

Data Mining – Installation of Weka and Python



Kyuseok Shim

Seoul National University

<http://kdd.snu.ac.kr/~shim>



Weka



Data Mining with Weka

- What's Weka?
 - *A bird found only in New Zealand?*
- Data mining tool
- Machine learning algorithms for data mining tasks
 - 100+ algorithms for classification
 - 75 for data preprocessing
 - 25 to assist with feature selection
 - 20 for clustering, finding association rules, etc



Getting started with Weka

- Install Weka
- Explore the “Explorer” interface
- Explore some datasets
- Build a classifier
- Interpret the output
- Use filters
- Visualize your data set



Download & Install Weka

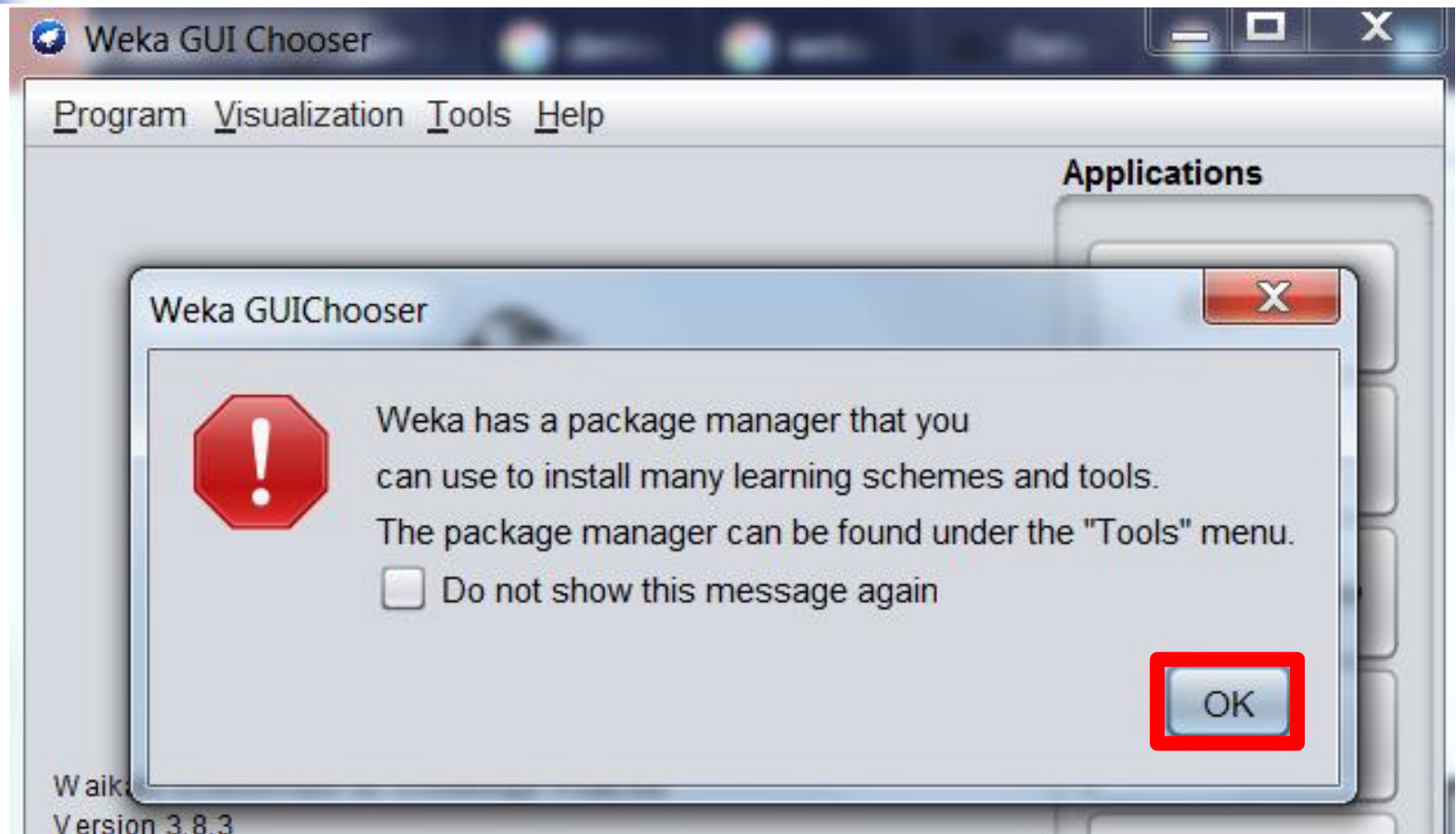
- Download Weka 3.8
 - <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>
 - For 64-bit Windows
 - <http://prdownloads.sourceforge.net/weka/weka-3-8-3jre-x64.exe>
 - For 32-bit Windows
 - <http://prdownloads.sourceforge.net/weka/weka-3-8-3jre.exe>
- The installation is simple
 - Just select 'Next', 'I Agree' or 'Install'



Download the Additional Dataset

- <http://kdd.snu.ac.kr/weka/>
 - [cluster2.arff](#)

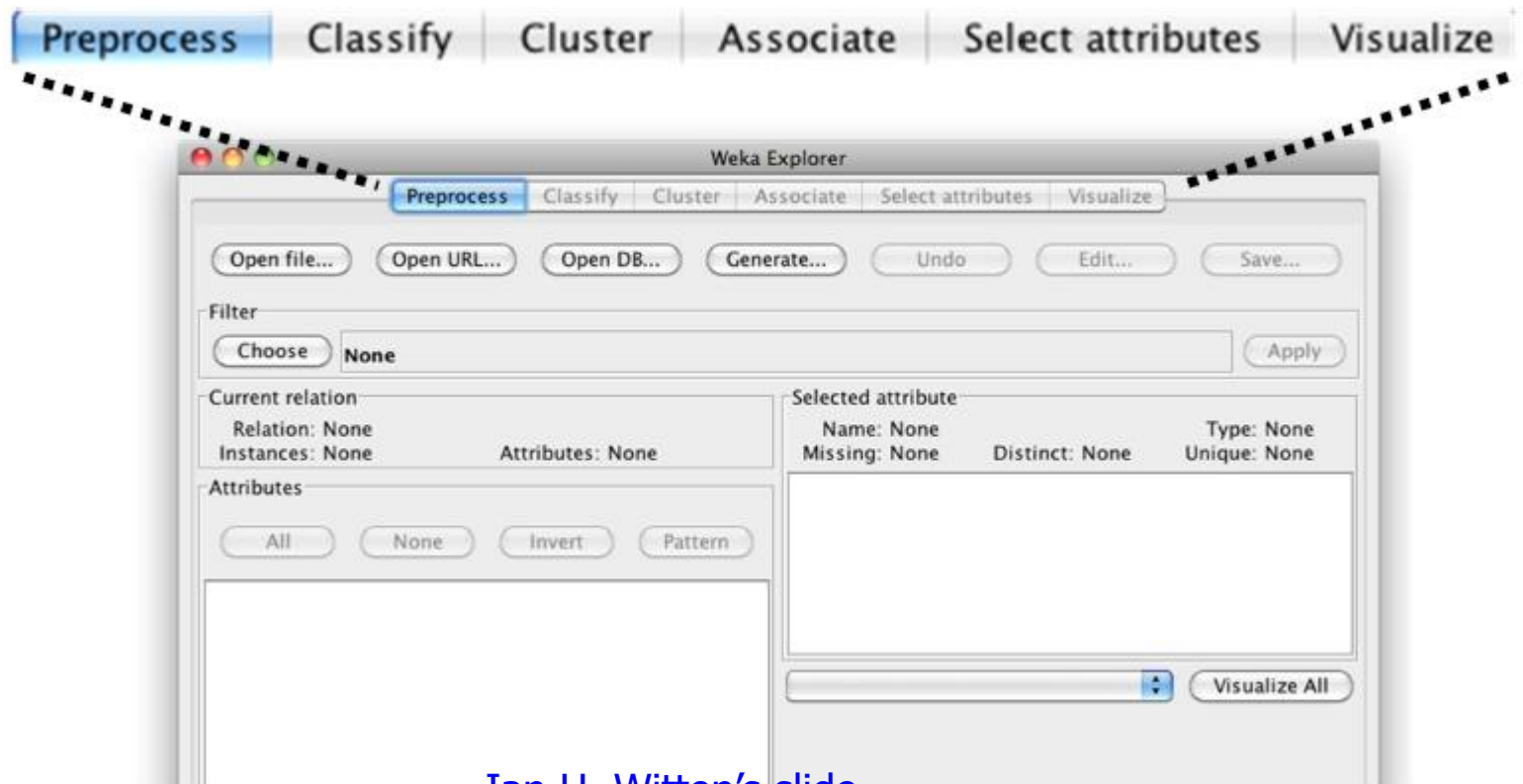
Getting Started with Weka



Exploring the Explorer



Exploring the Explorer



Ian H. Witten's slide



A Dataset: weather.nominal.arff

		attributes				
instances		Outlook	Temp	Humidity	Windy	Play
	1	Sunny	Hot	High	False	No
	2	Sunny	Hot	High	True	No
	3	Overcast	Hot	High	False	Yes
	4	Rainy	Mild	High	False	Yes
	5	Rainy	Cool	Normal	False	Yes
	6	Rainy	Cool	Normal	True	No
	7	Overcast	Cool	Normal	True	Yes
	8	Sunny	Mild	High	False	No
	9	Sunny	Cool	Normal	False	Yes
	10	Rainy	Mild	Normal	False	Yes
	11	Sunny	Mild	Normal	True	Yes
	12	Overcast	Mild	High	True	Yes
	13	Overcast	Hot	Normal	False	Yes
	14	Rainy	Mild	High	True	No



weather.nominal.arff

- C:\Program Files\Weka-3-8\data\weather.nominal.arff

@relation weather.symbolic

The name of the data

The name of an attribute

@attribute outlook {sunny, overcast, rainy}

@attribute temperature {hot, mild, cool}

@attribute humidity {high, normal}

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

The categories of the attribute

Generally, the last attribute represents the label to predict

@data

sunny,hot,high,FALSE,no

sunny,hot,high,TRUE,no

overcast,hot,high,FALSE,yes

.....

Data instances

Classification Problem

		attributes				
		Outlook	Temp	Humidity	Windy	Play
instances	1	Sunny	Hot	High	False	No
	2	Sunny	Hot	High	True	No
	3	Overcast			False	Yes
	4	Rainy			False	Yes
	5			Normal	False	Yes
	6			Normal	True	No
	7		Cool	Normal	True	Yes
	8		Mild			No
	9	Sunny	Cool	Normal	False	Yes
	10	Rainy	Mild	Normal	False	Yes
	11	Sunny	Mild	Normal	True	Yes
	12	Overcast	Mild	High	True	Yes
	13	Overcast	Hot	Normal	False	Yes
	14	Rainy	Mild	High	True	No

Classification problem:
predict the "class" value



weather.numeric.arff

- C:\Program Files\Weka-3-8\data\weather.nominal.arff

@relation weather

@attribute outlook {sunny, overcast, rainy}

@attribute temperature numeric

@attribute humidity numeric

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

A numeric attribute

@data

sunny,85,85,FALSE,no

sunny,80,90,TRUE,no

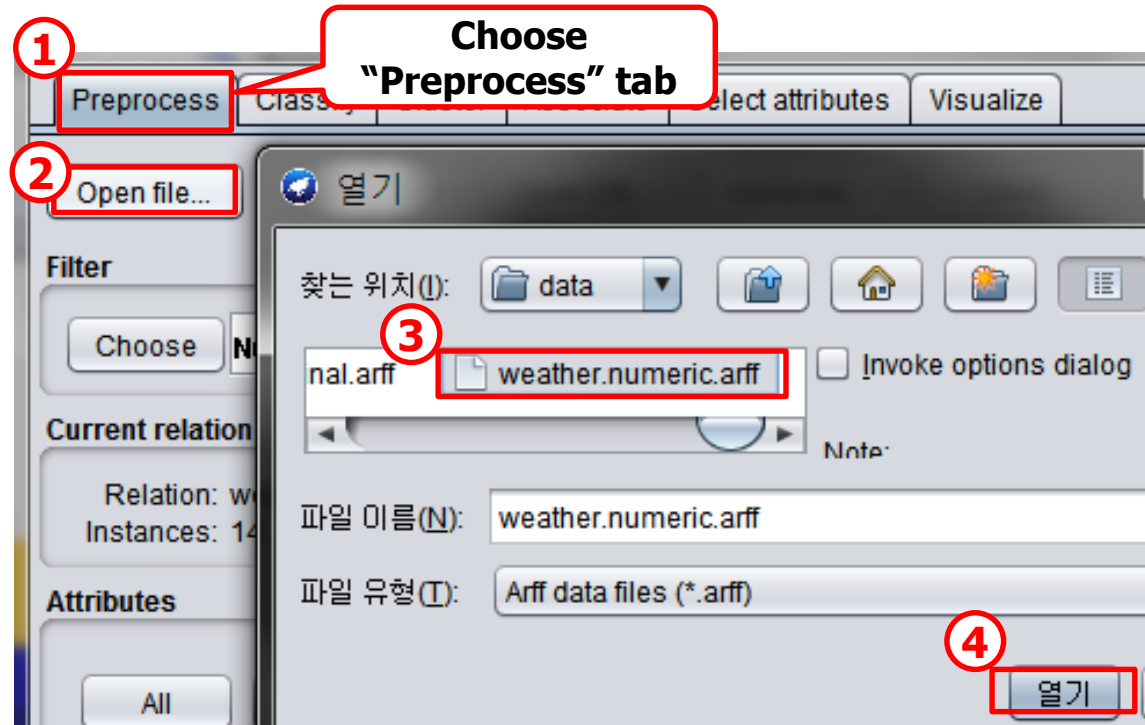
overcast,83,86,FALSE,yes

.....

outlook	temperate	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
...

Open the Dataset

- C:\Program Files\Weka-3-8\data\weather.numeric.arff



The Statistics of the Dataset

The sum of weights of instances, which is initially equal to the number of instances since the initial weight of each instance has 1

The name of the dataset

The number of attributes

The number of instances

Current relation

Relation: weather
Instances: 14

Attributes: 5
Sum of weights: 14

Selected attribute

Name: outlook
Missing: 0 (0%)

Attributes

All

None

Invert

Patte...

