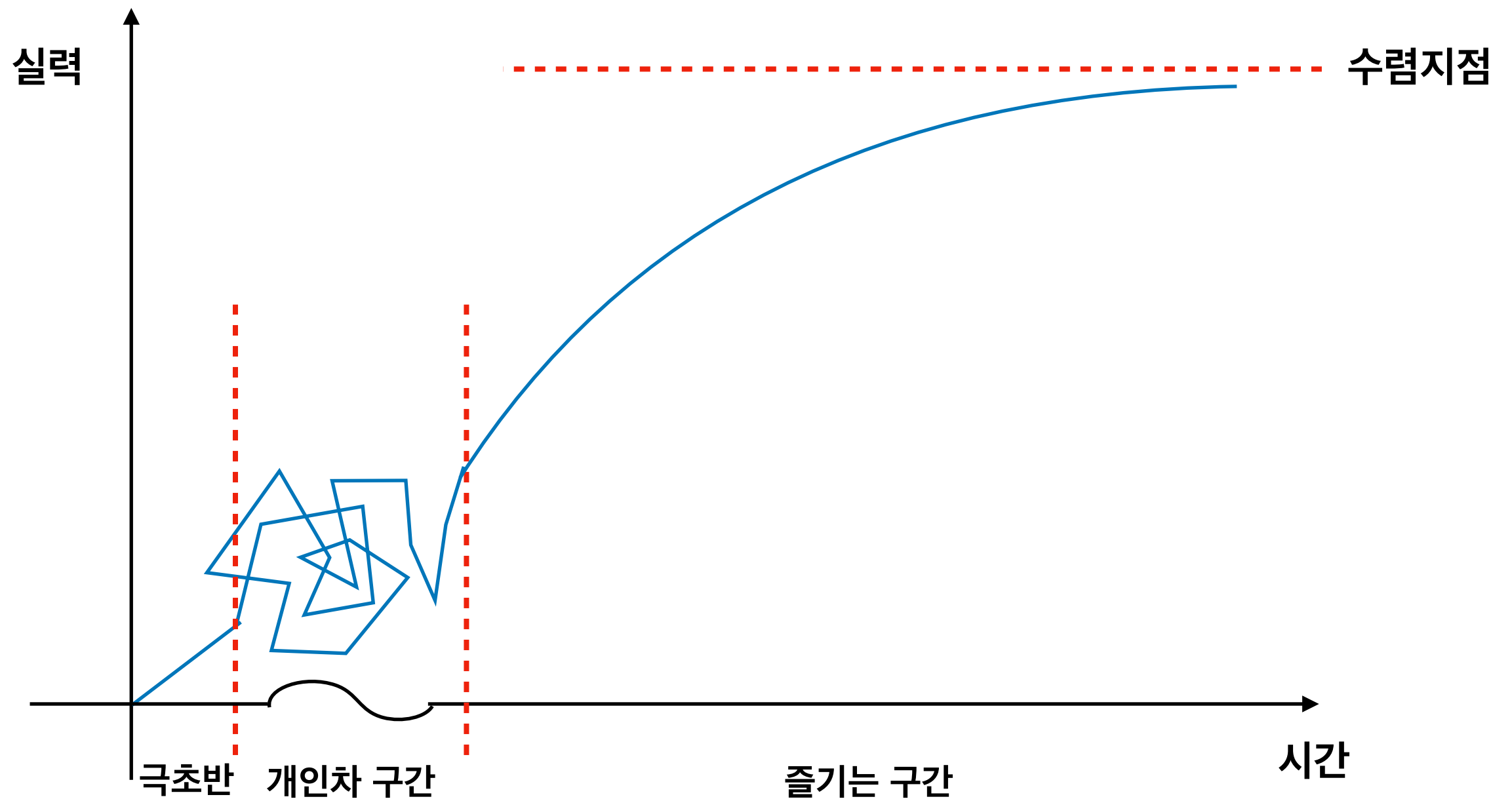


TEXT MINING for PRACTICE

Python을 활용한 비정형 데이터 분석 - APPENDIX
실습환경 소개

