

(540) BDA Open coaching

- Big Data concept
- Statistics, Data Mining, Machine learning basic
- R/Shiny 활용
- Python 활용

* 기술위원회 활동

빅데이터 분석 Open coaching

기본 (Free)

- R 프로그램으로 통계분석 모델 Programing 하여 수요기업에 무료 제공
 - ✓ 빅데이터 분석을 위한 기업 요구사항 협의
- 자주 사용하는 분석은 R/Shiny 을 이용 웹으로 개발 제공
- R/Markdown 이용 통계분석 각종 Report 개발 제공
- 통계분석을 위한 Open source 솔루션 활용법 교육

회원사 확보/협업

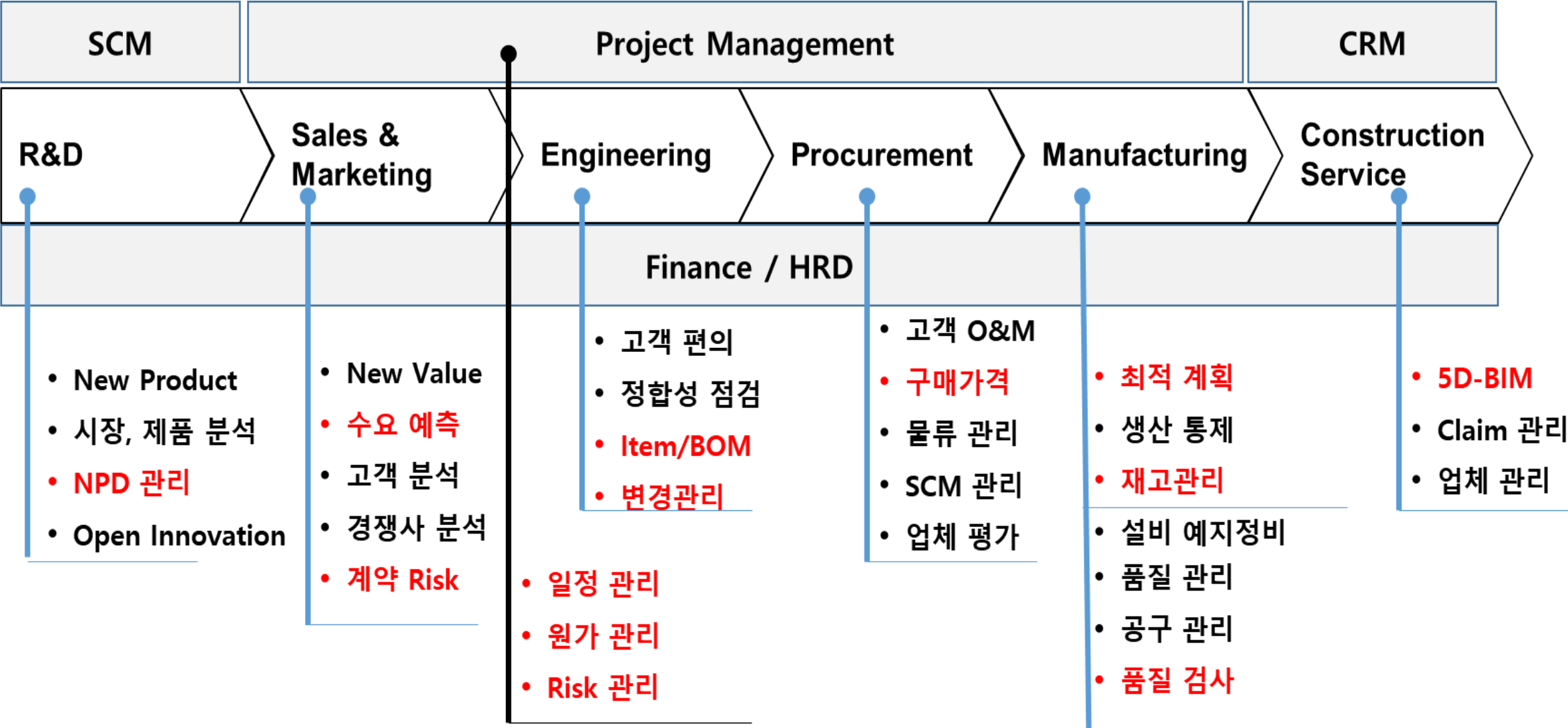
+

- BDA 에 대한 개념 전환 (video clip, Arduino 체험)
- BDA 솔루션 & Open source 학습과 체험 (이론개요, 분석, 결과 해석, 학습)
 - ✓ 통계기본, Data Mining & ML 개념 교육
 - ✓ R 통계분석, ML 을 위한 Python 조작 교육
 - ✓ 사용자 편의 Shiny, markdown 활용
 - ✓ 스스로 Programing
- 디지털 혁신의 현장적용을 위한 Design Thinking (R 응용, ML 활용 Idea)
 - ✓ 고객 가치 공감, Idea & Pilot
 - ✓ 디지털 기술 활용
 - ✓ 데이터 분석 활용을 위한 R 응용, ML 활용 Idea

현업 담당자 스스로 학습 방향 설정
디지털 분임조 활동

사용자의 동기부여 및 디지털 분임
조의 제대로 동작은 기업 고민

분석 대상 분야 (R, ML 동일)

