

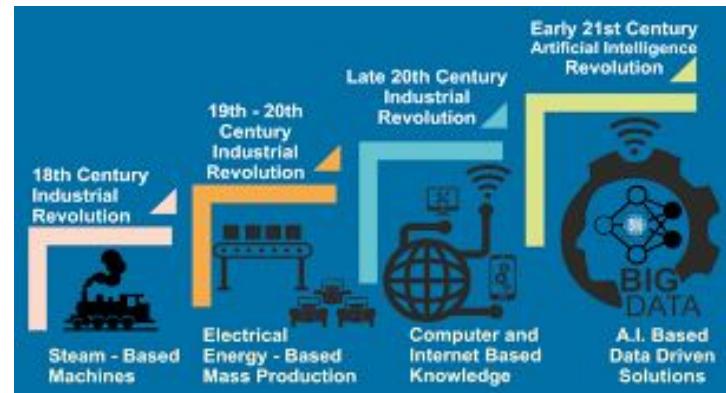
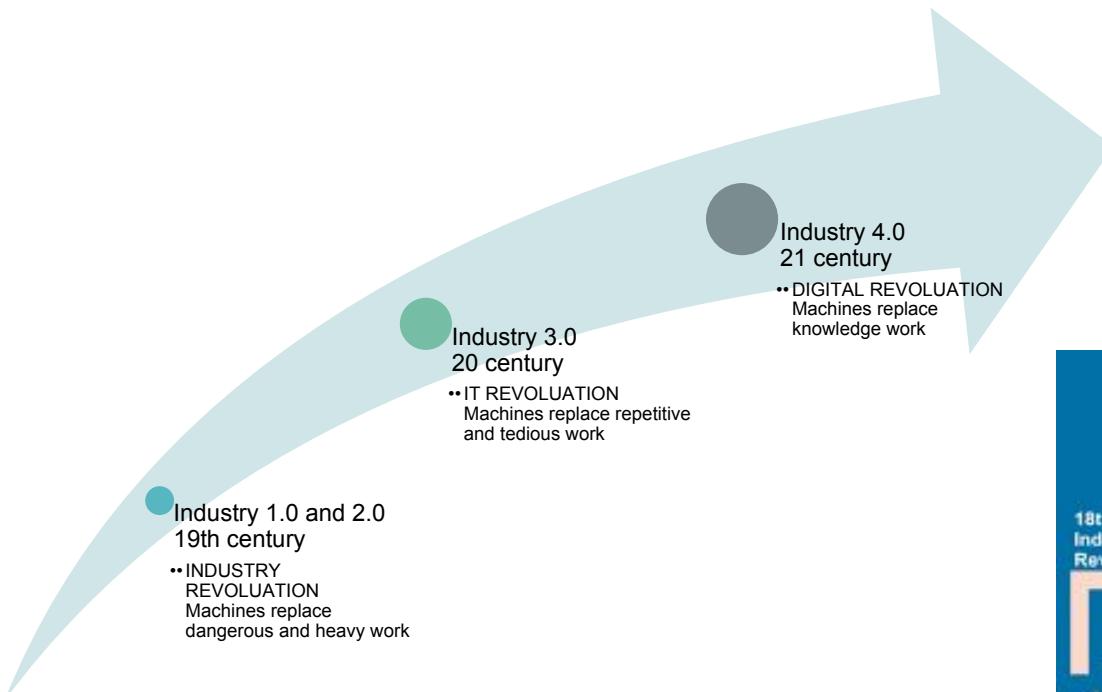
# DAY 1

Tableau

# Topics

- AI/Big Data
- Database
- Tableau

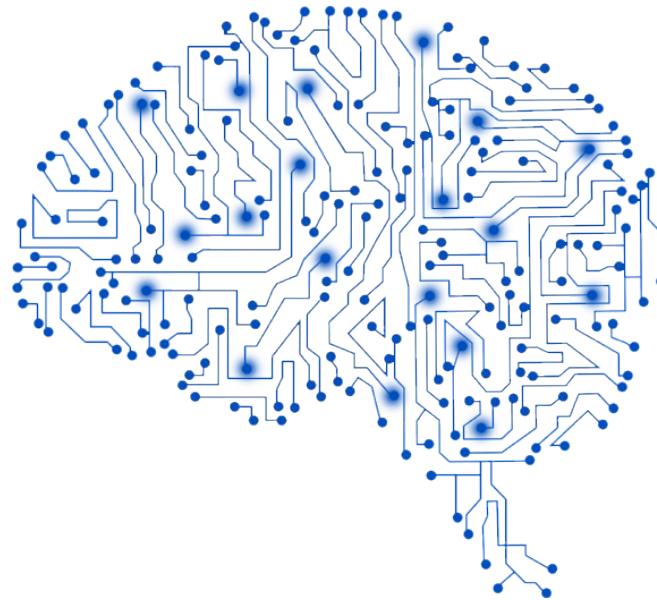
# The 4th Industrial Revolution



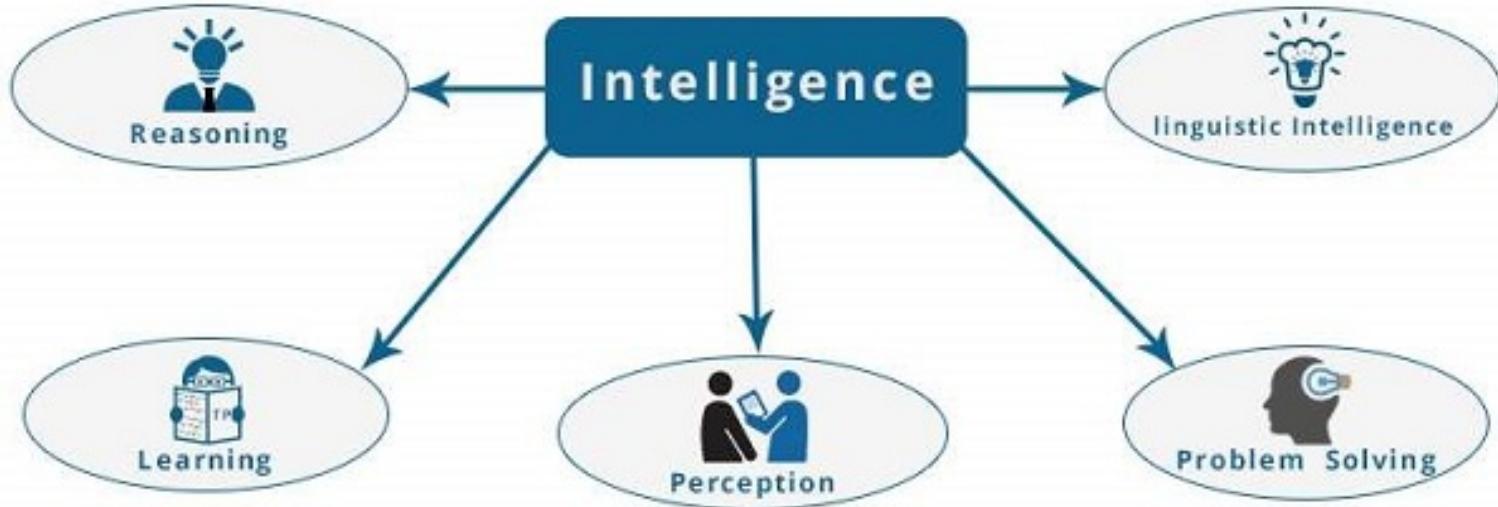
# Artificial Intelligence

# Artificial Intelligence

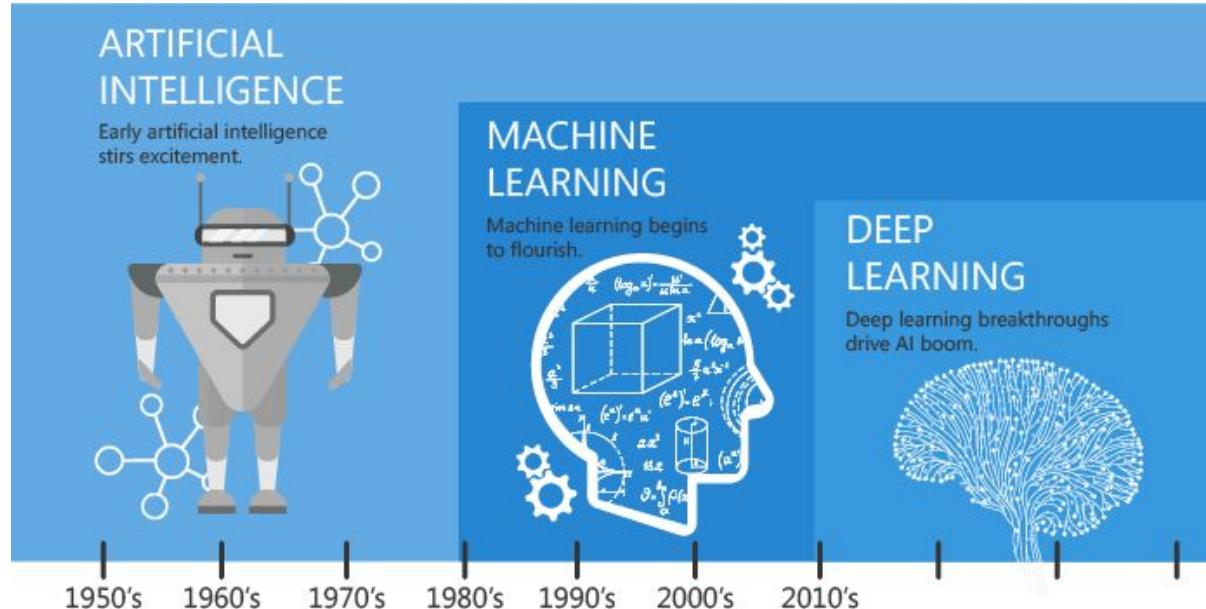
- The science of making machines imitate human thinking and behavior, can replace human decision making in some instances
- 기계가 인간의 생각, 행동을 모방하여 어느정도 인간의 의사결정을 대체할수 있게하는 과학



# Human Intelligence



# 인공지능의 발전



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

# Artificial Intelligence

## Artificial Intelligence

Any technique which enables computers to mimic human behavior.

컴퓨터가 사람의 행동을 흉내내는 모든 기술들

Fuzzy logic  
Genetic algorithm  
Expert system

## Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

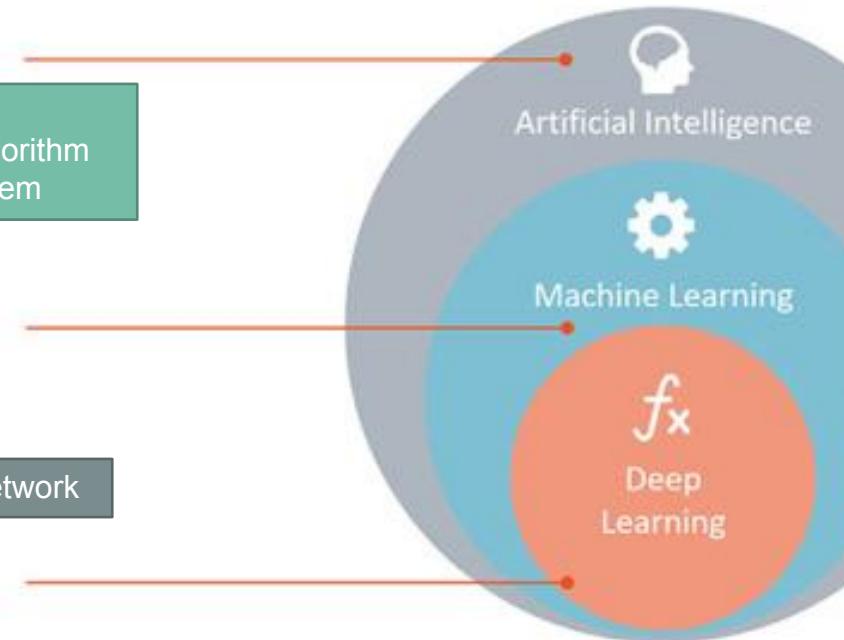
통계적 방법을 이용하여 기계가 학습하게 하는 인공지능 기술

Neural network

## Deep Learning

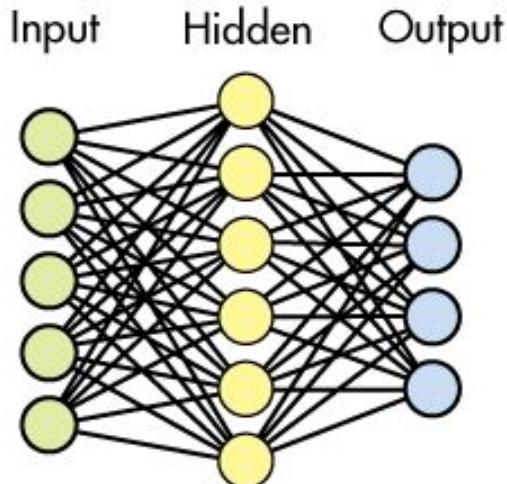
Subset of ML which make the computation of multi-layer neural networks feasible.

