



# 1주차: 데이터 사이언스 개요

**ChulSoo Park**

School of Computer Engineering & Information Technology

Korea National University of Transportation

E-Mail : pcs8321@naver.com



## 박 철 수 대한민국산업현장 교수/ 공학박사

### ❖ 경력

- 2019~현재 한국정보통신기술사협회 사업단장
- 2017~현재 대한민국산업현장교수(정보통신, 고용노동부)
- 2018~현재 국가기술자격정책 심의위원(과기부)
- 2020~현재 고경력 과학기술인(과기부)
- 2019~현재 4차 산업혁명,스마트팩토리(전문위원)
- 2020~현재 스마트팜 컨설턴트(ICT, 데이터 활용 분야)
- 2019~현재 과학기술분야 진로 컨설턴트(과기부,한국창의재단,노동부)
- 2018~2018 대학생 정보통신( ICT) 멘토(과학기술정보통신부)
- 1987~2020 농협(서울우유협동조합) 정보관리최고책임자(CIO/CISO)

### ❖ 학력

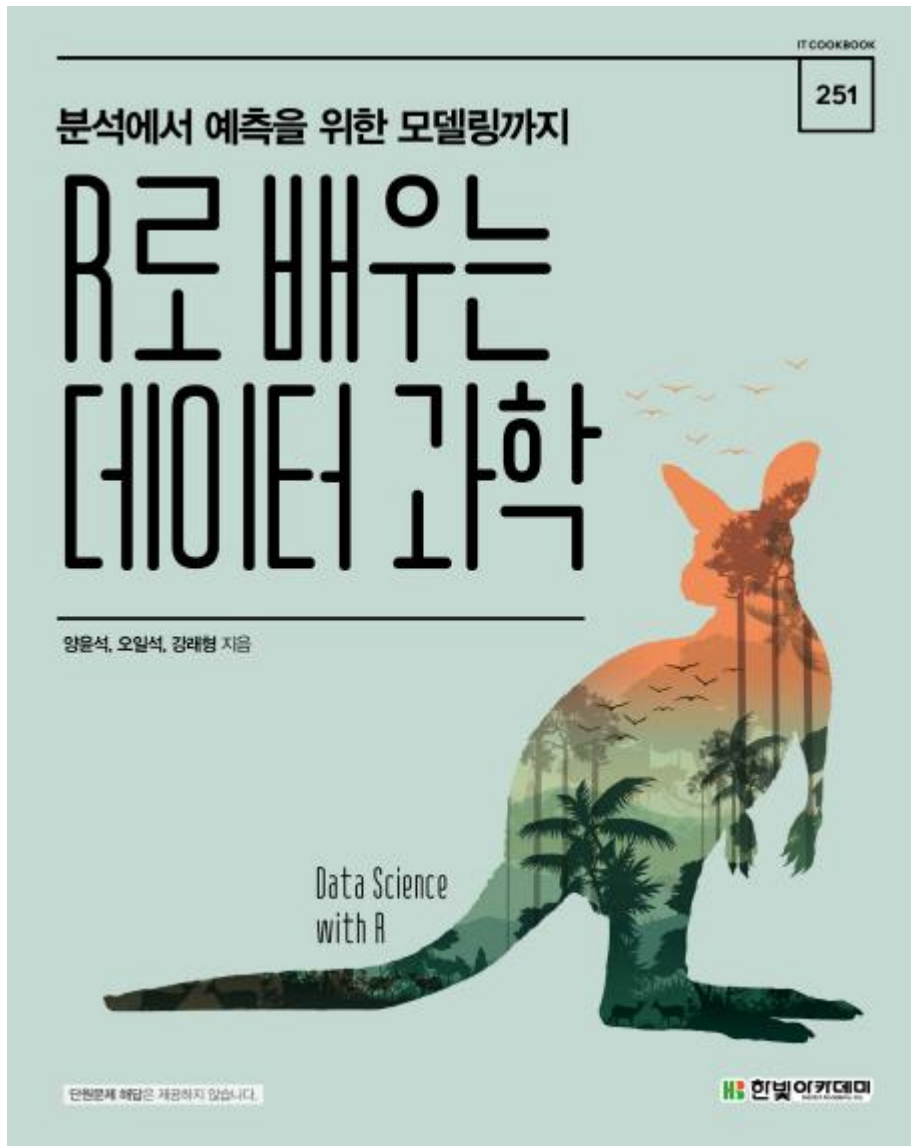
- 2011~2016 서울과학기술대학교 공학박사

### ❖ 저서

- 2019년 NCS학습모듈(빅데이터,인공지능(AI) 집필 책임)
- 2020년 NCS학습모듈(지능형영상정보처리,생체인식 집필 책임)



