

Media & Data Science

Lecturer: Changjun Lee
changjunlee@hanyang.ac.kr

College of Communication
Media Informatics



About Lecturer

Changjun Lee

- **Affiliation:** Asso. Prof. Hanyang Univ. ERICA. College of Comm. Media & Social Informatics
- **Research Area:** Media & Innovation, Technology Management, Public Policy, Economic Geography
 - Home: <https://cjleelab.netlify.app/>
- **Teaching:** Data Science
- **Other jobs:** Father of three kids
- **Things I love:**
 - Coffee #Beer #Whiskey
 - TV-shows #헤어질결심 #더글로리 #오징어게임 #사랑의이해
 - Research #연구에진심 #코딩덕후
 - Chat #의외로수다맨 #ENFP
 - Play #GUITAR #TENNIS #포켓몬고
 - Travel #방랑벽 #낮선곳에서오는기쁨

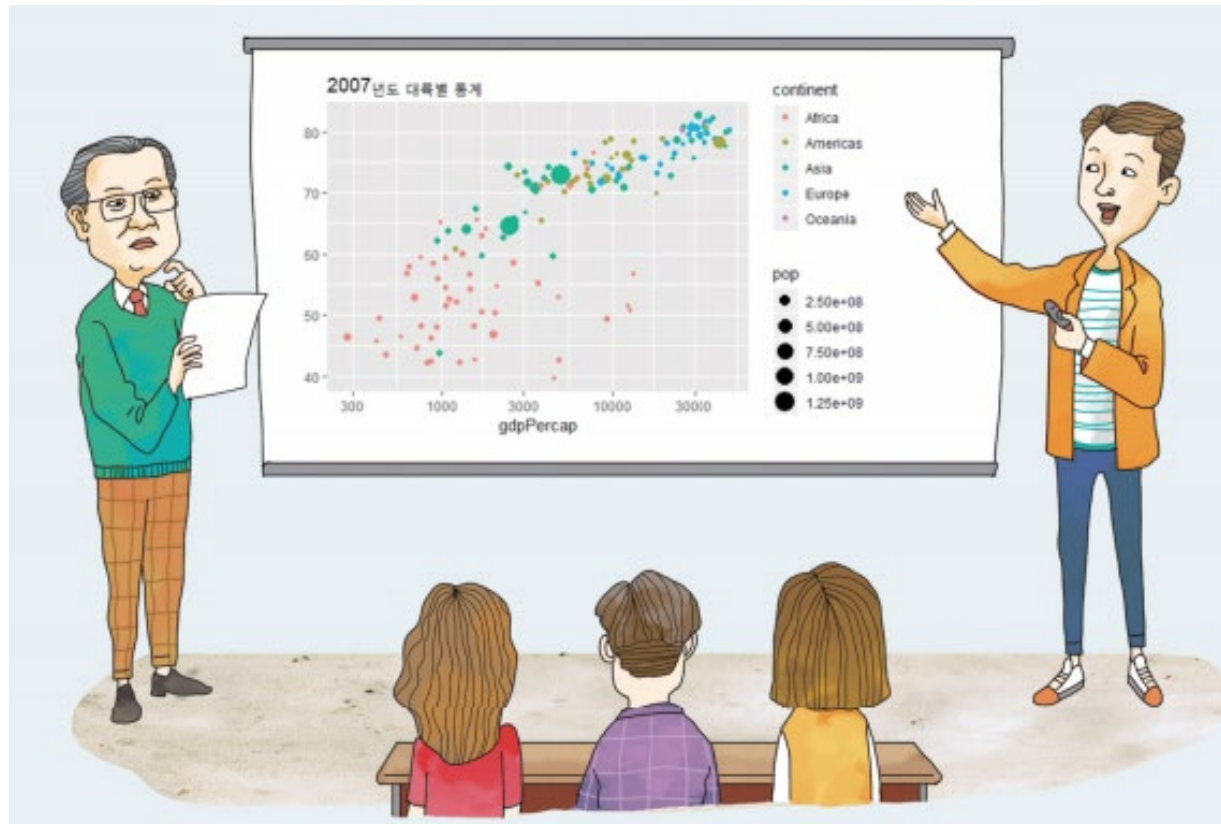


How about you?