다중층 신경망을 이용하는 머신러닝기술

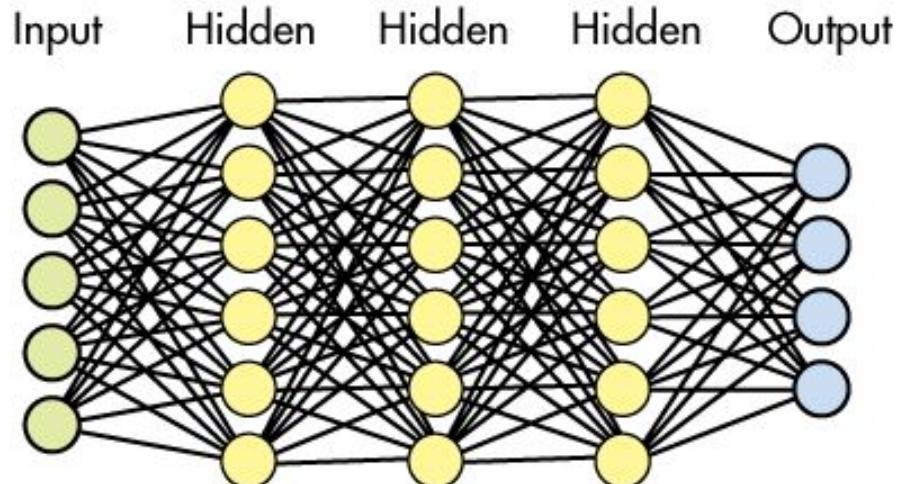


# Deep Learning using Neural Network

Neural network



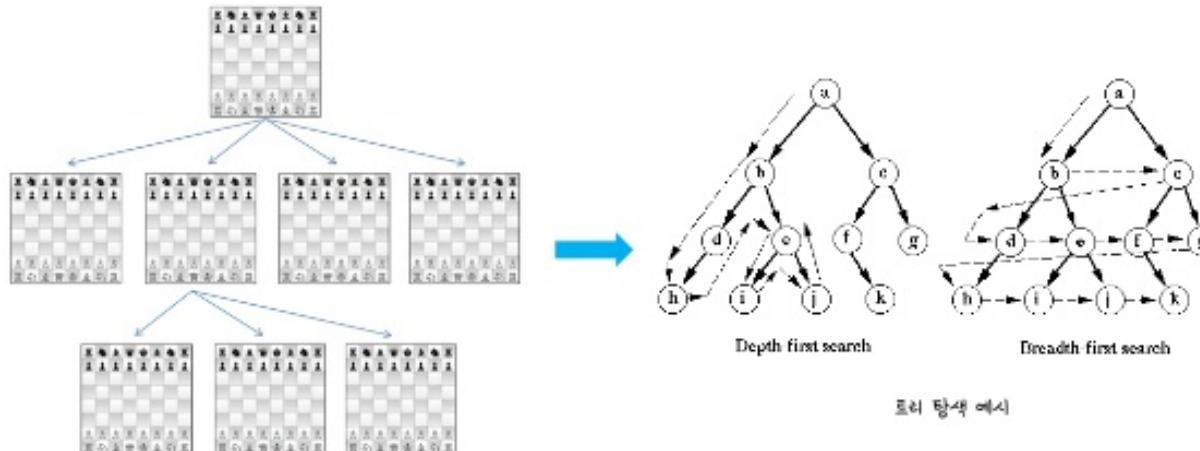
Deep neural network



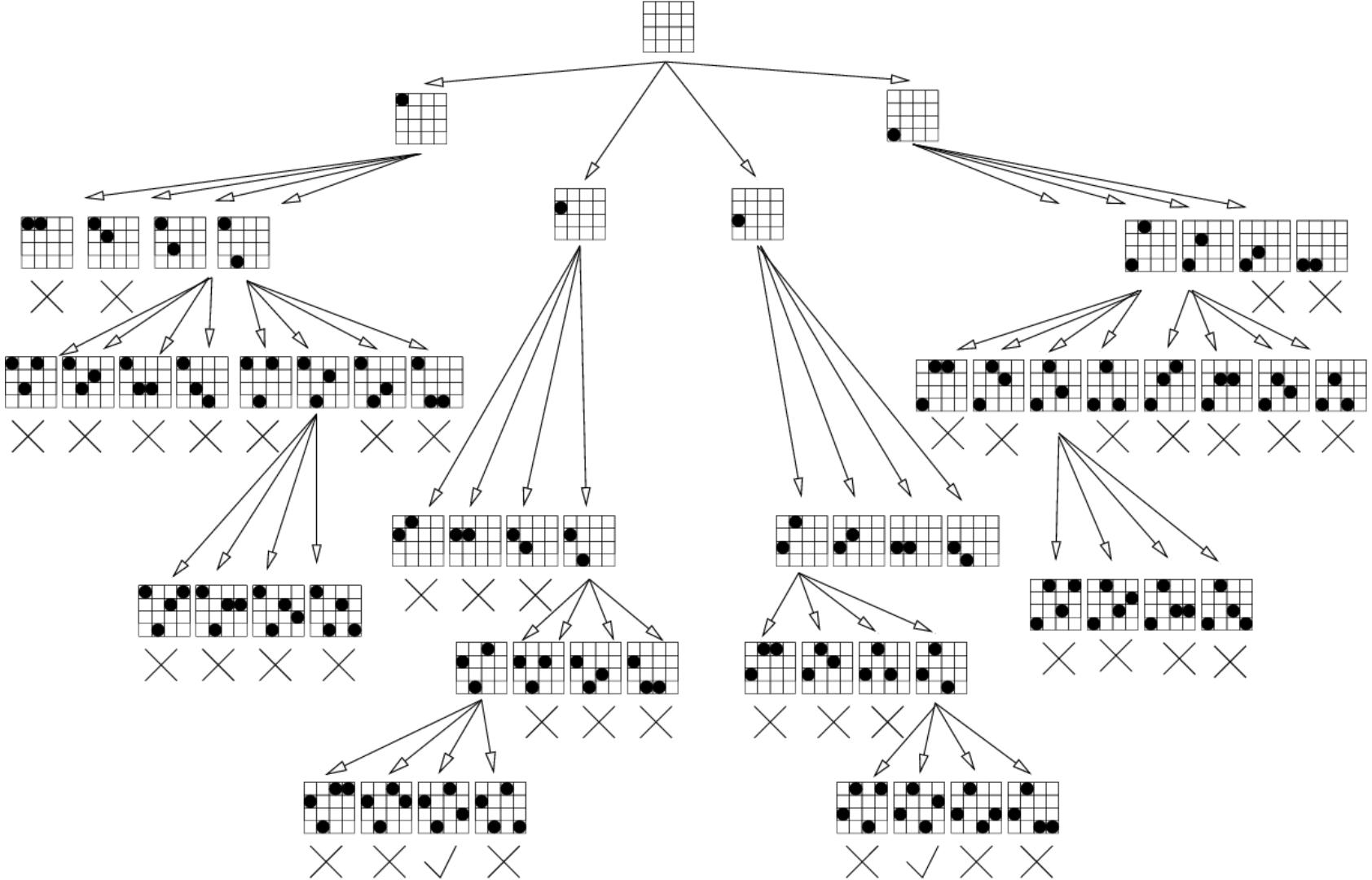
# Deep Learning Algorithm

## 바둑(or 체스) 인공지능 프로그램 기본

- 게임 경우의 수 탐색 → 트리(tree) 탐색 알고리즘 (e.g. 깊이 우선 탐색, 넓이 우선 탐색 알고리즘)



체스의 게임 트리 예시



# Exercise #1



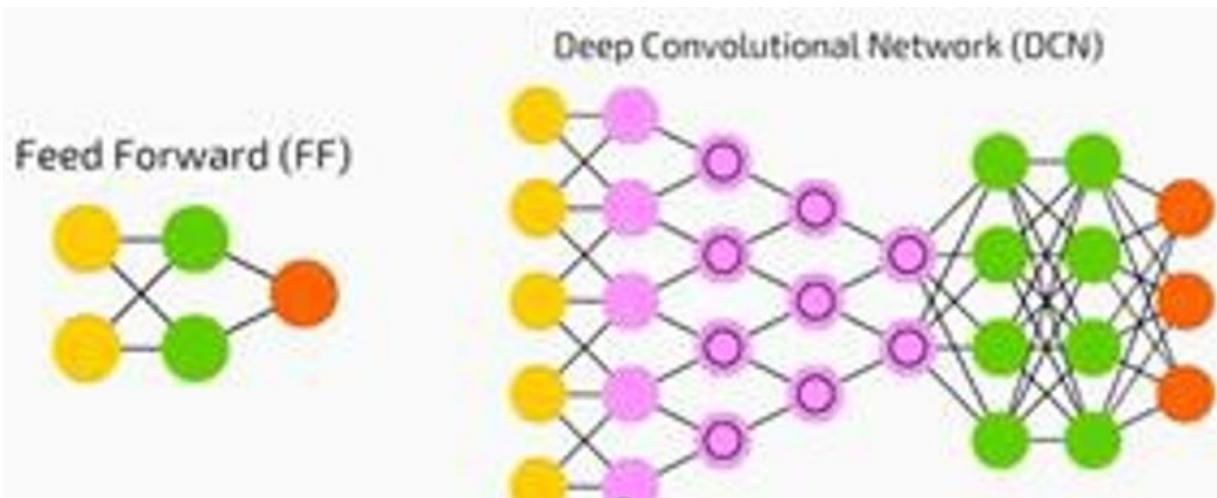
- Read the article (<https://interestingengineering.com/7-ways-ai-will-help-humanity-not-harm-it>) and discuss the difference between human and AI, and the ways that AI will help humanity, not harm it  
글을 읽고 AI와 사람의 다른점을 찾고, 어떻게 AI가 해를 끼치지 않고 사람을 도울수 있는지 토론

# Neural Network

- An artificial intelligence system that is capable of finding and differentiating patterns  
패턴을 구별하고 찾아내기 위한 인공지능시스템
- Take part in massive parallel processing  
대량 병렬처리를 요구함
- Analyze nonlinear relationships  
비선형적 문제를 분석할수 있음

# CNN (Convolutional Neural Network) 합성곱 신경망

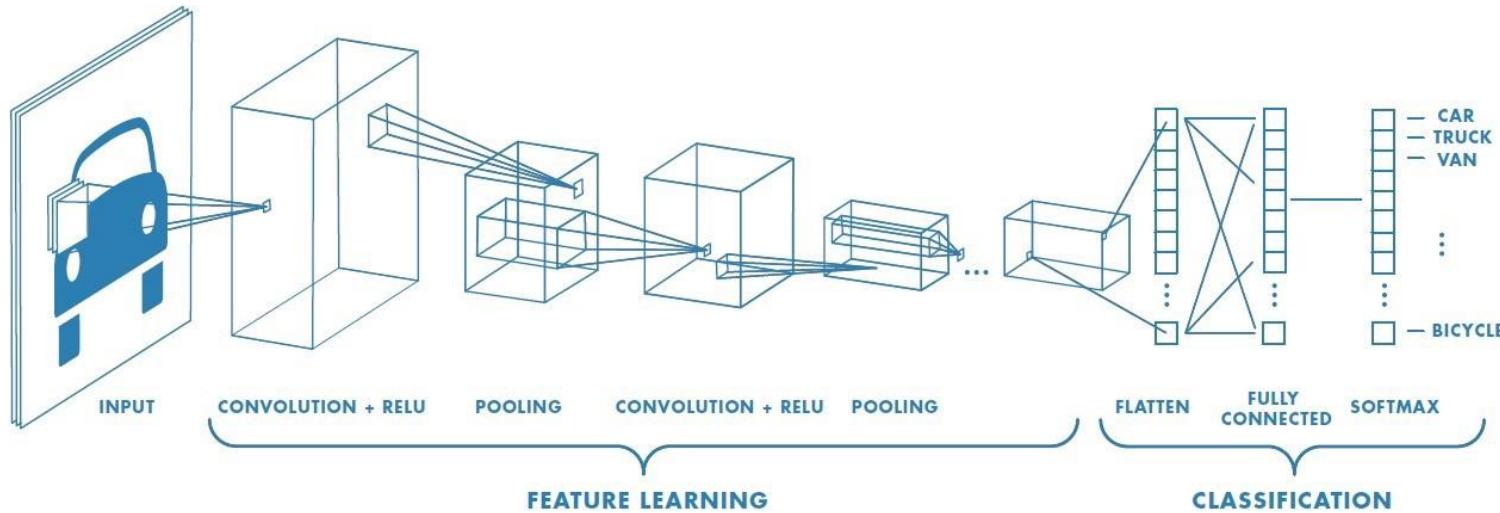
- A neural network with some convolutional layers (and some other layers), e.g.,  
Image processing  
시각적 이미지를 분석하는 데 사용되는 깊고 피드-포워드적인 인공신경망



# Convolutional Layers

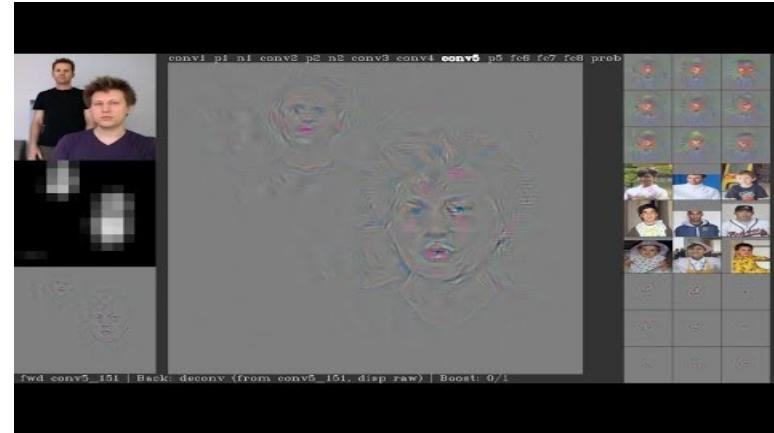
Responsible for the convolutional operation in which feature maps identifies features in the images

