

데이터 분석 및 시각화 개요

컴퓨터정보공학과
강 환수 교수

데이터 분석 및 시각화 수업 개요

- 2020-6/18(목), 6/19(금)에서 7/10(금)까지 4주 동안 매주 목 금
 - 총 48시간
 - 2시간 수업, 1시간 질의 및 자기주도학습
- 교재
 - 파이썬 라이브러리를 활용한 데이터 분석(Python for Data Analysis)
 - 사례 참조:
 - 파이썬으로 데이터 주무르기
 - 모두의 데이터 분석 with python
- 일정(변경될 수 있음)

			오전	오후
1주	2020-06-18(목)	1일차	수업 소개와 데이터 처리 개발환경 설정 1장 시간하기 전에	2장 파이썬 언어의 기본, ipython, 주피터 노트북 3장 내장 자료구조, 함수, 파일
1주	2020-06-19(금)	2일차	4장 넘파이 기본: 배열과 벡터 연산(약 48쪽)	4장 넘파이 기본: 배열과 벡터 연산(약 48쪽)
2주	2020-06-25(목)	3일차	5장 판다스 준비하기(약 55쪽)	6장 텍스트 파일 읽기(6.1.3까지)
2주	2020-06-26(금)	4일차	7장 데이터 정제 및 준비(약 37쪽)	7장 데이터 정제 및 준비(약 37쪽)
3주	2020-07-02(목)	5일차	8장 데이터 준비하기: 조인, 병합, 변형(약 39쪽)	8장 데이터 준비하기: 조인, 병합, 변형(약 39쪽)
3주	2020-07-03(금)	6일차	9장 그래프와 시각화(약 37쪽)	10장 데이터 집계와 그룹 연산(약 37쪽)
4주	2020-07-09(목)	7일차	10장 데이터 집계와 그룹 연산(약 37쪽)	12장 고급 판다스(일부)
4주	2020-07-10(금)	8일차	14장 데이터 분석 예제	14장 데이터 분석 예제

파이썬 기반 데이터 분석 및 시각화 소개

- 데이터 분석 라이브러리
 - Numpy
 - 행렬과 벡터 연산
 - Pandas
 - 시리즈와 데이터프레임 처리
- 시각화 라이브러리
 - Matplotlib, seaborn
 - Bokeh
 - 브라우저에서 뷰, html 파일 생성
- 데이터 분석 사례
 - 영화 평점 등
 - 교재의 14장
 - 서울 CCTV 분석
 - 파이썬으로 데이터 주무르기, 1장, BJ퍼블릭