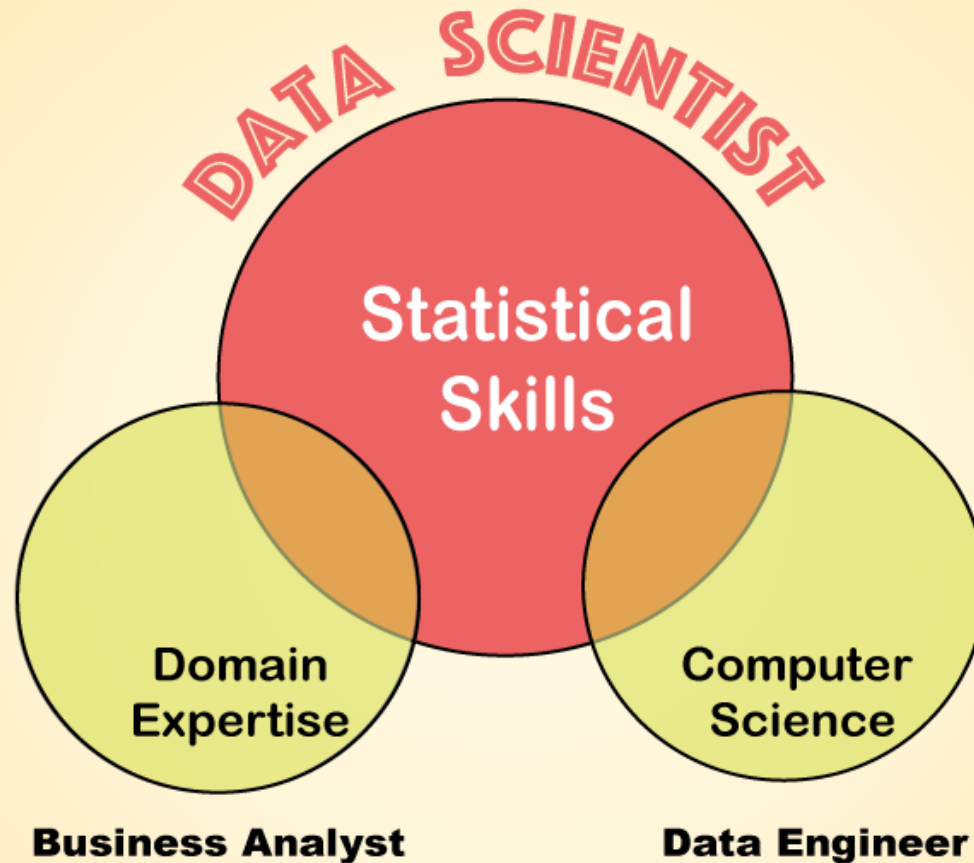
- Things you love
- Majors
- MBTI
- Goal for this course (Motivation) if u have any
 - 예) 석사 학위 논문에 데이터 사이언스 응용

Data-Science

Modern academic area helping people to draw useful information and intuition so that making them to make reasonable decisions



Requirement?



비전공자(#머글)이 데이터사이언티스트(#마법사)가 되면 더 좋은 이유?



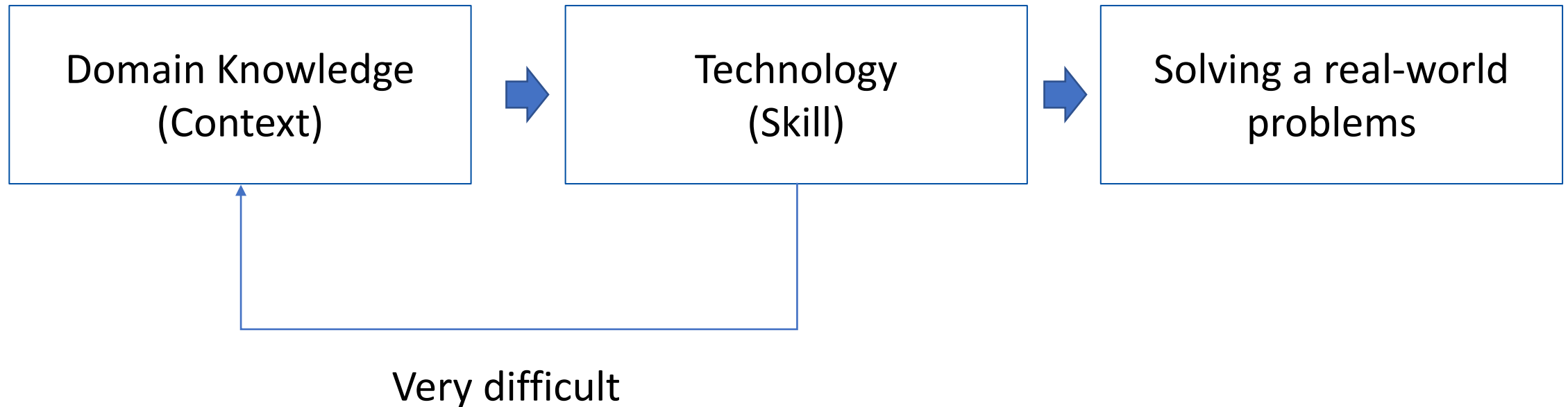
그런데 출연사건 해명기

신과 해커 이두희 (ENG CC)

의심 17 정공 부작을지가

진재 해커가 도17까지

Non-CS major being a data-scientist has an advantage



야, 너두! 할 수 있어.