이미지안에 있는 특성을 찾아내기 위해 컨볼루션 계산을 하는 층

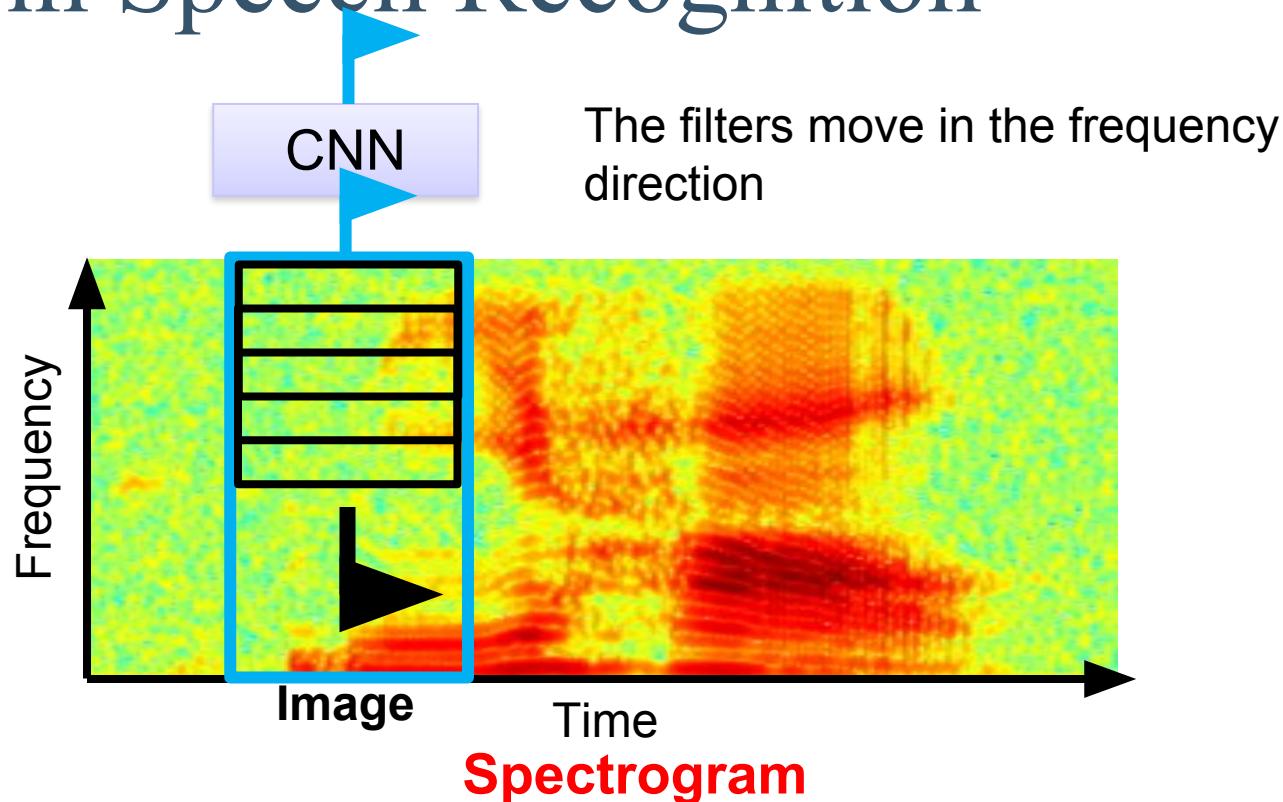


# Deep Visualization Toolbox

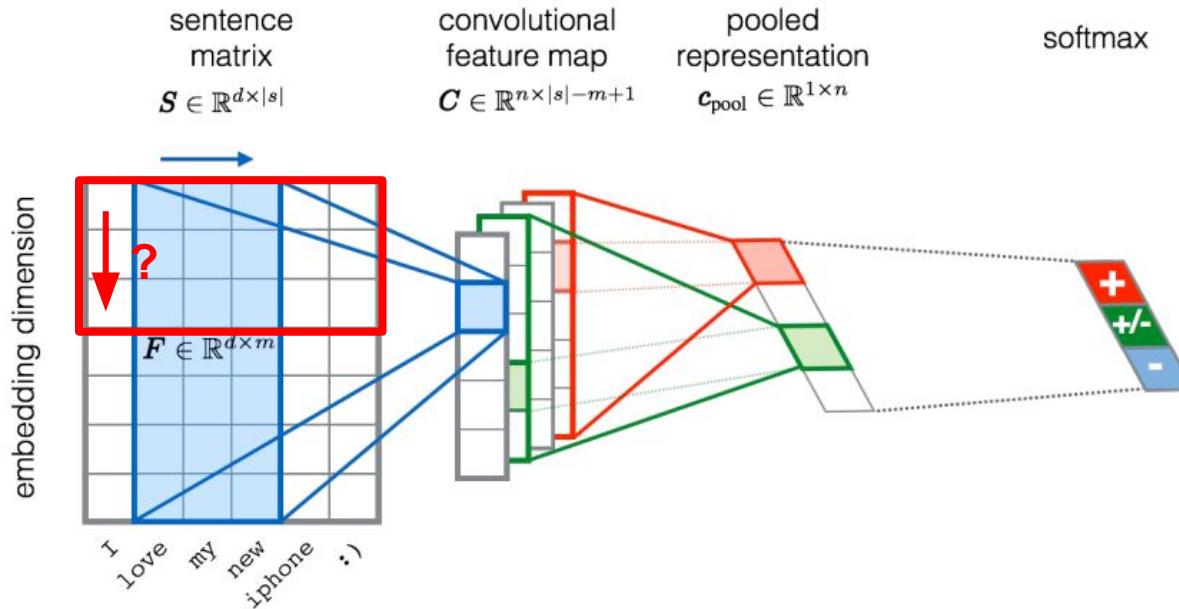
- Generate the neuron-by-neuron visualizations using regularized optimization 정규최적화를 상용해서 각각의 뉴런을 시각화함



# CNN in Speech Recognition



# CNN in Text Classification



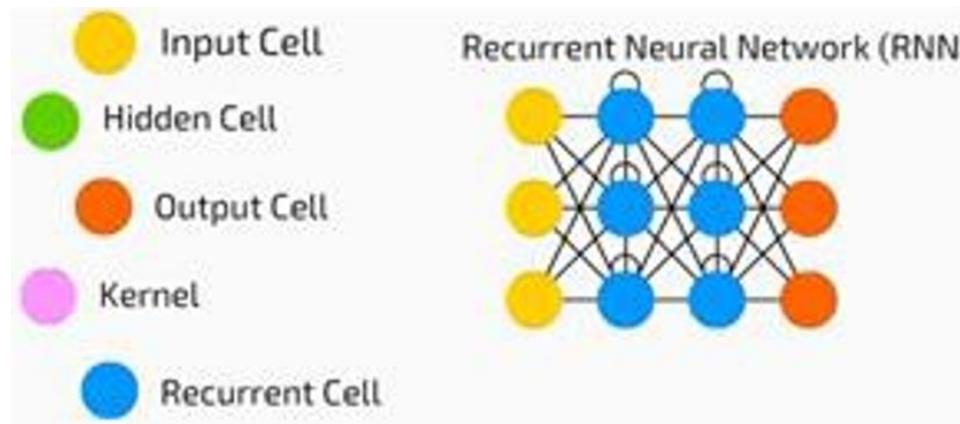
# RNN (Recurrent Neural Network) 순환 인공 신경망

- A type of Neural Network where the output from previous step are fed as input to the current step

신경망의 종류로 이전스텝의 결과물이 현재스텝의 입력값이 됨

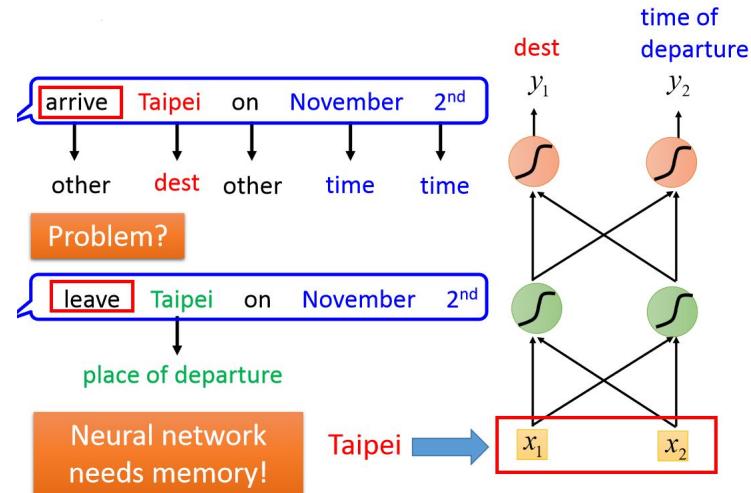
- Language modelling or Natural Language Processing (NLP)

자연어처리



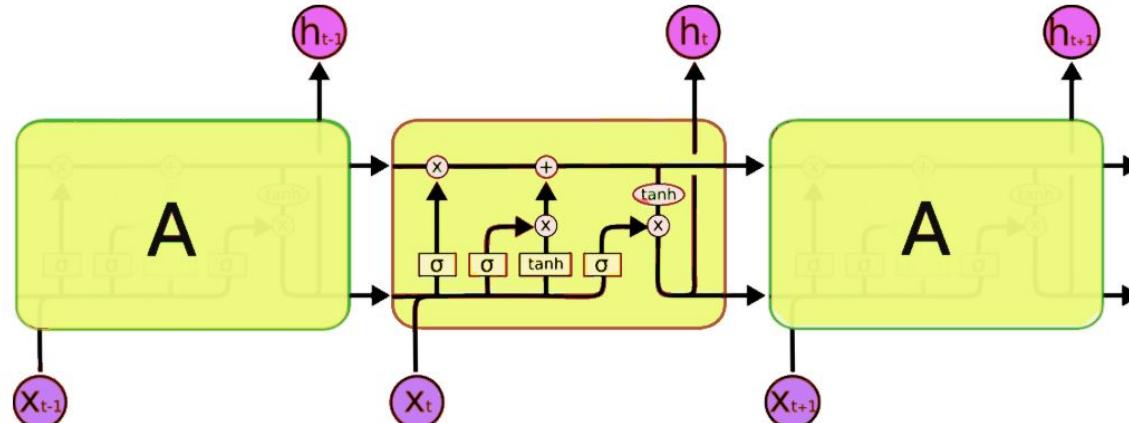
# Memory

- RNN has a “memory” which captures information about what has been calculated so far.
- 메모리를 이용하여 지금까지 계산한 결과를 저장해서 예측에 사용



# LSTM (Long Short Term Memory)

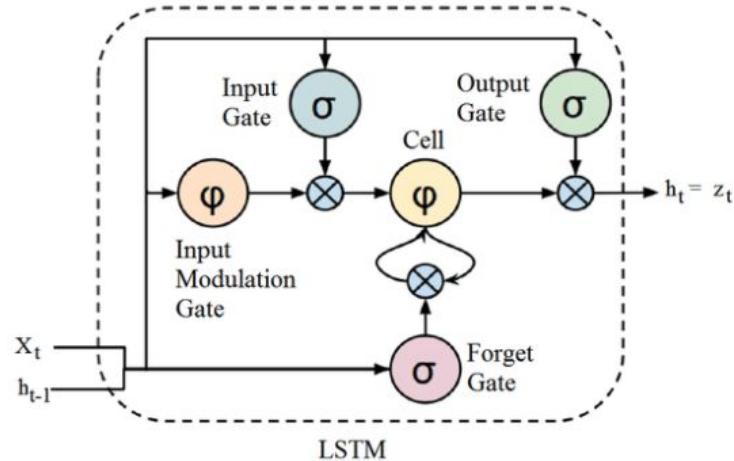
- A type of RNN that introduces the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network
- RNN의 한종류로 은닉층안에 있는 뉴런 대신 기억셀을 소개



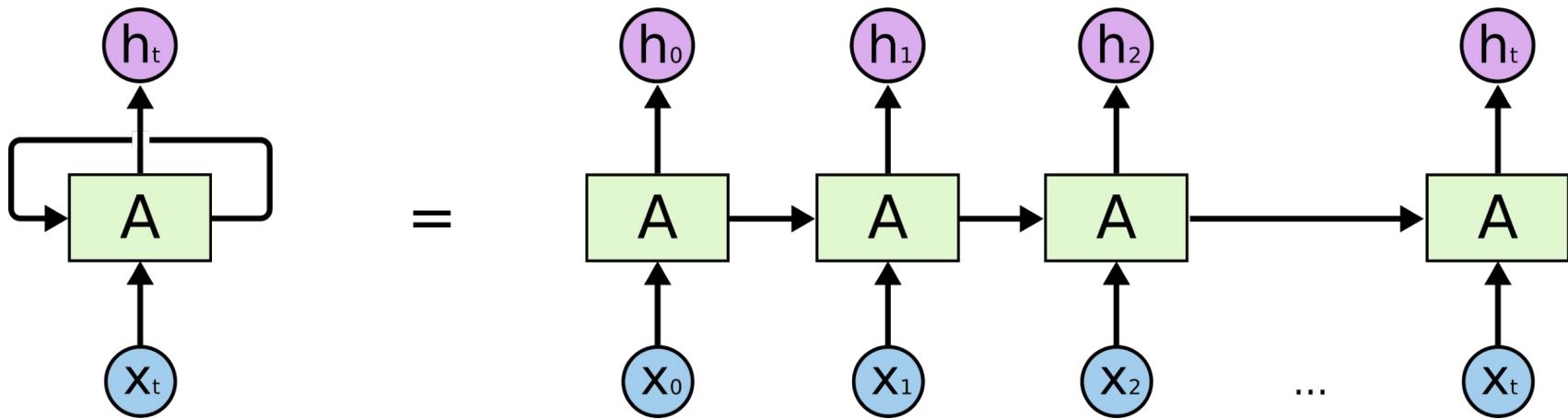
# Forget Gate

- Also called remember vector
- The output of the forget gate tells the cell state which information to forget by multiplying 0  
0을 곱해서 기억을 지우거나..
- If the output of the forget gate is 1, the information is kept in the cell state  
1을 곱해서 기억을 남기던가..

LSTM uses a set of gates to control the flow of information  
게이트를 이용하여 정보의 흐름을 조절함

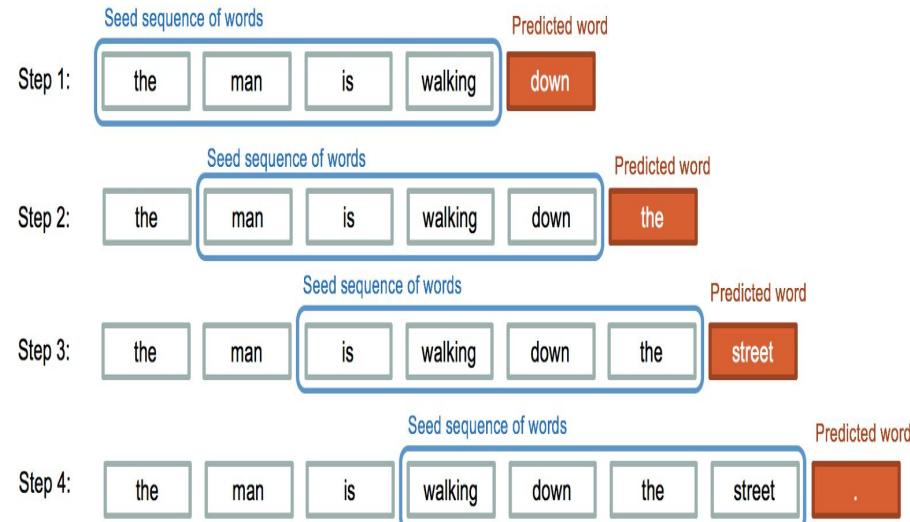


# Recurrent Cells



# Text Generation

- A most powerful models for processing sequential data  
연속된 데이터를 처리하기 가장 좋은 모델
- For example, if you want to predict the next word in a sentence, it is better know which words came before that word 문장에서 다음 단어를 예측할 때 그전에 오는 단어를 알면 더 잘 예측할 수 있음