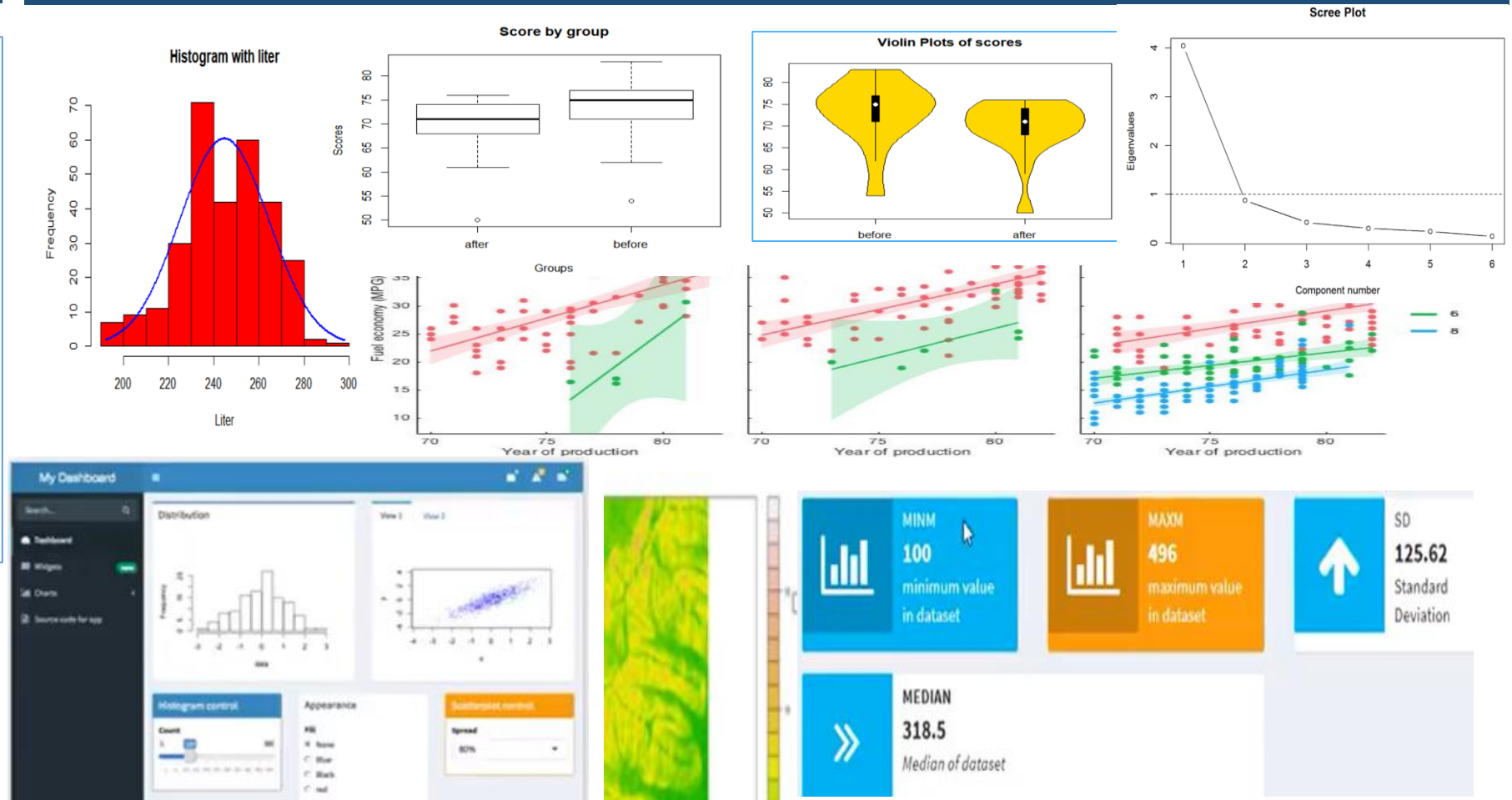


R 통계 분석 및 결과 시각화 영역 : 현재 수행 중인 업무에 활용 중심

R 통계 분석

- 공분산
- t-Test (일표본, 대응, 독립)
- ANOVA (one-way, two-way, MANOVA)
- 요인분석 (PCA/FA)
- 상관분석, 신뢰도 분석
- 단순/다중 회귀분석
- 로지스틱
- 판별분석, 군집분석