문과 출신의
비전공자가
데이터 업계에
취업할 수 있을까?

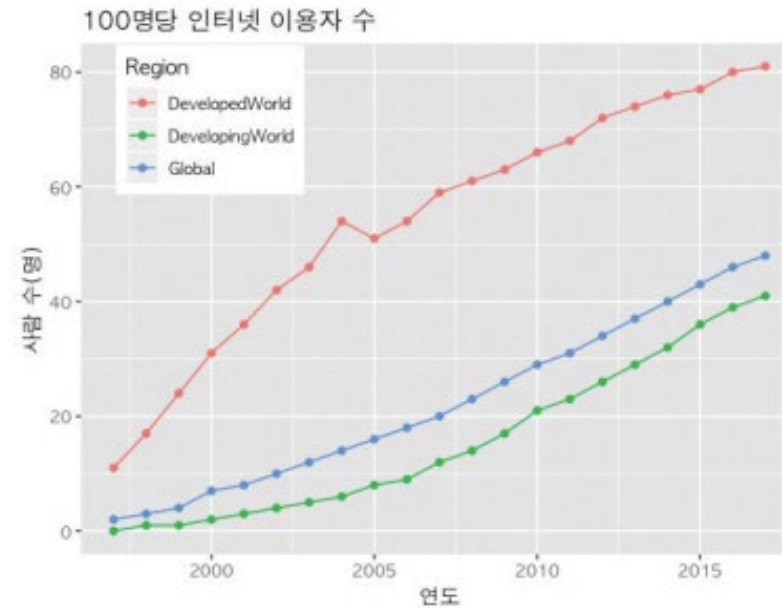
“비전공자들이 오히려 두각을 나타내는 사례들이 나오고 있음.”

비전공자로서 데이터 사이언티스트가 갖춰야 할 최소한의 기술

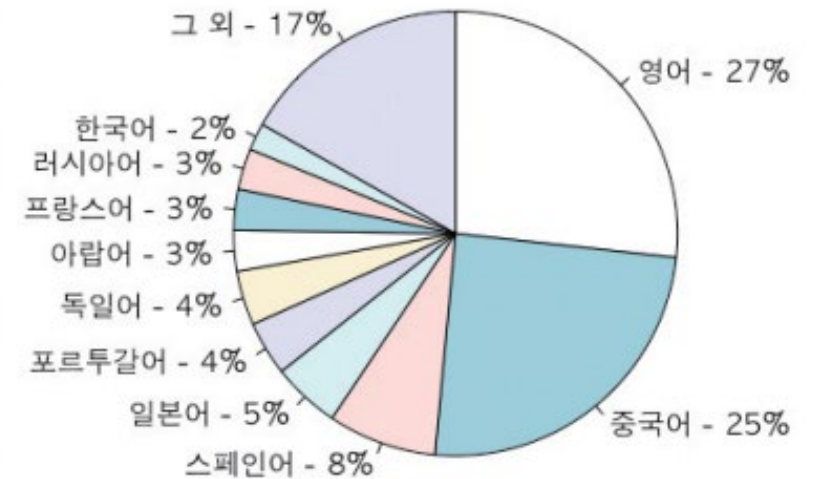
데이터 사이언티스트를 전공하거나 아직 전문지식을 갖추지 않았다고 해서 데이터 사이언스를 다루지 못하는 건 아닙니다. 하지만 최소한의 기술적 지식을 갖추고 있어야 합니다. 빅데이터 전문 업체의 현직 데이터 사이언티스트의 말에 따르면 최소한 원하는 데이터를 분석하고 결론을 내릴 정도의 기술을 갖춰야 합니다. 따라서 데이터베이스에서 사용하는 언어인 SQL 언어나, R, 파이썬과 같은 스크립트 언어를 알아야 합니다. 스스로 비즈니스 지표를 만들고 활용할 수 있어야 합니다.