# RNN Applications

- Language modeling and prediction

랭귀지예측

- Speech recognition

언어인식

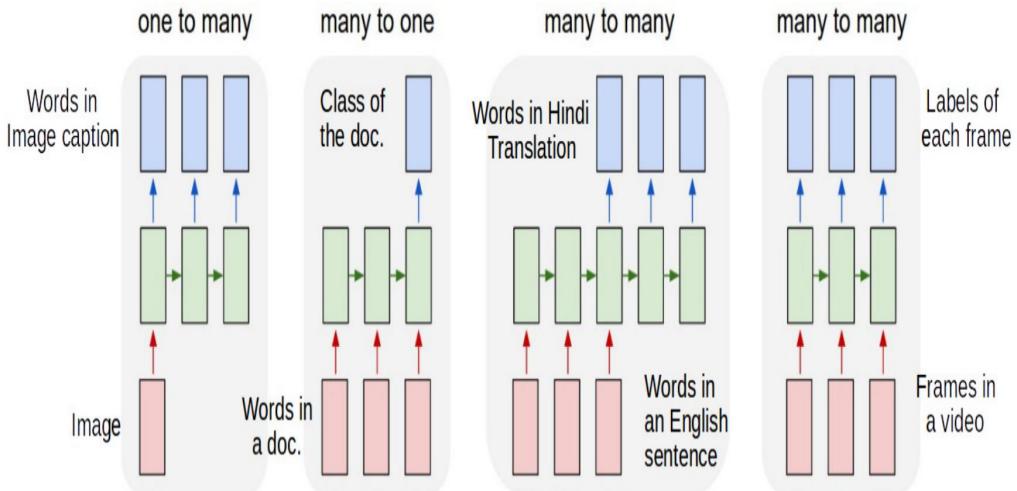
- Machine translation

기계통역

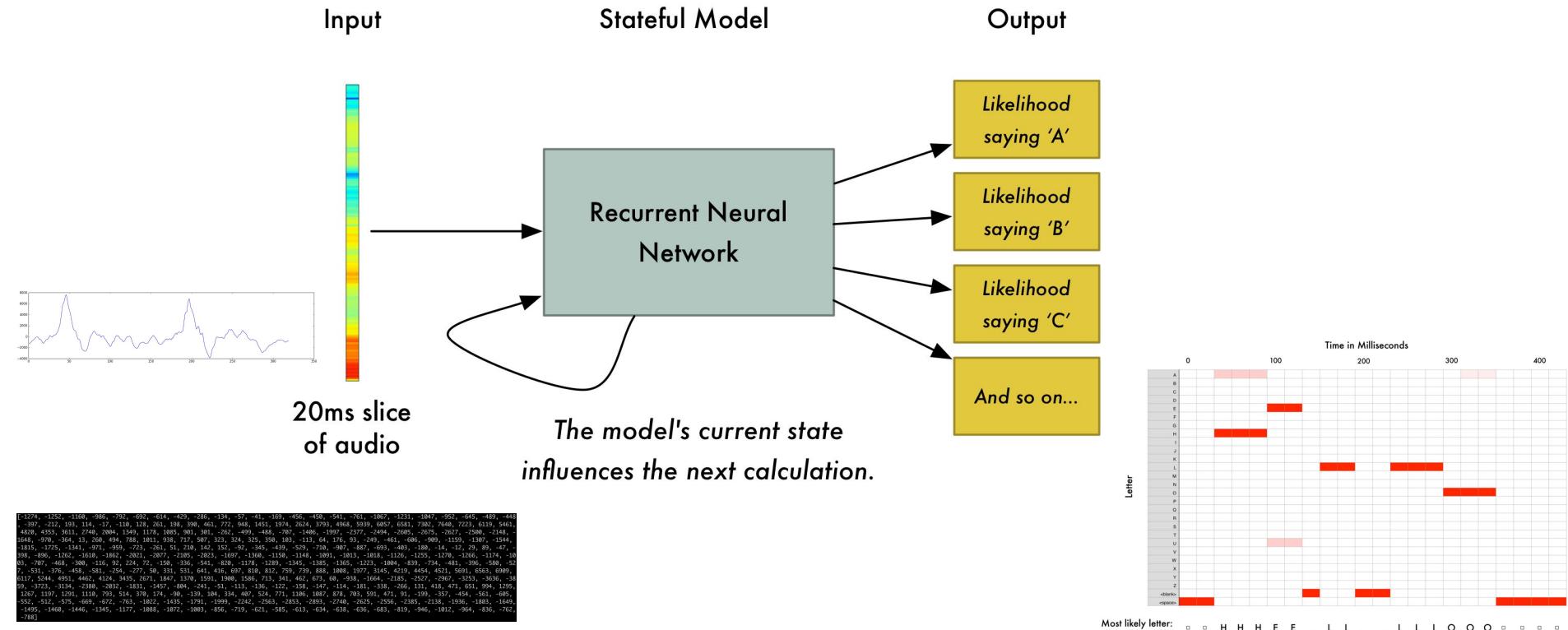
- Image recognition and

characterization

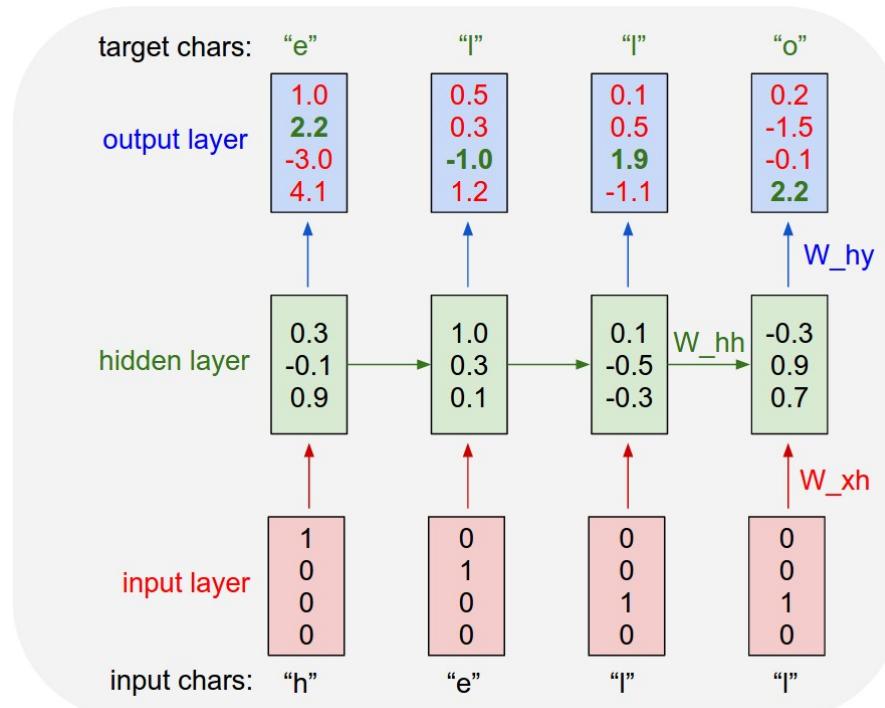
이미지인식



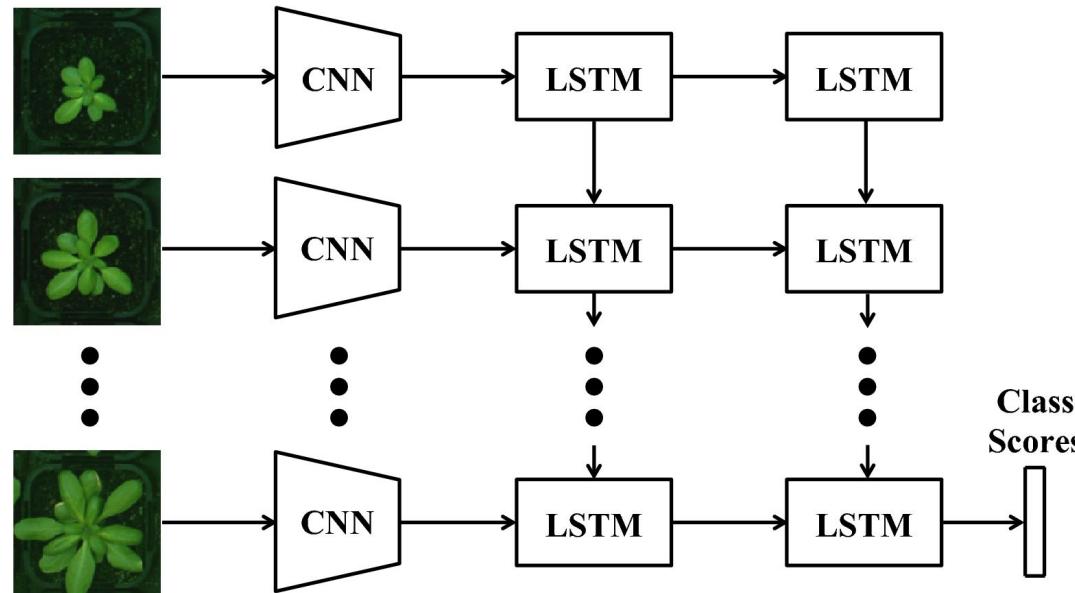
# Speech Recognition



# Machine Translation

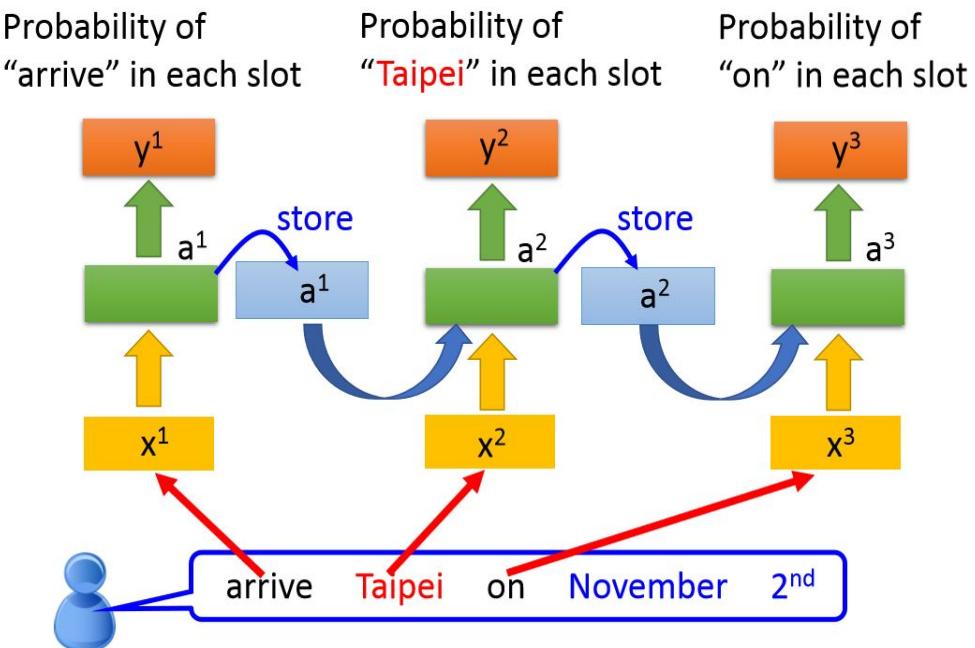


# Image Recognition and Characterization



# Solving Slot Filling

- Solving slot filling using feedforward network 빈칸채우기
  - Input: a word (Each word is represented as a vector) 단어를 벡터로 저장
  - Output: Probability distribution that the input word belonging to the slots 각변수가 속할 위치에 대한 확률분포



# Python Libraries for Deep Learning

Theano

Keras

TensorFlow

# Expert Systems (Knowledge-Based System)

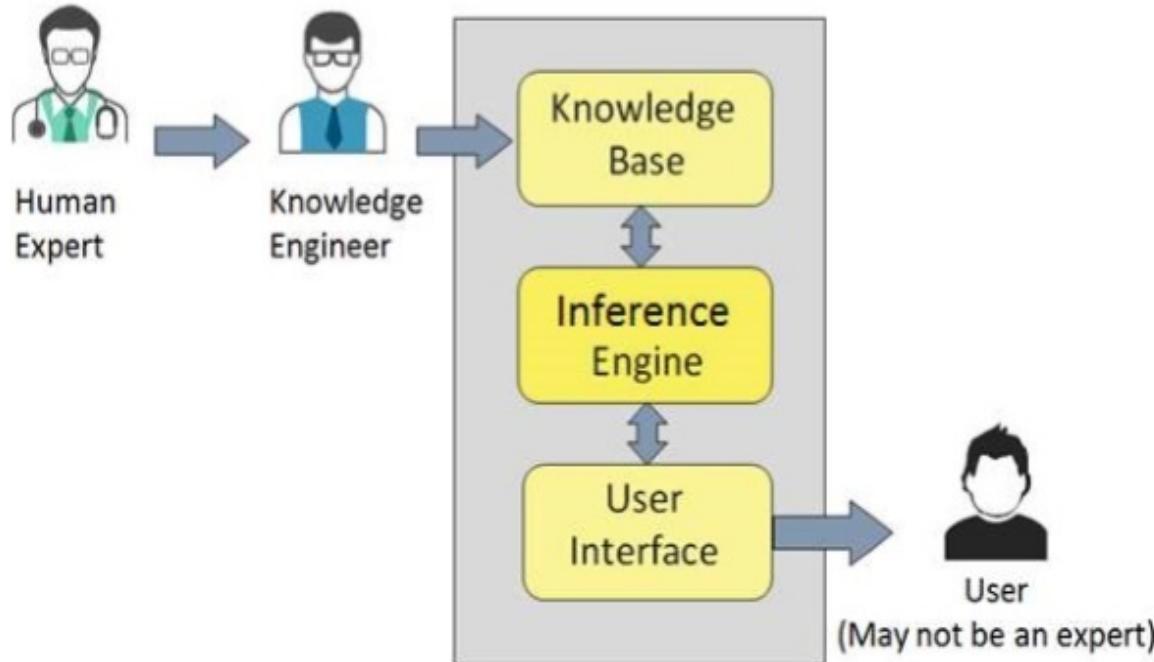
- An artificial intelligence system that applies reasoning capabilities to reach a conclusion

결론에 이르는 추리를 가능하게 하는 인공지능시스템, 전문가의 기능을 대체

- Used for:

- Diagnostic problems (what's wrong?) 진단문제
- Prescriptive problems (what to do?) 처방문제

# Architecture of Expert System



# Fuzzy Logic

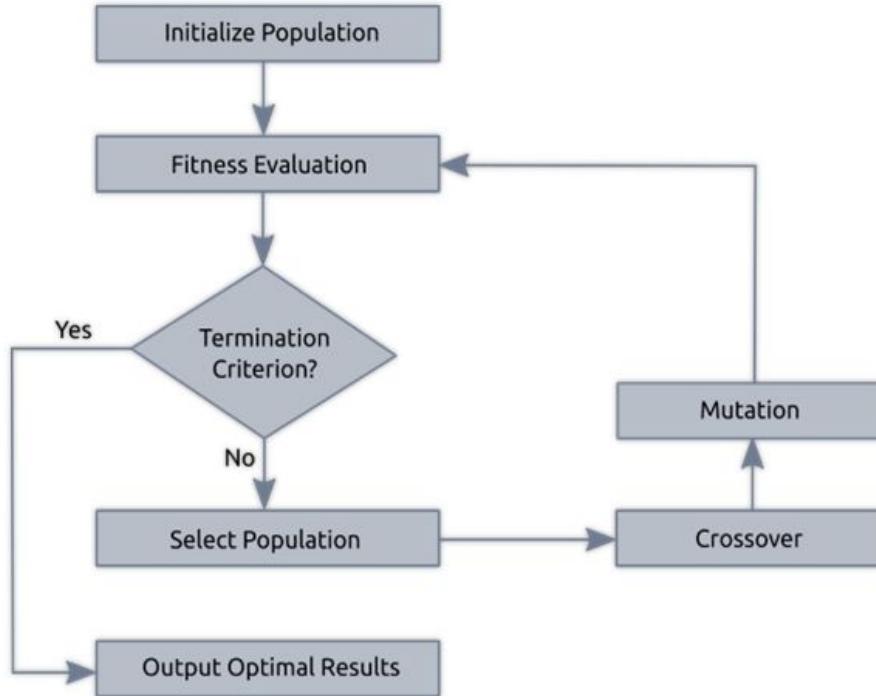
- A mathematical method of handling imprecise or subjective information  
부정확하고 주관적인 정보를 처리
- Used to make ambiguous information such as “short” usable in computer systems  
애매모호한 정보를 처리할 때 사용
- Applications
  - Google’s search engine
  - Washing machines



# Genetic Algorithms

- An artificial intelligence system that mimics the evolutionary, survival-of-the-fittest process to generate increasingly better solutions to a problem  
더 나은 해결책을 제시하기 위해 진화, 적자생존의 방법을 씀
- Take thousands or even millions of possible solutions and combine and recombine them until it finds the optimal solution  
최적의 해답을 찾을때까지 여러개의 답을 조합
- Work in environments where no model of how to find the right solution exists **답을 위한 적절한 모델이 없는 경우 사용**

# Basic Structure of Genetic Algorithm



# Genetic Algorithm Applications

- Staples – determine optimal package design characteristics  
최적화된 패키지 디자인을 결정
- Boeing – design aircraft parts such as fan blades  
비행기의 부품 디자인
- Many retailers – better manage inventory and optimize display areas  
재고관리와 디스플레이 장소 최적화

# Exercise #1

- Read the articles below and introduce any business that applies AI techniques.

- <http://news.zum.com/articles/57470988>

- <https://www.mk.co.kr/news/it/view/2020/01/42536/>

아래 글을 읽고 AI기술을 활용한 비지니스를 소개하시요.