연세대학교 | 서중원

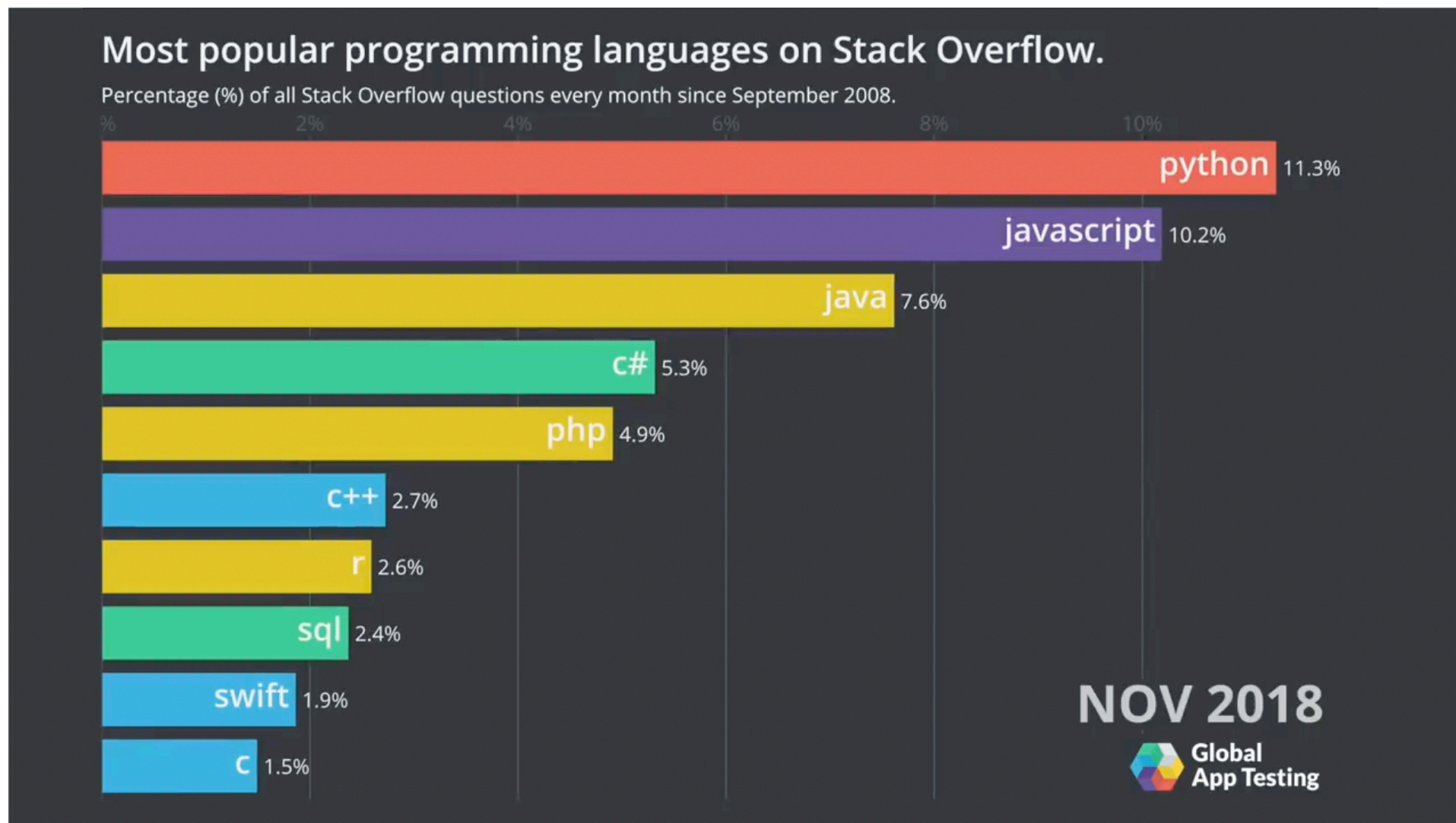
시간에 따른 프로그래밍 실력



프로그래밍 언어

Python