분석 결과 시각화 영역

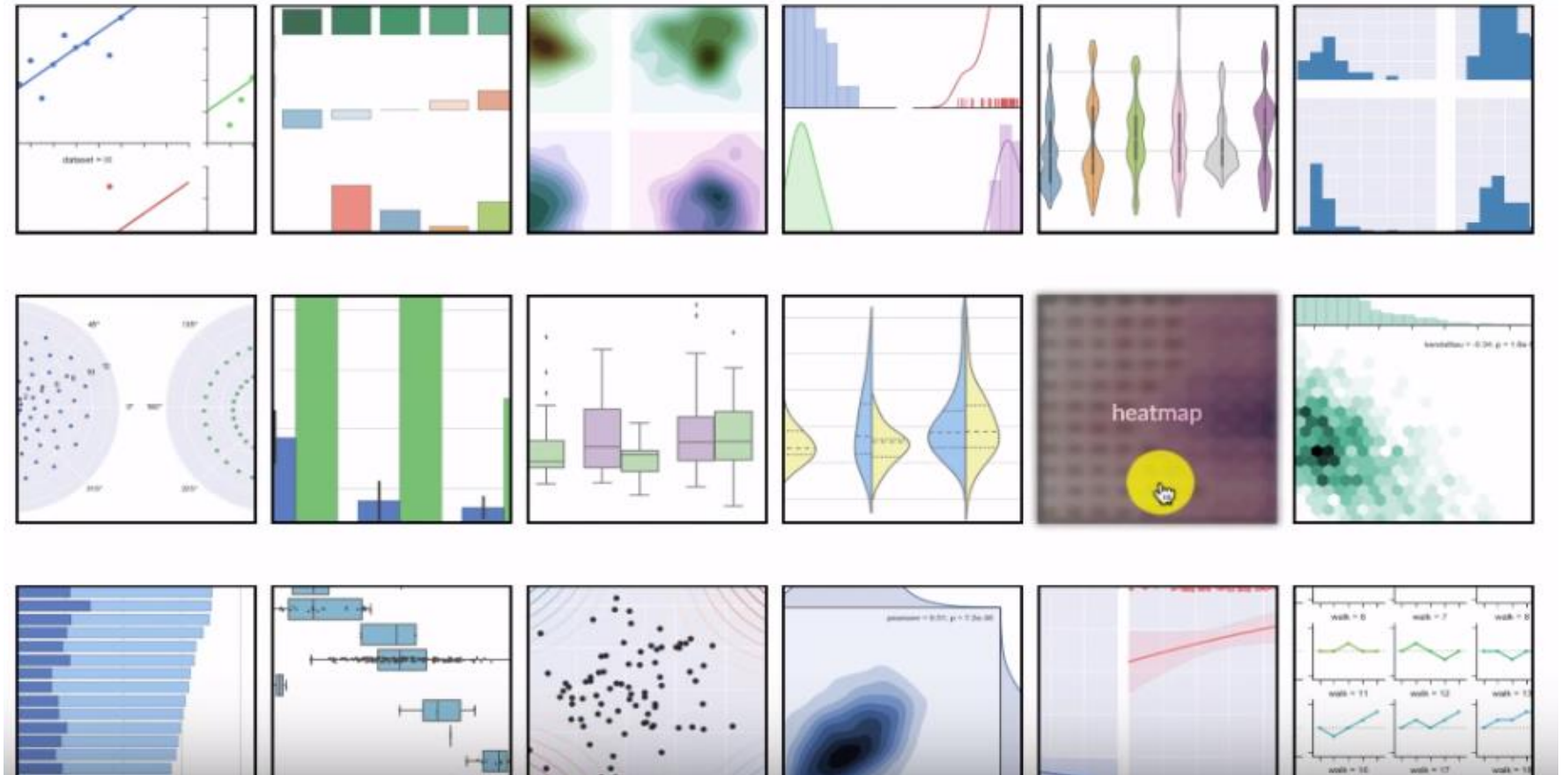


Python ML 분석 및 결과 시각화 영역 : 향후 활용 가능한 AI/ML 적용 Idea 발굴

ML 분석

- Linear Regression
- Logistic classification
- Softmax
- Deep learning
- CNN
- RNN