# 딥러닝을 활용한 자율주행



# Big Data

Data sets that are so voluminous and complex that traditional data processing application software are inadequate to deal with them –

Wikipedia

데이터셋이 너무 크고  
복잡해서 전통적인  
데이터처리 방식으로는  
처리가 안되는 데이터



# Life Cycle for Data Analytics

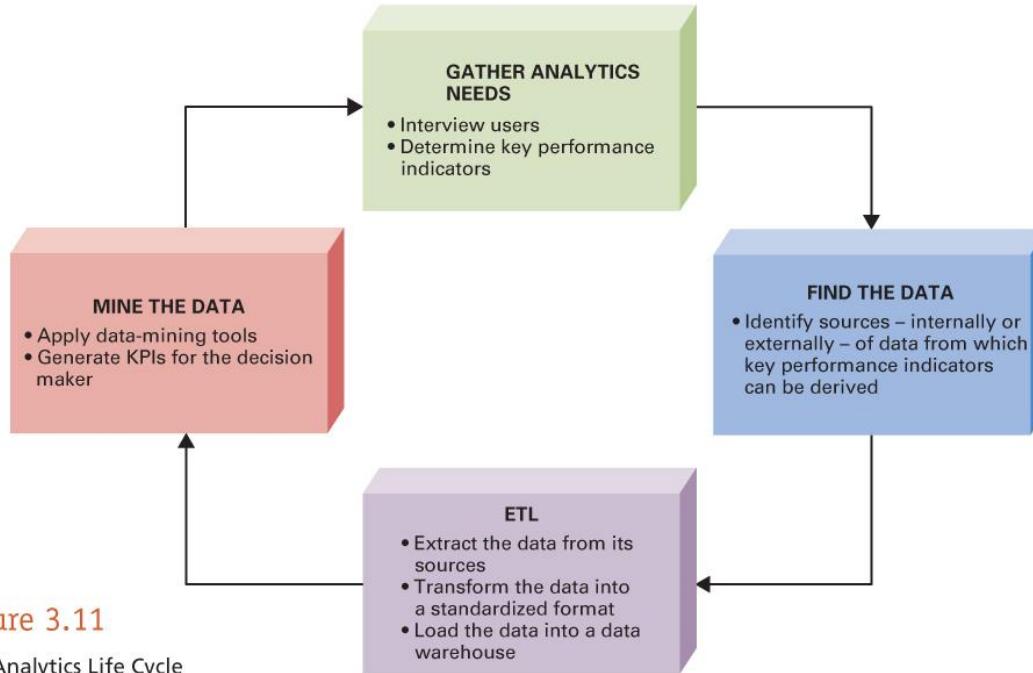
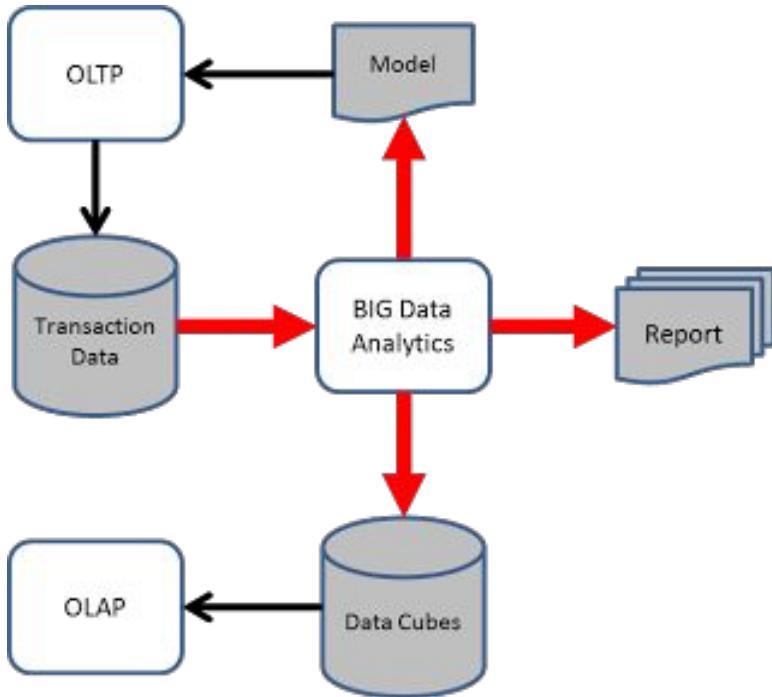


Figure 3.11

The Analytics Life Cycle

# Typical Data Processing

## 기존 데이터 프로세싱

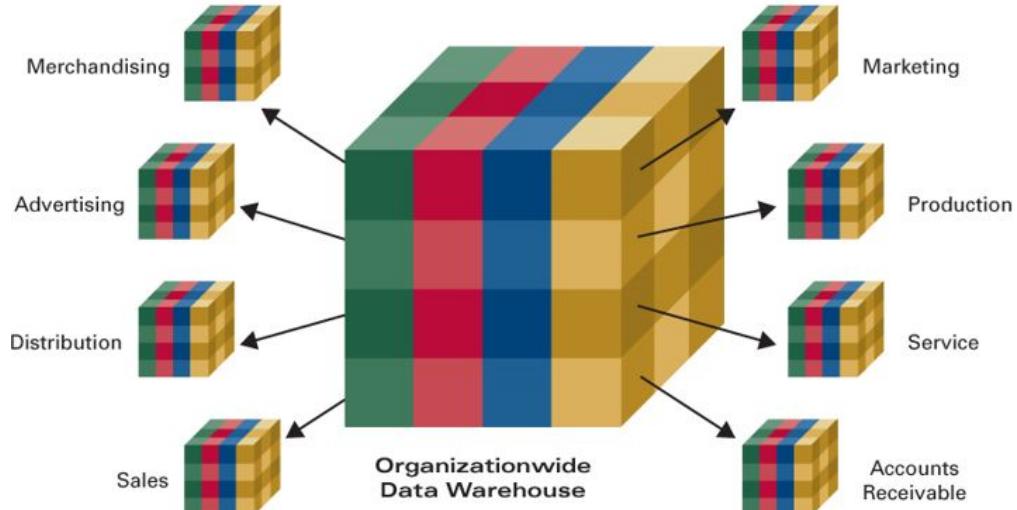


- Data is created from the OLTP (On-Line Transaction Processing) system,  
OLTP로부터 데이터가 들어와서
- Flowing into the BIG Data Analytics system,  
빅 데이터분석 시스템으로 들어가
- Which produced various outputs; including data mart/cubes for OLAP (On Line Analytic Processing),  
reports for the consumption of business executives,  
and predictive models that feedback decision support  
for OLTP  
데이터마트, 큐브, 보고서, 예측모델등 다양한  
결과물을 생산함.

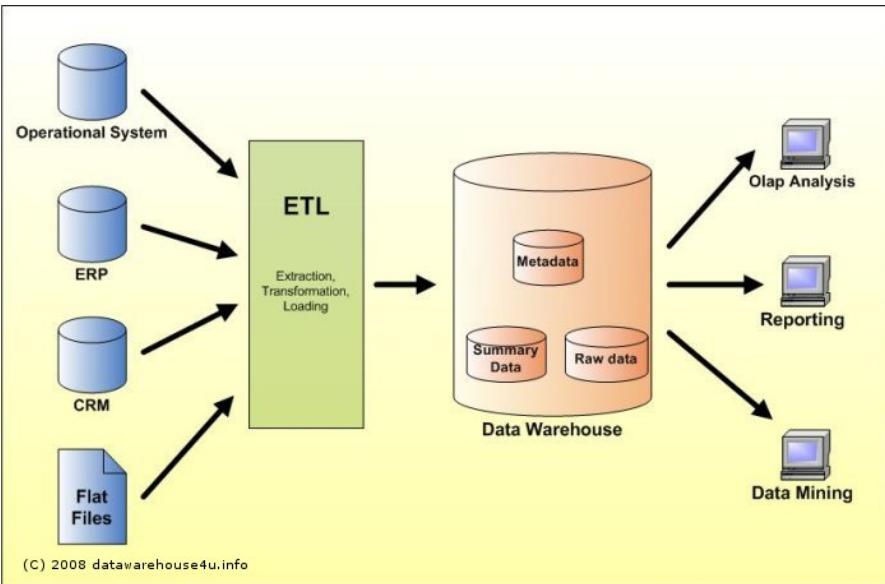
# Data Marts

- Subset of a data warehouse in which only a focused portion of the data warehouse information is kept

데이터창고 정보의  
일부에 집중한  
데이터창고의 부분집합



# Data Warehouse

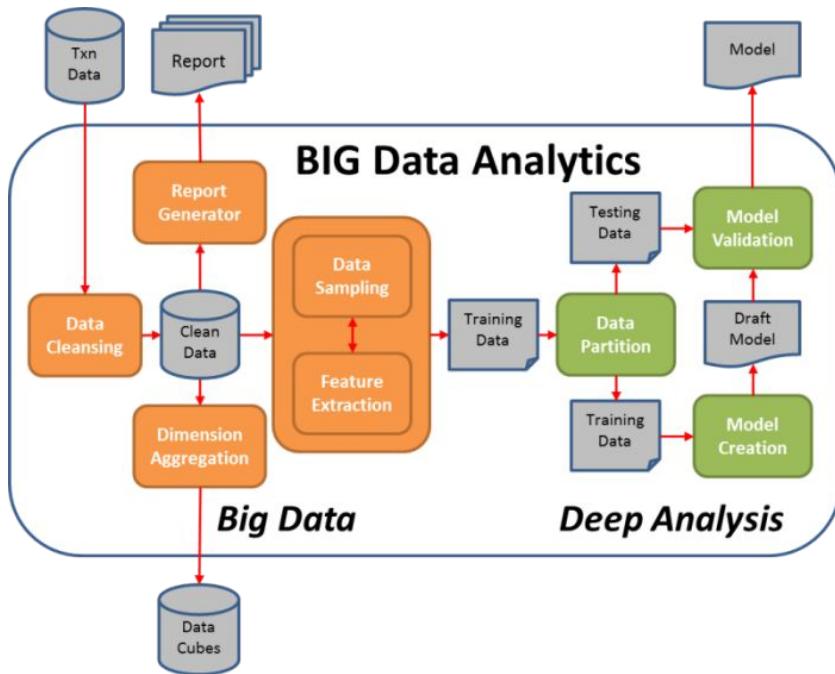


- A relational database that is designed for query and analysis  
쿼리와 분석을 위해 디자인된 관계형 데이터베이스
- Usually contains historical data derived from transaction data and includes data from other sources.  
거래에서 발생된 데이터들과 다른 소스에서 들어온 데이터들을 합침

# Extraction, Transformation, and Loading (ETL)

- ETL is a three-step process
  - Extract needed information from its source  
데이터를 소스로부터 추출해서 가져옴
  - Transform the data into a standardized format  
표준화된 모양으로 변환시킴
  - Load the transformed data into a data warehouse  
데이터창고로 변환된 데이터를 로드함

# Big Data & Deep Analysis



- The BIG data analytics box is usually done in a batch fashion (e.g. once a day), usually we see that big data processing and deep data analysis happen at different stages of this batch process.  
빅 데이터를 처리하는 부분과 분석하는 부분으로 나뉘어 있는데, 한꺼번에 처리됨 (예를 들어 하루에 한번)

# Big Data Processing Part

- Usually done using Hadoop/PIG/Hive technology with classical ETL logic implementation.  
전통적인 ETL로직 설치와 함께 하둡/피그/하이브을 이용하여 빅 데이터를  
프로세싱
- By leveraging the Map/Reduce model that Hadoop provides, we can linearly scale up the processing by adding more machines into the Hadoop cluster.  
하둡이 제공하는 맵리듀스를 이용하여 더 많은 머신을 추가할수 있음
- Drawing cloud computing resources (e.g. Amazon EMR) is a very common approach to performing this kind of tasks.  
클라우드 컴퓨팅을 이용하여 처리할수도 있음

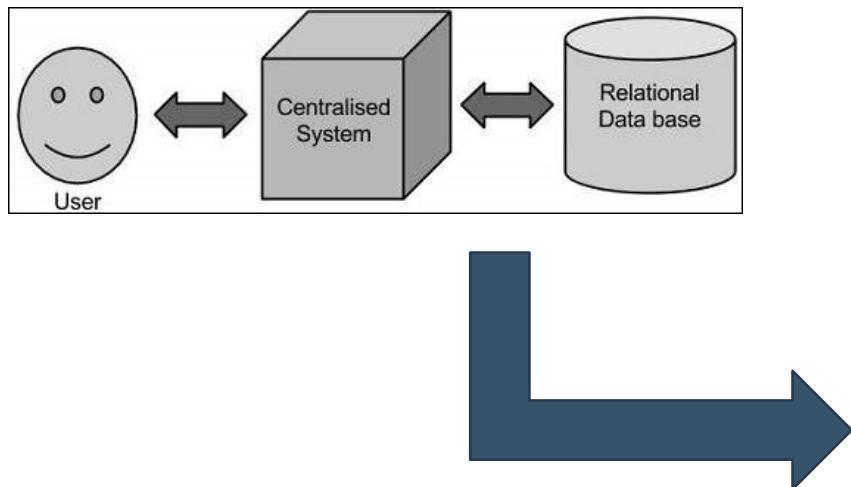
# Hadoop

- An open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models.  
빅데이터를 저장, 처리하기 위해 여러대의 컴퓨터에 분산처리하는 오픈소스 프레임워크
- It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.  
한대의 서버에서 수많은 기계로 분산하여 지역계산, 저장을 함

# MapReduce

- Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel with others.

병렬처리



- MapReduce divides the task into small parts and assigns them to many computers, and collects the results from them which when integrated, form the result dataset.

