



수업 운영 계획 및 개요

AI빅데이터프로그래밍

Prof.Hyunsil Moon (hsmoon@kookmin.ac.kr)

1.

수업 운영 계획

강의 개요

- 과목명 : AI빅데이터프로그래밍
- 강의시간 : 토 09:00-12:00
- 강의실 : 경영관 505호
- Office Hour: 사전에 이메일 연락 후 일정 조율
 - E-mail Address: hsmoon@kookmin.ac.kr
- 수업 관련 질의 응답 : e-mail

2 강의 목표

- 빅데이터 처리 및 인공지능 구현을 위한 Python 프로그래밍을 학습
- Pandas와 Numpy를 활용하여 실제 데이터의 처리 및 분석 과정을 실습
- 비즈니스 데이터 처리와 인공지능을 위한 데이터 준비 과정에 필요한 기능을 학습
- 데이터 시각화를 통해 빅데이터 처리 및 분석 결과의 표현을 실습

실습 환경

- Windows 10 64bit
 - MAC 사용자는 질의시 운영체제 언급 필수
- Anaconda
 - Python 3.10
 - Jupyter Notebook with VSCODE
- Codes & Lecture Notes & Homework
 - e-campus : <http://ecampus.kookmin.ac.kr>

평가 방법

- 중간고사 : 20%
 - 데이터 처리 프로그래밍 관련 문제
- 기말고사 : 20%
 - EDA 구현 프로그래밍
- 프로젝트 : 20%
- 과제 : 20%
- 출석 : 10%
- 수업참여도 : 10%

- 프로젝트
 - 팀 프로젝트
 - 프로젝트 용도로 주어진 데이터 셋 중 1가지를 선택하여 데이터 처리 및 분석을 수행하고 그 결과를 발표
- 주요일정
 - 10주차 : 데이터셋 공개
 - 15주차 발표

- 과제물
 - 실습 후 제시되는 Wrap-up Exercise를 해결
 - 과제당 10점 : 총 8회 (80점 만점 -> 20%로 환산)
 - 과제 기한 후 과제 수행 결과 발표 : 수업 참여도 반영

- Lecture Notes

Week	내용	비고
1	- AI빅데이터프로그래밍 수업 운영계획 - 수업 개요	
2	- Python Basic Review I: 자료구조의 이해와 활용	16주차 수업 진행
3	- Python Basic Review II: 제어함수와 프로그래밍	
4	- Python Basic Review III: 모듈과 파일다루기	
5	- Numpy 기초 문법 - Pandas 기초 문법	
6	- 데이터프레임의 이해 - 슬라이싱과 인덱싱	
7	- 그룹화/필터링 - 데이터 변환과 재구성	
8	- 중간고사 : 데이터 처리 프로그래밍	

Week	내용	비고
9	- Pandas 객체 병합 - Tidy Data의 이해	
10	- 데이터 시각화 개요 - Matplotlib/Seaborn 기초	
11	- 동적시각화와 데이터 앱(Data App) - 데이터 앱 개발을 위한 StreamIt 기초 사용법	
12	- StreamIt 템플릿 소개 - StreamIt을 사용한 동적시각화 기반 데이터 앱 구현 사례	
13	- 부처님오신날	16주차 수업 진행
14	- 탐색적데이터분석(EDA) 절차 및 주요 과업 - 시각화를 활용한 EDA 사례	
15	- 팀프로젝트 발표 : 빅데이터 처리 및 분석/시각화 결과	
16	- 기말고사 : EDA 구현 프로그래밍	

2.