Data flooding era

- **Data flood**
 - Clicking, Watching, Buying... Any kinds of activities online are creating data
 - If you just walk around with your smart-phone → Also DATA created → Saved in the server
- **Main reason 1: Internet**



(a) 1996~2018년 인터넷 이용자 추이

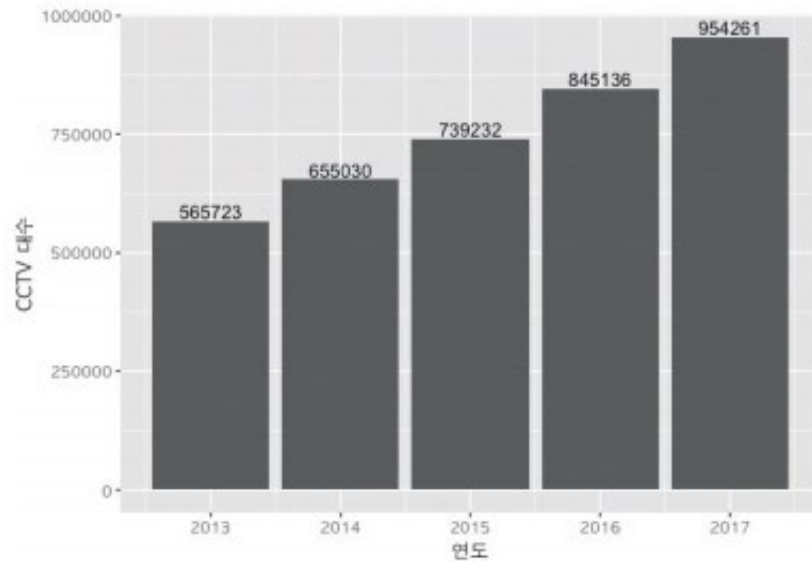


(b) 언어별 인터넷 이용자 비율

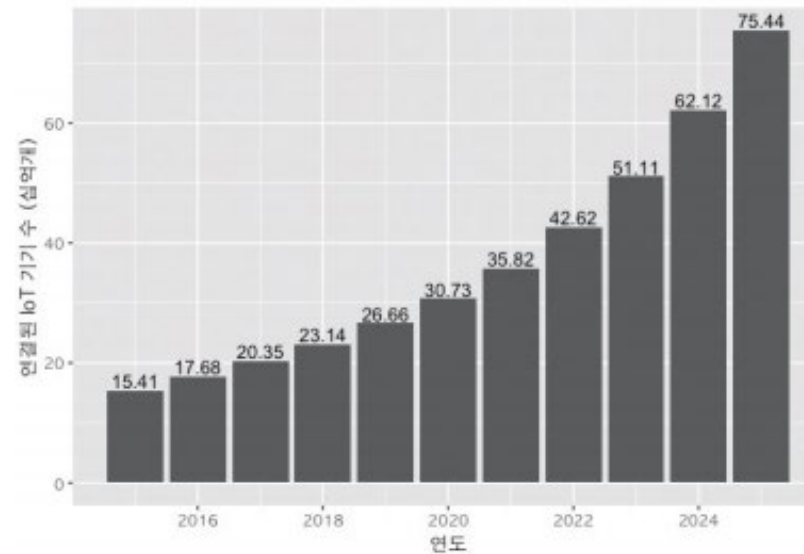
그림 1-1 인터넷 관련 세계 통계

Data flooding era

- Main reason 2: Sensor
 - CCTV
 - Wearable devices
 - Smart-home system
 - Car → Auto-driving cars (TESLA)



(a) 대한민국 CCTV 대수



(b) 전 세계 IoT 연결 장치 대수

그림 1-2 센서 대수 추이 통계

Data flooding era

- **Main reason 3: Smart-phone**

- Smart-phone has Mobile USIM, Internet, Sensors, GPS and so on
- Being used in a variety of areas like government, armies, hospitals, schools...
→ Quality and Quantity of DATA increase