한 업무를 작은 파일로 잘라서 여러 컴퓨터에 나눠서 처리하고 다시 모아서 하나의 데이터셋을 만듬

# Apache Hadoop Ecosystem

Management & Monitoring  
**(Ambari)**

Scripting  
**(Pig)**

Machine Learning  
**(Mahout)**

Query  
**(Hive)**

Coordination  
**(ZooKeeper)**

Workflow & Scheduling  
**(Oozie)**

Distributed Processing  
**(MapReduce)**

Distributed Storage  
**(HDFS)**

NoSQL Database  
**(HBase)**

Data Integration  
**(Sqoop/REST/ODBC)**

# Hadoop Process

- Data is initially divided into directories and files. Files are divided into uniform sized blocks of 128M and 64M (preferably 128M). 데이터를 폴더들과 파일들로 자름
- These files are then distributed across various cluster nodes for further processing. 파일들을 분산시킴
- HDFS, being on top of the local file system, supervises the processing. HDFS에서 이 과정을 관리
- Blocks are replicated for handling hardware failure. 하드웨어실패에 대비해 복사
- Checking that the code was executed successfully. 코드가 제대로 실행되는지 체크
- Performing the sort that takes place between the map and reduce stages. 정렬을 수행
- Sending the sorted data to a certain computer. 정렬된 데이터를 컴퓨터에 보냄
- Writing the debugging logs for each job. 각 업무에 대한 로그파일 작성

# Deep Analysis Part

- Usually done in R, SPSS, or SAS using a much smaller amount of carefully sampled data that fits into a single machine's capacity (usually less than couple hundred thousands data records)

샘플링한 데이터를 R, SPSS, SAS등을 이용하여 분석

- Usually involves data visualization, data preparation, model learning (e.g. Linear regression, K-nearest-neighbour/Support vector machine/Bayesian network/Neural network, Decision Tree), model evaluation.  
데이터 시각화, 데이터 전처리, 모델구축, 평가등을 포함

# Big Data Skills for Data Analysis

## 데이터분석을 위한 빅데이터 기술들

- Learn basic mathematics and statistics required for data science

데이터분석을 위한 기본 수학, 통계

- Develop a basic understanding of machine learning algorithms and solving real life problems from them

머신러닝 알고리즘에 대한 기본적 이해와  
실제문제에 적용할수 있는 능력

**MODERN DATA SCIENTIST**

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

**MATH & STATISTICS**

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

**PROGRAMMING & DATABASE**

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g. R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

**DOMAIN KNOWLEDGE & SOFT SKILLS**

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

**COMMUNICATION & VISUALIZATION**

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

# Basics of Mathematics and Statistics

- Descriptive Statistics 기술통계
- Probability 확률
- Inferential Statistics 추론통계
- Linear Algebra 선형대수
- Structured Thinking 조직적 사고

# R or Python

- Tools 툴들
- Exploration and Visualization 탐색, 시각화
- Feature Selection/ Engineering 속성선택

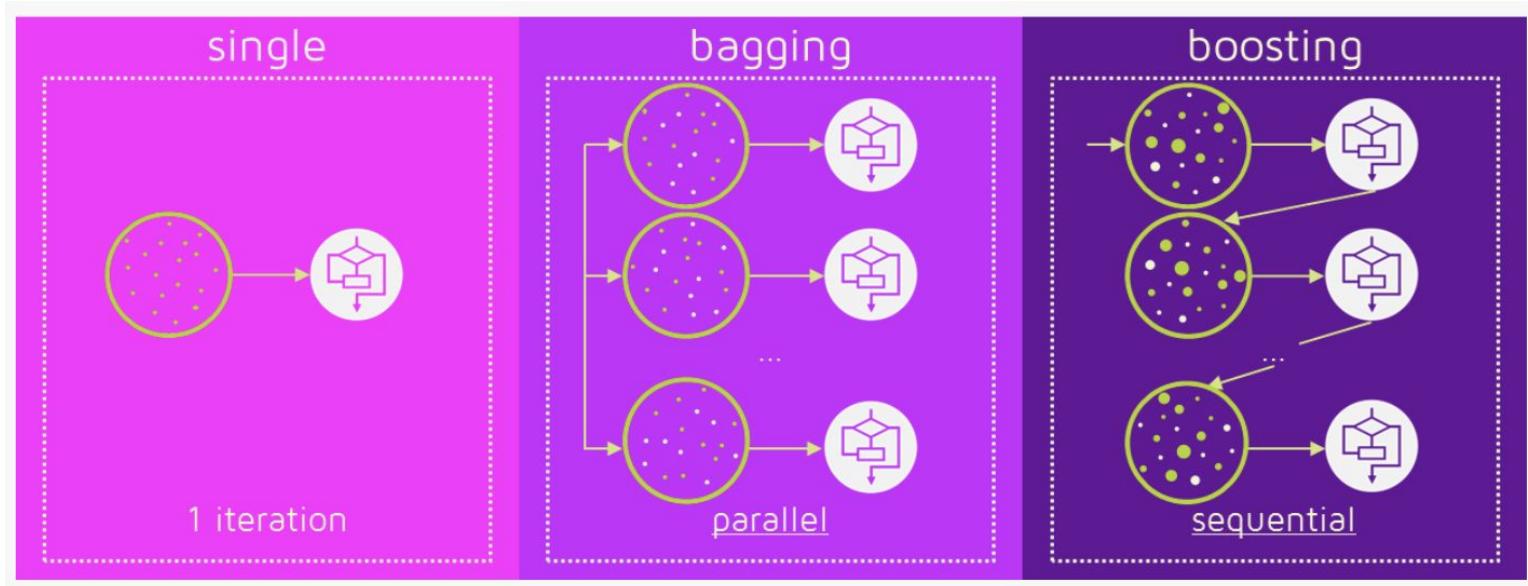
# Basic Machine Learning Algorithms

- Linear Regression 선형회귀
- Logistic Regression 로지스틱회귀
- Decision Trees 의사결정트리
- kNN (K-Nearest Neighbours)
- K-Means 군집분석 방법중 하나
- Naïve Bayes 나이브베이지안
- Dimensionality Reduction 차원축소

# Advanced Machine Learning Algorithms

- Random Forests 랜덤포레스트
- Dimensionality Reduction Techniques 차원축소기술
- Support Vector Machines 서포트 벡터머신
- Gradient Boosting Machines
- XGBOOST
- Neural Network 신경망

# Bagging and Boosting



# Exercise #1

- Read the article on

<https://news.naver.com/main/read.nhn?mode=LSD&mid=shm&sid1=105&oid=022&aid=0003346568>

[사이언스프리즘] 빅데이터 플랫폼 확보 전쟁  
링크에 있는 글을 읽으시요

- Form a team with 3-4 people to discuss the following questions.

3-4명의 학생들과 한팀이 되어 다음의 질문들에 대해 토의하시요

- How do you feel that the companies such as Amazon, Google, Apple, Microsoft collect your personal data?  
아마존, 구글, 애플, 마이크로소프트와 같은 회사들이 개인의 정보를 수집하는 것에 대해 어떻게 느끼십니까?
- Are there any other ways to collect customers' data without violating their privacy?  
개인의 사생활을 침해하지 않고 고객의 정보를 수집할수 있는 방법은?
- How should Korean companies react to this battle to occupy the big data platform?  
빅데이터 플랫폼 확보전쟁에 대해서 한국회사들은 어떻게 대비해야 할까요?

# Tableau

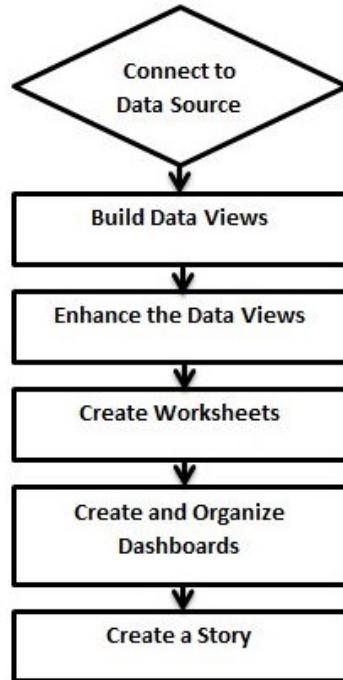
# Tableau

- A Business Intelligence tool for visually analyzing the data. 데이터를 시각적으로 분석할 수 있는 BI 툴
- Users can create and distribute an interactive and shareable dashboard, which depict the trends, variations, and density of the data in the form of graphs and charts. 그래프와 차트를 이용하여 상호작용적인 대시보드를 만들 수 있음
- Tableau can connect to files, relational and Big Data sources. 다양한 데이터를 연결할 수 있음
- The software allows data blending and real-time collaboration. 데이터 블렌딩과 실시간 협업이 가능함

# Data Type

Data Type	Example
STRING	'Hello' 'Quoted' 'quote'
NUMBER	3 142.58
BOOLEAN	TRUE FALSE
DATE & DATETIME	"02/01/2015" "#3 March 1982"

# Design Flow

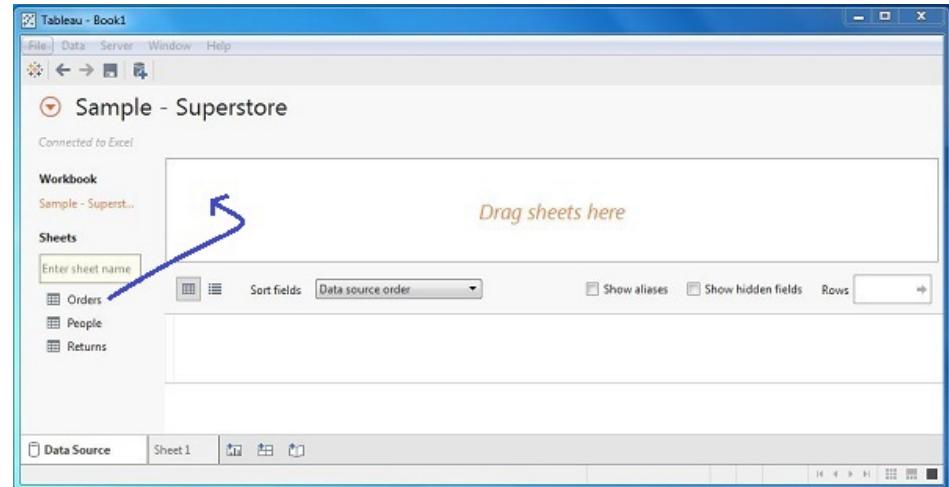


- Views which are traditionally known as Reports
- You can apply filters, aggregations, labeling of axes, formatting of colors and borders, etc.

# Exercise #1

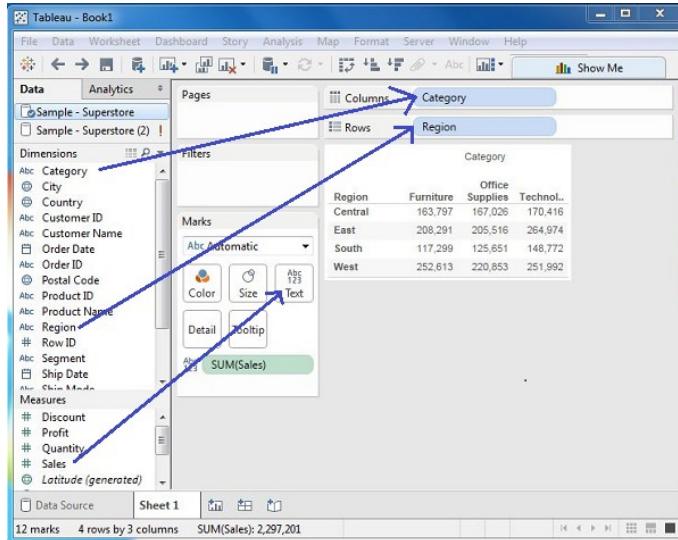
- Tableau를 다운로드 받아서 설치하세요.
- C:\Users\Kim\Documents\내 Tableau 리포지토리\데이터 원본\2019.4\en\_US-US 디렉토리에 있는 Sample – Superstore.xls 파일을 여시요.
- Orders를 선택하세요.

[https://www.tutorialspoint.com/tableau/tableau\\_get\\_star\\_ted.htm](https://www.tutorialspoint.com/tableau/tableau_get_star_ted.htm)

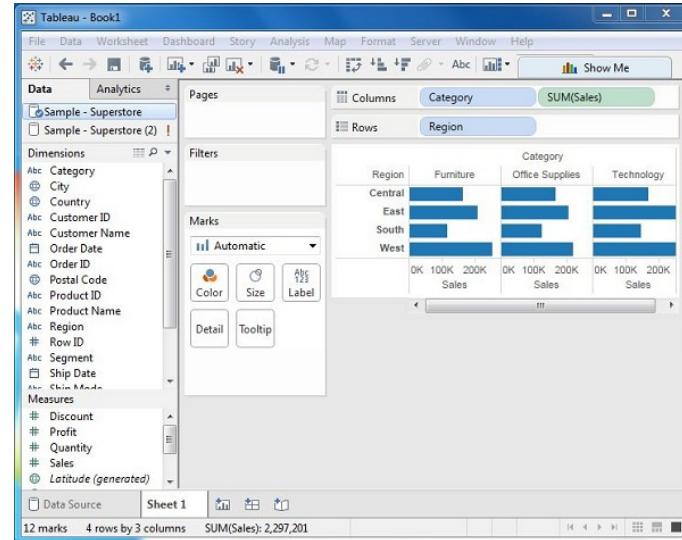


# Exercise #1

- Category를 열에 Region을 행에  
집어 넣고, Sales를 텍스트에  
넣으시요.

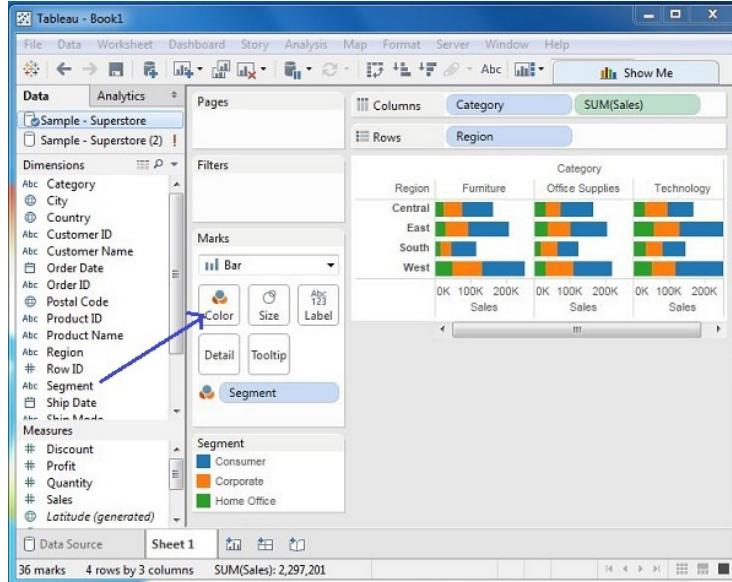


- Marks에 있는 Sales값을 행으로  
이용하시요.



# Exercise #1

- Segment를 Color위에 드래그하시요.