교재

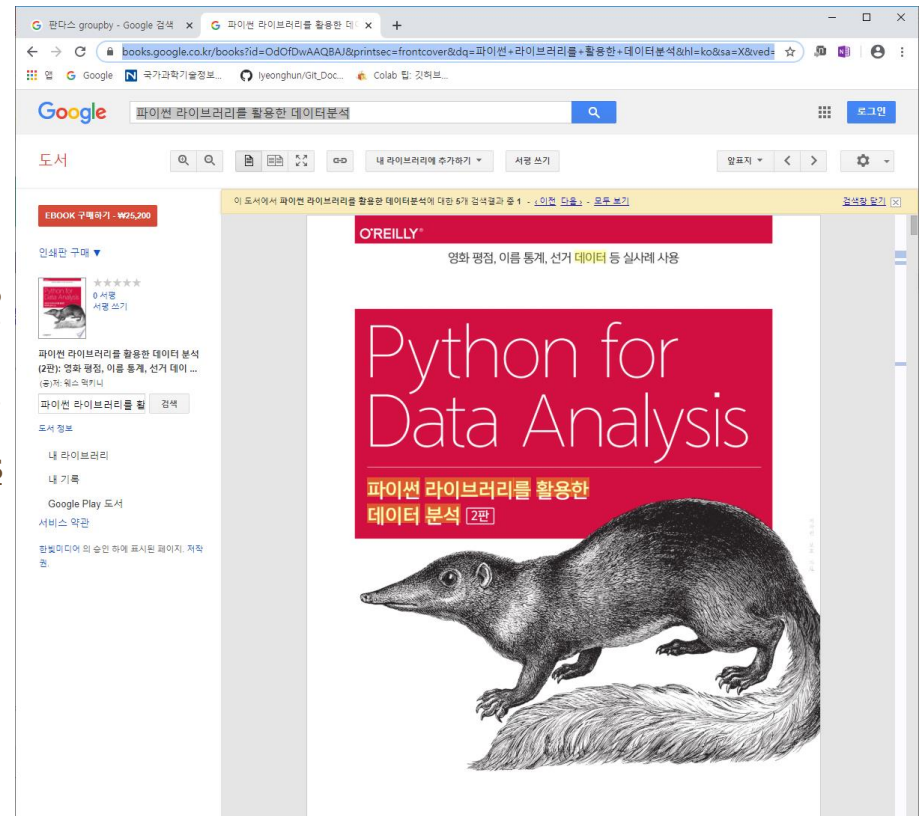
- 파이썬 라이브러리를 활용한 데이터 분석 2판

- 데이터 분석의 바이블
- Python for Data Analysis
 - **영문 pdf 파일**
- Ebook, 한글로 검색
 - books.google.com
 - https://books.google.co.kr/books?id=OdOfDwAAQBAJ&pg=PA10&pg=PA10&dq=%EC%9B%8C%EC%8A%A4+%EB%A7%A5%ED%82%A4%EB%8B%88&source=bl&ots=8sOi-IQE7S&sig=ACfU3U1C0CVJtntxZC6zJKo6z_3o1sQBpQ&hl=ko&sa=X&ved=2ahUKEwisn72CxP3pAhWOBlgKHcJBCXQQ6AEwA3oECAoQAQ#v=onepage&q=%EC%9B%8C%EC%8A%A4%20%EB%A7%A5%ED%82%A4%EB%8B%88&f=false

- 1판은 2012년

- 머리말

- 권장



주요 사이트

- **교재**

- 저자
 - **Wes Mckinney**
 - 판다스 개발자
 - <http://wesmckinney.com>
- 깃허브 저장소
 - <https://github.com/wesm/pydata-book>



- **주요 사이트**

- Numpy
 - <https://numpy.org/>
- Pandas
 - <https://pandas.pydata.org/>
- 우리 수업 저장소
 - <https://github.com/lee7py/Pydata-ANS-VIS>