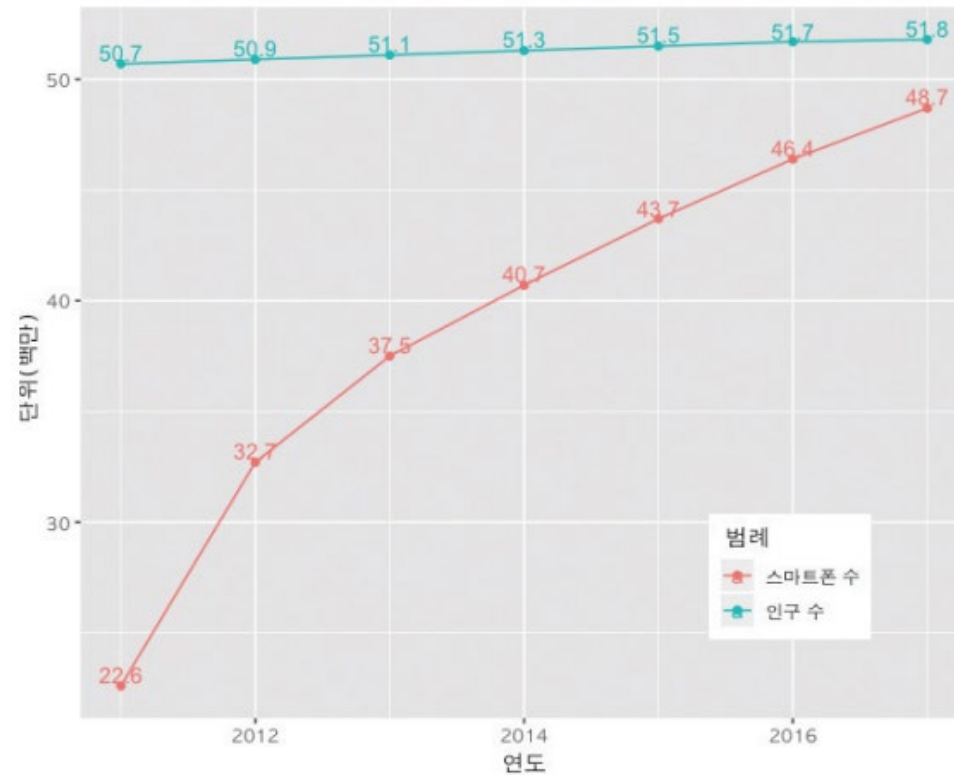
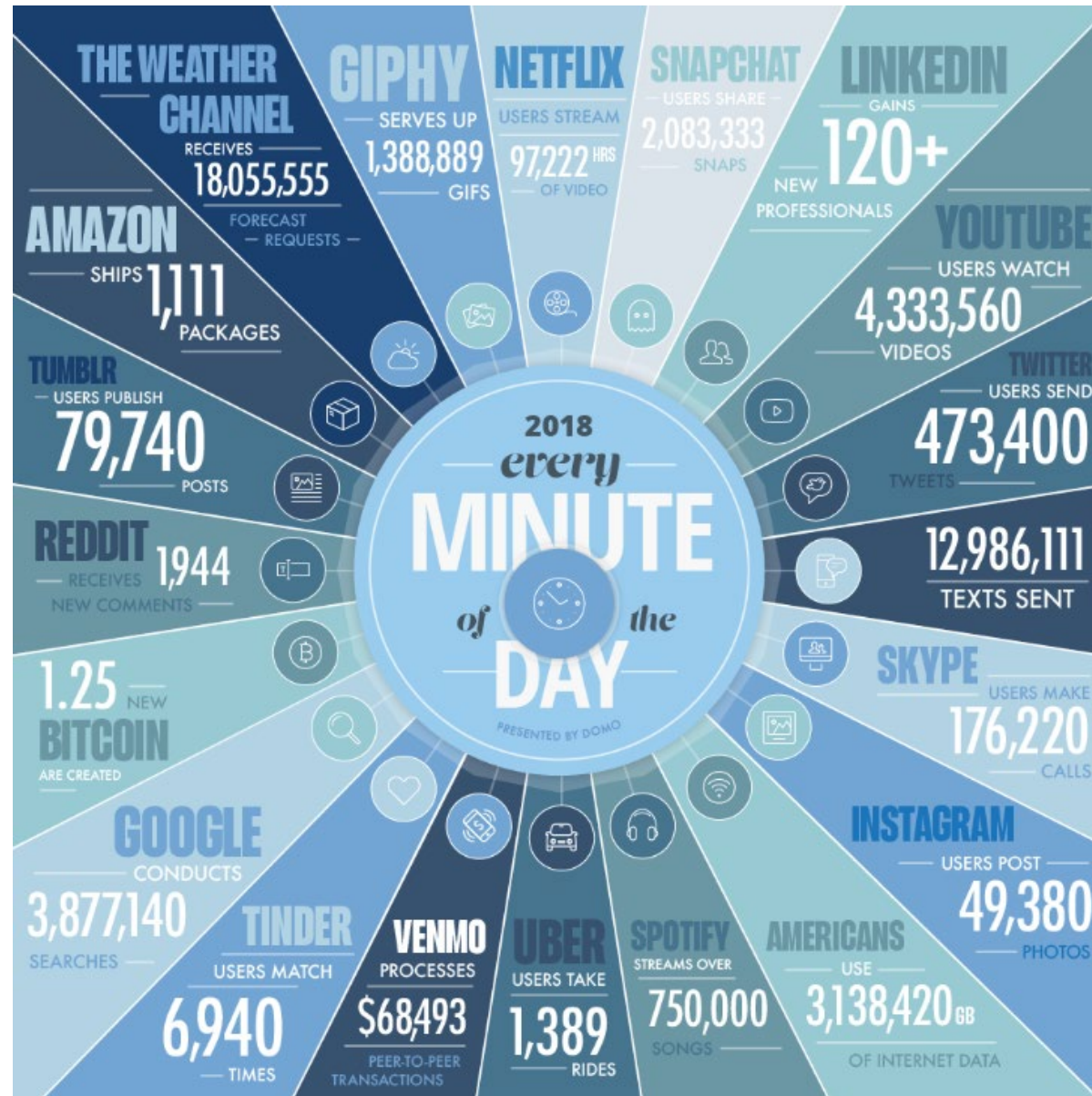