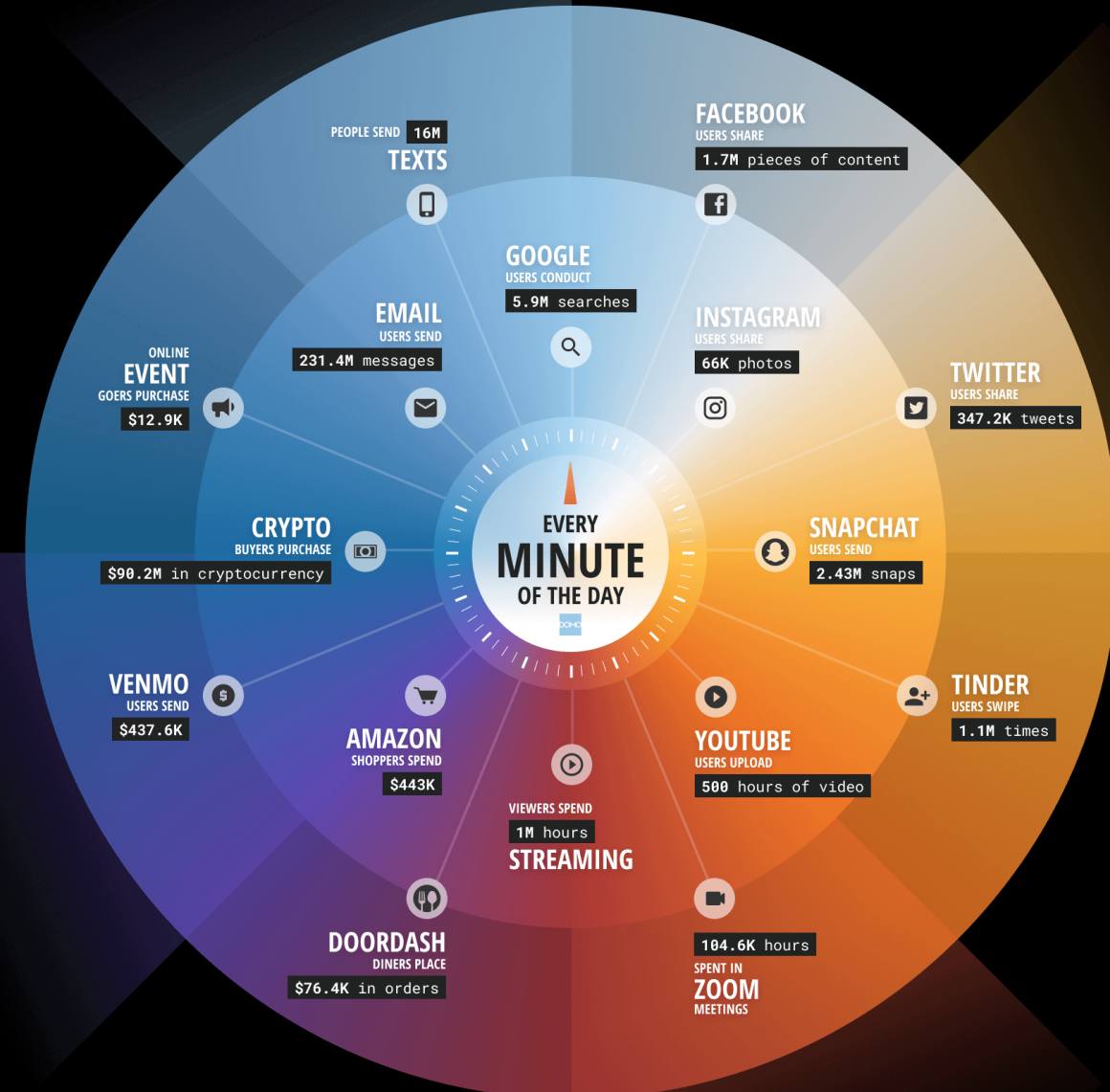
인공지능과 빅데이터,
그리고 프로그래밍

Introduction

- Data Never Sleeps
- <https://www.domo.com/data-never-sleeps#>

Over the last ten years, digital engagement through social media, streaming content, online purchasing, peer-to-peer payments and other activities has increased hundreds and even thousands of percentage points. While the world has faced a pandemic, economic ups and downs, and global unrest, there has been one constant in society:

our increasing use of new digital tools to support our personal and business needs, from connecting and communicating to conducting transactions and business. In this 10th annual "Data Never Sleeps" infographic, we share a glimpse at just how much data the internet produces each minute from some of this activity, marveling at the volume and variety of information that has been generated.