No.		Name
1	<input checked="" type="checkbox"/>	outlook
2	<input type="checkbox"/>	temperature
3	<input type="checkbox"/>	humidity
4	<input type="checkbox"/>	windy

No.	Label
1	sunny
2	overcast
3	rainy

Class: play (Nom)

Select the Class

Current relation
Relation: weather
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

AllNoneInvertPattern

No.		Name
1	<input type="checkbox"/>	outlook
2	<input type="checkbox"/>	temperature
3	<input type="checkbox"/>	humidity
4	<input type="checkbox"/>	windy
5	<input checked="" type="checkbox"/>	play

Remove

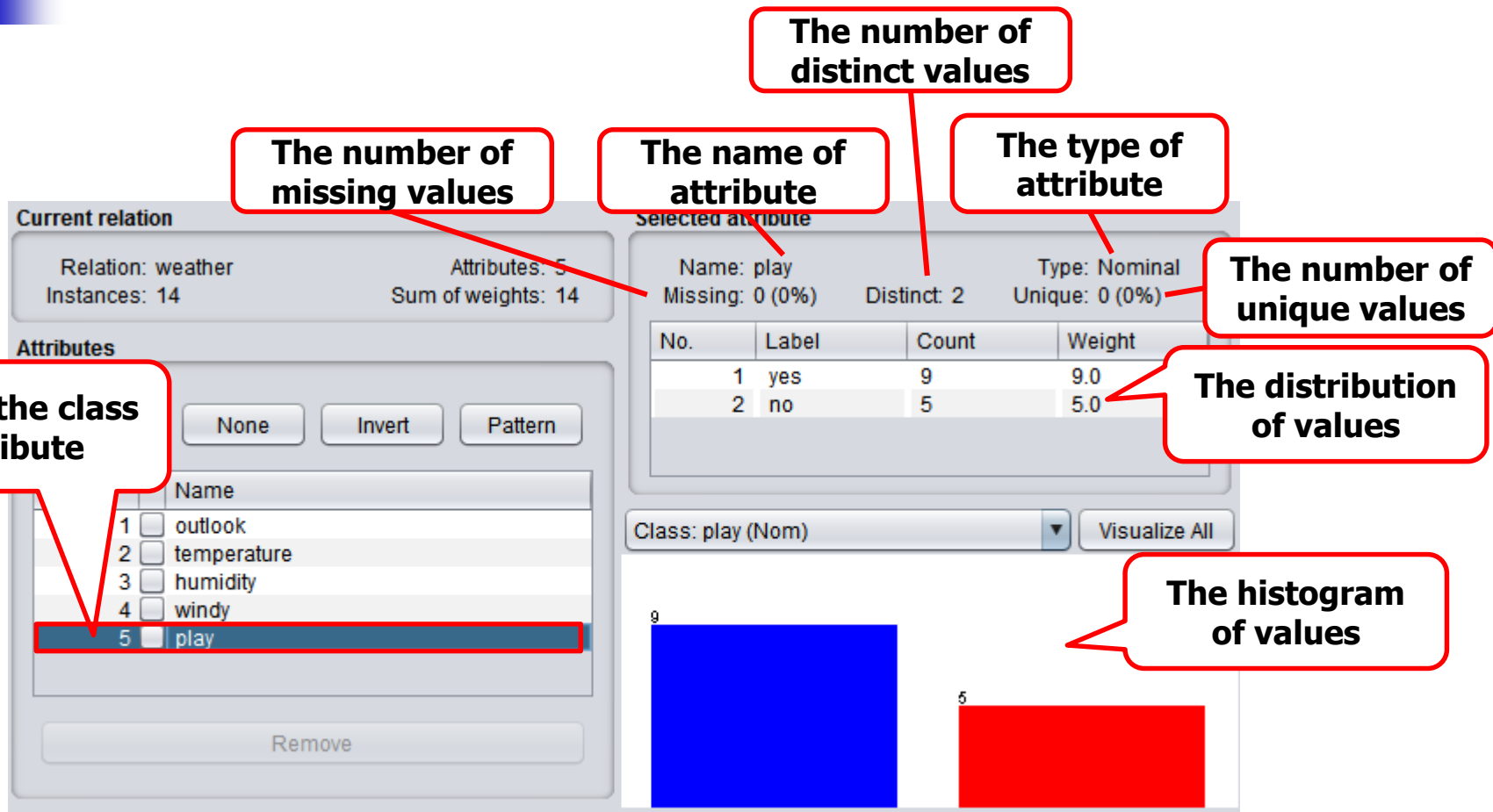
Selected attribute
Name: play
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	yes	9	9.0
2	no	5	5.0

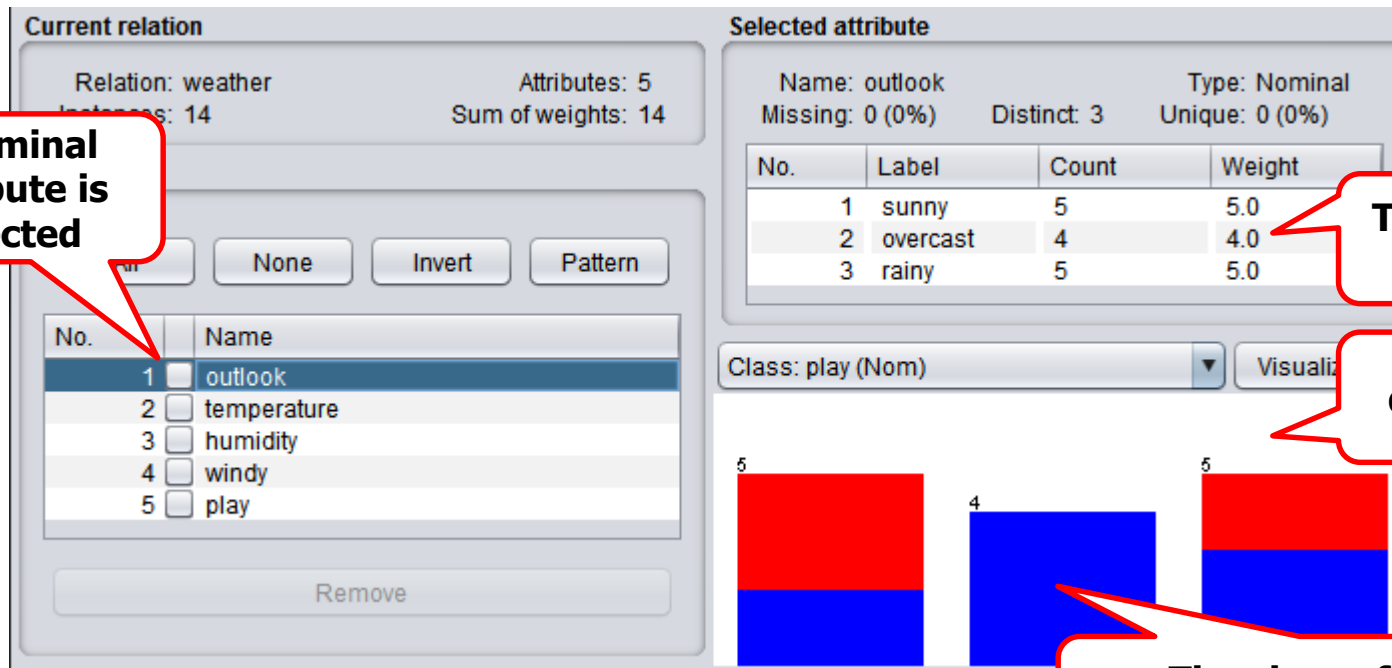
Class: play (Nom)
No class
Class: outlook (Nom)
Class: temperature (Num)
Class: humidity (Num)
Class: windy (Nom)
Class: play (Nom)

The last attribute is selected as default

The Statistics of the Class



The Statistics of an Attribute



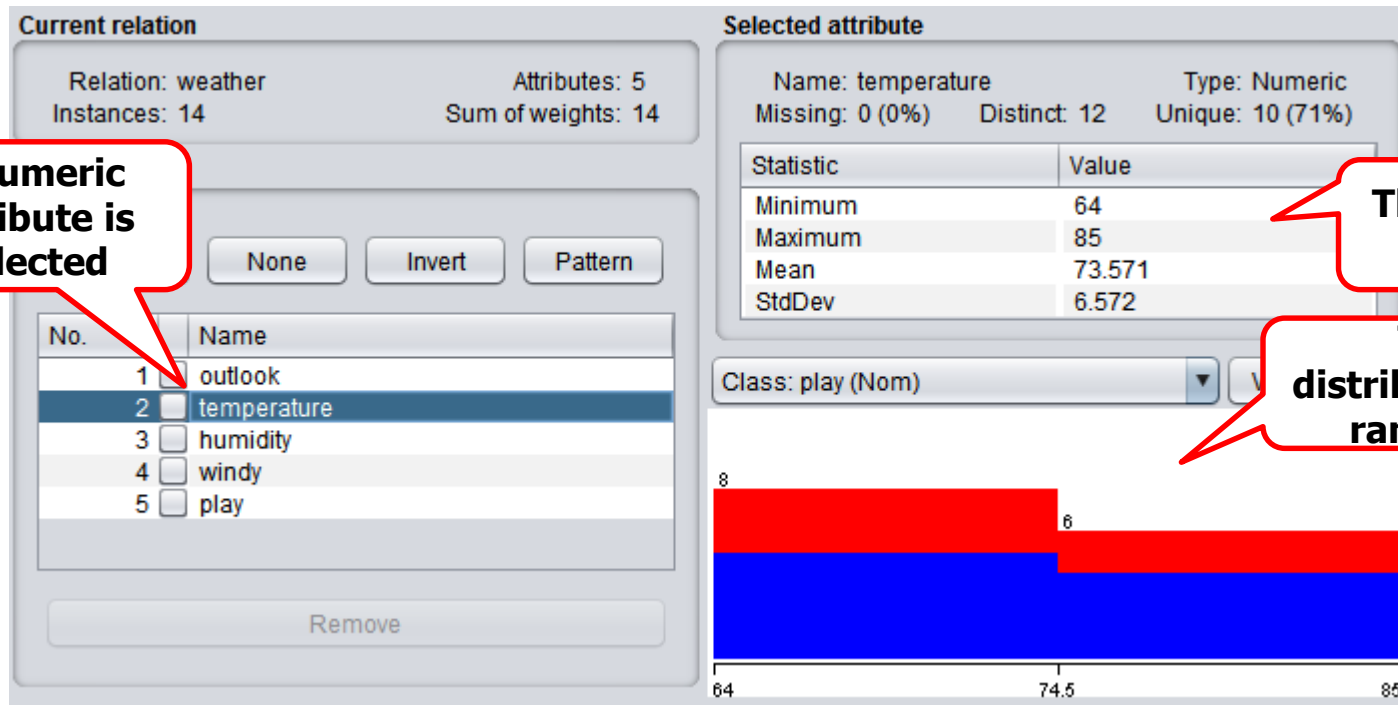
A nominal attribute is selected

The distribution of values

The class distribution of each value

The class of all the 'overcast' instances are 'yes'

The Statistics of an Attribute



A numeric attribute is selected

The distribution of values

The class distribution for each range of value



Weka – Using CSV Files



CSV Files

- CSV files could be easily exported from Microsoft Excel or Google Spreadsheet

	A	B	C	D	E
1	outlook	temperature	humidity	windy	play
2	sunny	85	85	FALSE	no
3	sunny	80	90	TRUE	no
4	overcast	83	86	FALSE	yes
5	rainy	70	96	FALSE	yes
6	rainy	68	80	FALSE	yes
7	rainy	65	70	TRUE	no

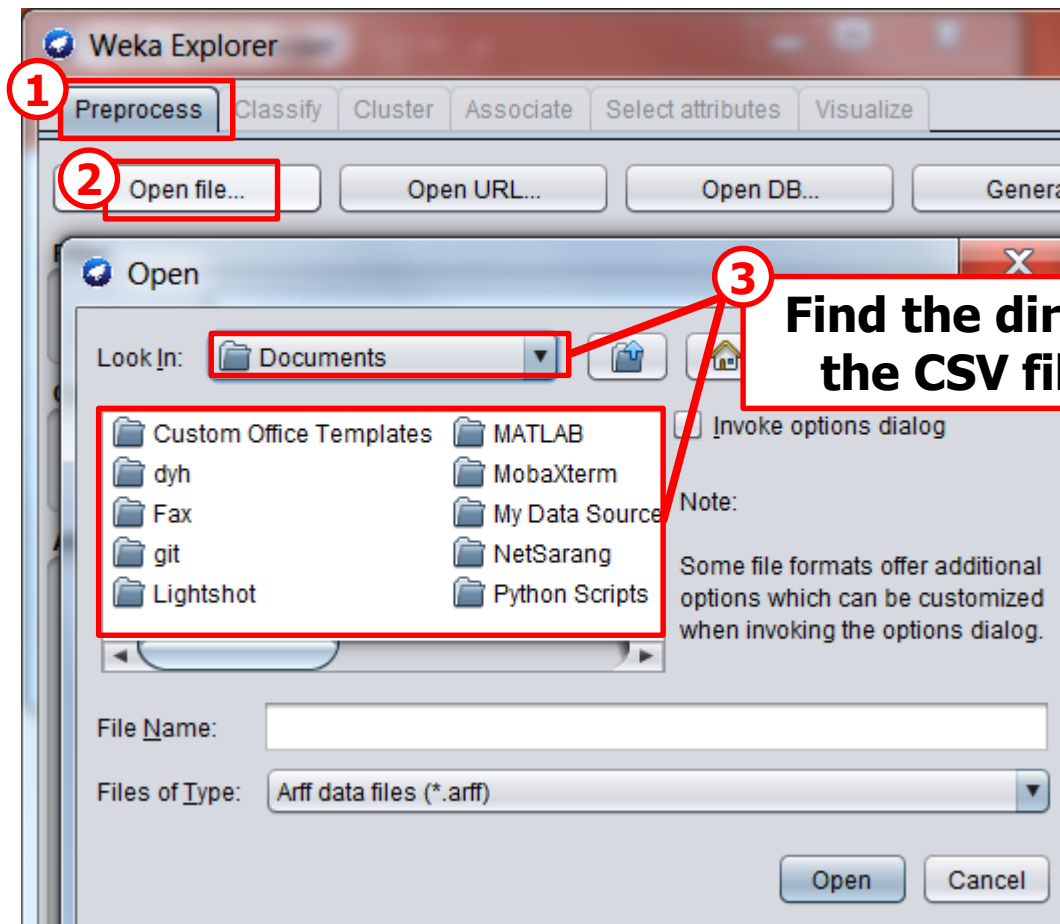
```
outlook,temperature,humidity,windy,play
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
```