그림 1-3 대한민국 스마트폰 통계

Data flooding era

- Data Never Sleeps



Data Science RULES!

- **Growth of Kaggle (The biggest data science community)**
 - Members: over 1 Million in June 2017 (194 countries)
 - Open data over 13 thousands in 2019
 - Data science competition
 - Many competitions are going on
 - The hugest jackpot winning was 100,000\$ “using news to predict stock movements”
- **Data science area: Strong Opening Sprit**
 - Kaggle: DATA, Source code, Mentoring
 - DATA science language: R and Python
 - Public data portal in Korea (data.go.kr)

Data Science RULES!

데이터셋

제공신청

활용사례

정보공유


이용안내

데이터를 검색해보세요!



5.인천광역시북구도서관

최근 사회현안 및 이슈

미세먼지 

데이터를 확인해 보세요!

국가중점데이터

데이터 카테고리



건축정보



교통사고정보



국민건강정보



상권정보



수산정보



실시간 수도정보



농수축산가격정보



등산로 정보



부동산종합정보



통합재정정보



지방행정정보



부동산거래정보



식의약품종합정보



지방재정정보



법령정보



+ 더보기

「데이터1번가」

여러분이 필요로 하는 공공데이터를 자유롭게 신청할 수 있는 데이터소통창구

[신청하기](#) [자세히보기](#)

이슈 데이터 미세먼지 정보

미세먼지 데이터

미세먼지와 관련된 공공데이터 결합정보를 쉽게 확인하실 수 있습니다.



- 노도별위(PM10)
- 서울(0-30) 나경(81-150)
- 부산(31-80) 대구나경(151-)



나랑 데이터 하자!

2019 제7회 범정부 공공데이터 활용 창업 경진대회

24개 예선리그 참가 접수중

예선 → 통합본선 → 최종합선

www.startupidea.kr

주최: 행정안전부, 중소벤처기업부, 주관: NIA 한국정보진흥원, 협찬: 한국데이터산업진흥원



|Data-Scientist?

