http://en.wikipedia.org/wiki/DIKW>

What do I need to learn in DS?

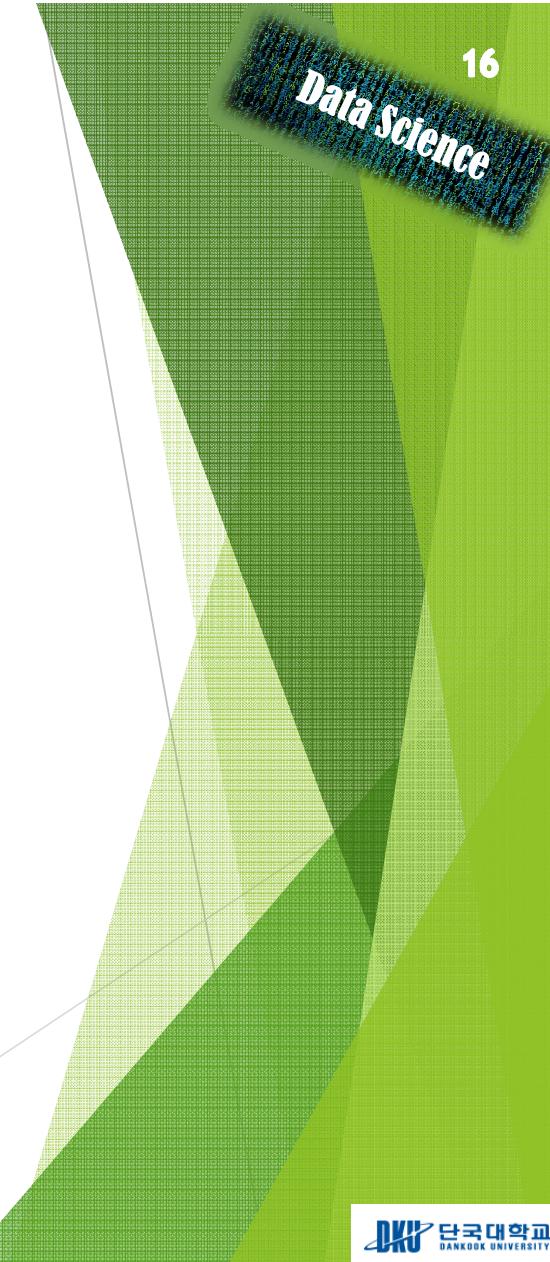
- ▶ **Math:** Statistics & Probability, Linear Algebra, Discrete Math
- ▶ **Computer Science:**
 - ▶ Data Structures and Algorithms
 - ▶ Machine Learning
 - ▶ Coding Skills: Excel, R, Python, SQL, Hadoop, Tensorflow, ...
- ▶ **Domain Knowledge**

eg, To make a good Glaucoma model, we need to learn how doctors diagnose Glaucoma
- ▶ **Communication Skills**



How to become a good D.S.tist (1)

- ▶ Start as “Nerd” and end as “History Maker”
- ▶ Patience, Challenging to learn new things,
Enthusiasm (“No pain, No gain”)
- ▶ Understand Society and Business
- ▶ Learn Ethics
- ▶ Read Good (Text)Books



Programming 101 (Ch1-1)

- ▶ Why do I need to learn Computer Programming?
 - ▶ Tool
 - ▶ Computational Thinking
- ▶ Programming vs. Programming Languages
 - ▶ Algorithms