분석 결과 시각화 영역

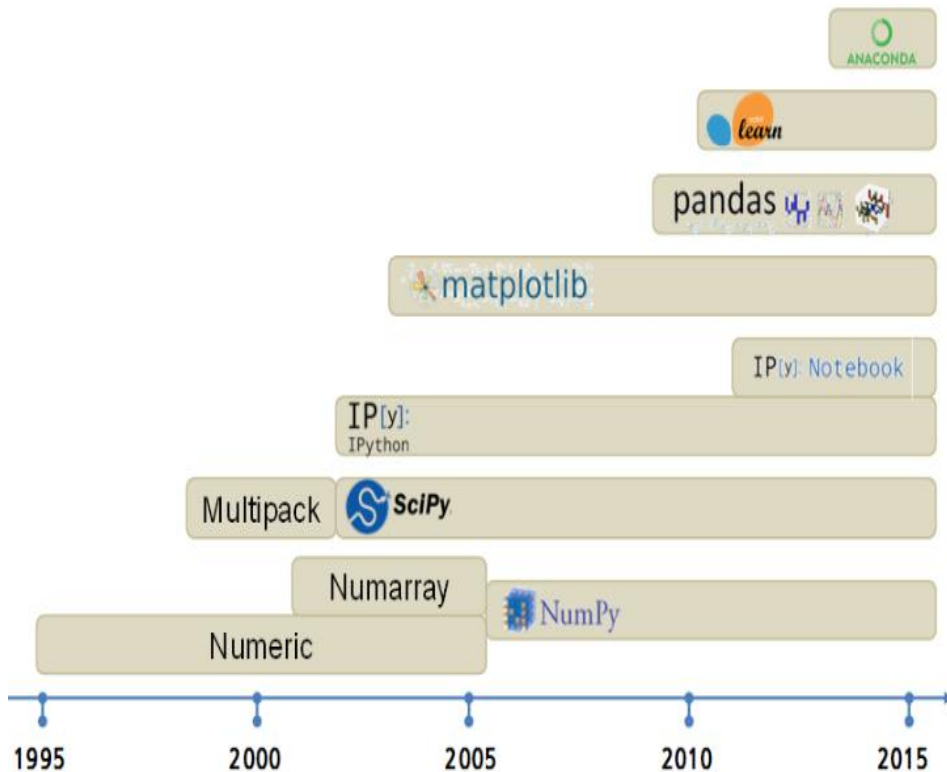


<https://www.youtube.com/watch?v=D8yHdrrdNxFw>

계속 추가 발전하는 Library를 다룰 수 있는 유연성 역량 필요

Shiny

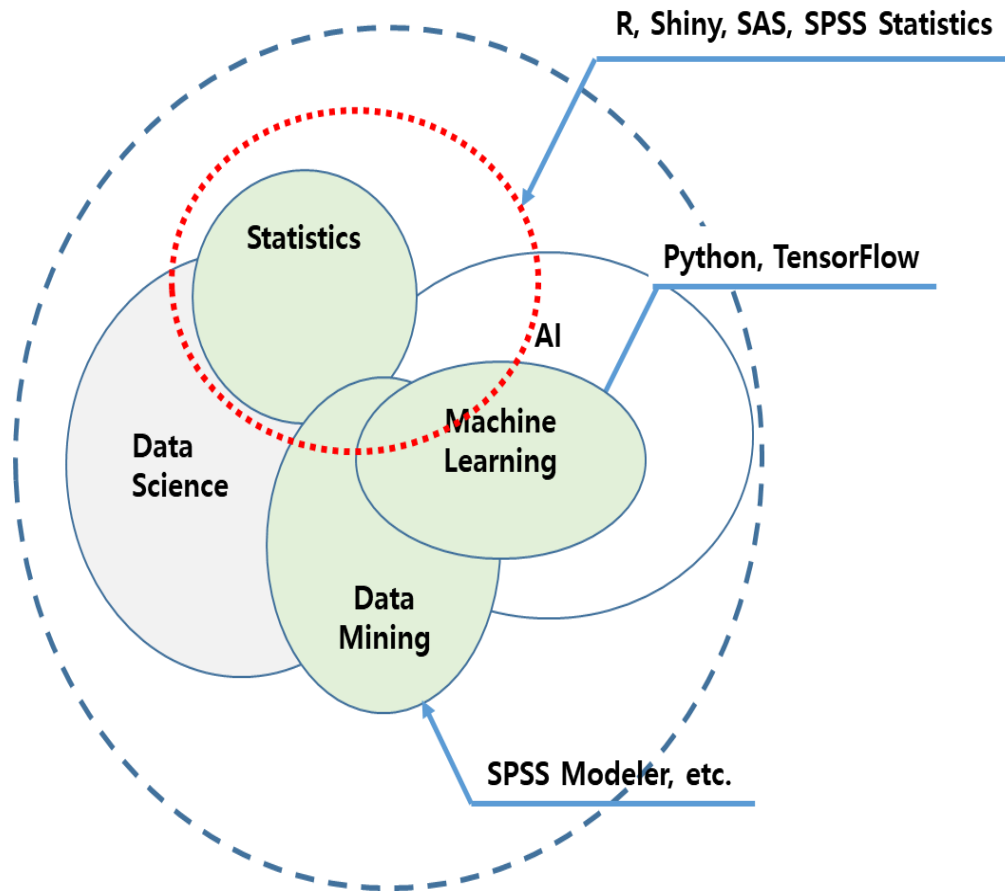
scikit-learn



- 파이썬 기계학습(machine learning) 패키지
- 데이터 마이닝과 자료분석 도구
- NumPy, SciPy, matplotlib 을 기반
 - ✓ 분류: SVM, random forest, nearest neighbors, ...
 - ✓ 회귀분석 : SVR, Ridge regression, Lasso, ...
 - ✓ 군집분석 : K-Means, spectral clustering, mean-shift, ...
 - ✓ 차원축소 : PCA, feature selection, non-negative matrix factorization
 - ✓ 모형선택 : grid search, cross validation, metrics
 - ✓ 전처리 : preprocessing, feature extraction