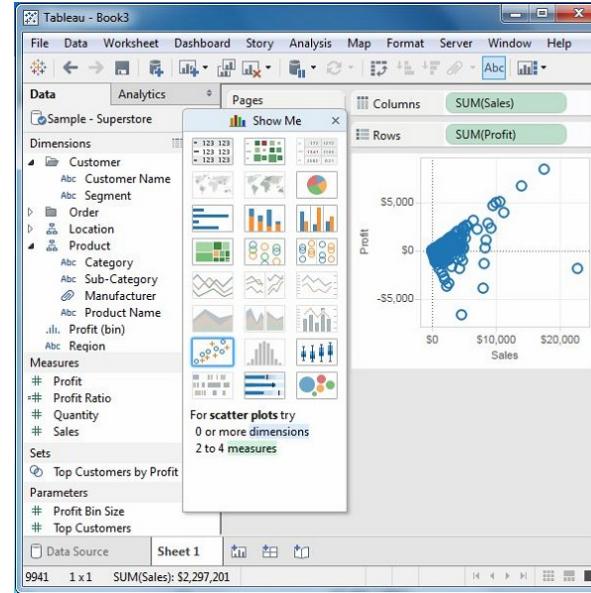
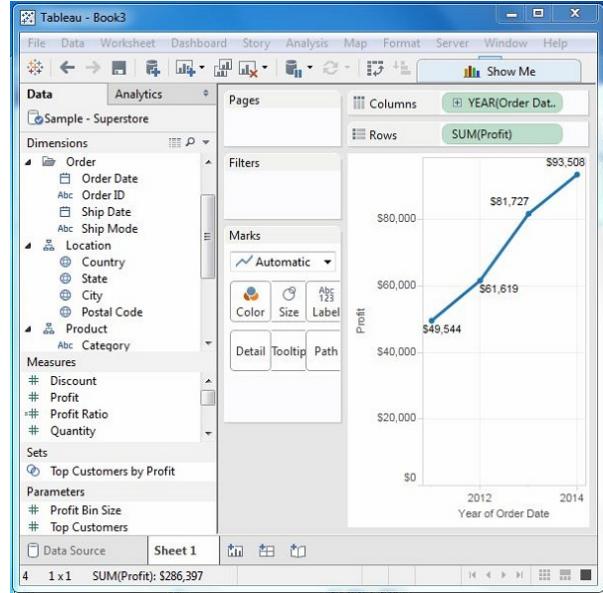
# Exercise #1

- Complete the Tableau Interface tutorial on

[https://help.tableau.com/current/pro/desktop/en-us/getstarted\\_buildmanual\\_ex1basic.htm](https://help.tableau.com/current/pro/desktop/en-us/getstarted_buildmanual_ex1basic.htm)

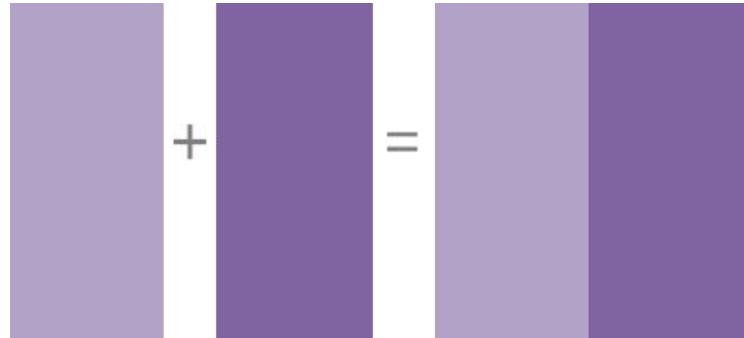
# Exercise #1

[https://www.tutorialspoint.com/tableau/tableau\\_show\\_me.htm](https://www.tutorialspoint.com/tableau/tableau_show_me.htm)



# Join

- A method for combining tables related by common fields (that is, common columns). 공통필드로 합치는 방법



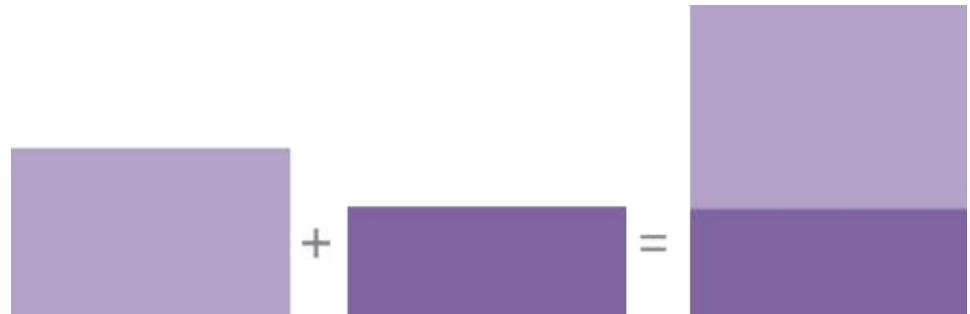
# Blending

- A method that lets you combine data.  
데이터를 합치는 방법
- When you use data blending to combine your data, you combine data in what is called a primary data source with common fields from one or more secondary data sources.  
공통되는 것이라 상관없이 있는 레벨에 따라 합치는 방법



# Union

- A method for appending values  
(that is, rows) to tables. 데이터를  
행으로 붙이는 방법



# Exercise #1

- Complete the union tutorial on

<https://help.tableau.com/current/pro/desktop/en-us/union.htm>

- Union tables manually
- Union tables using wildcard search (Tableau Desktop)

# Arithmetc Operators

Operator	Example
+(addition)	$7 + 3$ Profit + Sales <code>'abc' + 'def' = 'abcdef'</code> <code>#April 15, 2004# + 15 = #April 30, 2004#</code>
-(subtraction)	$-(7+3) = -10$ <code>#April 16, 2004# - 15 = #April 1, 2004#</code>
*(Multiplication)	$23 * 2 = 46$
/(Division)	$45/2 = 22.5$
%(modulo)	$13 \% 2 = 1$
^(power)	$2^3 = 8$

# Comparison Operators

Operator	Example
<b>= = or = (Equal to)</b>	'Hello' = 'Hello' 5 = 15 / 3
<b>!= or &lt;&gt; (Not equal to)</b>	'Good' <> 'Bad' 18 != 37 / 2
<b>&gt; (Greater than)</b>	[Profit] > 20000 [Category] > 'Q' [Ship date] > #April 1, 2004#
<b>&lt; (Less than)</b>	[Profit] < 20000 [Category] < 'Q' [Ship date] < #April 1, 2004#

# Logical Operators

Operator	Example
AND	[Ship Date] > #April 1, 2012# AND [Profit] > 10000
OR	[Ship Date] > #April 1, 2012# OR [Profit] > 10000
NOT	NOT [Ship Date] > #April 1, 2012#

# Exercise #1

- Complete the Join tutorial on

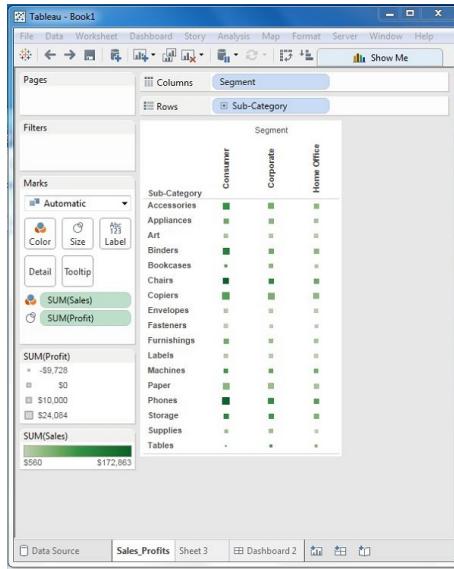
[https://help.tableau.com/current/pro/desktop/en-us/joining\\_tables.htm](https://help.tableau.com/current/pro/desktop/en-us/joining_tables.htm)

- Create Table1 and Table2 and inner-join them
- Inner-join Orders and Returns
- Inner-join Orders\_June and Orders\_July, then select Data > Join null values to null values.
- String mismatch
- Date mismatch

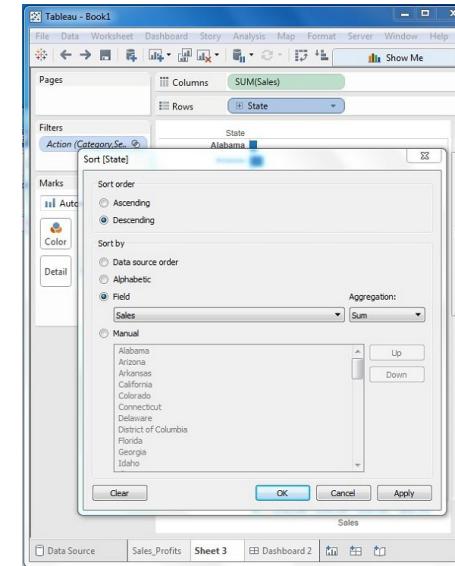
[https://www.tutorialspoint.com/tableau/tableau\\_dashboard.htm](https://www.tutorialspoint.com/tableau/tableau_dashboard.htm)

# Exercise #1

## ▪ Sales\_Profits Worksheet



## ▪ Sales\_state Worksheet



# Exercise #1

- Profit\_Dashboard



- 필터를 이용하여 어떻게 데이터가 연결되는지 확인하시요

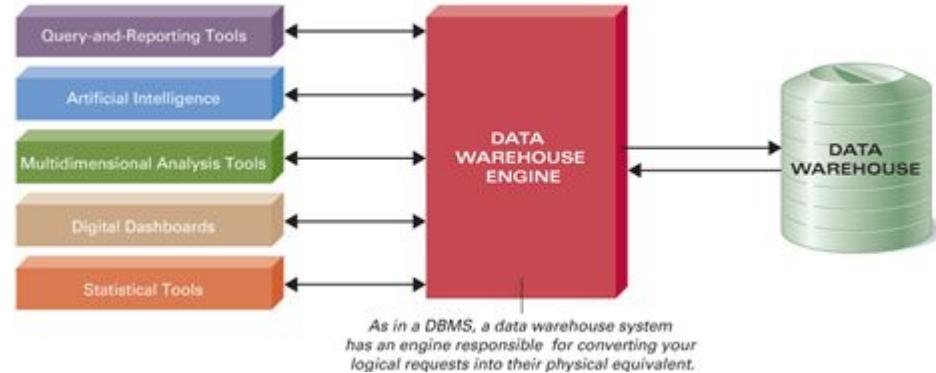
# Exercise #1

- Create worksheets and dashboards for:
  - Chipotle
  - Occupation
  - World Food Facts (<https://www.kaggle.com/openfoodfacts/world-food-facts/data>)

# Data Mining Tools

# Data-Mining Tools

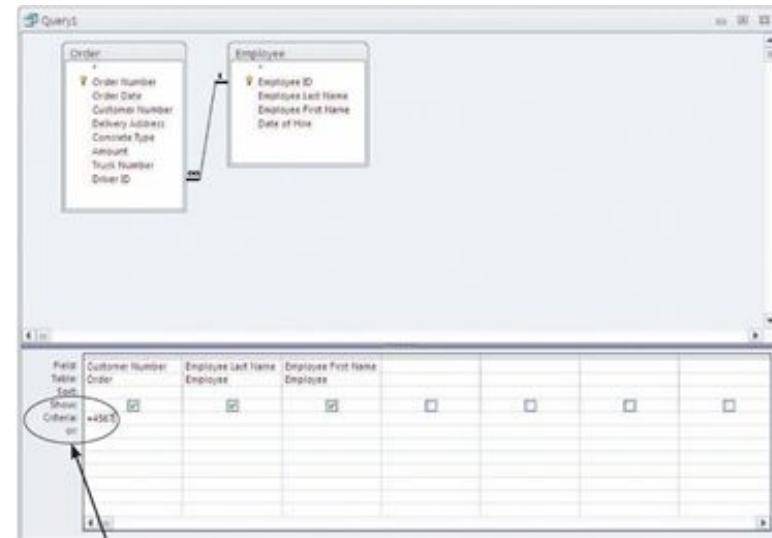
- Software tools you use in a data warehouse environment
  - Query-and-reporting tools
  - Artificial intelligence
  - Multidimensional analysis tools
  - Digital dashboards
  - Statistical tools



# Query-by-Example (QBE) Tool

- You can graphically design the answer to a question

질의어를 만들어  
데이터マイ닝



Customer Number	Employee Last Name	Employee First Name
4567	Evaraz	Antonio
4567	Robertson	John
4567	Robertson	John
4567	Robertson	John

# Report Generator

- You can quickly define formats of reports and what information you want to see in a report

리포트를 만들어  
데이터마이닝

The image shows two windows of the Report Wizard. The left window, titled 'Report Wizard - Select Fields', lists 'Selected Fields' from the 'Order file': Customer Number, Order Number, Order Date, and Amount. The right window, titled 'Report Wizard - Select Report Type', shows 'Report formats' including List, Grouped, Drag, and Query. Below these windows is a sample report titled 'CUSTOMER AND AMOUNT REPORT'.

Customer Number	Order Number	Order Date	Amount
1234	100000	9/1/2004	8
1234	100002	9/2/2004	6
1234	100006	9/5/2004	4
1234	100009	9/7/2004	8
1234	100015	9/12/2004	8
2345	100007	9/6/2004	5
2345	100012	9/9/2004	8
3456	100001	9/1/2004	3
4567	100003	9/3/2004	4
4567	100004	9/4/2004	8
4567	100011	9/9/2004	6
4567	100013	9/10/2004	4
5678	100005	9/4/2004	4
6789	100008	9/6/2004	8
6789	100010	9/9/2004	7
6789	100014	9/10/2004	6
4567	Robertson	John	
4567	Robertson	John	

# Structured Query Language (SQL)

- An international standard for creating, processing, and querying databases and their tables.

SQL을 이용하여  
데이터마이닝

```
SELECT    CUSTOMER.CustomerLastName,  
          CUSTOMER.CustomerFirstName,  
          CUSTOMER.Phone,  
          COURSE.CourseDate,  
          ENROLLMENT.AmountPaid,  
          COURSE.Course, COURSE.Fee  
FROM CUSTOMER, ENROLLMENT, COURSE  
WHERE    CUSTOMER.CustomerNumber  
        = ENROLLMENT.CustomerNumber  
        AND    COURSE.CourseNumber  
        = ENROLLMENT.CourseNumber;
```

# Artificial Intelligence

- Tools to help you  
“discover” information  
and trends

인공지능툴이나  
알고리즘을 이용하여  
데이터マイ닝

## Artificial Intelligence

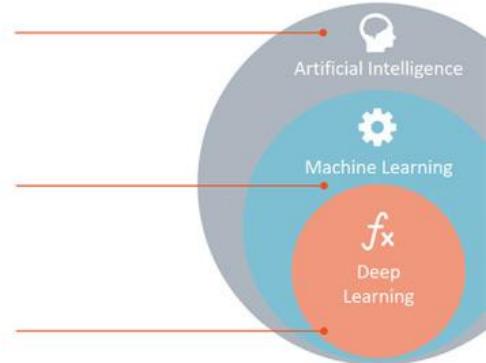
Any technique which enables computers  
to mimic human behavior.

## Machine Learning

Subset of AI techniques which use  
statistical methods to enable machines  
to improve with experiences.

## Deep Learning

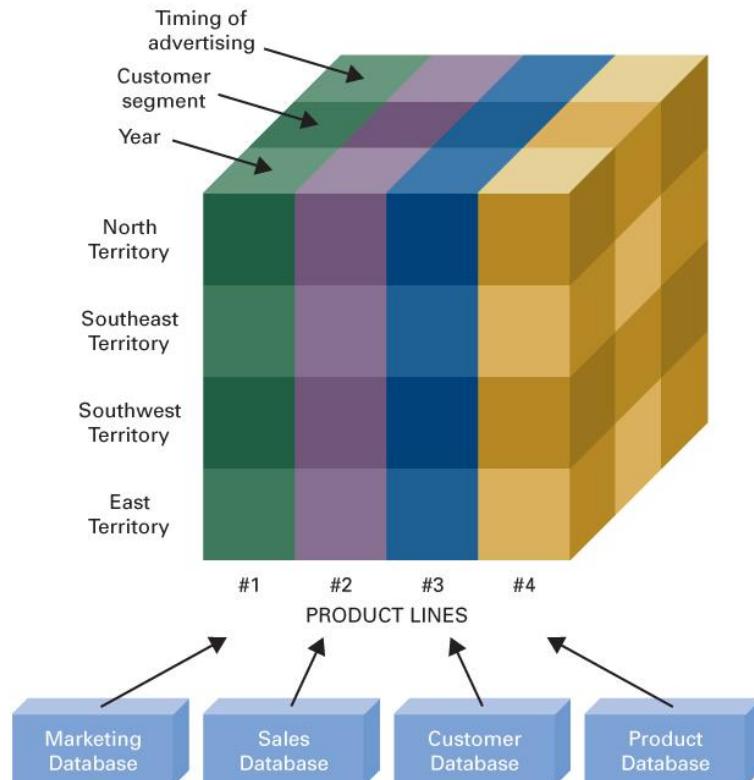
Subset of ML which make the  
computation of multi-layer neural  
networks feasible.