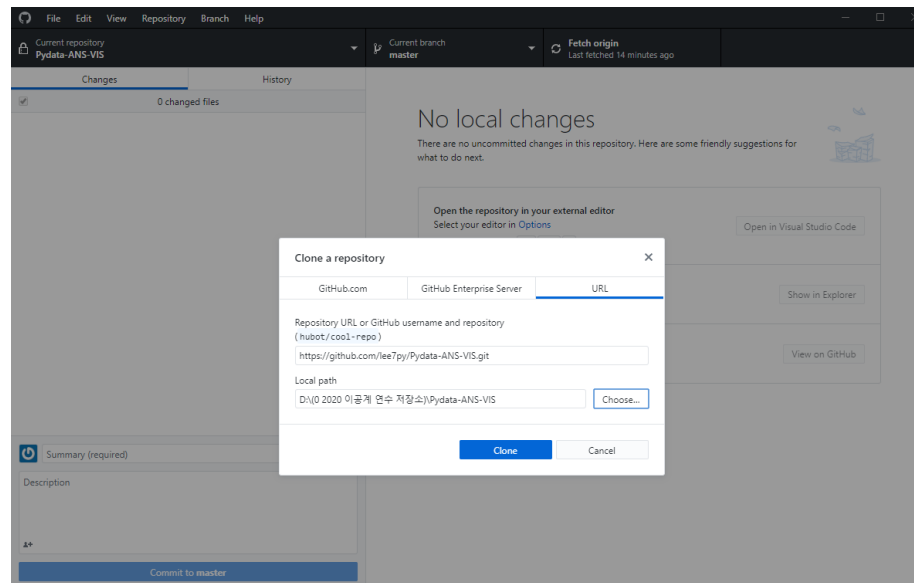
파이썬 개발환경에서 데이터 분석 및 시각화 개발환경 준비

컴퓨터정보공학과
강 환수 교수

1. 깃허브 저장소의 자료 복사

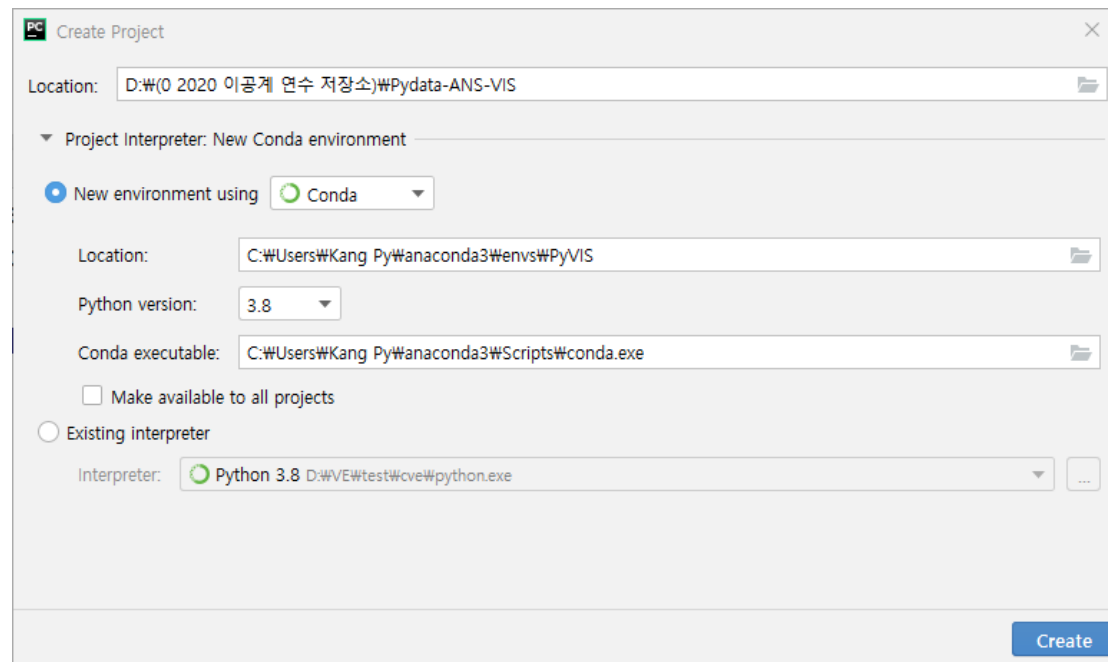
• 자신의 PC

- 적당한 폴더 하부
 - 폴더 '2020 Pydata' 하부
 - 깃허브 저장소의 소스 자료 복사 또는 깃허브 데스크탑으로 Clone 복사
- 다음 사이트에서 [Open in desktop] 선택
 - Clone a repository, URL
 - <https://github.com/lee7py/Pydata-ANS-VIS>