Download a Sample CSV File

- <http://kdd.snu.ac.kr/weka/>
 - Download the **test.csv**

Using CSV Files for Weka



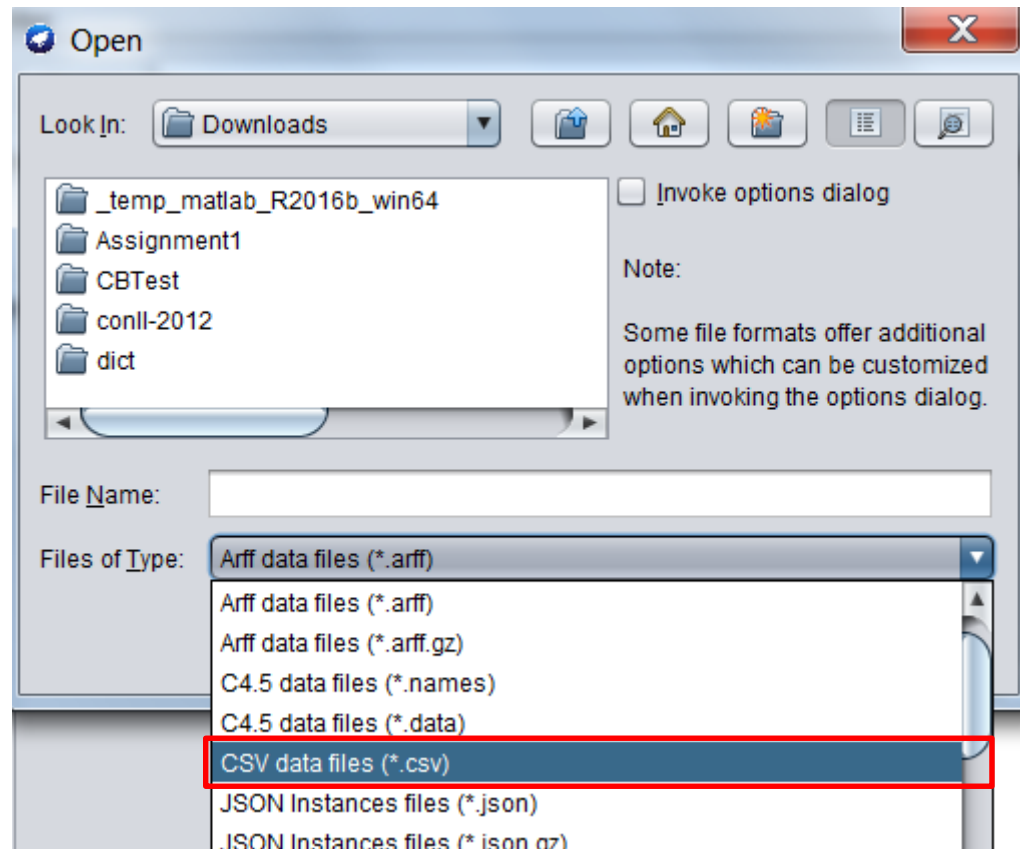
Using CSV Files for Weka

- IF the test.csv is downloaded in C:\Users\dyhong\Downloads



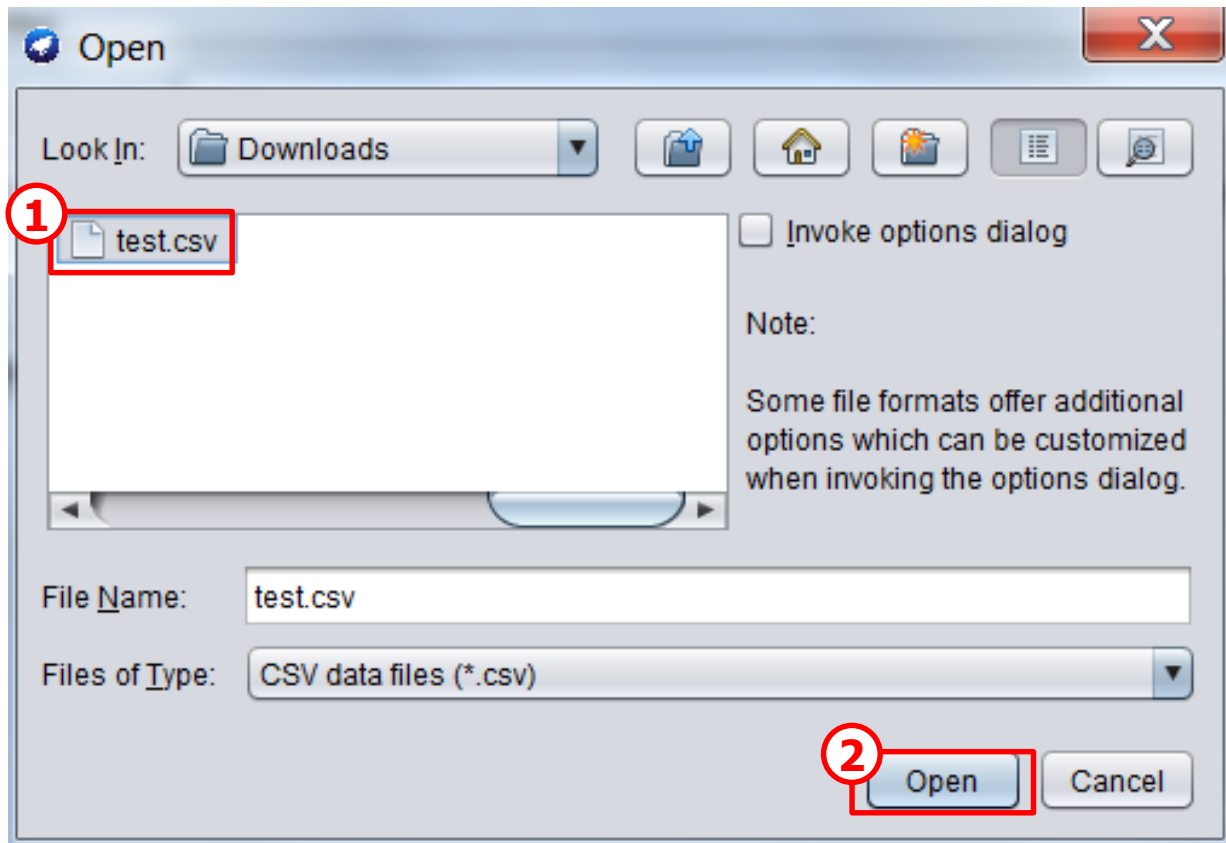
Using CSV Files for Weka

- IF the test.csv is downloaded in C:\Users\dyhong\Downloads



Using CSV Files for Weka

- IF the test.csv is downloaded in C:\Users\dyhong\Downloads





Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file...

Open UR...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

No

The data is loaded successfully

Apply

Stop

Current relation

Relation: test
Instances: 14

Attributes: 5
Sum of weights: 14

Selected attribute

Name: outlook
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

Attributes

All

None

Invert

Pattern

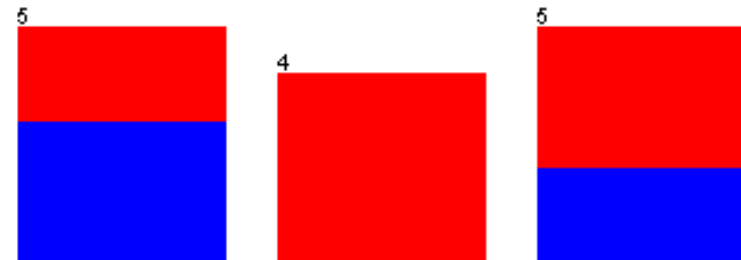
No.		Name
1	<input checked="" type="checkbox"/>	outlook
2	<input type="checkbox"/>	temperature
3	<input type="checkbox"/>	humidity
4	<input type="checkbox"/>	windy
5	<input type="checkbox"/>	play

Remove

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

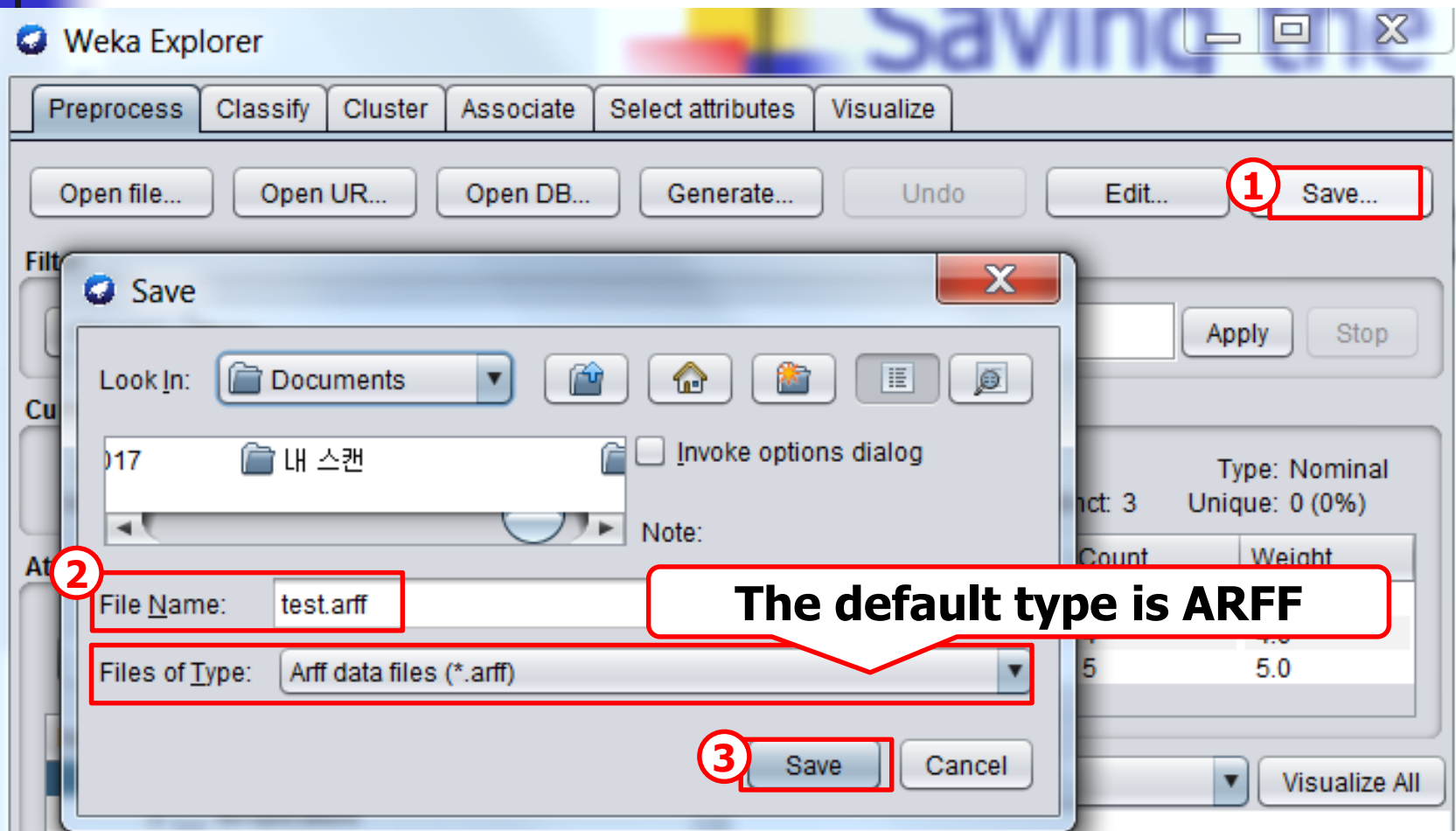
Class: play (Nom)

Visualize All



Status

Saving the CSV File as ARFF File





Basics of Python



Anaconda

- A free and open-source distribution of the Python
 - Includes useful scientific packages
 - Simplifies package management and deployment





Anaconda Installation

- <https://www.anaconda.com/distribution/>



Products

Why Anaconda?

Solutions

Resources

Company

Download

🔍 Search

Anaconda Distribution

Click

The World's Most Popular Python/R Data Science Platform

Download

Anaconda Installation



Windows



macOS



Linux

1. Click

Anaconda 2018.12 for Windows Installer

2. Click

Python 3.7 version

Download

64-Bit Graphical Installer (614.3 MB)

32-Bit Graphical Installer (509.7 MB)

Python 2.7 version

Download

64-Bit Graphical Installer (560.6 MB)

32-Bit Graphical Installer (458.6 MB)



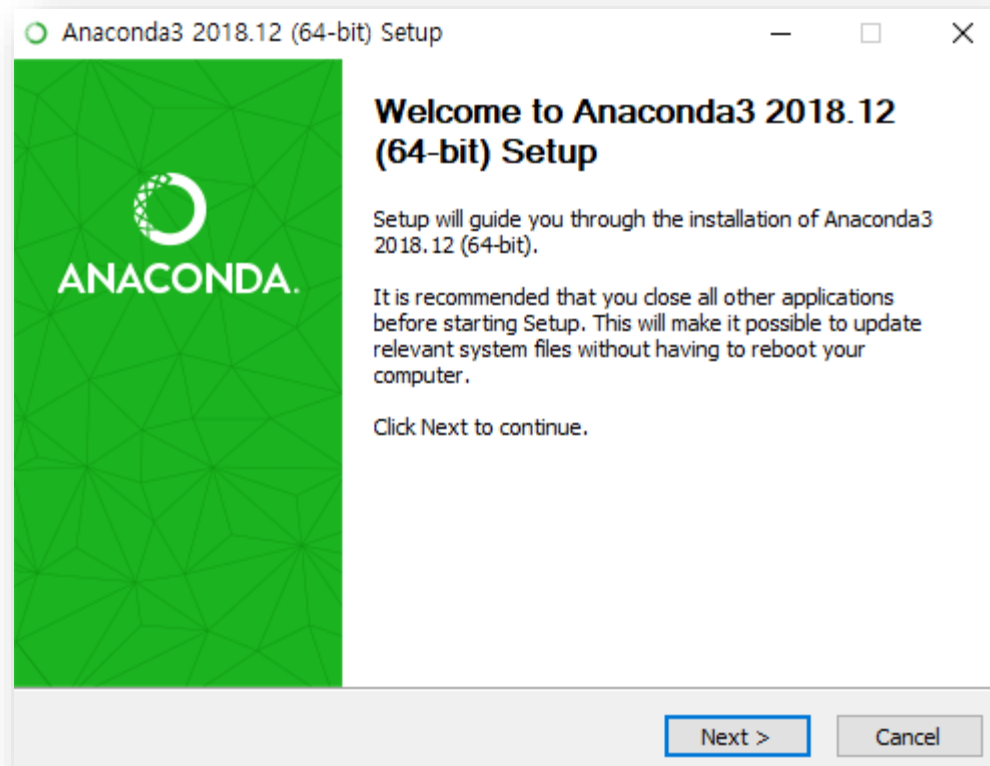
Anaconda Installation

- When

<https://www.anaconda.com/distribution/> is not available, please download the file from <http://kdd.snu.ac.kr/python/>