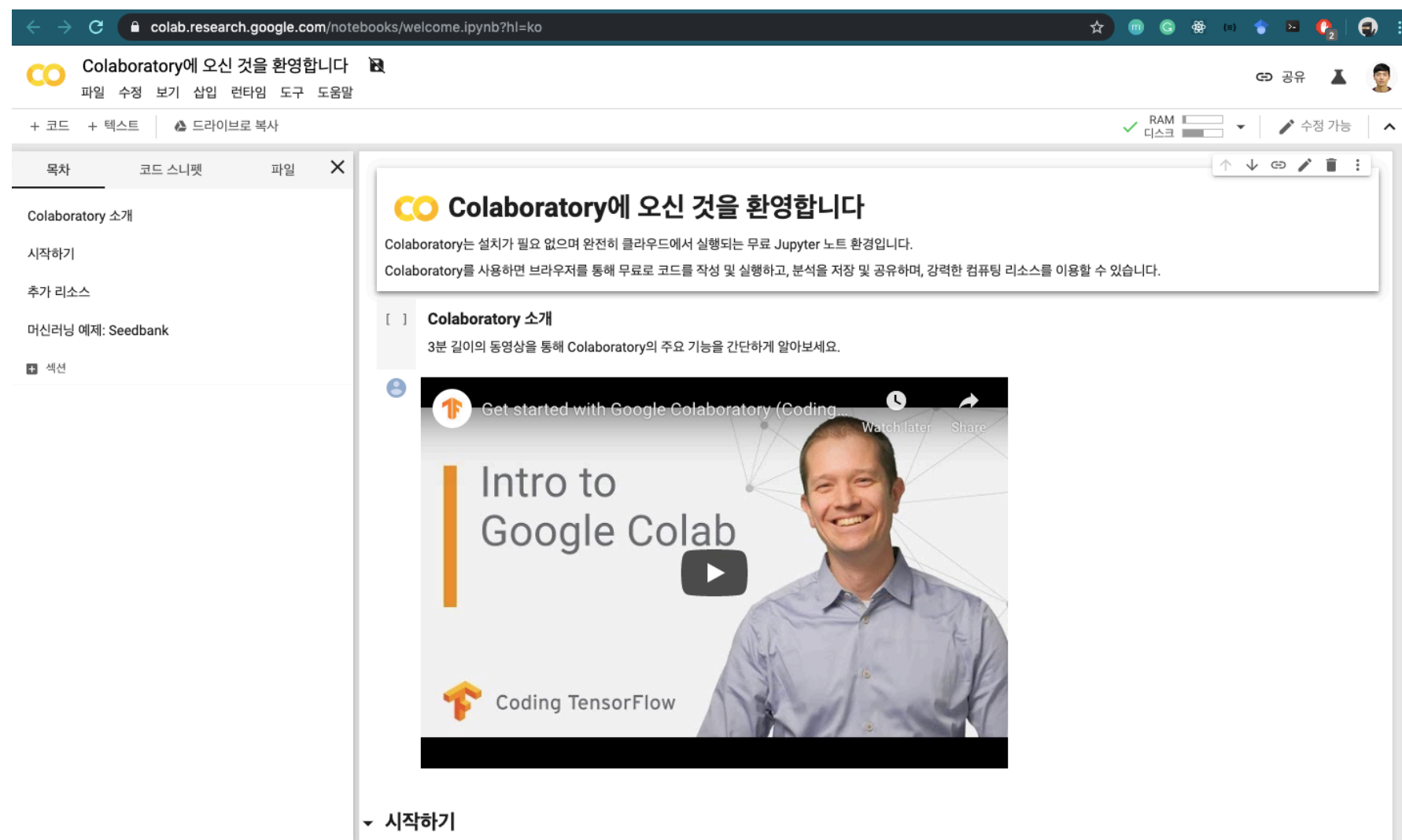
- ▶ 데이터 사이언스에서 가장 널리 쓰이는 언어로, 쉽게 배울 수 있으며 다양한 라이브러리를 제공
- ▶ 실습에서는 주로 Python3.x 버전을 사용