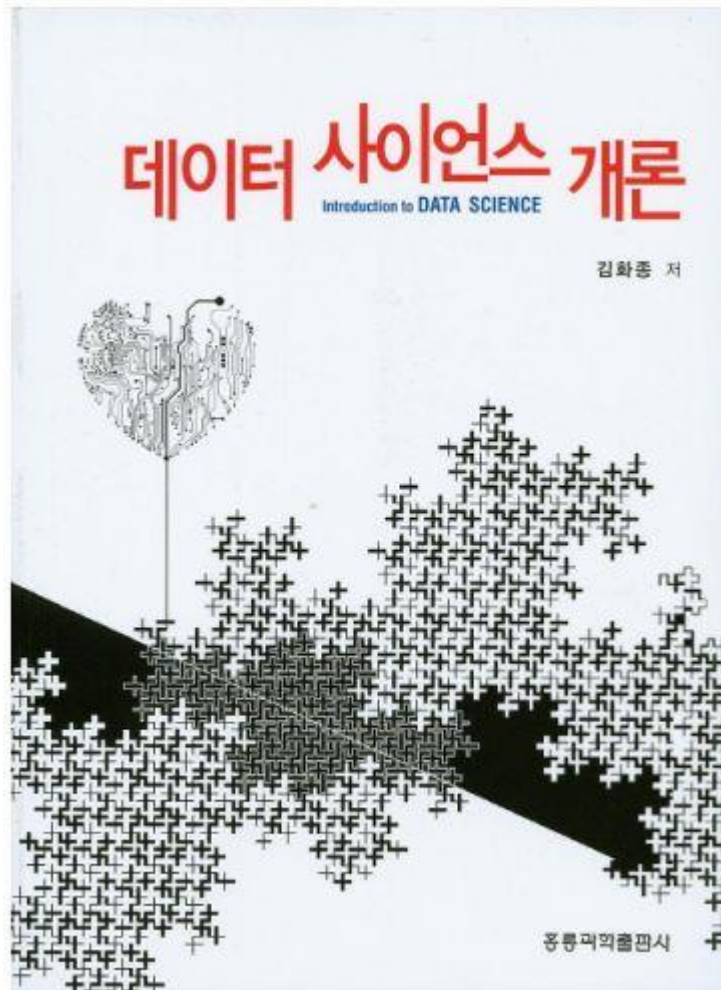
# 교재 소개



저 자 : 양윤석, 오영일, 강래형  
출판사 : 한빛아카데미  
발행일 : 2020.03.30



# 참조 자료



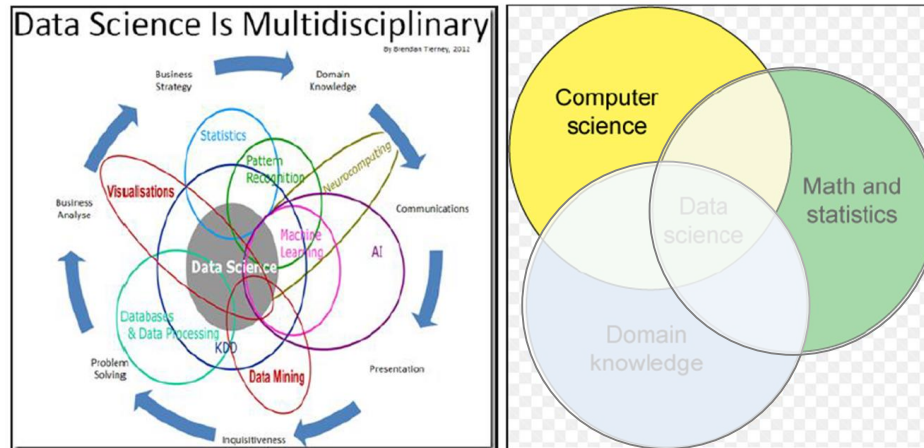
저 자 : 김화종  
출판사 : 홍릉과학출판사  
발행일 : 2014.08.31



# 학습목표 (교과목)

인터넷, IoT, 스마트, 웨어러블 시대의 도래로 빅데이터 시대에서

1. 데이터과학 분야의 전문가로서 갖추어야 할 다양한 학문적 이론과 실무지식을 익히고
2. 데이터 과학자로서 요구되는 창의성, 통찰력, 올바른 윤리의식 등을 갖추며
3. 데이터의 가치를 창출하기 위한 융합적 사고를 기반으로 데이터 분석, 설계 및 예측 능력을 갖춘 데이터 시대를 선도하는 전문가의 능력 함양