# Multidimensional Analysis (MDA tools)

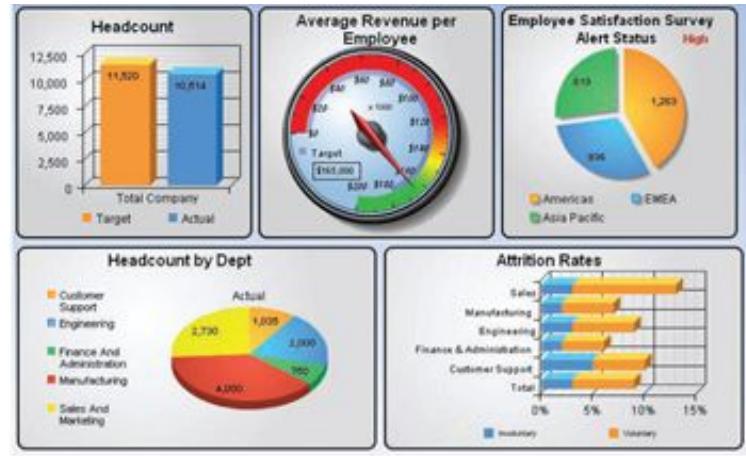
- Slice-and-dice techniques for viewing multidimensional information

다차원정보를 잘라서  
데이터마이닝



# Digital Dashboard

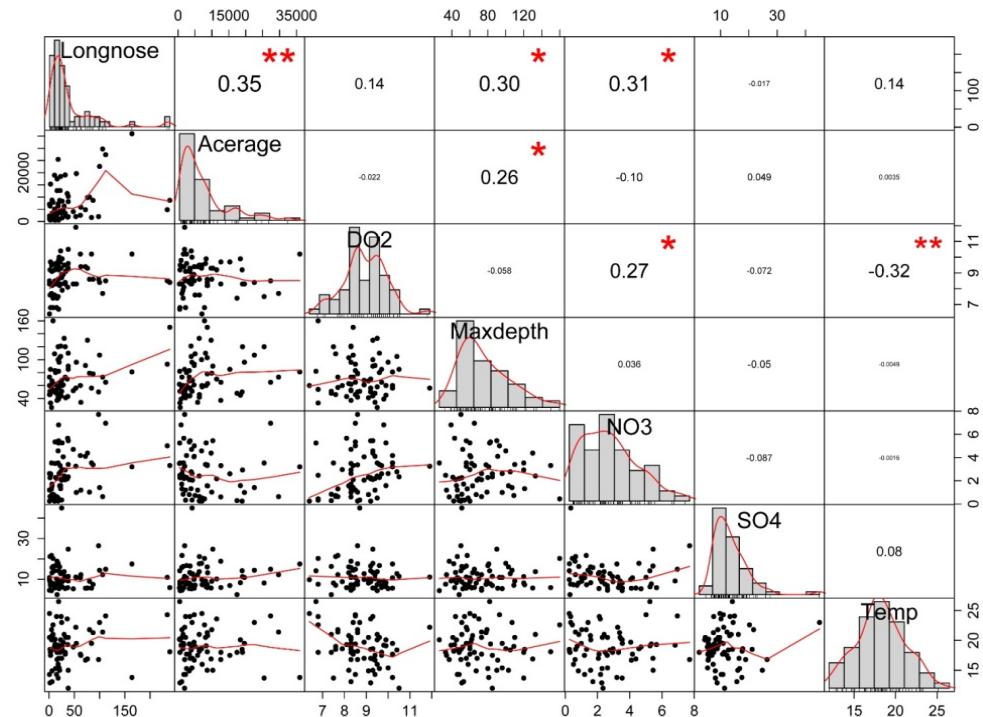
- Displays key information on a computer screen tailored to the needs and wants of an individual 필요에 따라 주요성과지표를 요약
- Key performance indicator (KPI) – most essential information used in any analytics initiative 주요성과지표를 시각화



# Statistical Tools

- Applying mathematical models to data warehouse information.  
e.g., linear regression

회귀분석 같은  
수학모델을 이용하여  
데이터マイ닝

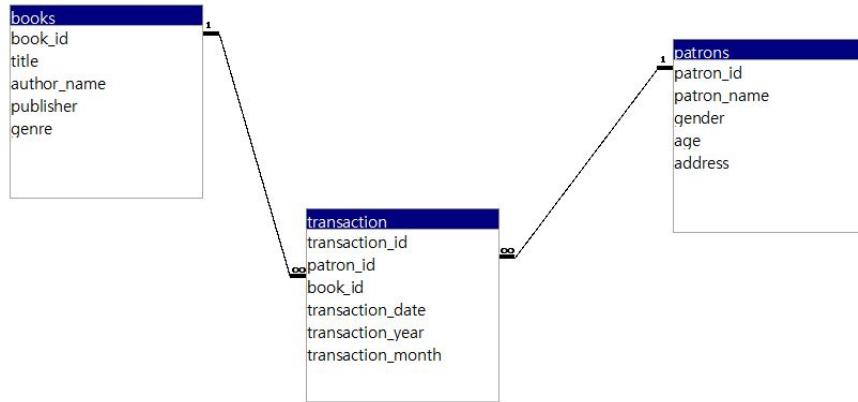


# Exercise #1

- Create a simple library database with three tables: books, patrons, and transaction

세개의 테이블을 가진 도서관 데이터베이스를 만드시요

- You can import the given text files into the tables. If you can't directly import the text files into Access, copy and paste into Excel then import the Excel file
- 데이터를 텍스트파일로 만들어서 임포트하던가  
엑셀파일로 만들어서 임포트할수 있음



# Sample SQL Code for books and patrons

```
create table books (
    book_id char(13),
    title varchar(100),
    author_name varchar(50),
    publisher varchar(50),
    genre varchar(50),
    constraint pk_book_id
primary key (book_id));
```

```
create table patrons (
    patron_id char(5),
    patron_name varchar(80),
    gender char(1),
    age number,
    address varchar(200),
    constraint pk_patron_id
primary key (patron_id));
```

# Sample SQL Code for transaction

```
create table transaction(
    transaction_id char(5),
    patron_id char(5),
    book_id char(13),
    transaction_date char(50),
    transaction_year char(4),
    transaction_month varchar(2),
    constraint pk_transaction_id primary key(transaction_id),
    constraint fk_patron_id foreign key(patron_id) references patron(patron_id),
    constraint fk_book_id foreign key(book_id) references book(book_id));
```

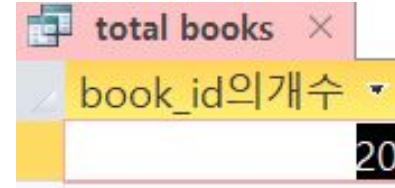
# Exercise #1

- Create relationships among tables and produce a ER Diagram report.

데이터 사이의 관계를 만들고  
ER다이어그램 리포트를 만드시요

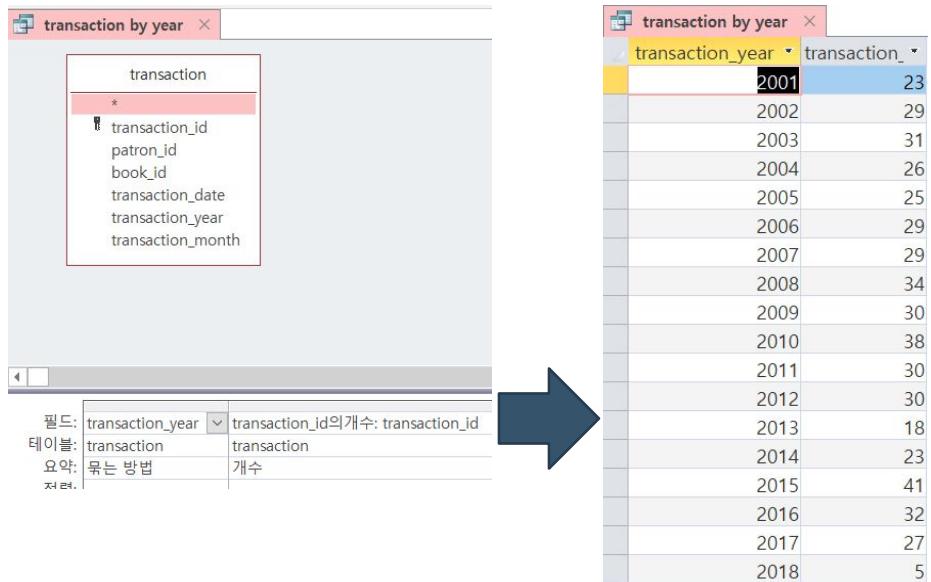
- Create queries to answer the following questions 다음의 질문을  
답할 수 있는 질의어를 만드시요

- Total number of books?
- Total number of patrons?
- Total number of transactions?



# Exercise #1

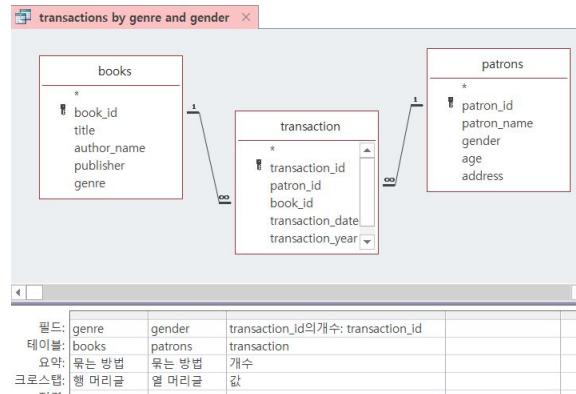
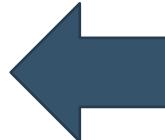
- Number of transactions by year?
- Number of transactions by genre?
- Number of transactions by month?



# Exercise #1

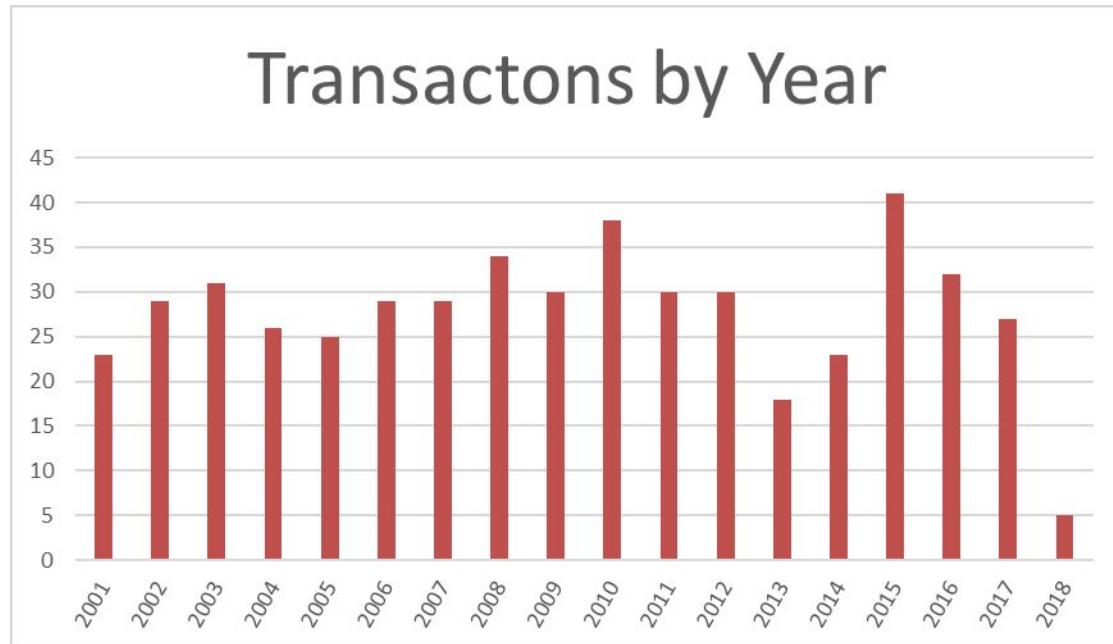
- Number of transactions by genre and gender?
- Number of transactions by month and gender?

genre	F	M
소설	47	53
아동도서	27	23
에세이	91	109
학습서	73	77



# Number of Transactions by Year

transaction_year	transaction_
2001	23
2002	29
2003	31
2004	26
2005	25
2006	29
2007	29
2008	34
2009	30
2010	38
2011	30
2012	30
2013	18
2014	23
2015	41
2016	32
2017	27
2018	5



# Exercise #1

- Build a “Dashboard” that summarize information about the patrons, books, and transactions

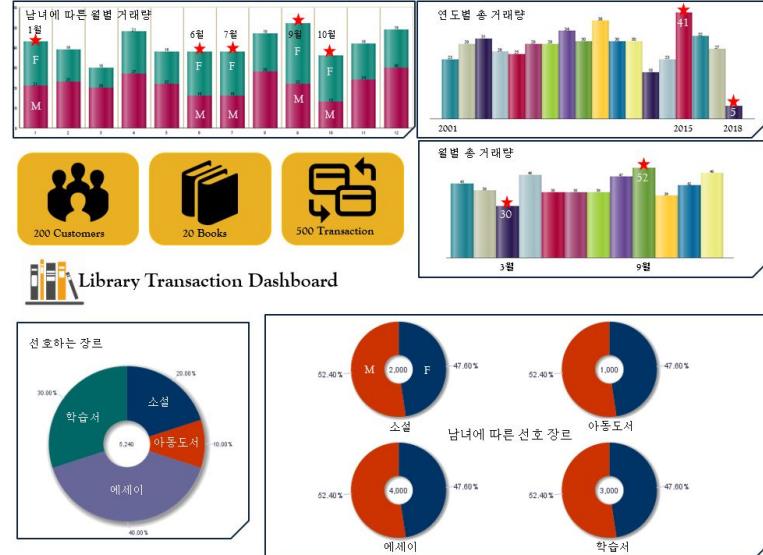
고객, 책, 거래내역에 대한 정보를  
요약하는 대시보드를 만드시요

- Export the result of queries to Excel and create appropriate charts for your data

쿼리결과를 엑셀로 엑스포트해서 적절한  
그래프를 그리시요

- Summarize the result and charts using a PowerPoint slide

파워포인트를 이용해서 결과와 차트를  
요약하시요



# Exercise #1

- Complete the chart tutorials on

[https://www.tutorialspoint.com/tableau/tableau\\_bar\\_chart.htm](https://www.tutorialspoint.com/tableau/tableau_bar_chart.htm)

- Complete the map tutorial on

[https://help.tableau.com/current/pro/desktop/en-gb/maps/howto\\_simple.htm](https://help.tableau.com/current/pro/desktop/en-gb/maps/howto_simple.htm)

# Exercise #1

## Chipotle Example

- Analyze the Chipotle data using graphs
  - Scatter plot
  - Histogram
  - Bar chart, horizontal bar chart, stacked bar chart
  - Box plot
  - Area chart
  - Heat map

# Exercise #1

## Occupation Example

- Analyze the Occupation data using graphs
  - Scatter plot
  - Histogram
  - Bar chart, horizontal bar chart, stacked bar chart
  - Box plot
  - Area chart
  - Heat map

# Exercise #1

## World Food Facts

- Analyze the world food fact data using graphs
  - Scatter plot
  - Histogram
  - Bar chart, horizontal bar chart, stacked bar chart
  - Box plot
  - Area chart
  - Heat map