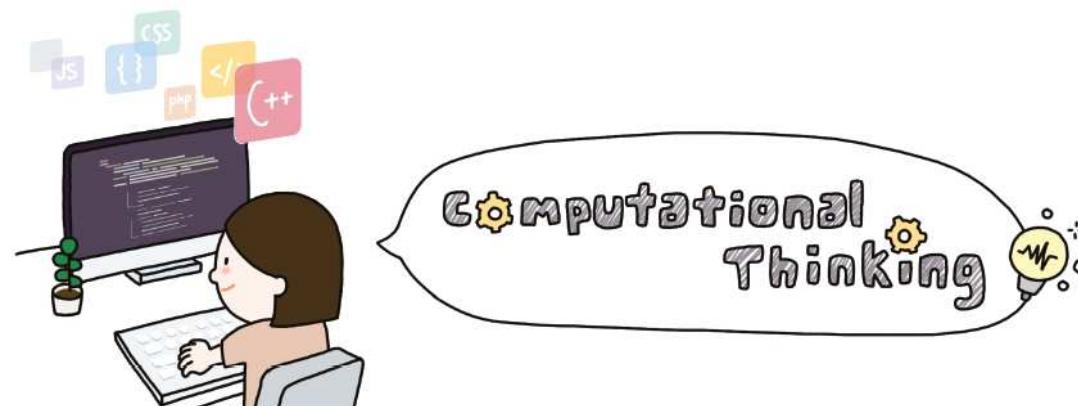


그림 1-1 컴퓨팅 사고를 통한 프로그래밍

Programming 101 (Ch1-1)

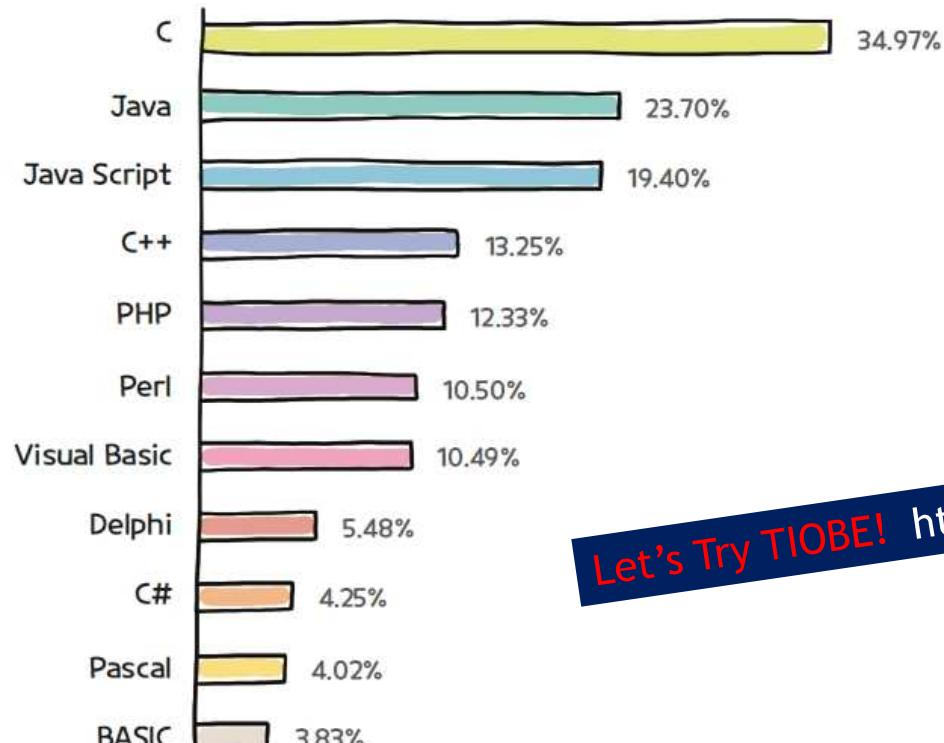


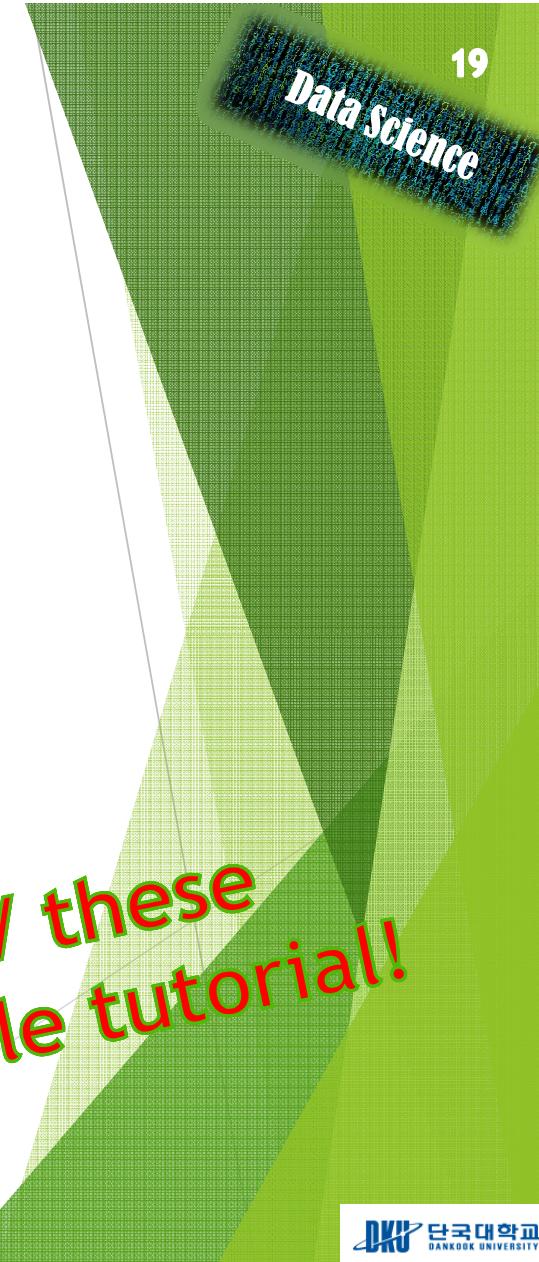
그림 1-6 프로그래밍 언어 ©Popular programming languages 1965–2019