# 강의 계획표

주	해당 장	주제
1	1장	데이터 과학 소개
2	2장	데이터 과학 학습을 위한 도구, 데이터 시각화 맛보기
3	3장	R 기본 문법
4	4장	데이터 취득과 정제
5	5장	데이터 가공
6	(참고서적:7장)	데이터 분석 방법, 데이터 마이닝의 이해
7	6장	데이터 시각화
8	필기	중간고사(30%)
9	7장	모델링과 예측 : 선형 회귀
10	8장	일반화 선형 모델
11	9장	분류를 위한 모델
12	10장	모델의 성능 평가
13	11장	텍스트 마이닝
14	12장	구글 플레이 스토어 앱 데이터를 이용한 실전 프로젝트
15		보강 주간
16	필기	기말고사(40%)



# 학습목표 (1주차)

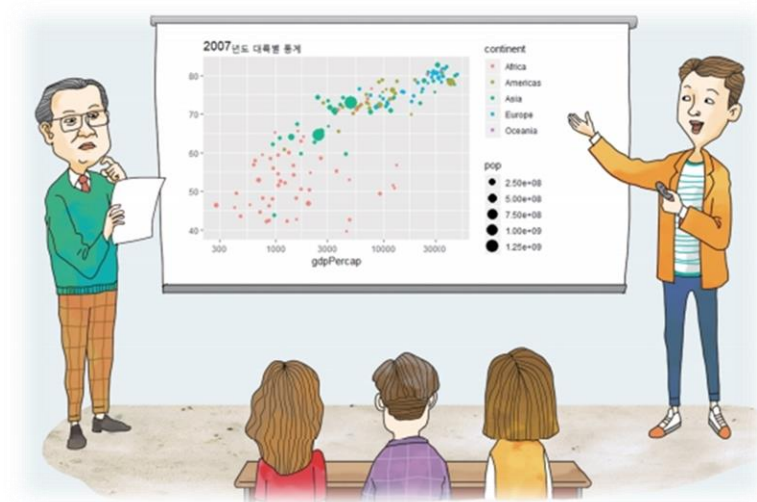
- ❖ 빅데이터 시대의 이해
- ❖ 데이터 사이언스의 정의
- ❖ 데이터 사이언스 절차
- ❖ 데이터 사이언스 분야
- ❖ 데이터 사이언티스트 정의



# 01

## CHAPTER

# 데이터 과학 개요



## CONTENTS

- 1.1 데이터 홍수 시대
- 1.2 데이터 과학 열풍
- 1.3 데이터 과학이란?
- 1.4 데이터 과학의 절차
- 1.5 데이터 과학 관련 분야
- 1.6 데이터 과학 자원
  - 요약 및 역사 속의 데이터 과학





# 4차산업혁명 : 사물인터넷(IoT),빅데이터(Big Data),인공지능(AI) → 의사결정,현업활용

출처: 제4차 산업혁명, 김진호,최용주, p25



## 4차산업혁명 환경하의 디지털 경영혁신

이성열 · 강성근 · 김순신 지음



기업은 변화의 혁신을 통하여 성장한다. 4차 산업혁명과 디지털 혁신으로 바뀌고 있는 경영환경의 급격한 변화에서 전통기업들이 일어나 신속하게 디지털 경영혁신을 통하여 변화에 나갈 수 있는 가는 향후 기업생존의 핵심이 될 것이다.



	제1차 산업혁명	제2차 산업혁명	제3차 산업혁명	제4차 산업혁명
시기	18세기	19~20세기 초	20세기 후반	20세기
특징	증기기관 기반의 '기계화 혁명'	전기 에너지 기반의 '대량생산 혁명'	컴퓨터와 인터넷 기반의 '디지털 혁명'	사물인터넷(IoT)과 빅데이터, 인공지능(AI) 기반의 '만물 초지능 혁명'
영향	수공업 시대에서 증기기관을 활용한 기계가 물건을 생산하는 기계화 시대로 변화	전기와 생산조립 라인의 출현으로 대량생산 체계 구축	반도체와 컴퓨터, 인터넷 혁명으로 정보의 생성·가공·공유를 가능케 하는 정보기술시대의 개막	사람, 사물, 공간을 연결하고 자동화 지능화되어 디지털·물리적·생물학적 영역의 경계가 사라지면서 기술이 융합되는 새로운 시대