데이터 사이언티스트는
Data-Scientist is

데이터를 가지고 **길**을 만드는 사람
who makes a **Road** with data

첨단 기술이 바꿀 미래

키워드로 보는 CES 10년

벤츠 첫 전시, 기조연설

자동차 전시 확산

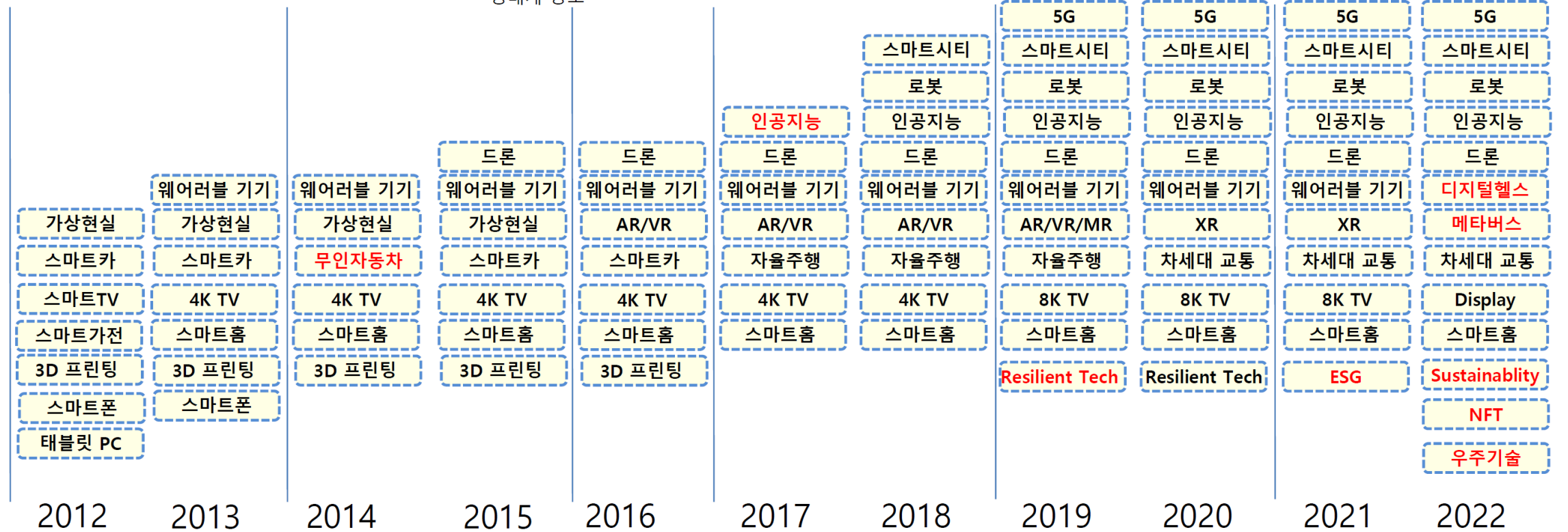
CTA로 변화 서비스/콘텐츠 생태계 강조

인공지능 등장 자율주행 강조

5G 본격 등장

코로나19 팬데믹 확산 온라인/비대면 기술 강조

온/오프라인 하이브리드 코로나19와 기술의 발전



↓
선박업체 카니발 Co. 기조연설 큰 반향 없음

↓
세빗 개최 중단 참고: 컴덱스 2003년 중단

↓
항공업체 델타 기조연설 DT의 빠른 변화

↓
온라인 진행

↓
하이브리드 진행 메타버스/NFT 의료융합(Abott) 우주기술

| 첨단 기술이 바꿀 미래

- 메타 버스 (Meta-verse)
- 인공지능 (Artificial Intelligence)

**왜 코딩으로 데이터를 가공하고
시각화하고 분석해야 할까?**

Toward reproducible research
Human forget!

- Which language is better for data science?
- What should I learn for?



Data Science with R

- **Basic elements & grammar (syntax) in R**
- **Data communication with R**
 - Importing DATA
 - Data preparation (dplyr)
 - Data visualization (ggplot)
- **Basic math & Statistics for data-science**

- **Data analytics with R**
 - Machine learning modelling
 - Supervised learning
 - Classification
 - Regression
 - Unsupervised learning
 - Clustering
 - Pattern finding

- **Advanced data analysis with R**
 - Web-crawling
 - Data mining
 - Text mining
 - Econometrics
 - Geo-spatial analysis
 - Causal inference
 - Network analysis

- Web-publishing with **Shiny R & Quarto**
- Data communication with **Shiny R & Quarto**

What we R doing is ...

- **Basic elements & grammar (syntax) in R**
- **Data communication with R**
 - Importing DATA
 - Data wrangling (전처리)
 - Data visualization
- **Basic math & Statistics for data-science**

- **Data analytics with R**
 - **Machine learning modelling**
 - Supervised learning
 - Classification
 - Regression
 - Unsupervised learning
 - Clustering
 - Pattern finding

- **Advanced data analysis with R**
 - Web-crawling
 - Data mining
 - **Text mining**
 - Econometrics
 - Geo-spatial analysis
 - Causal inference
 - Network analysis

- Web-publishing with Shiny R & Quarto
- Data communication with Shiny R & Quarto

What you can do with Quarto

- Quarto는 Pandoc에 구축된 오픈 소스 과학 및 기술 출판 시스템
- Python, R, Julia 및 Observable을 사용하여 동적 콘텐츠를 만들어 줌



What you can do with R Shiny

- R Shiny: R 프로그래밍 언어를 사용하여 데이터 시각화와 웹 애플리케이션 개발을 지원하는 프레임워크
- Shiny를 사용하면 R 코드를 기반으로 다양한 그래픽 요소와 상호작용을 포함한 웹 애플리케이션을 개발할 수 있음
- 데이터 분석가나 통계학자, 과학자, 업무 전문가 등 다양한 전문가들이 데이터를 시각적으로 전달하고 상호작용하는 방식으로 제공할 수 있는 강력한 도구