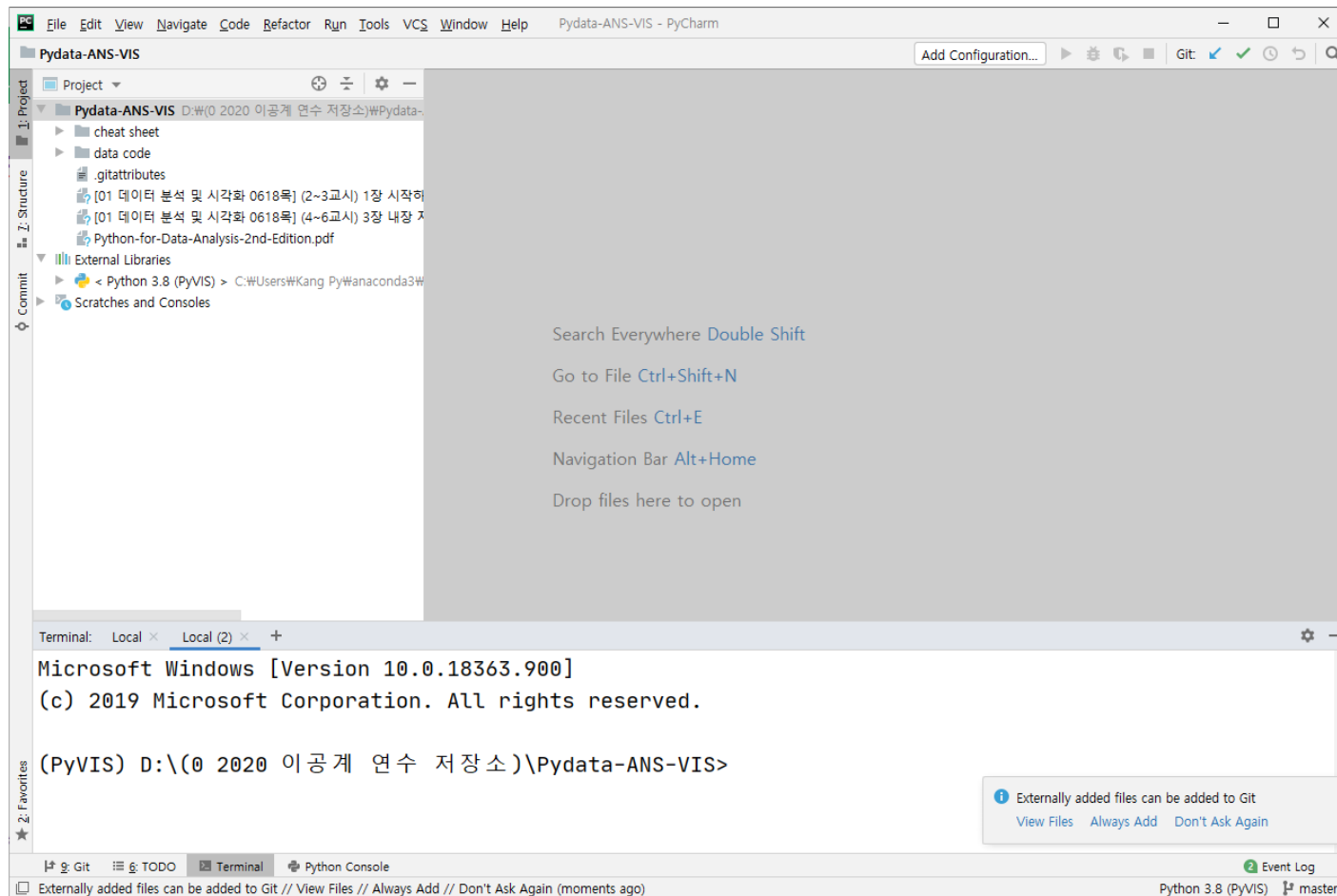
2. 파이참 프로젝트 생성

- 폴더 지정
- New environment using
 - 콘다로 생성
 - C:\Users\사용자명\anaconda3\envs\PyVIS



3. 프로젝트의 가상환경 확인

- 터미널 열어서 가상환경 이름 확인



4. 필요 패키지 설치

- jupyter, matplotlib, pandas, seaborn, bokeh

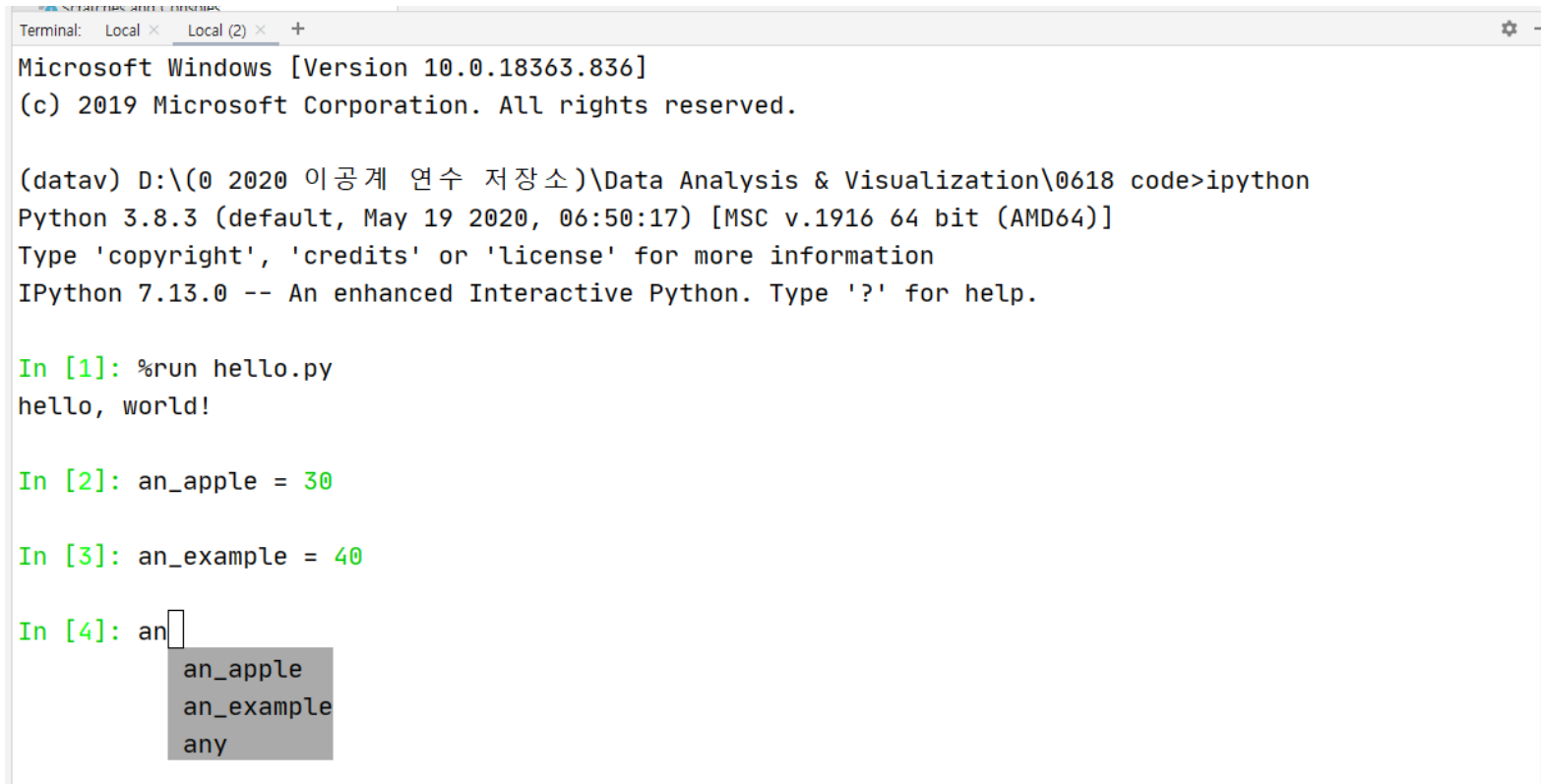
```
(datav) D:\(0 2020 이공계 연수 저장소)\Data Analysis & Visualization>conda list
# packages in environment at C:\Users\Kang Py\anaconda3\envs\datav:
#
# Name          Version          Build      Channel
ca-certificates 2020.1.1         0
certifi         2020.4.5.1       py38_0
openssl         1.1.1g           he774522_0
pip             20.0.2           py38_3
python          3.8.3            he1778fa_0
setuptools      46.4.0           py38_0
sqlite          3.31.1           h2a8f88b_1
vc              14.1             h0510ff6_4
vs2015_runtime  14.16.27012      hf0eaf9b_2
wheel           0.34.2           py38_0
wincertstore    0.2              py38_0
zlib            1.2.11           h62dcd97_4
```