자료 : 미래에셋대우 글로벌투자전략부



**危** 위험

Coronavirus Economy Impacts

**경제 戰時 상황**

COVID-19

**포트코로나 시대로의 전환기**

'보이지 않는 바이러스가 세상을 송두리째 바꾸어 우리의 일상을 근본적으로 변화'

K-생역은 국제적 모범사례

**제4차 산업혁명의 가속화**

**機** 기회

**코로나-19로 인한 지능화의 가속화**

[진단] AI기반 의료영상진단속 판독

[역학조사] 자가격리자 안전보호 앱

[치료] AI기반 치료제/백신 조기개발

[원격수업] 온라인 개학, 교육공백 최소화

[무인시스템] 비대면, 무인화 시스템 확산

**디지털 뉴딜**

5G 인프라 조기 구축

데이터 인프라 구축

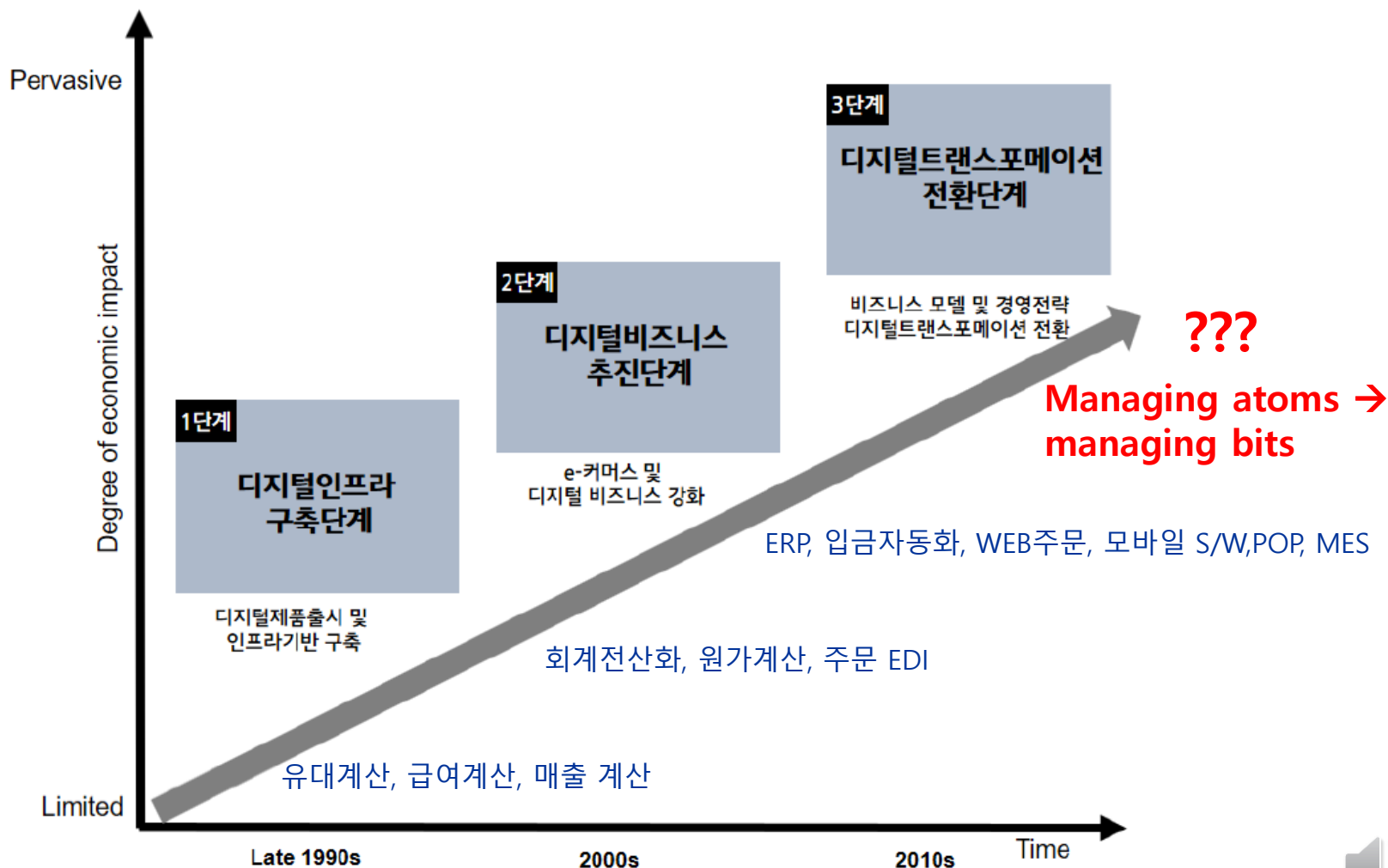
**비대면 산업 육성**

국가 기반시설 지능화



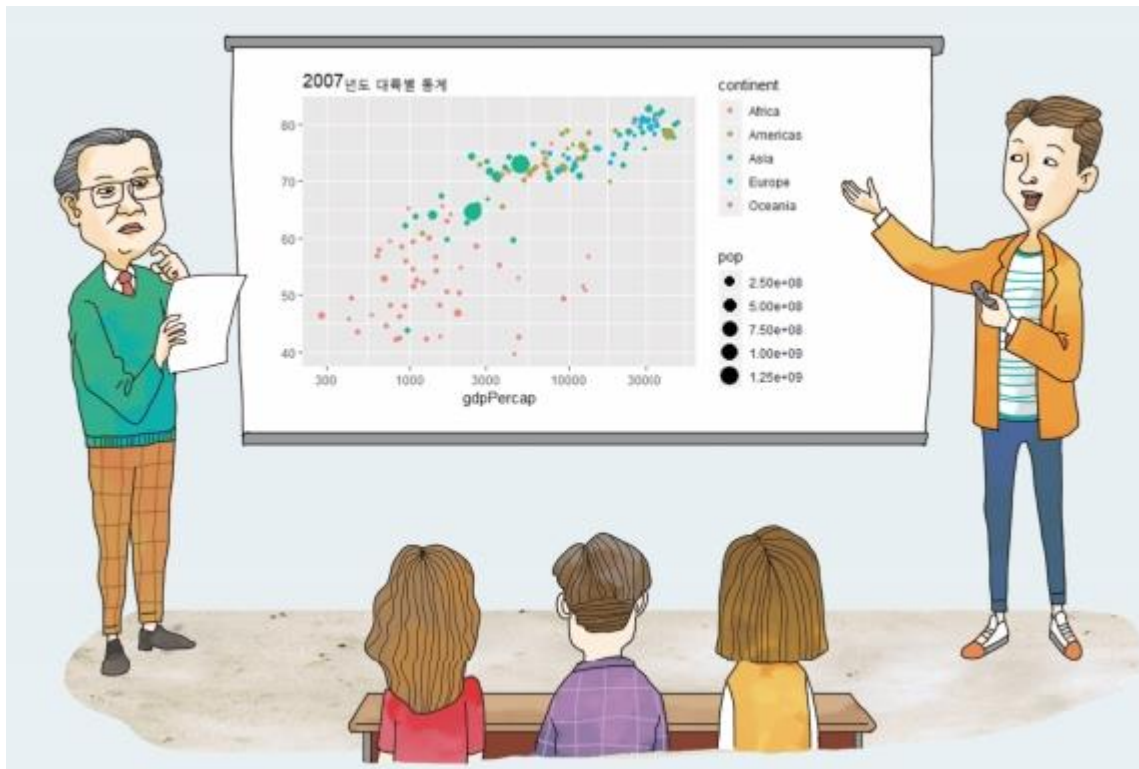