- 빅데이터의 정의
 - 위키피디아 : 데이터양이 매우 크고 데이터 구조가 복잡하여 일반적인 데이터베이스 관리도구나 기존 데이터 분석 툴로는 처리하기 어려운 데이터의 집합
 - 맥킨지 : 데이터양이 기존 일반적인 데이터베이스 도구를 통해 기록, 저장, 관리, 분석의 범위를 넘어서는 데이터
- 협의의 정의
 - 기존의 데이터베이스 기술로는 저장, 관리, 분석할 수 없는 대용량 데이터 집합
- 광의의 정의
 - 이러한 대규모 데이터를 통해서 가치를 창출할 수 있게 하는 기술들, 분석 기법들, 아키텍처 그리고 거버넌스

Introduction

- 빅데이터의 특성

- 규모(Volume)

- 기존의 DBMS가 관리하던 데이터가 기가바이트(GB)수준이었다면 테라바이트(TB)나 페타바이트(PB)이상
- 오늘날 데이터의 90%는 지난 2년간 생성된 데이터
- 하루 2.5EB가 생산 : 블루레이디스크 1,000만 개 분량

- 다양성(Variety)

- 전통적인 데이터와 달리 구조적인 데이터뿐만 아니라 비구조적인 데이터를 포함
- 생성되는 데이터의 80%가 비디오, 이미지, 문서 등과 같은 비정형 데이터
- unstructured 'dark' data

- 속도(Velocity)

- 데이터 생성에서 유통되어 소비되는 기간이 대폭 단축되었음을 의미



1 Introduction

- IBM에서는 3V 중 두가지만 만족시키면 빅데이터 기술이라고 언급
 - 최근에는 빅데이터 특성을 3V에서 4V, 6V, 6V+1C로 확장하여 제시
 - 3가지 특성에 가치(Value), 변동성(Variability), 정확성(Veracity), 복잡성(Complexity) 등 추가

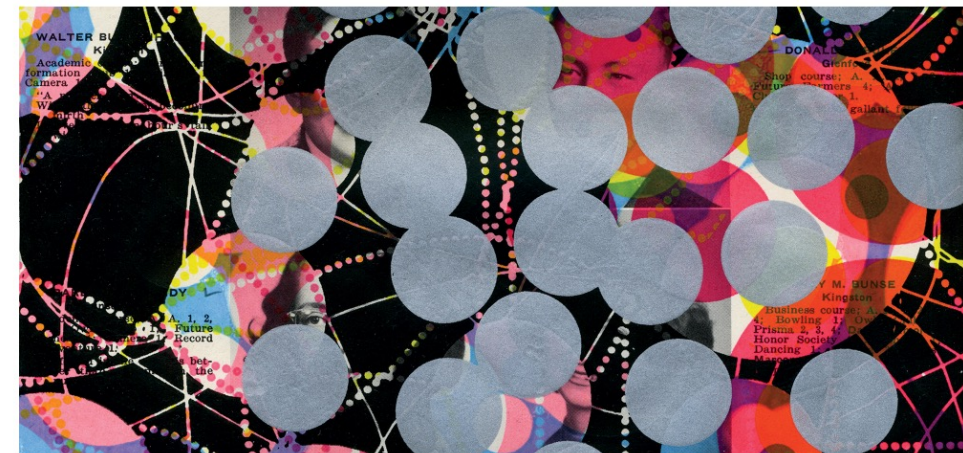
2 Data Science & Big data

- Unique blend of skills that can both unlock the insights of data and tell a fantastic story via the data – D.J. Patil
 - What data scientists do is make discoveries while swimming in data
 - A data scientist is someone who is better at statistics than any S/W engineer and better at S/W engineering than any statisticians

Data Scientist: The Sexiest Job of the 21st Century

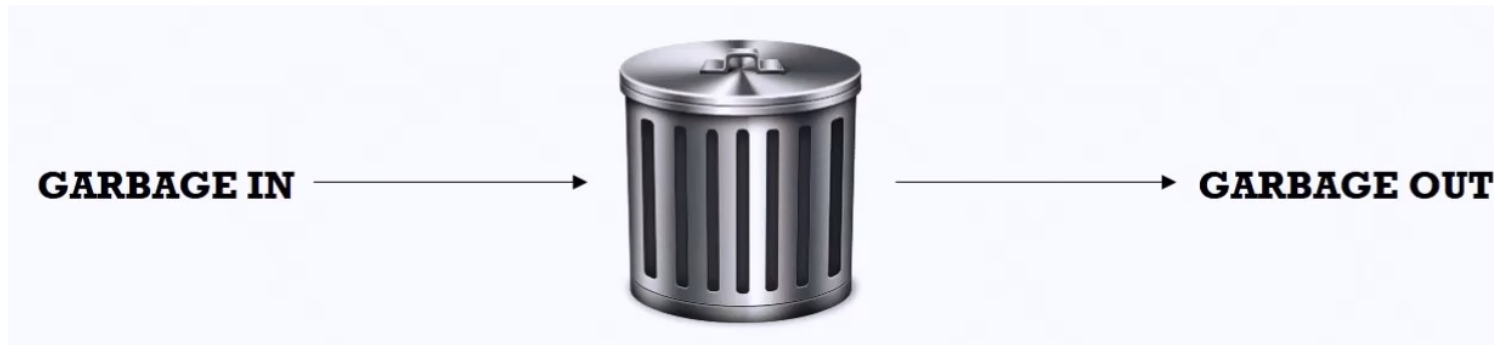
Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)