Anaconda Installation

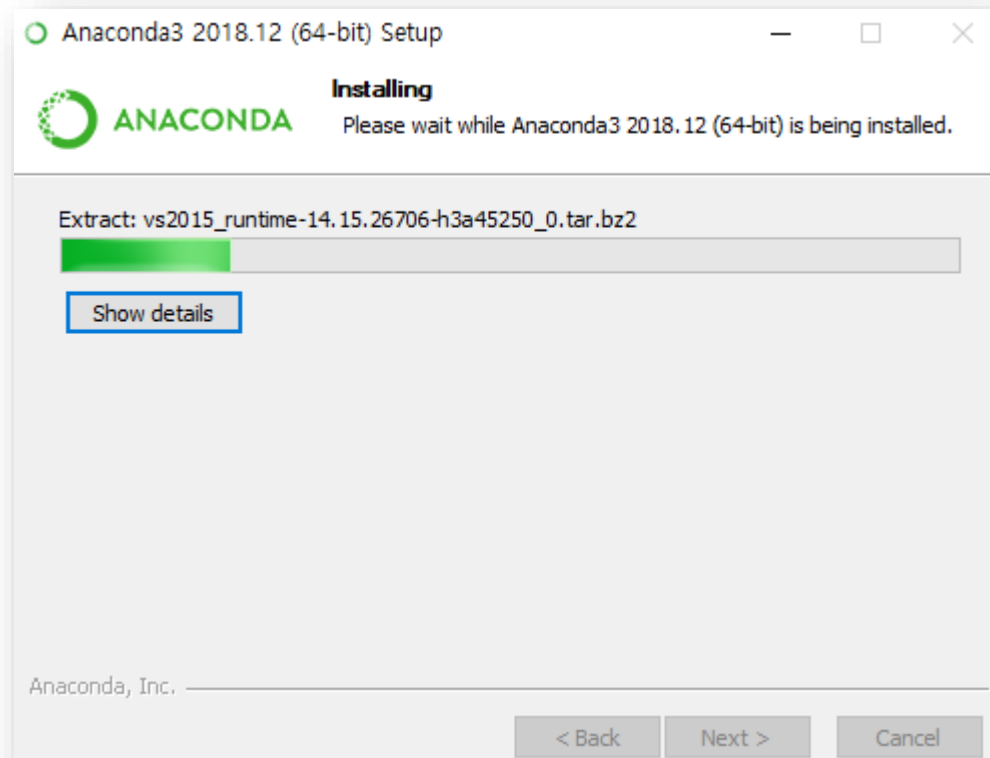
- 1. Run the downloaded file
- 2. Select 'Next' and 'I Agree' and 'Install' for all





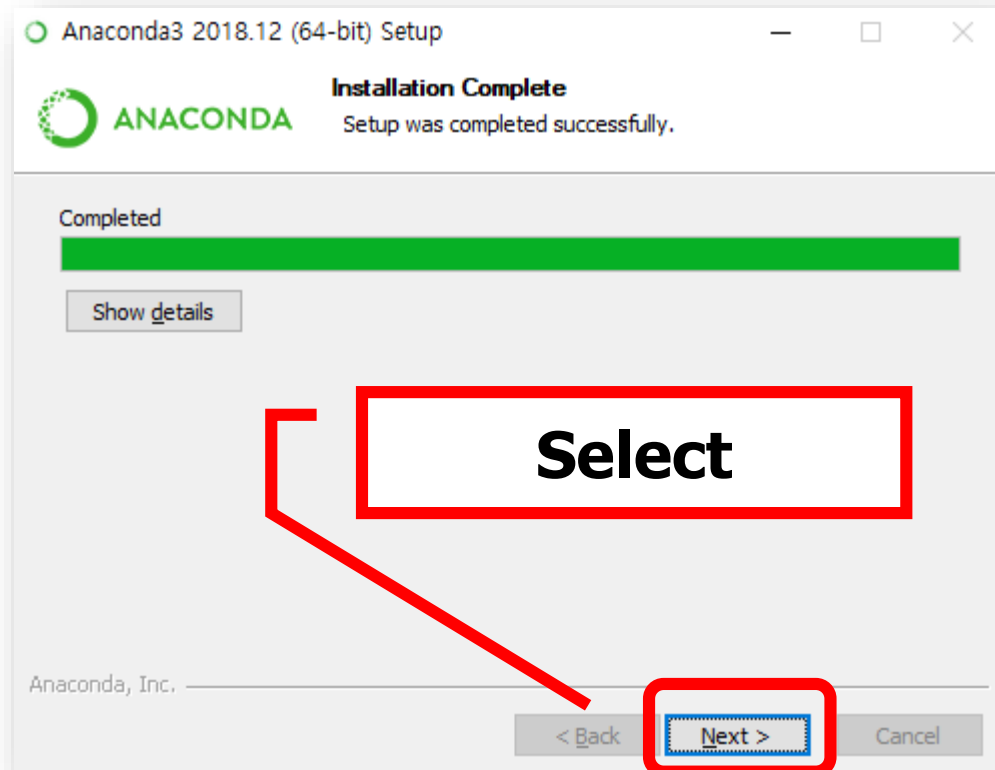
Anaconda Installation

- Installation is in progress

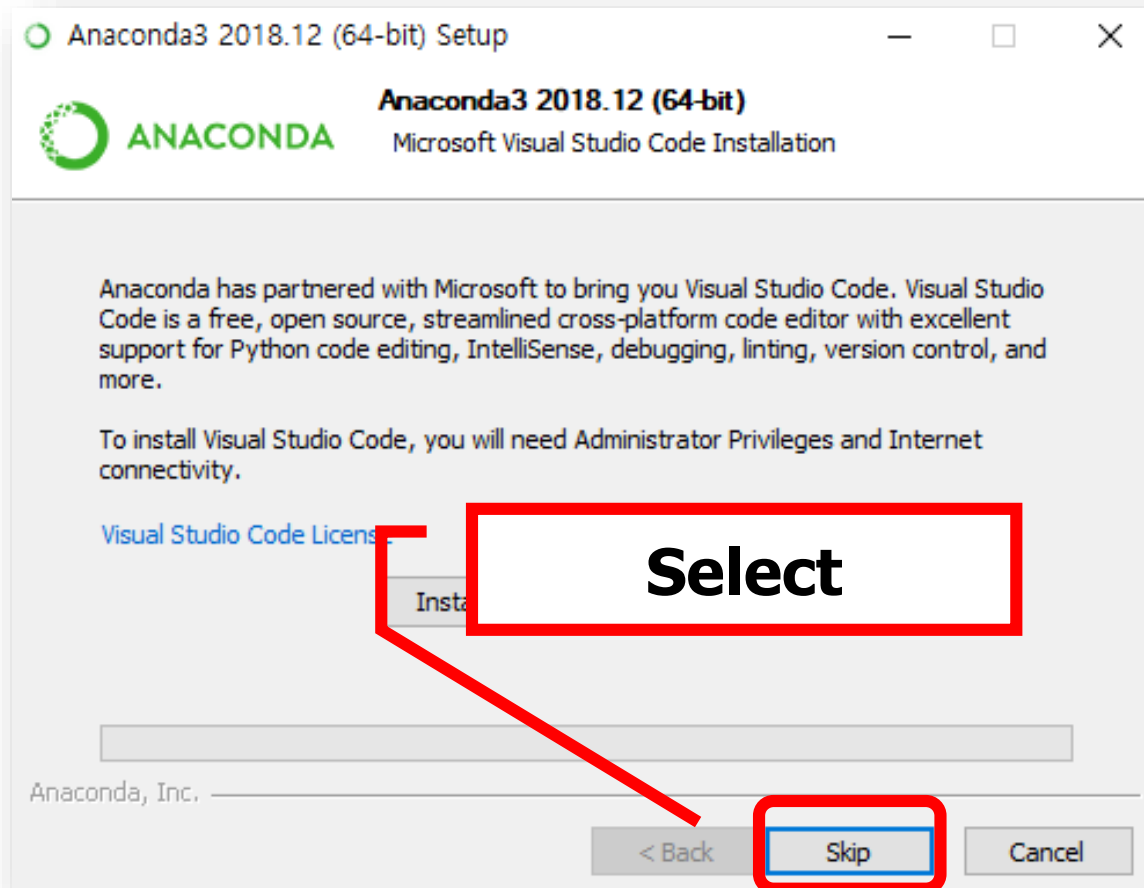


Anaconda Installation

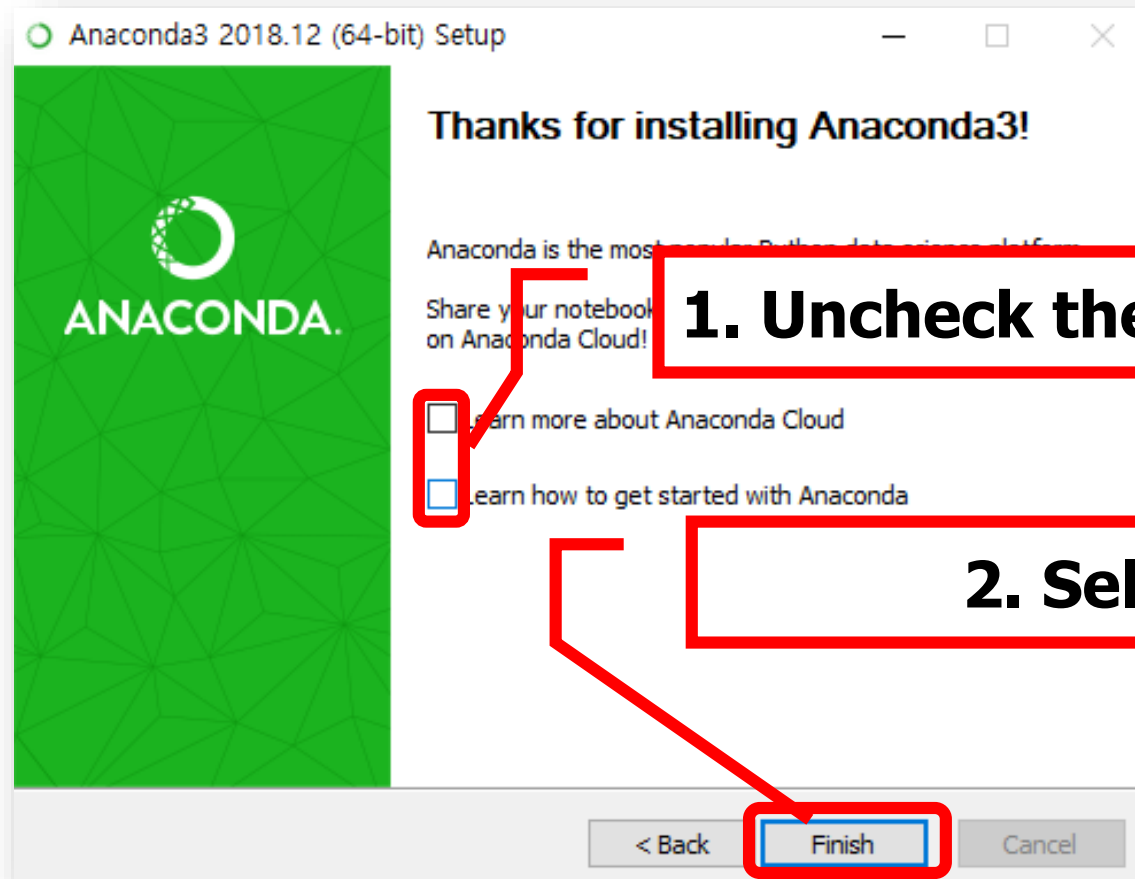
- Installation is complete



Anaconda Installation

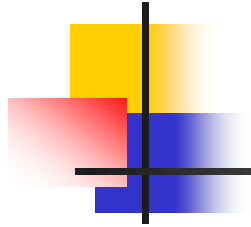


Anaconda Installation



1. Uncheck these options

2. Select



GRAPHVIZ INSTALLATION



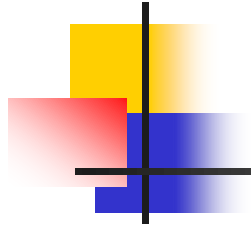
Graphviz

- An open source graph visualization tool
- Used to visualize decision trees in this course