What you can do with **R** Shiny

<https://shiny.rstudio.com/gallery/>

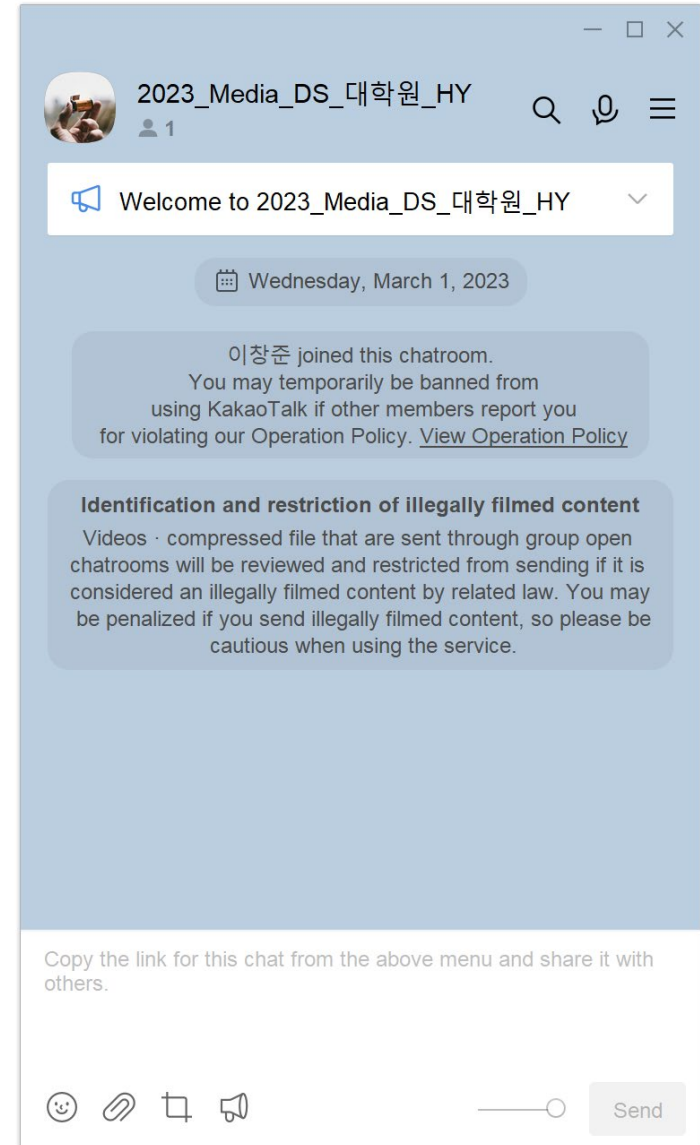
What is your goal for this course?

Whatever it is, it's really just up to you

For our communication during this semester



<https://open.kakao.com/o/gomSif7e>



Course Intro

There is a web page for this course

A screenshot of a web page. At the top, there is a blue header bar with the text 'CJL Lab' in white. Below the header, the main content area has the title 'Media & Data Science' in a large, black, sans-serif font. Underneath the title is the Korean text '미디어와 데이터사이언스 (대학원)'. There are three blue, underlined links: 'Notice', 'Syllabus', and 'Weekly_content', arranged vertically.

CJL Lab

Media & Data Science

미디어와 데이터사이언스 (대학원)

[Notice](#)

[Syllabus](#)

[Weekly_content](#)

https://cileelab.netlify.app/teaching/media_ds/

Course design

https://cjleelab.netlify.app/teaching/media_ds/notice/

Weekly design

https://cjleelab.netlify.app/teaching/media_ds/about/

Weekly design

- **Two 5 mins presentations**

- 데이터 사이언스를 본인 전공 분야에 활용한 논문 소개
- 데이터 사이언스를 본인 전공 분야에 활용한 논문 프로포절
 - 1. 연구 배경 (문제 분석)
 - 2. 연구 질문
 - 3. 연구 질문을 해결하기 위한 방법론 제시
 - 4. 예상 결과

- **Final output**

- 데이터 사이언스를 본인 전공 분야에 활용한 논문 프로포절
 - 분량 제한 없지만 5페이지 넘지 않게 (짧고 굵게 핵심만)
 - 프로포절이 관철은 경우에 한해서 수업 이후에 논문 진행 가능 (본인이 원할 시)

한번 코드 쳐보면서 따라해 보는 것이 백 번 보는 것보다 낫다

1. 15주 수업에 모든 것을 담을 수는 없습니다.
(개인 과외 수업이 아닙니다. 스스로 학습하셔야 해요)
2. 수업 내용 안에서 목표치를 정하고 얻어가고 싶은 부분을 잘
발췌해서 학습하시기 바랍니다.

Install your gadgets

R, Rstudio, Rtools

https://cjleelab.netlify.app/teaching/media_ds/notice/