사용할 빅데이터 분석 Tool

Excel, R, Python을 상호 보완적으로 활용



분석
Tool



- R, dplyr
- Shiny
- Markdown



- Python, pandas / numpy
- tb, TensorFlow
- seaborn, bokeh, folium/matplotlib

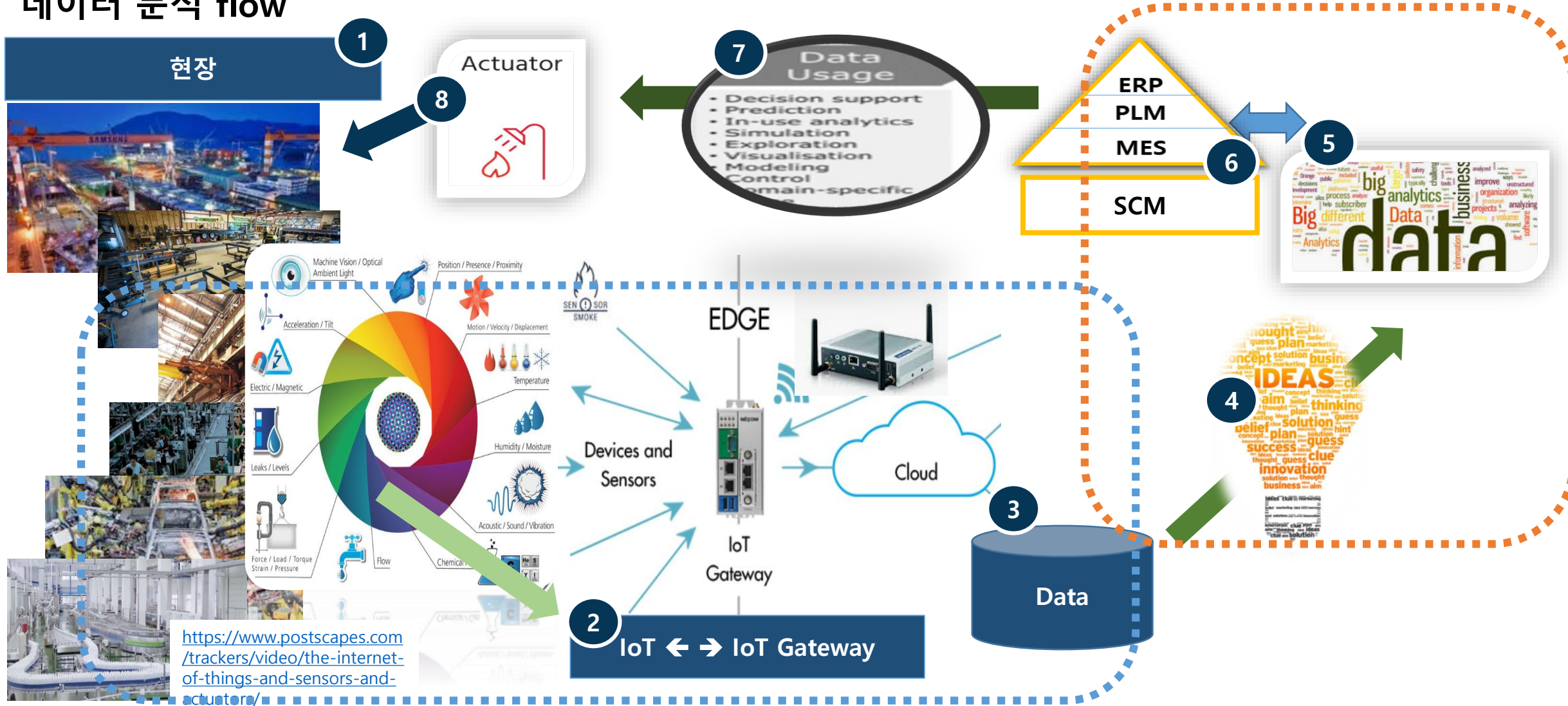
AI / 머신러닝 개념

통계 이론 / 사회조사 분석

우리가 관심을 갖는 곳은 Idea 도출 !

별첨

데이터 분석 flow



<https://www.postscapes.com/trackers/video/the-internet-of-things-and-sensors-and-actuators/>

(312) 통계분석 w/ R

- t 검정, 분산분석, 타당성과 신뢰성
- 요인분석, 회귀분석
- 군집분석

(420) 통계분석에 R/ Shiny 활용

- 통계분석에 R/Shiny 활용 사례 및 Source code
- Shiny web 시스템 활용
- 확대 적용

(440) R Markdown report

- 통계분석 Report (HTML, PDF, WORD)
- Shiny 연결 웹 시스템으로 활용

(312) Data Mining

- Data Mining basic
- 주요 분석 및 모형의 평가
- 활용 방안

(321) ML Concept

- AI, xor, ML, Deep Learning
- Python, TensorFlow, TensorBoard
- Matplotlib

(322) ML Algorithm

- Linear Regression
- Logistic Classification
- Softmax
- AI, ML & Deep Learning
- CNN, RNN