## 디지털 트랜스포메이션 단계별 진화

## 경쟁력 제고



## ■ 데이터 과학

- 데이터로부터 유용한 정보와 통찰을 끄집어내고 합리적인 의사결정을 돕는 현대적인 학문 분야





## 1-1 데이터 수집 및 정리<사양분야> 컨설팅 수행모듈

낙농 사양

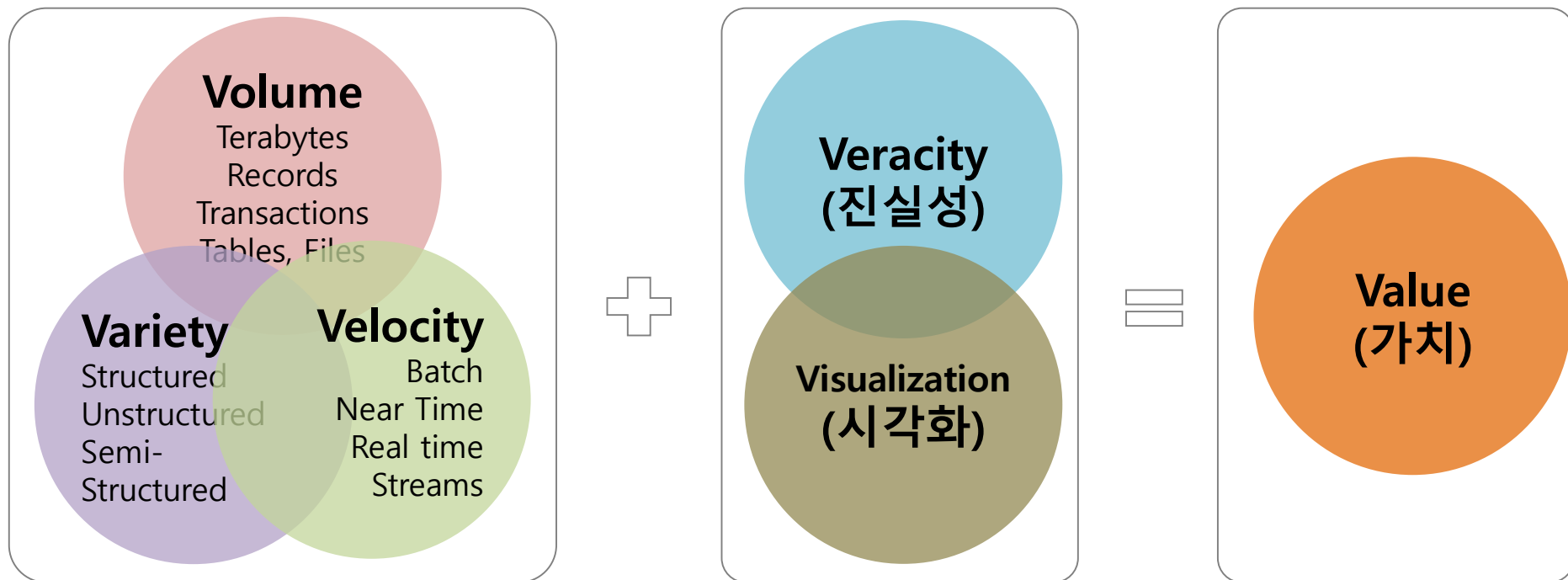
컨설팅 수행모듈 안내

컨설팅 카테고리		컨설팅 수행 모듈					
분류	중분류	기본모듈		선택모듈			
공통	사적진단	축사/시설 진단	추가 사용현황				
	종합진단	변경 컨설팅 수행설계	변경 성과 분석				
	장비SW/Data 분석	Data활용 기초교육	장비활용 교육	제어SW 활용교육	Data활용 중급	Data활용 심화	
수의	질병 분석	임상진단	질병역학분석 /방역,위생	처방관리 /백신표프로그램	유방염분석	혈액검사	발급진단
	변식 분석	변식현황	변식장애 진단	임신감정	난소진단	자궁진단	MPT재형 /BCS
사양	사료/급이/급수 분석	급이진단	급이계획	급이/BCS분석 (사료성분분석)	TMR배합비	P/F분석 MLN분석	