Let's Try TIOBE! <https://www.tiobe.com/tiobe-index/>

Programming 101

- ▶ Variable vs. Constant & Scopes
- ▶ Data Types:
 - ▶ Int, Float, T/F, Array, Object
- ▶ If,else : _____ statement?
(Boolean Algebra: **Truth Table**)
- ▶ for, while: _____?
- ▶ function: _____

If you are not sure w/ these topics, watch a simple tutorial!



Types of Data (Ch4-1): 1D, 2D

▶ 1D Data:



Fig 4-4 Population of Cities/Provinces

R: Vector, List, Factor

▶ 2D Data:

도시	전체인구	도시지역면적	농림지역면적	택시업체수
서울	972	606	0	255
부산	341	940	0	97
대구	243	797	37	92
인천	295	580	263	60
광주	145	481	4	76
대전	147	495	28	76
울산	114	755	283	43
세종	34	140	148	5

Fig 4-5 City Statistics

R: Matrix, Data Frame

Categorical vs. Numerical

Categorical data is the statistical data type consisting of **categorical variables** (qualitative variables) or of data that has been converted into that form.. (Wiki)
(Cannot do Numeric Operations)

법안 찬성여부	YES, YES, NO, YES, NO, NO, YES, ...
선호하는 색상	blue, red, orange, blue, green, red, ...
음식 맛 평가	good, bad, bad, good, good, bad, ...
성별	M, F, F, M, M, F, F, F, M,

Fig 4-6 Categorical Data Example

Nominal vs. Ordinal ?

Numerical data is the data type consisting of **numeric variables** (quantitative variables) with a sense of size (Can do Numeric Operations)

키	169, 173, 158, 185, 169, 176, 181, ...
시력	1.5, 1.2, 0.7, 0.6, 0.1, 1.3, 2.0 ...
온도	10, 14, 26, -5, -20, 16, 31, -10, ...

Fig 4-6 Numerical Data Example

Discrete vs. Continuous ?

Installing R & R-Studio

- ▶ Installing R: <https://www.r-project.org/>
- ▶ Installing R-Studio: <https://www.rstudio.com/>

The screenshot shows the homepage of the R Project for Statistical Computing. At the top left is the R logo. To its right are links for [Home], Download, CRAN, R Project, and About R. The main title "The R Project for Statistical Computing" is centered above a "Getting Started" section. Below this, a paragraph explains that R is a free software environment for statistical computing and graphics, compiled and running on various platforms. It includes a link to download R from CRAN mirrors. Further down, another paragraph provides information about R's license terms and a link to frequently asked questions. A banner at the bottom of the page announces the "RStudio R&R Holiday" from Thursday, September 2 to Monday, September 6, 2021, with a "LEARN MORE" button.

Next Week Topics

► Using R studio
(Ch1-2,3, Ch2)

► Big Data

To-dos for Next Class (Flipped Learning)

- ▶ **Install R & R-Studio**
- ▶ Try some lines of code
- ▶ Learn ahead if you will

- ▶ **Get Textbook**
- ▶ Read ch1 and 2

24
Data Science