Artwork: Tamar Cohen, Andrew J. Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

2 Data Science & Big data



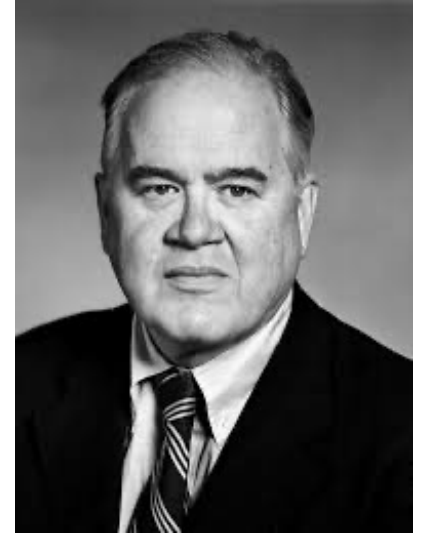
3 Data Wrangling(Data Munging)

- The process of transforming and mapping data from one 'raw' data form into another format with the intent of making it more appropriate and valuable for a variety downstream purposes



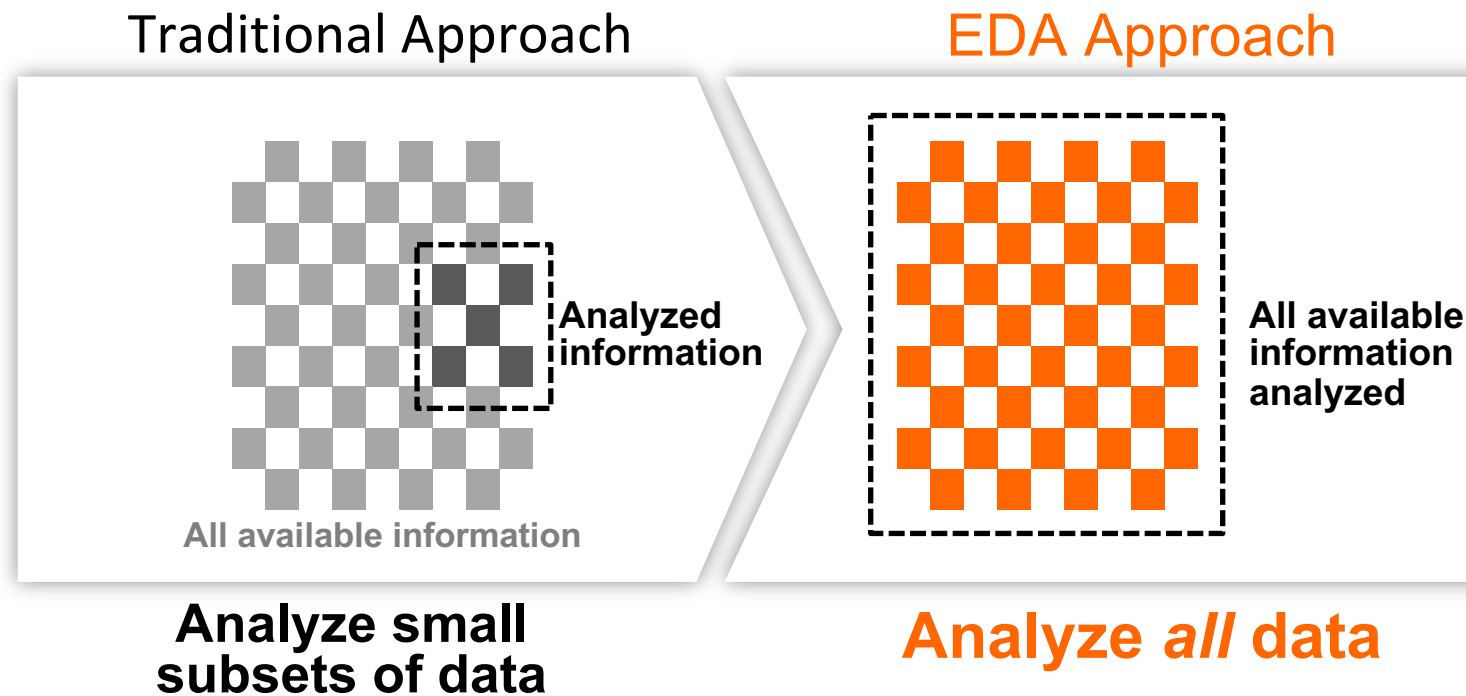
탐색적 데이터 분석(Exploratory Data Analysis; EDA)