	환경 분석		MPT분석 (대사판정)	위충만/분변 /자초진황지수	도/폐사 분석	분뇨처리/ 수질분석	
경영	손익/원가 분석	손익 분석	원가분석 (유사비)	분리 육성비분석	투자효율성 분석		
	증장기 경영예측		자산관리/현금흐름	작업계획			



## 축산 빅데이터 컨설팅 플랫폼 개요도





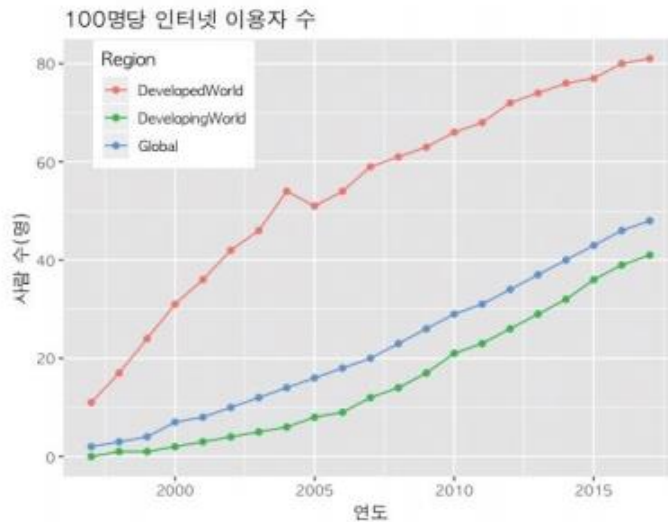
- ❖ 크기(Volume) : 방대한 양의 데이터(테라, 페타바이트 이상의 크기)
- ❖ 다양성(Variety) : 정형(DBMS 등)+비정형(SNS, 동영상, 사진, 음성텍스트 등)
- ❖ 속도(Velocity) : **실시간으로 생산되며, 빠른 속도로 데이터를 처리 분석**
- ❖ 진실성(Veracity) : 주요 의사결정을 위해 **데이터의 품질과 신뢰성 확보**
- ❖ 시각화(Visualization) : 복잡한 대규모 데이터를 시각적으로 표현
- ❖ 가치(Value) : 비즈니스 효율을 실현하기 위해 궁극적 가치 창출