개발환경

Google Colaboratory

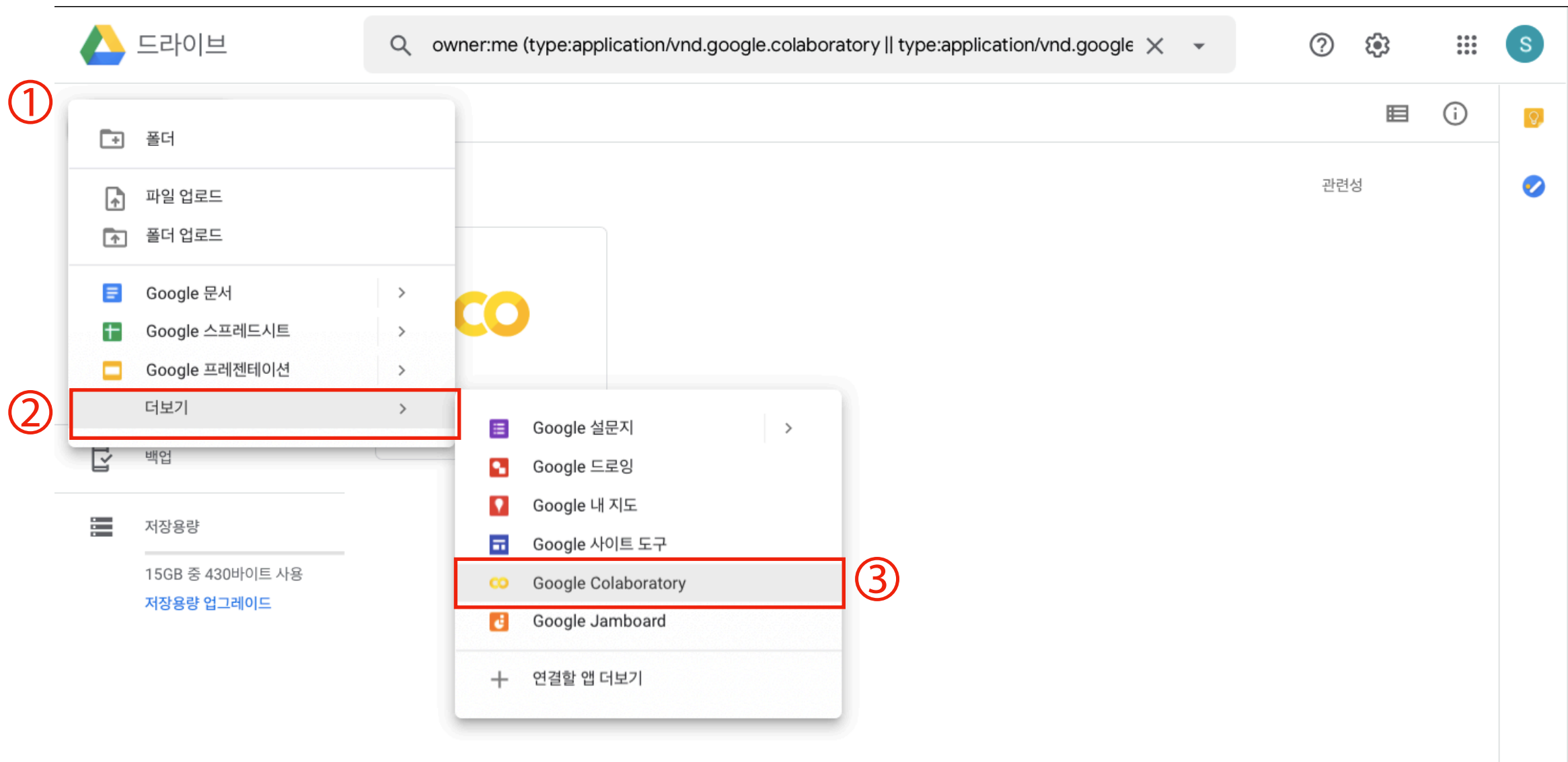
- ▶ 설치 없이, 웹 브라우저 환경에서 바로 작업 할 수 있는 환경
- ▶ 중간 중간 값을 확인해야하는 데이터 사이언스의 작업 특성상, 줄단위로 코드를 실행 시킬 수 있는 **대화형** 환경이 필요