- 통계학
 - 데이터 분석과 모델링을 다루는 일종의 응용과학
- EDA는 통계학에서 비교적 새로운 영역
 - 전통적인 통계학에서는 추론이 주 영역
 - 적은 표본(sample)을 가지고 더 큰 모집단에 대한 결론을 도출하기 위한 일련의 복잡한 과정
- 1962년 존 투키(Tukey, John W., The Future of Data Analysis)
 - 통계적 추론을 하나의 구성 요소로 보는 데이터 분석(Data Analysis)이라는 새로운 과학적 학문 제안
 - 1977년 탐색적 데이터 분석 책을 통해 기본 개념 정립

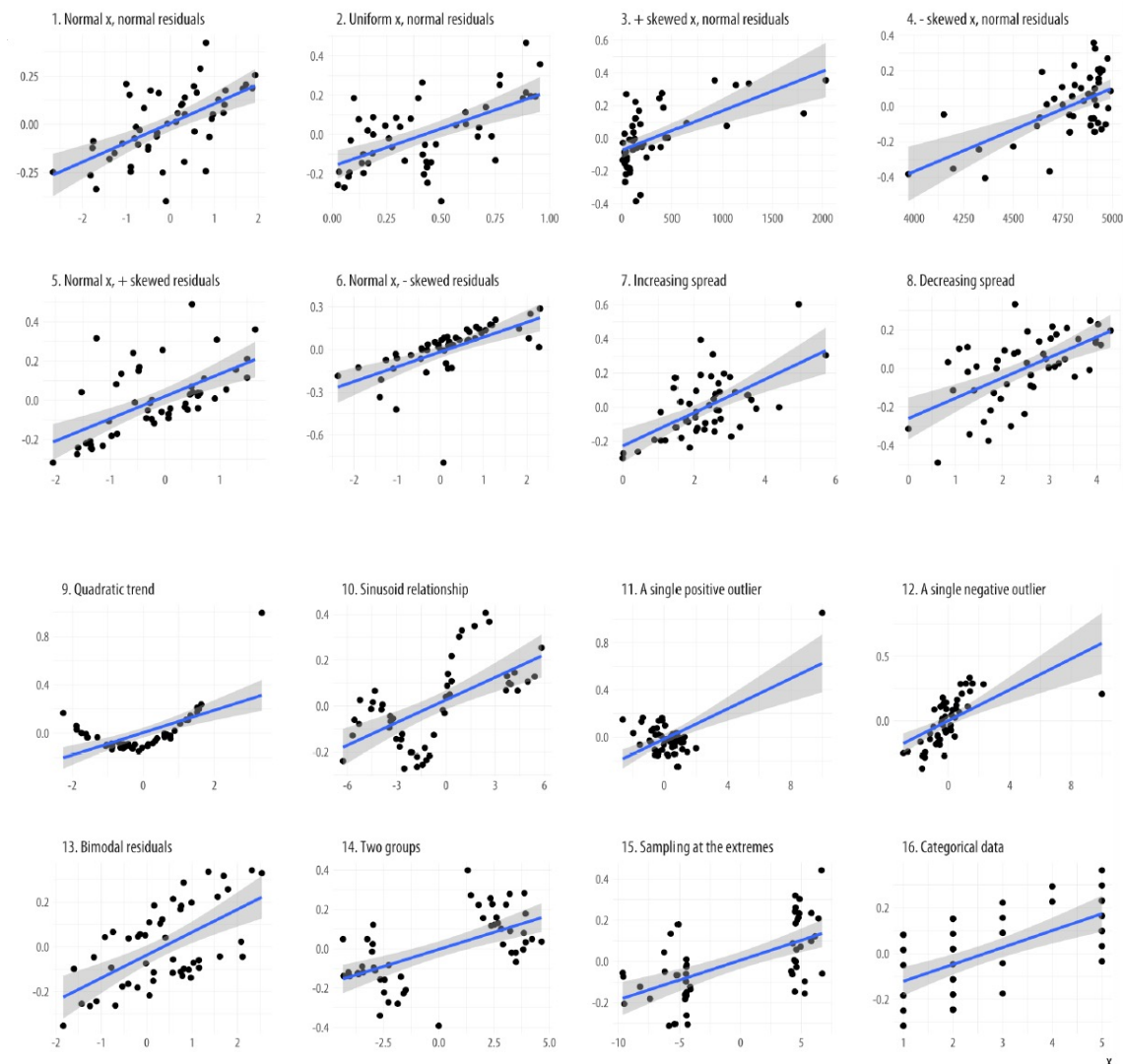
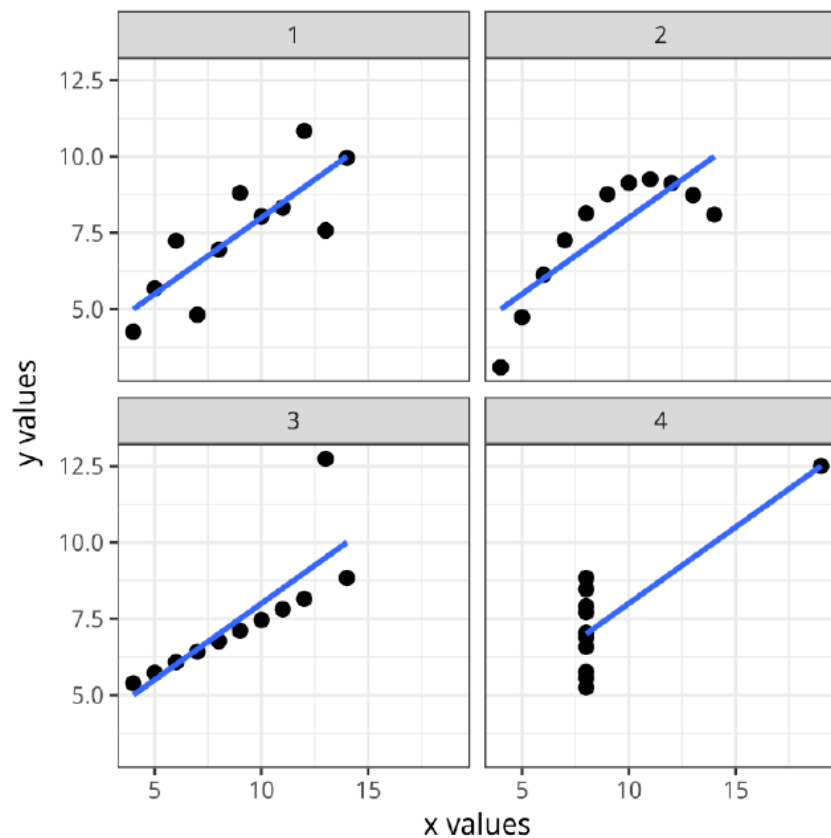


탐색적 데이터 분석(Exploratory Data Analysis; EDA)