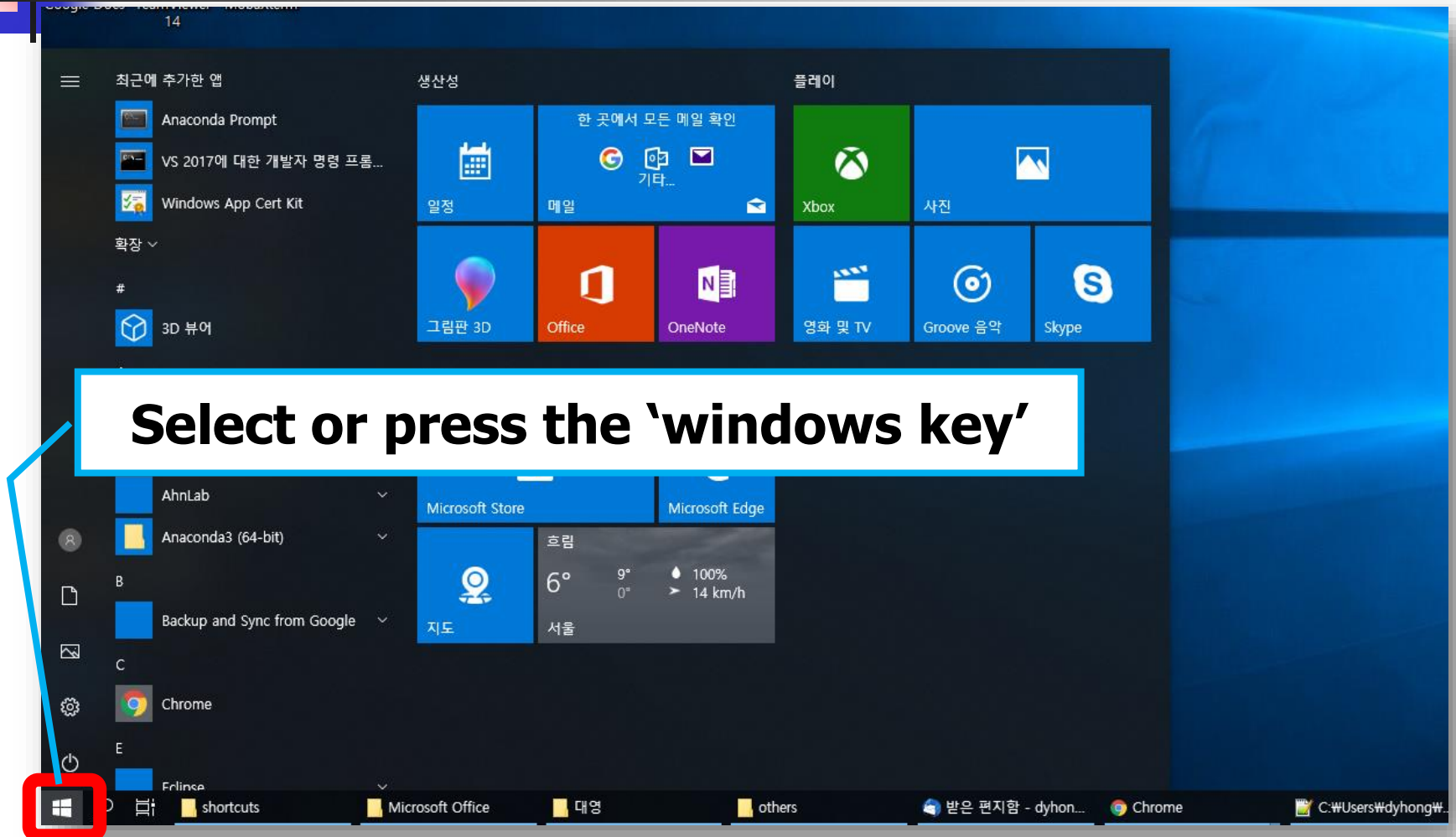
Graphviz Installation Methods

- Installation with Conda
- Manual Installation with Conda
- Installation with pip



GRAPHVIZ INSTALLATION WITH CONDA

Graphviz Installation with Conda



1. Uncheck these options

모두 앱 문서 설정 사진 더보기

모두 앱 문서 설정 사진 더보기

사용자 의견 ...

가장 정확

Anaconda Prompt
데스크톱 앱



Anaconda Prompt
데스크톱 앱

앱

Anaconda Navigator

Uninstall-Anaconda3.exe

폴더

2. Select

anaconda_navigatorsite-packages의 -

anaconda_navigatorsite-packages의 -

anaconda_projectsite-packages의 -

anaconda_projectsite-packages의 -

문서

anaconda-navigator-script.pyw

anaconda

anaconda project 0.0.2 py37_ojson

1. Type "anaconda"

anaconda



Graphviz Installation with Conda

**Type this command and press
Enter**

Anaconda Prompt

(base) C:\Users\Wdyhong>conda install python-graphviz

Anaconda Prompt - conda install python-graphviz

```
(base) C:\Users\Wdyhong>conda install python-graphviz
Collecting package metadata: done
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: C:\Users\Wdyhong\Anaconda3
```

```
added / updated specs:
- python-graphviz
```

Type "y" and press Enter

```
The following NEW packages will be INSTALLED:
```

```
python-graphviz      pkgs/main/win-64::python-graphviz-0.8.4
```

```
Proceed ([y]/n)? 
```



Graphviz Installation with Conda

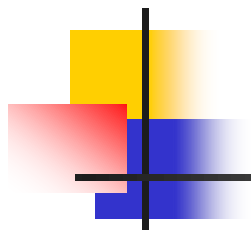
- The installation is complete



```
Anaconda Prompt

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(base) C:\Users\Wdyhong>
```



MANUAL GRAPHVIZ INSTALLATION WITH CONDA



Manual Graphviz Installation with Conda

- An alternative installation method when the previous method does not work

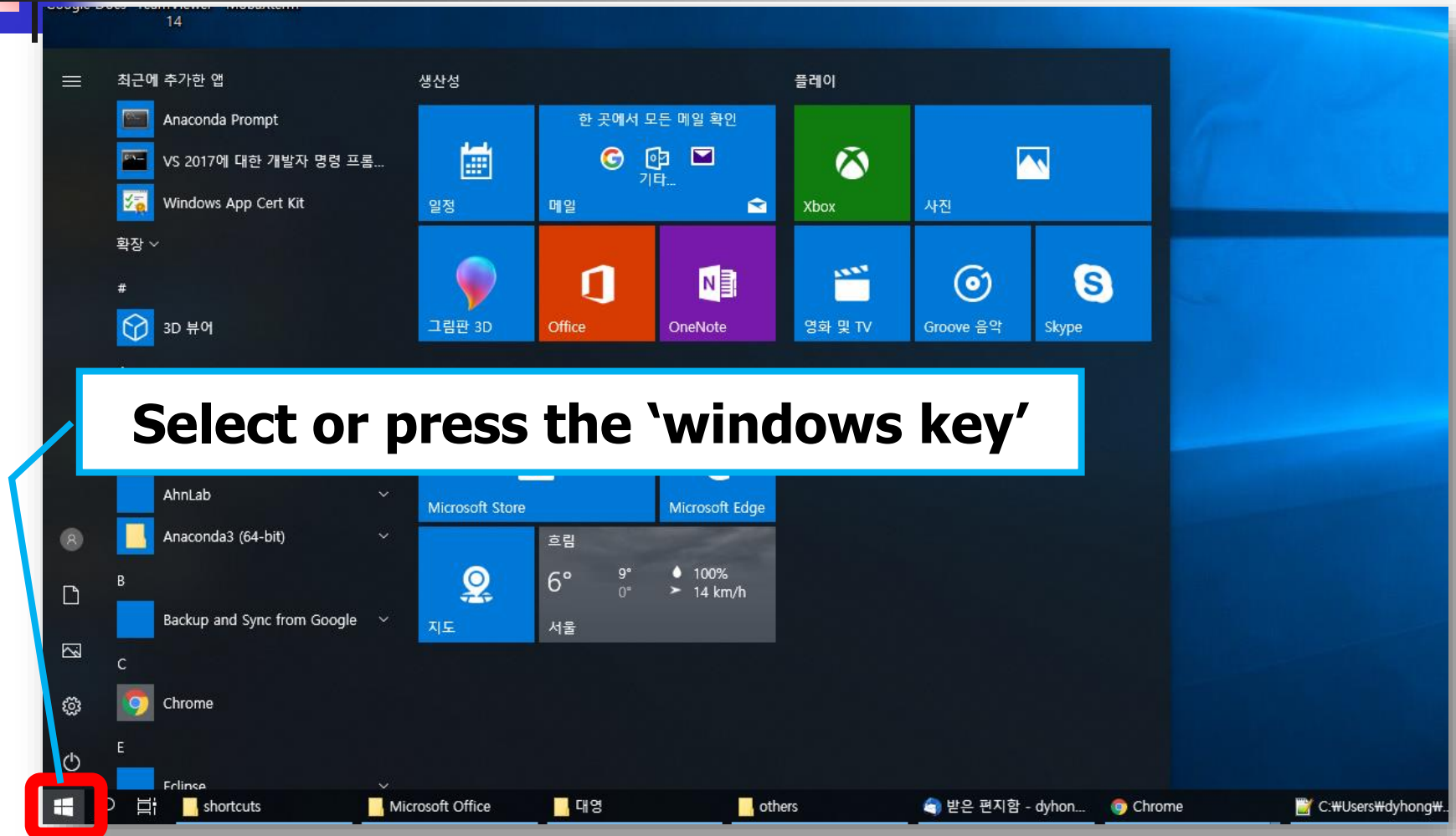


Manual Graphviz Installation with Conda

- Graphviz
 - conda
 - Windows 64-Bit
 - [conda](#)
 - [graphviz](#)
 - [python-graphviz](#)
 - OSX 64-Bit
 - [conda](#)
 - [graphviz](#)
 - [python-graphviz](#)
 - Linux 64-Bit
 - [conda](#)
 - [graphviz](#)
 - [python-graphviz](#)

- Download the following files for your OS from <http://kdd.snu.ac.kr/python/>
 - conda
 - graphviz
 - python-graphviz

Manual Graphviz Installation with Conda



1. Uncheck these options

모두 앱 문서 설정 사진 더보기

모두 앱 문서 설정 사진 더보기

사용자 의견 ...

가장 정확

Anaconda Prompt
데스크톱 앱



Anaconda Prompt
데스크톱 앱

앱

Anaconda Navigator

Uninstall-Anaconda3.exe

폴더

2. Select

anaconda_navigatorsite-packages의 -

anaconda_navigatorsite-packages의 -

anaconda_projectsite-packages의 -

anaconda_projectsite-packages의 -

문서

anaconda-navigator-script.pyw

anaconda

anaconda project 0.0.2 py37_ojson

1. Type "anaconda"

anaconda

Manual Graphviz Installation with Conda

- Type "conda install [downloaded conda file path]"

Downloaded conda file path

Anaconda Prompt

```
(base) C:\Users\kddlab0>conda install Downloads\conda-4.6.8-py37_0.tar.bz2
```



Manual Graphviz Installation with Conda

- The *conda* is installed

Anaconda Prompt

```
(base) C:\Users\kddl\lab0>conda install Downloads\conda-4.6.8-py37_0.tar.bz2
```

```
Downloading and Extracting Packages
```

```
#####
```

```
Preparing transaction: done
```

```
Verifying transaction: done
```

```
Executing transaction: done
```



Manual Graphviz Installation with Conda

- Repeat for 'graphviz' and 'python-graphviz'

```
(base) C:\Users\kddlab0>conda install Downloads\graphviz-2.38-hfd603c8_2.tar.bz2
```

```
Downloading and Extracting Packages
```

```
#####
```

```
Preparing transaction: done
```

```
Verifying transaction: done
```

```
Executing transaction: done
```

```
(base) C:\Users\kddlab0>conda install Downloads\python-graphviz-0.8.4-py37_1002.tar.bz2
```

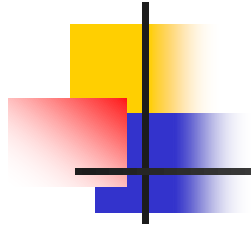
```
Downloading and Extracting Packages
```

```
#####
```

```
Preparing transaction: done
```

```
Verifying transaction: done
```

```
Executing transaction: done
```



GRAPHVIZ INSTALLATION WITH PIP



Graphviz Installation with pip

- An alternative installation when Conda is not available

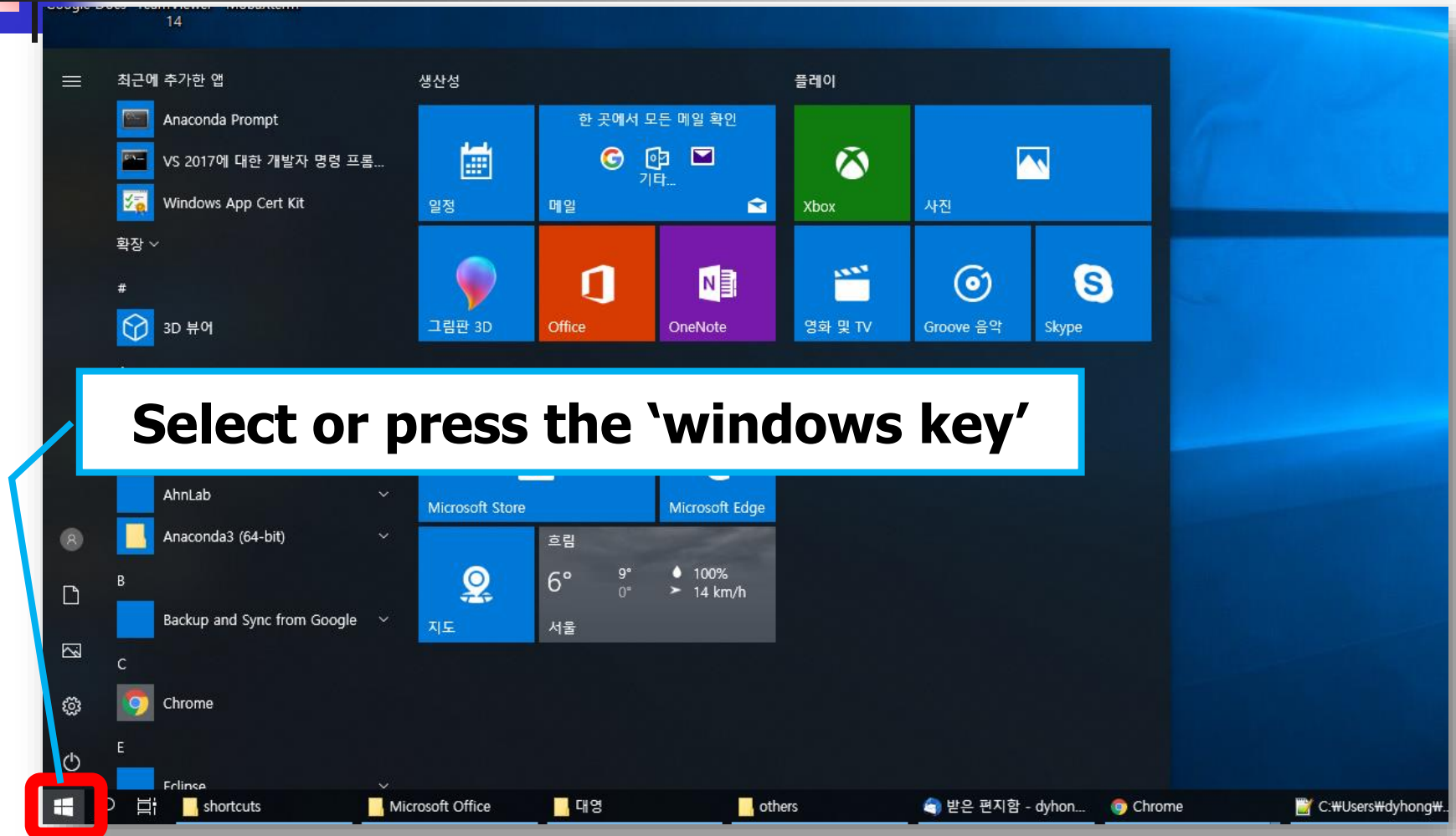


Graphviz Installation with pip

- Download from <http://kdd.snu.ac.kr/python/>
 - [whl file](#)
 - [Windows installer](#)

- pip
 - [whl file](#)
 - Installer
 - [Windows](#)

Graphviz Installation with pip





1. Uncheck these options

모두 앱 문서 설정 사진 더보기

모두 앱 문서 설정 사진 더보기

사용자 의견 ...