- conda install matplotlib pandas jupyter

```
(datav) D:\(0 2020 이공계 연수 저장소)\Data Analysis & Visualization>conda list
# packages in environment at C:\Users\Kang Py\anaconda3\envs\datav:
#
# Name          Version          Build      Channel
attrs           19.3.0           py_0
backcall        0.1.0            py38_0
blas            1.0              mkl
bleach          3.1.4            py_0
bokeh           2.0.2            py38_0
ca-certificates 2020.1.1         0
certifi         2020.4.5.1       py38_0
colorama        0.4.3            py_0
cyclor          0.10.0           py38_0
decorator       4.4.2            py_0
defusedxml      0.6.0            py_0
entrypoints     0.3              py38_0
freetype        2.9.1            ha9979f8_1
icc_rt          2019.0.0         h0cc432a_1
icu             58.2             ha925a31_3
importlib-metadata 1.6.0           py38_0
importlib-metadata 1.6.0           0
intel-openmp    2020.1           216
ipykernel       5.1.4            py38h39e3cac_0
ipython         7.13.0           py38h5ca1d4c_0
ipython_genutils 0.2.0            py38_0
ipywidgets      7.5.1            py_0
jedi            0.17.0           py38_0
jinja2          2.11.2           py_0
jpeg            9b               hb83a4c4_2
jsonschema      3.2.0            py38_0
jupyter         1.0.0            py38_7
jupyter_client  6.1.3            py_0
jupyter_console 6.1.0            py_0
jupyter_core    4.6.3            py38_0
kiwisolver      1.2.0            py38h74a9793_0
libpng          1.6.37           h2a8f88b_0
libsodium       1.0.16           h9d3ae62_0
libtiff         4.1.0            h56a325e_0
m2w64-gcc-libgfortran 5.3.0           6
m2w64-gcc-libls 5.3.0           7
m2w64-gcc-libstdc++ 5.3.0           7
m2w64-gmp       6.1.0            2
m2w64-libwinpthread-git 5.0.0.4634.697f757 2
markupsafe      1.1.1            py38he774522_0
matplotlib      3.1.3            py38_0
matplotlib-base 3.1.3            py38h64f37c6_0
mistune         0.8.4            py38he774522_1000
mkl             2020.1           216
mkl-service     2.3.0            py38hb782985_0
```

IPython 실행

- 자기주도학습
 - P44~50



```
Terminal: Local × Local (2) × +
Microsoft Windows [Version 10.0.18363.836]
(c) 2019 Microsoft Corporation. All rights reserved.

(datav) D:\(0 2020 이공계 연수 저장소)\Data Analysis & Visualization\0618 code>ipython
Python 3.8.3 (default, May 19 2020, 06:50:17) [MSC v.1916 64 bit (AMD64)]
Type 'copyright', 'credits' or 'license' for more information
IPython 7.13.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: %run hello.py
hello, world!

In [2]: an_apple = 30

In [3]: an_example = 40

In [4]: an
an_apple
an_example
any
```

주피터 실행

- 소스가 있는 폴더에서 터미널 열어
 - jupyter notebook
- 노트북 열기
 - ch02-study.ipynb

책과 함께 수업

- 한글 버전
- 원 버전
 - Python-for-Data-Analysis-2nd-Edition.pdf