- 기존 분석과의 차별점

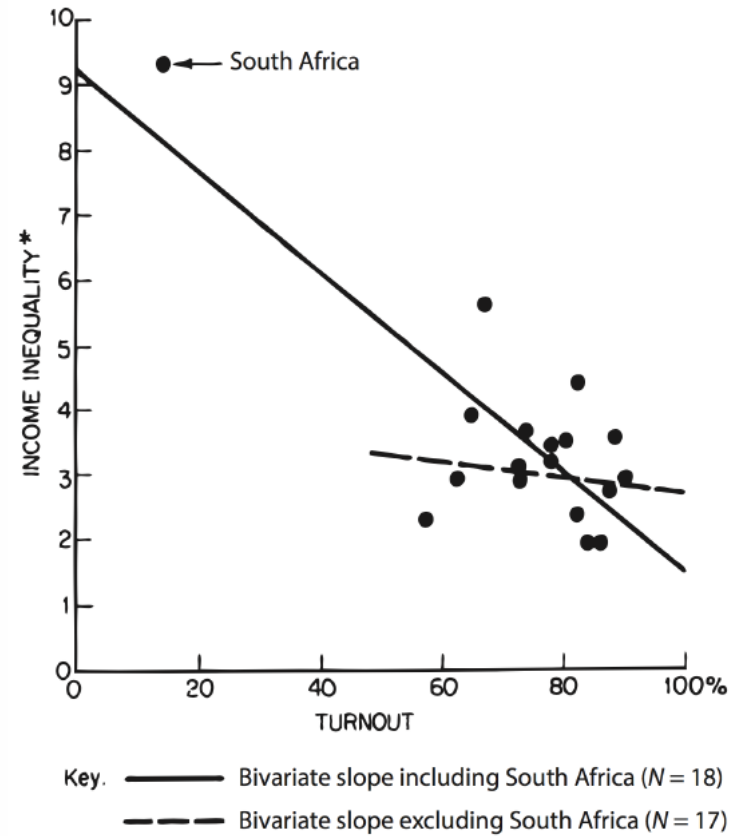


- 같은 상관 관계?



5 EDA에서 시각화의 역할

- 이상치(Outlier)?



3.