가장 정확



Anaconda Prompt

데스크톱 앱



Anaconda Prompt

데스크톱 앱

앱



Anaconda Navigator



Uninstall-Anaconda3.exe



폴더

2. Select



anaconda_navigatorsite-packages의 -



관리자 권한으로 실행



anaconda_navigatorsite-packages의 -



파일 위치 열기



anaconda_projectsite-packages의 -



시작 화면에 고정



anaconda_projectsite-packages의 -



작업 표시줄에 고정



제거

문서



anaconda-navigator-script.pyw



anaconda

1. Type "anaconda"



anaconda project 0.0.2 py37_ojson



anaconda



파일 탐색기

받은 편지...

Chrome

clustering ...

dyhong - ...

*C:\#Users...

kdd - dyh...

Excel 20



Graphviz Installation with pip

Type this and press Enter

**The path of the
downloaded whl file**

Anaconda Prompt

```
(base) C:\Users\Wdyhong>pip install Downloads\graphviz-0.10.1-py2.py3-none-any.whl
```



Graphviz Installation with pip

- Successfully installed the whl file

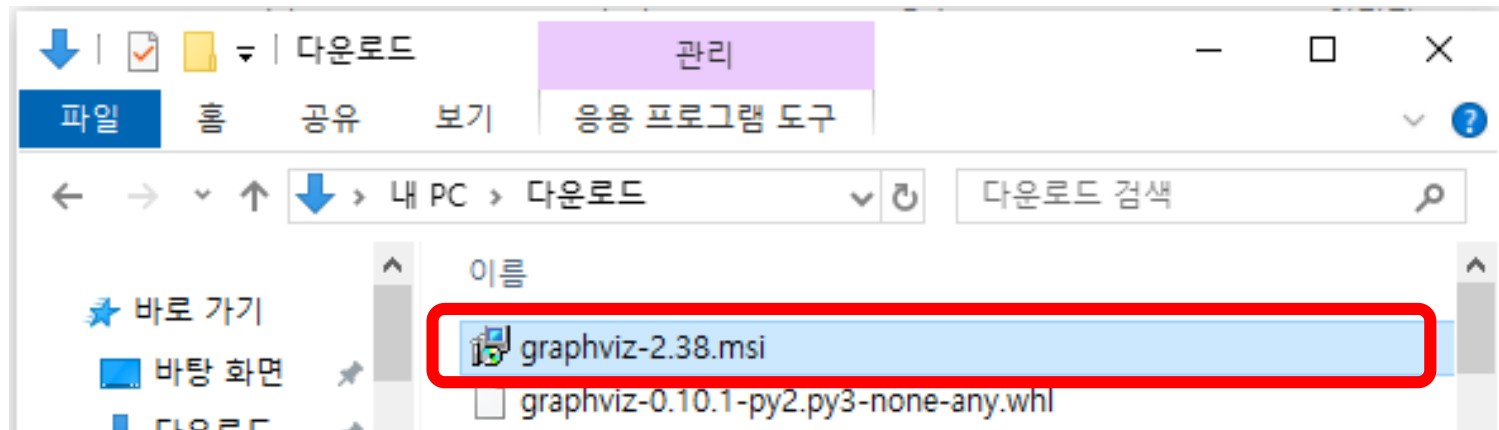
Anaconda Prompt

```
(base) C:\Users\Wdyhong>pip install Downloads\graphviz-0.10.1-py2.py3-none-any.whl
Processing c:\Users\Wdyhong\downloads\graphviz-0.10.1-py2.py3-none-any.whl
Installing collected packages: graphviz
Successfully installed graphviz-0.10.1

(base) C:\Users\Wdyhong>_
```