Colab 사용법

1. 구글 계정 준비

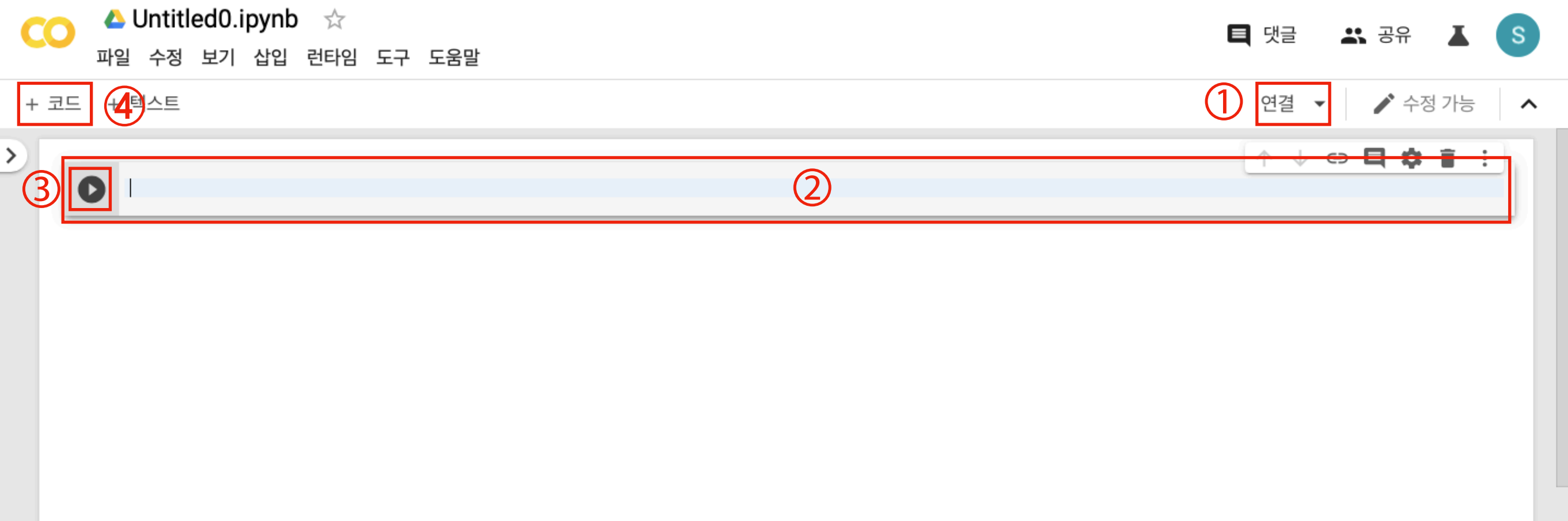
- ▶ 구글 드라이브에 로그인 후(<https://drive.google.com>), Google Colabortory 파일 생성
- ▶ 새로만들기 -> 더보기 -> Google Colabotory 선택



Colab 사용법

2. 주요 구성 요소

- ① 연결 버튼 : 필요한 리소스 (RAM, CPU) 를 갖춘 서버와 연결하기 위한 버튼
- ② 코드 셀 : 코드의 한 묶음 단락을 '셀'이라고 표현함. 각각의 셀은 독립적으로 실행되지만, 메모리상의 변수는 공유 할 수 있음
- ③ 실행 버튼: 해당 셀을 실행시키는 버튼
- ④ 셀 추가 버튼 : 코드 셀을 추가 할 수 있음