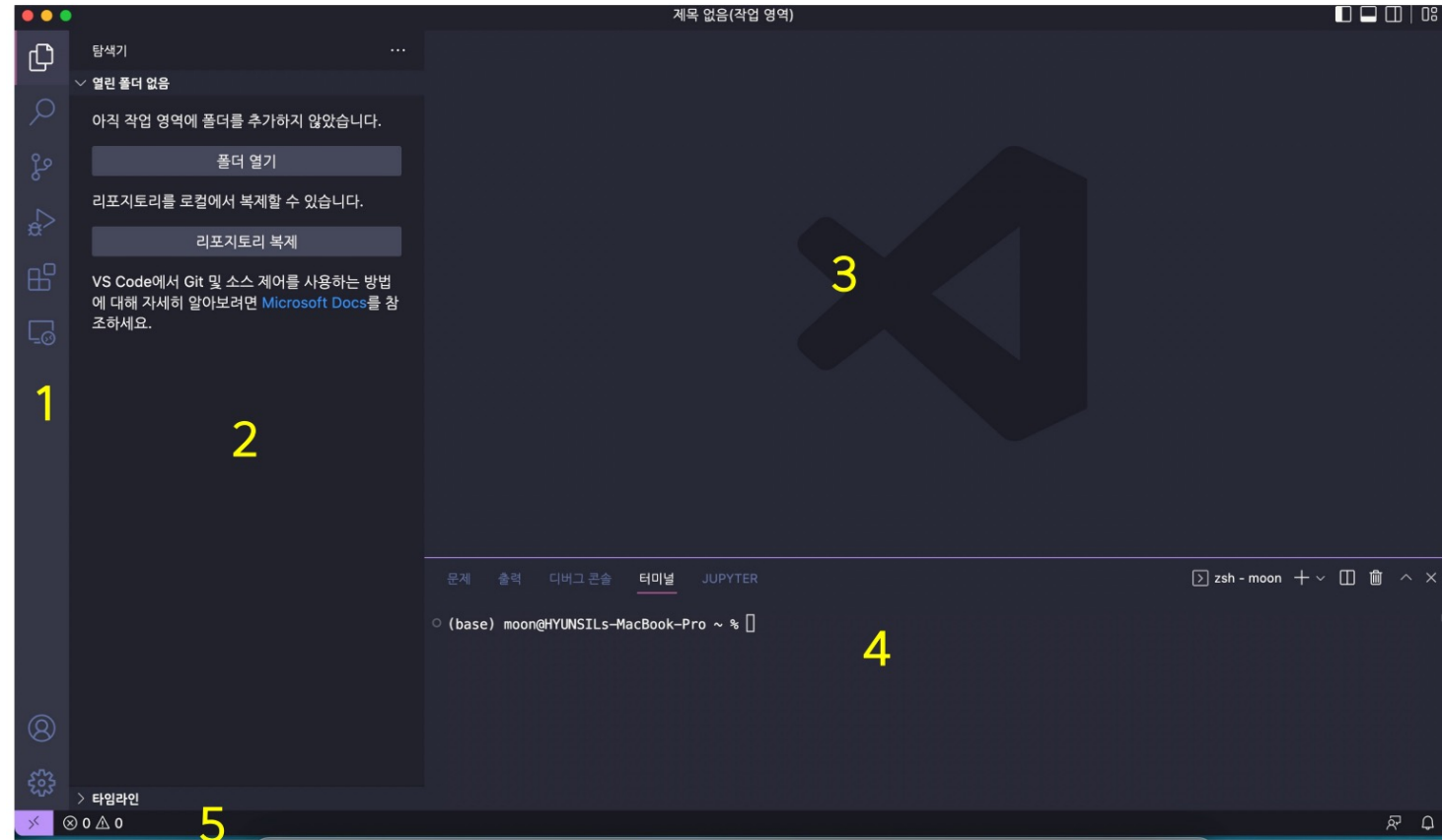
VSCODE

1 Visual Studio Code 개요

- Microsoft에서 개발한 무료 코드 편집기 (<https://code.visualstudio.com/>)
 - 윈도우, 맥, 리눅스 모두 지원
 - 자체 터미널 기능
 - 프로그래밍 언어에 따른 syntax 컬러링
 - Jupyter Notebook 클라이언트 내장: Variable Explorer 등
 - 확장 기능을 통한 기능 지원: 가벼운 용량 + 강력한 기능
 - 빌더가 내장된 IDE가 아님에 따라 프로그래밍 언어별 확장을 통해 IDE로 사용 가능
 - MS의 GitHub인수로 Electron으로 개발한 Atom을 대체(Atom은 2022년 지원 종료 예정)

2 Visual Studio Code 설치

- VS CODE 화면 구성
 - 1: 활동 표시줄(Activity Bar)
 - 2: 사이드바(Side bar)
 - 3: 편집기(Editor)
 - 4: 패널(Panel)
 - 5: 상태표시줄(Status Bar)



2 Visual Studio Code 설치

- 한글 언어팩 설치
 - 초기 실행시 나타나는 언어팩 설치로 설치
 - 확장 프로그램 -> Korean 검색 -> Korean Language Pack 설치
- Code runner 및 python 확장 설치(필수)
 - 확장 프로그램 -> code runner 검색 -> 설치
 - 확장프로그램 -> python 검색 -> 설치
- 테마 팩 설치