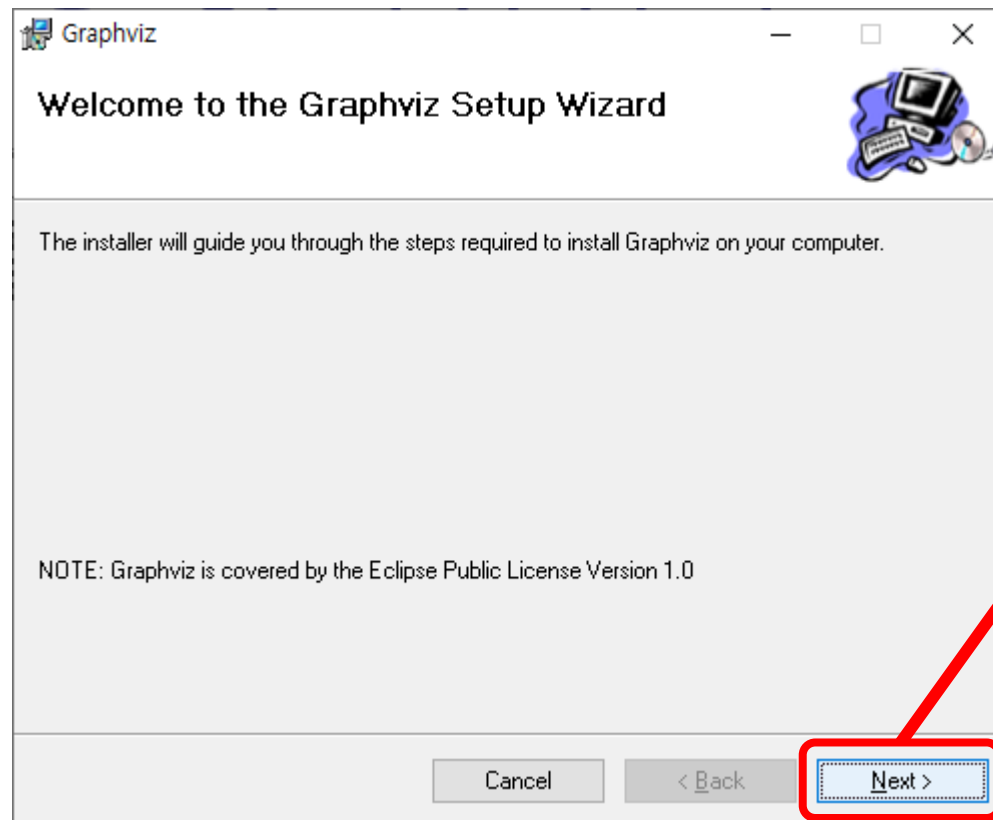
Graphviz Installation with pip

- Run the downloaded Windows installer of Graphviz



Graphviz Installation with pip

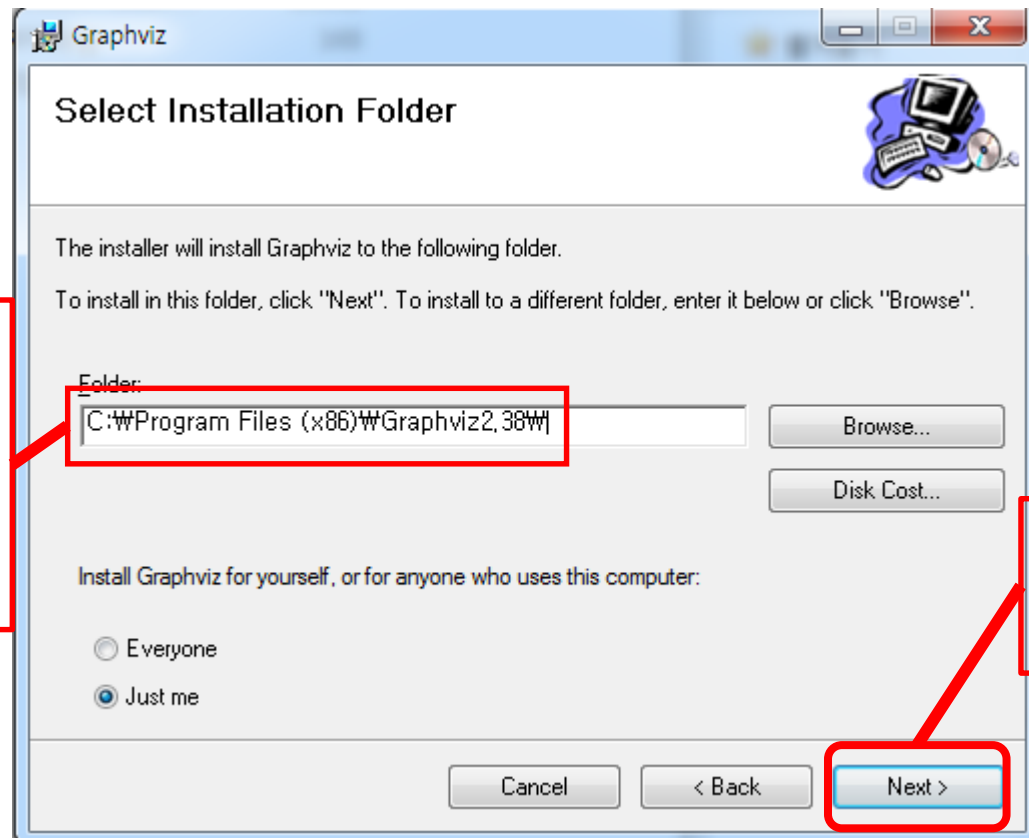
- Press Next



**2. Press
Next**

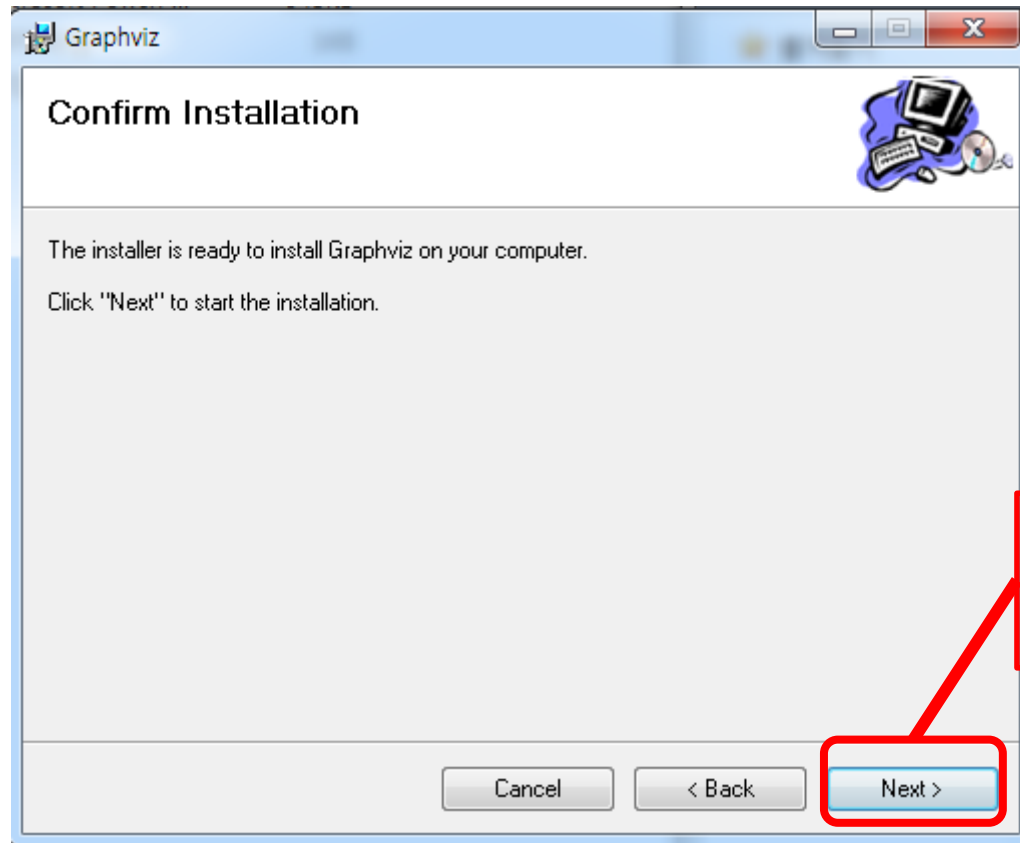
Graphviz Installation with pip

1. Copy and save the path for further usage



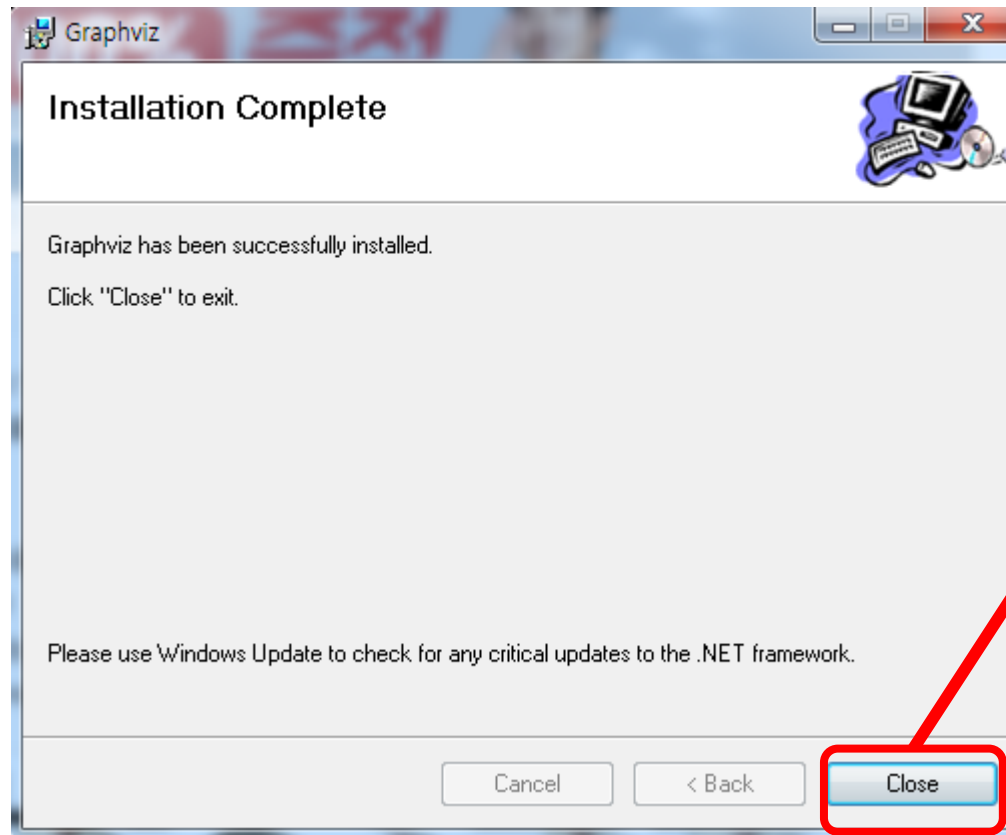
2. Press Next

Graphviz Installation with pip



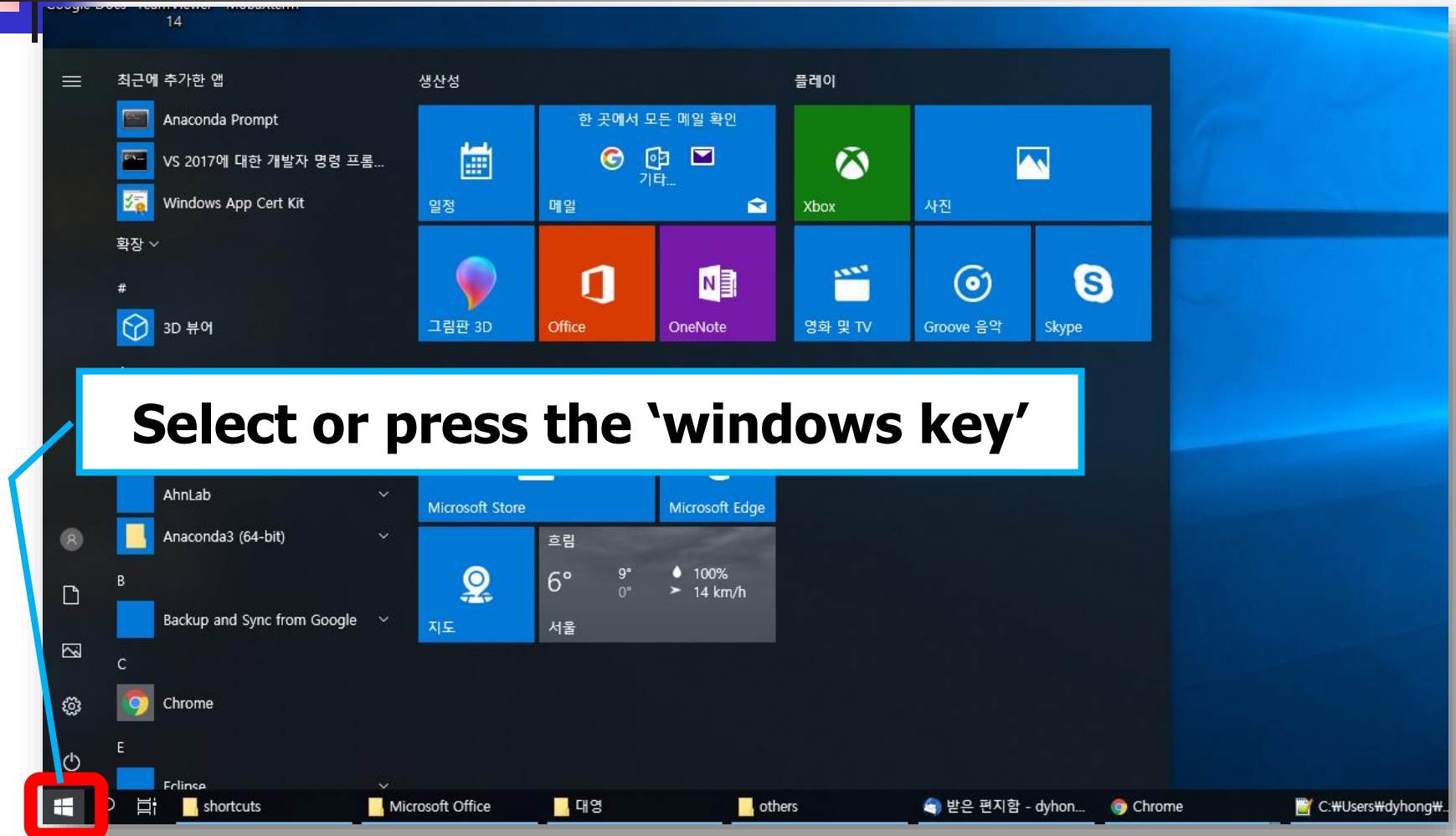
**2. Press
Next**

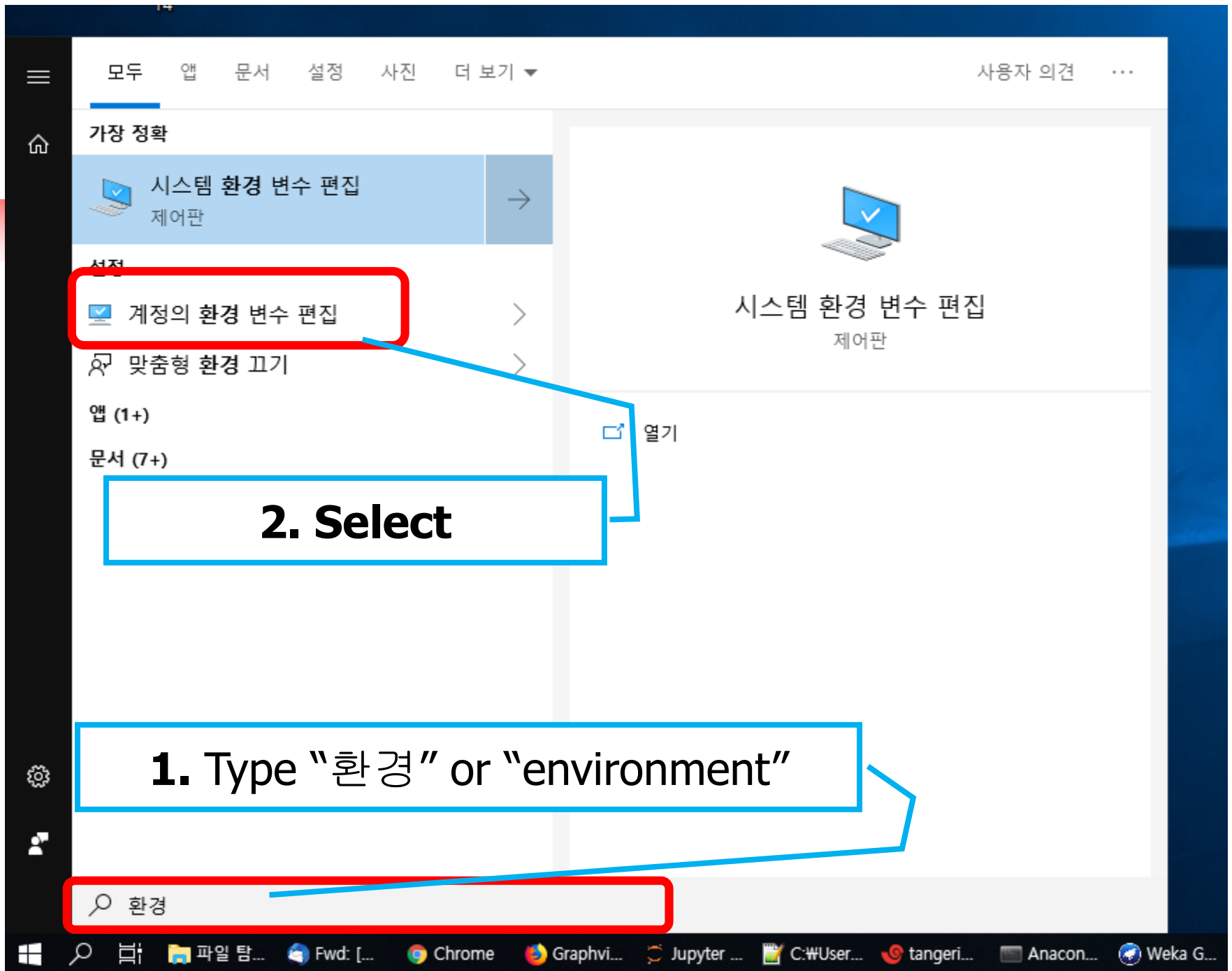
Graphviz Installation with pip



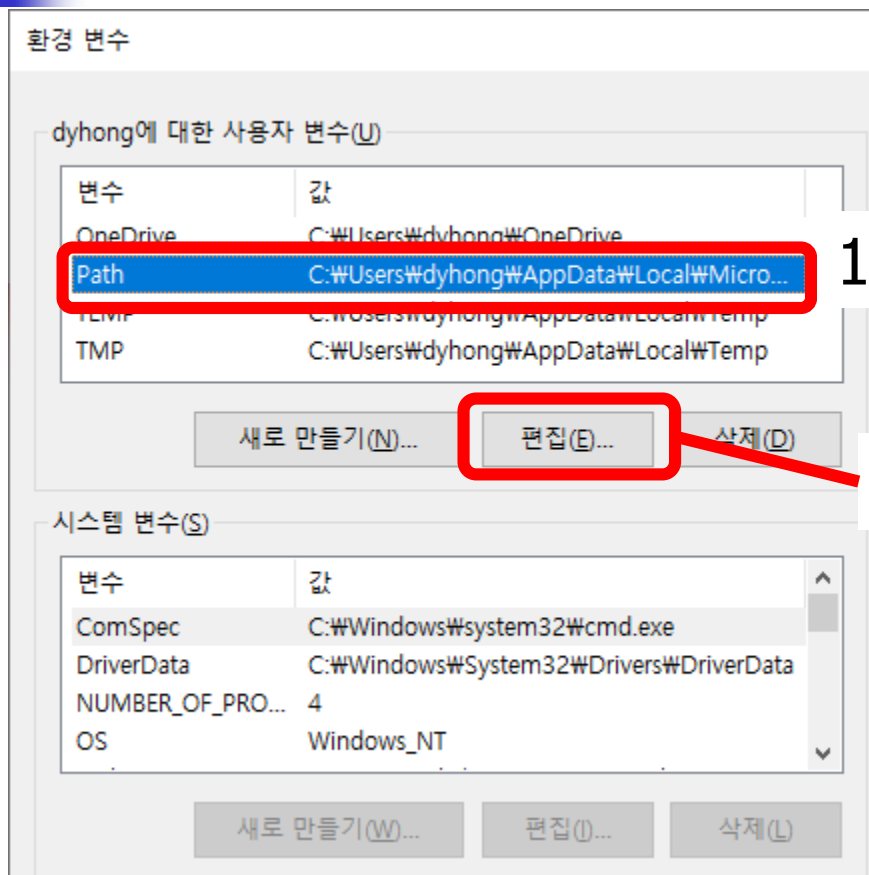
**2. Press
Close**

Graphviz Installation with pip





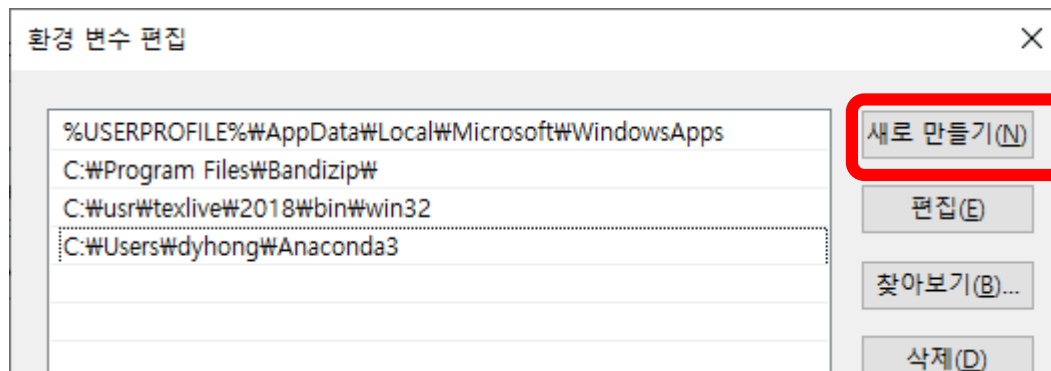