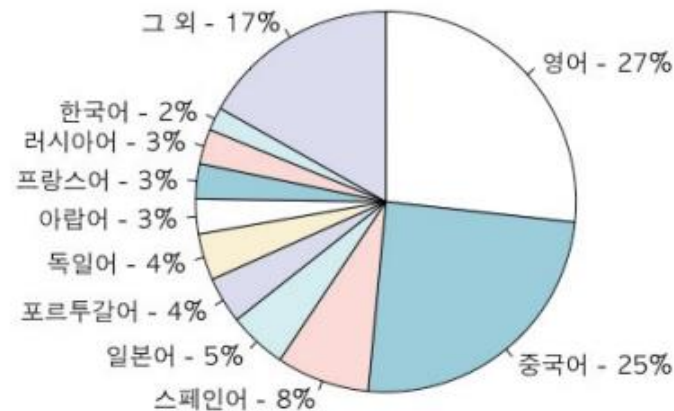
## ■ 데이터 홍수 시대

- 키보드를 두드리거나 마우스를 움직이는 순간 데이터 발생
- 스마트폰을 주머니에 넣고 여기저기 돌아다니기만 해도 데이터 발생 → 서버에 저장

## ■ 주요 원인 1: 인터넷



(a) 1996~2018년 인터넷 이용자 추이



(b) 언어별 인터넷 이용자 비율

그림 1-1 인터넷 관련 세계 통계

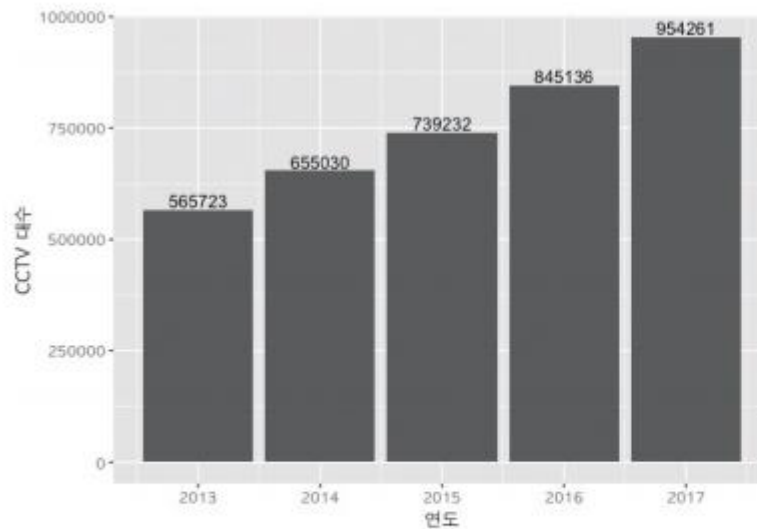