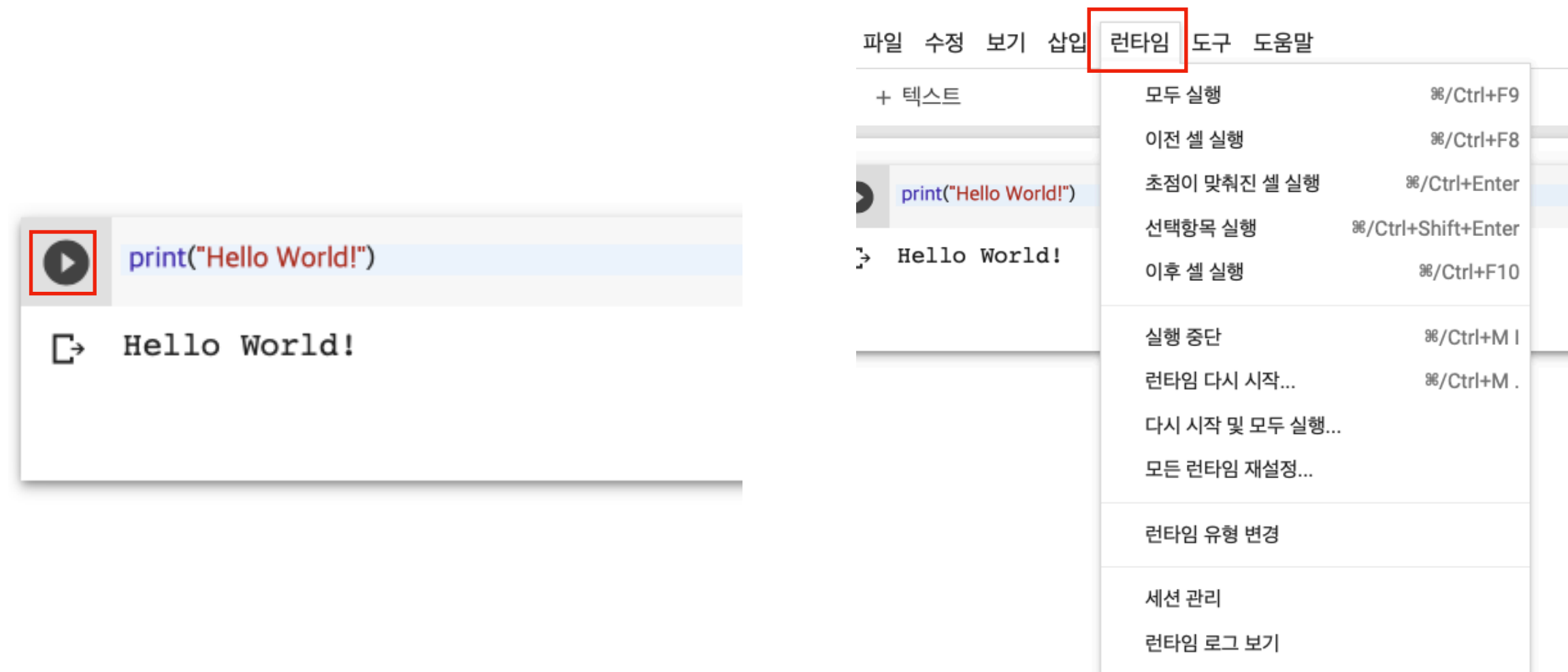


Colab 사용법

3. 실행

단일 셀 실행 : 셀에 코드 작성후 실행 버튼 클릭

여러 셀 실행 : 상단의 런타임 메뉴에서 여러 옵션들을 제공



Colab 사용법

4. 파이썬 패키지 설치

- 대부분의 패키지는 이미 설치 되어 있지만, 많은 텍스트 마이닝 관련 패키지들은 추가적으로 설치 필요
- 기본적으로 pip install 명령어로 설치를 하나, 제일 앞에 느낌표(!)를 사용하므로써 파이썬 명령어가 아닌 시스템 명령어로 인식 시킬 수 있음

```
[8] import numpy
    numpy.zeros(10)

array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])

[9] import konlpy

-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-9-45c72386164e> in <module>()
----> 1 import konlpy

ModuleNotFoundError: No module named 'konlpy'

-----

NOTE: If your import is failing due to a missing package, you can
manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the
"Open Examples" button below.
```