 - Dracula
 - Sublime Text
 - Atom One Dark

3 Visual Studio Code 주요 단축키

- 단축키 목록
 - Ctrl + k + s
- 편집 창 관련
 - 현재 에디터 창 닫기: Ctrl+w
 - 이전에 닫힌 창 다시 열어 주기: Ctrl+shift+t
 - 화면 창 분할 : Ctrl + \
 - 포커싱 : Ctrl + 숫자

Visual Studio Code 주요 단축키

- 소스 편집 관련 단축키
 - 들여쓰기: Tab, 내어쓰기: Shift+Tab
 - 현재 라인 이동: Alt + 위/아래 화살표
 - 현재 라인 복사: Shift+Alt+위/아래 화살표
 - 현재 라인 잘라내기/삭제: Ctrl+x, Ctrl+Shift+k
 - 원하는 라인으로 이동: Ctrl+g
 - 이전/이후 작업 부분으로 이동: Alt+왼쪽/오른쪽 화살표
 - 한줄 주석처리: Ctrl+/

Visual Studio Code 주요 단축키

- 열 모드 관련 단축키
 - 특정 문자열을 수정: Ctrl+d
 - 특정 문자열 모두를 한번에 수정: Ctrl+Shift+L
 - 블록 열모드: Shift+Alt+드래그
- 검색/찾기/바꾸기
 - 오류 또는 경고로 이동: F8
 - 현재 페이지에서 검색: Ctrl+f
 - 현재 페이지에서 문자열 변경하기: Ctrl+h

3 Visual Studio Code 주요 단축키

- 명령어 검색/설정
 - F1 또는 Ctrl+Shift+P



Q&A

Thank you for your listening