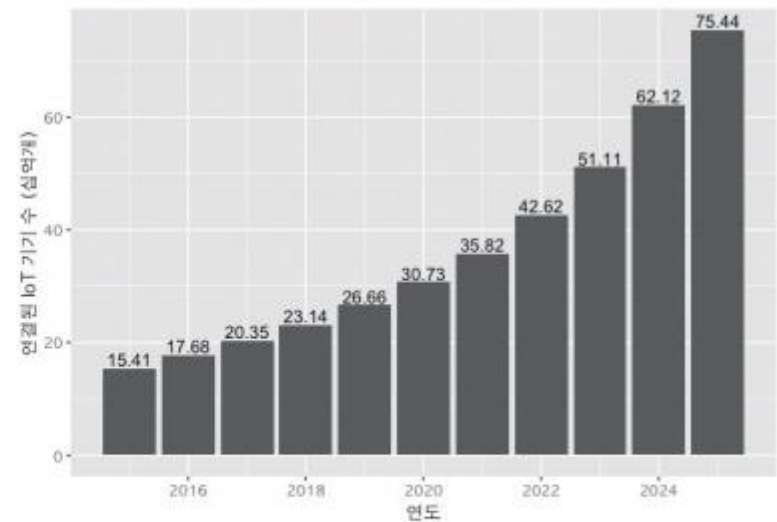
# 1.1 데이터 홍수 시대

## ■ 주요 원인 2: 센서

- 예) 자동차, 자율 주행 자동차는 주행하며 센싱 데이터를 서버로 전송



(a) 대한민국 CCTV 대수



(b) 전 세계 IoT 연결 장치 대수

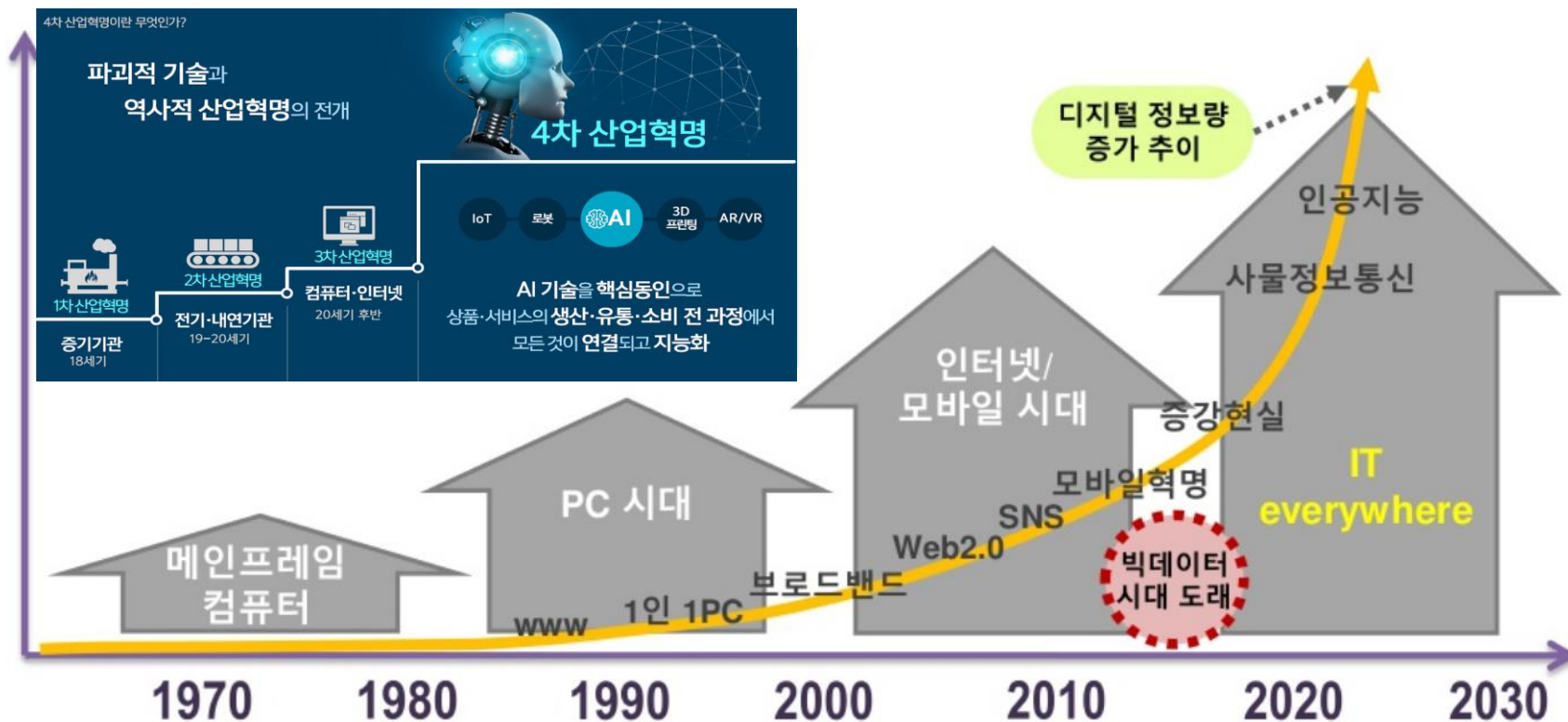
그림 1-2 센서 대수 추이 통계



## 1.1 데이터 홍수 시대



# 1.1 데이터 홍수 시대



# Thank you