OPEN EXAMPLES SEARCH STACK OVERFLOW

konlpy 패키지 설치 전

```
!pip install konlpy

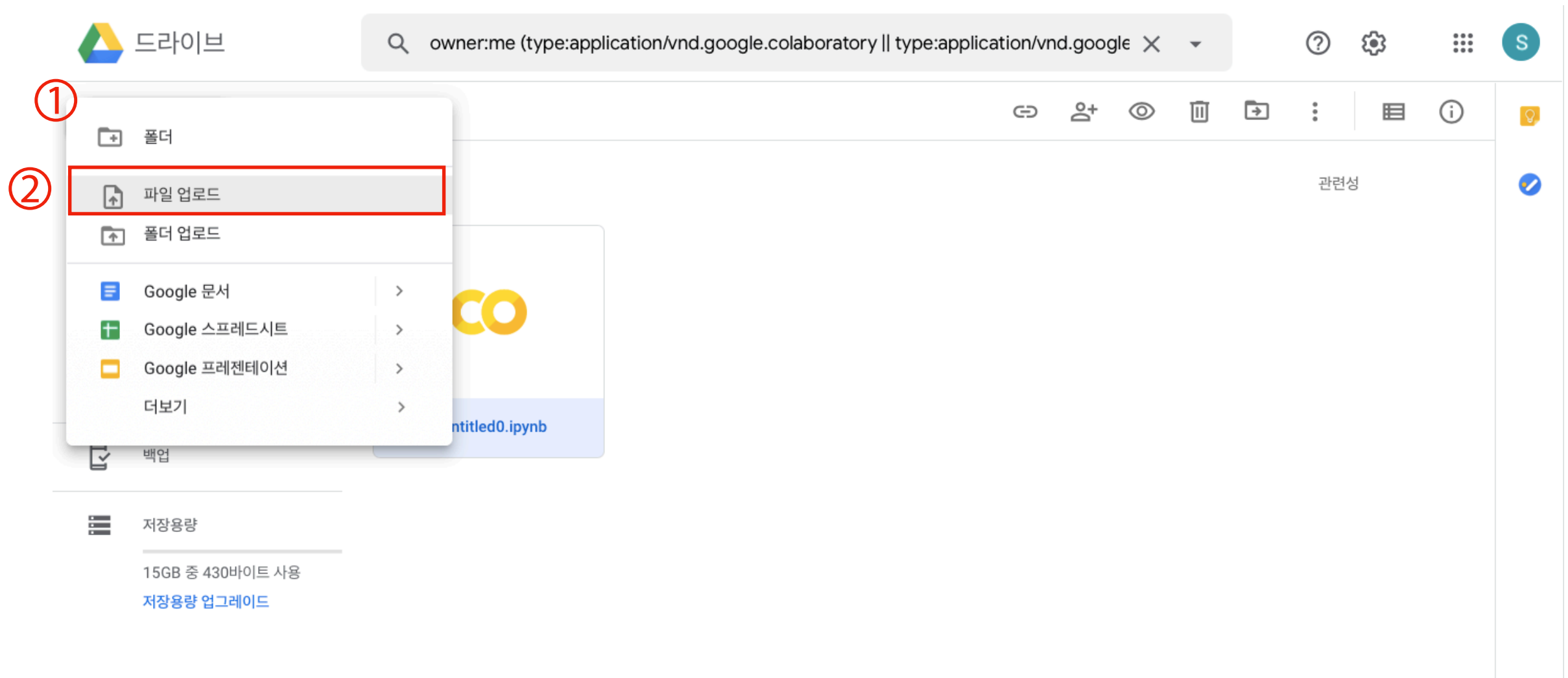
Collecting konlpy
  Using cached https://files.pythonhosted.org/packages/
Requirement already satisfied: JPype1>=0.5.7 in /usr/l
Installing collected packages: konlpy
Successfully installed konlpy-0.5.1
```

konlpy 설치 명령어

Colab 사용법

5. 이미 작성된 코드 사용

외부에서 다운 받은 ipynb 파일이 있다면, 구글 드라이브에 업로드 후 실행을 Google Colabortory로 설정



외부 파일 선택 (확장자 : ipynb)

Colaboratory에 오신 것을 환영합니다 - Colaboratory

검색결과 - Google 드라이브

Untitled0.ipynb - Colaboratory

드라이브

새로 만들기

내 드라이브

공유 문서함

최근 문서함

중요

휴지통

백업

저장용량

15GB 중 430바이트 사용

[저장용량 업그레이드](#)

Mac용 Backup and Sync 다운로드

week-01

검색

이름	수정일
W01_1_python_basic_1.ipynb	2019년 7월 26일 오전 12:16
W01_2_python_basic_2.ipynb	2019년 7월 26일 오전 12:16
W01_2_python_data_structure.ipynb	2019년 7월 20일 오후 9:43
W01_2_python_loop_condition.ipynb	2019년 7월 26일 오전 12:16


취소

선택





관련성


내 드라이브 > Colab Notebooks > Untitled0.ipynb


업로드 된 파일 클릭


 드라이브


드라이브 검색


   


 새로 만들기


 내 드라이브


 공유 문서함

 최근 문서함

 중요








 휴지통

 백업


 저장용량


15GB 중 18.6KB 사용
[저장용량 업그레이드](#)


내 드라이브 > Colab Notebooks

파일 이름 ↑

 Untitled0.ipynb

 Untitled1.ipynb

 W01_1_python_basic_1.ipynb

Google Colaboratory 선택

The screenshot shows the Google Drive web interface. At the top, a file named 'W01_1_python_basic_1.ipynb' is selected. A dropdown menu shows 'Google Colaboratory(으)로 열기'. A modal dialog is open in the center with the title '미리보기가 없음' (No preview). It contains two buttons: '다운로드' (Download) and '추가 앱 연결...' (Add app connection...). Below the modal, a message says '이 항목을 열거나 수정하려면 아래 앱 중 하나를 사용해 보세요.' (Try one of the following apps to open or edit this item). Under the heading '연결된 앱' (Connected apps), 'Google Colaboratory' is listed and highlighted with a red rectangle. The left sidebar shows navigation options like '내 드라이브' (My Drive), '공유 문서함' (Shared with me), '최근 문서함' (Recent), '중요' (Important), '휴지통' (Trash), '백업' (Backup), and '저장용량' (Storage). The bottom left corner has a notification for 'Mac용 Backup and Sync 다운로드' (Download Backup and Sync for Mac).

W01_1_python_basic_1.ipynb

드라이브 검색

Google Colaboratory(으)로 열기

내 드라이브 > Colab Notebooks

파일 이름 ↑

미리보기가 없음

다운로드 추가 앱 연결...

이 항목을 열거나 수정하려면 아래 앱 중 하나를 사용해 보세요.

연결된 앱

Google Colaboratory

저장용량

15GB 중 18.6KB 사용

저장용량 업그레이드

Mac용 Backup and Sync 다운로드

완료!



W01_1_python_basic_1.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말

댓글

공유



+ 코드 + 텍스트

연결 ▾

수정 가능



TEXT MINING for PRACTICE: Python Basic-1

기본 변수 타입

Numeric, Boolean, String 타입에 대해 알아보고 관련 연산자 및 함수들을 테스트해봅니다.

1. Numeric

가장 기본이 되는 숫자 관련 변수와 기본적인 연산에 대해 배워보도록 하겠습니다.

```
[ ] # 정수형 변수 선언
x = 1
print(x)
print(type(x))
# 유리수형 변수 선언
x = 1.0
print(x)
print(type(x))
```



```
[ ] # 사칙연산을 계산해봅니다.
x = 9
print(x+3)
```

실습자료

Github

모든 자료는 Github에 업로드 될 예정 (<https://github.com/thejungwon/text-mining-course>)

GitHub, Inc. [US] | github.com/thejungwon/text-mining-course

6 commits2 branches0 releases1 contributorApache-2.0

Branch: masterNew pull requestCreate new fileUpload filesFind FileClone or download

thejungwon Update README.mdLatest commit 8e1829e 7 minutes ago

LICENSE

Create LICENSElast month

README.md

Update README.md7 minutes ago

README.md

TEXT MINING COURSE

Curriculum

WEEK 01. Python 기초문법 알아보기

- 강의개요 소개
- 비정형 텍스트 데이터 분석 기본개념

E.O.D