Graphviz Installation with pip



1. Select 'Path'

2. Select 'Edit'

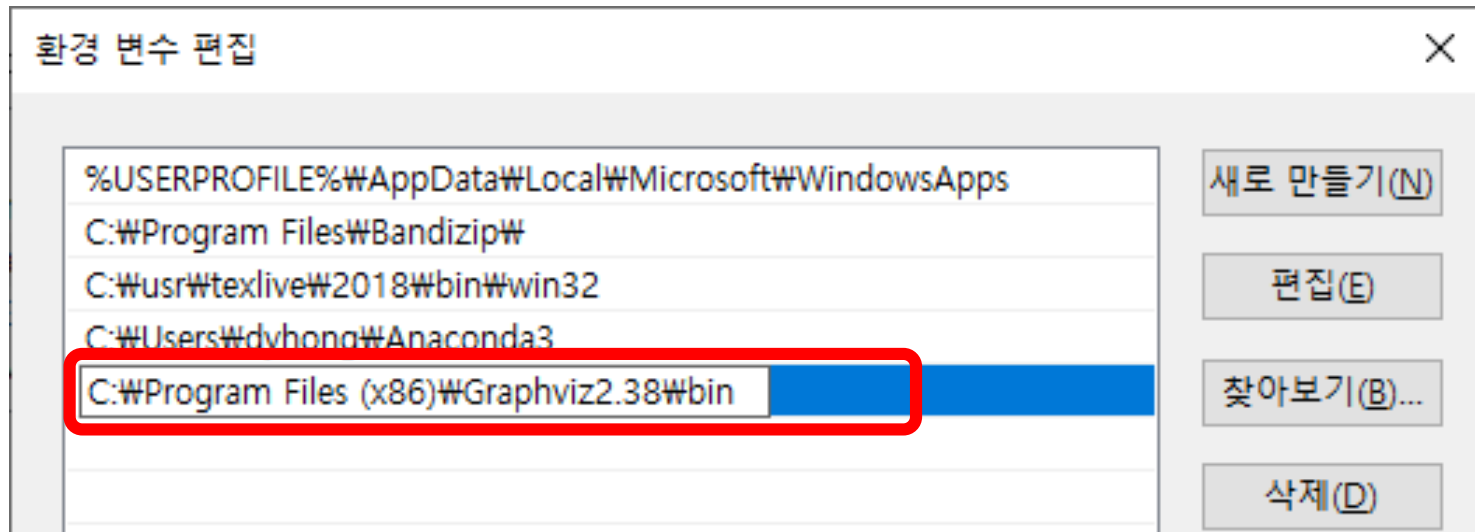
Graphviz Installation with pip



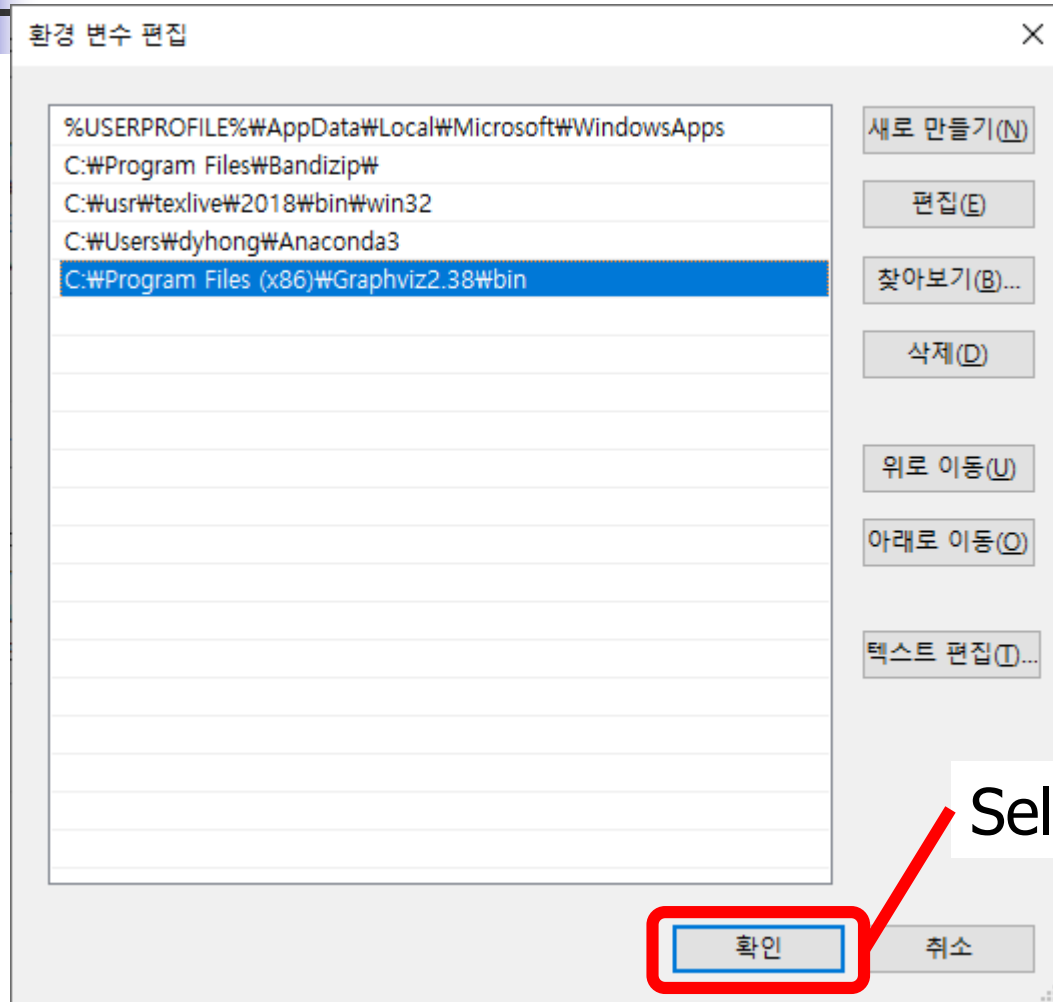
Select

Graphviz Installation with pip

- Write down
"[the installed Graphviz path]\bin"
 - In this example,
C:\Program Files (x86)\Graphviz2.38\bin

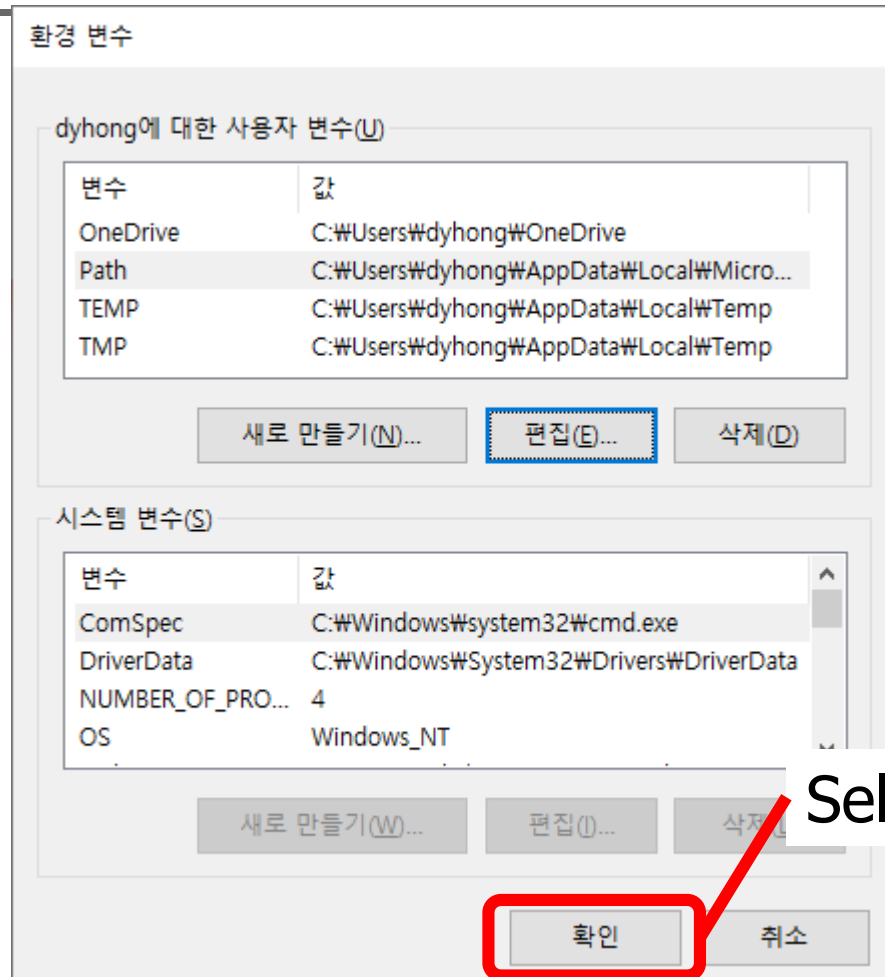


Graphviz Installation with pip



Select

Graphviz Installation with pip





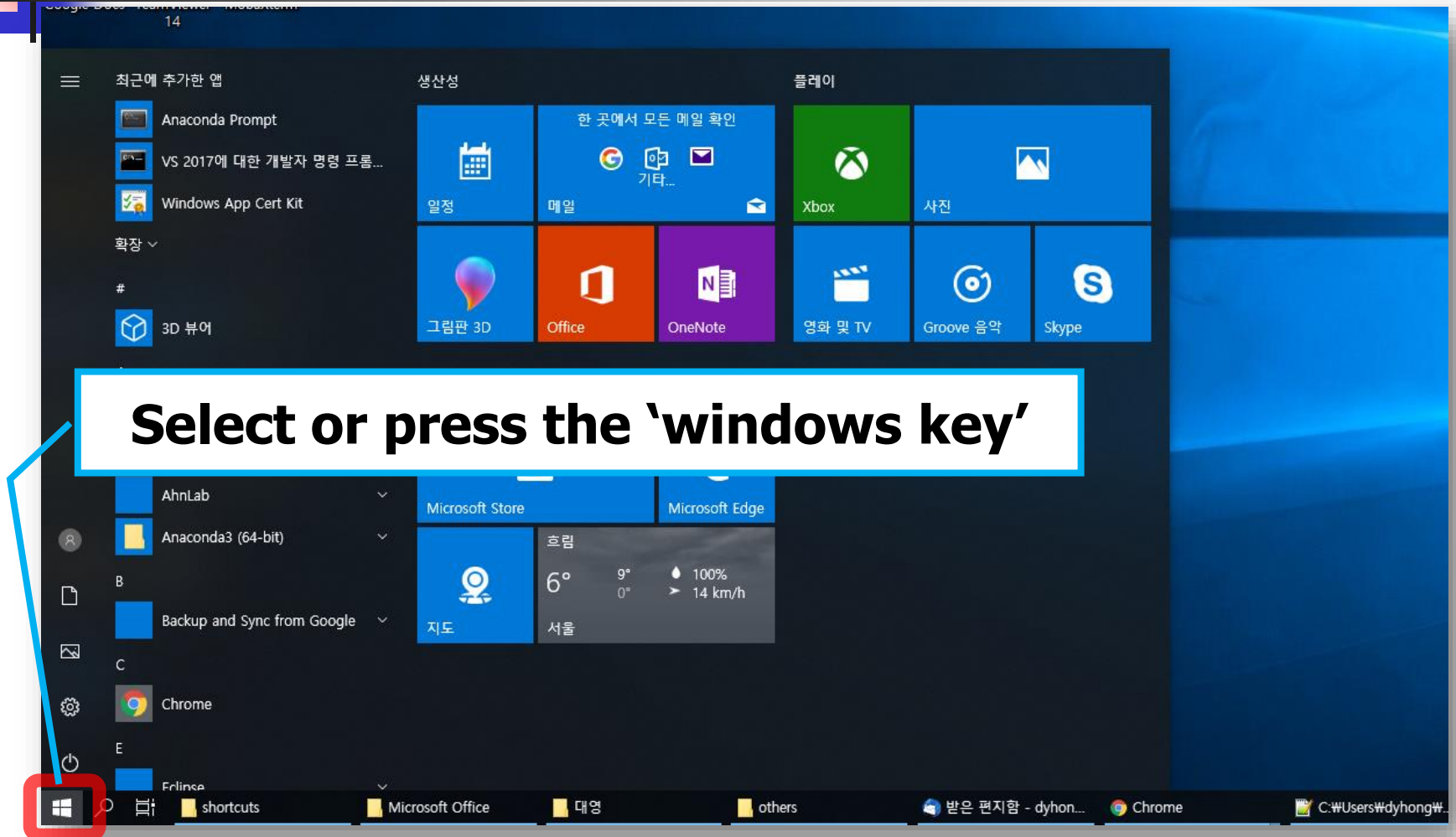
JUPYTER NOTEBOOK



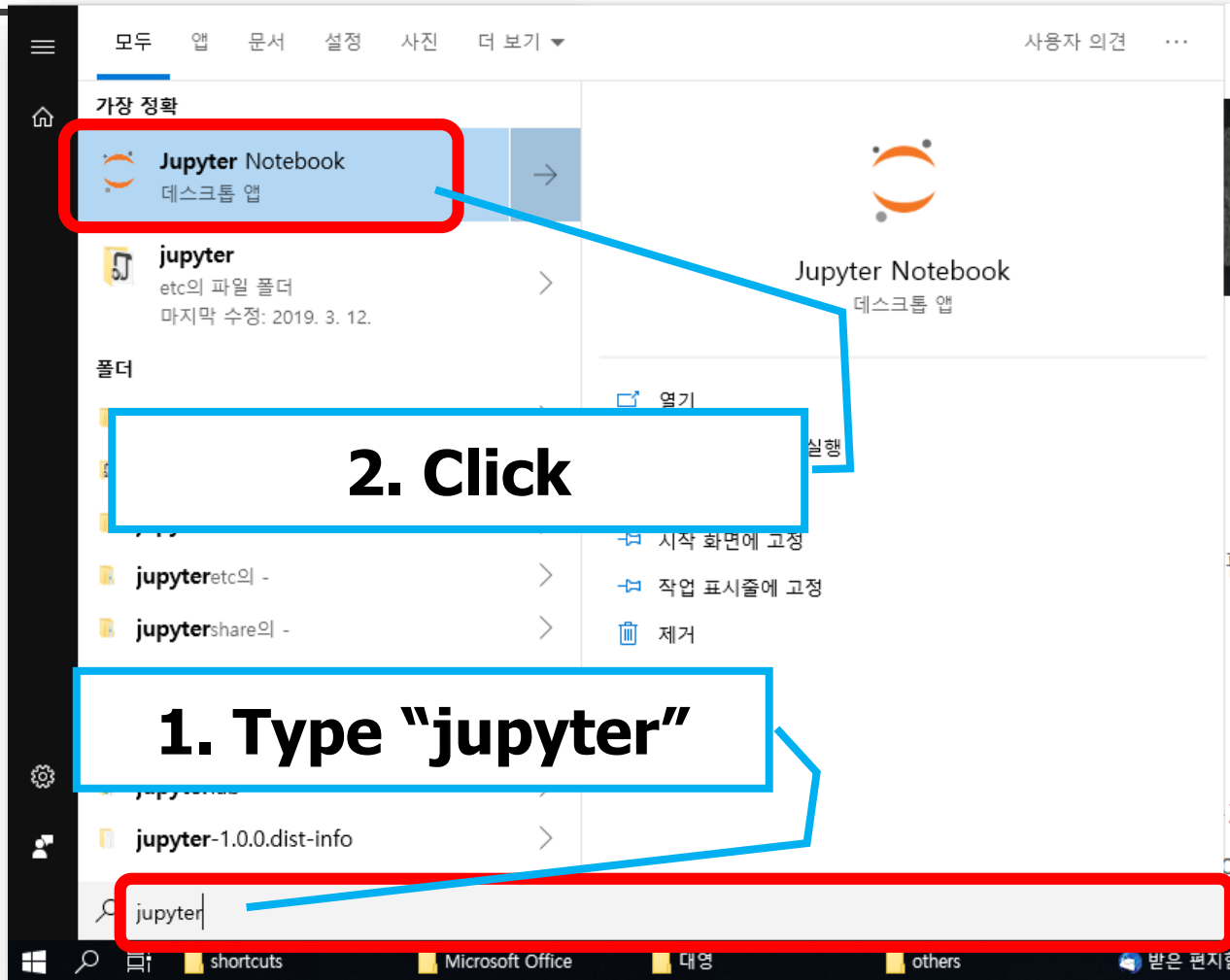
Jupyter Notebook

- An open-source web application that allows you to edit and run python code
- Already installed since it is contained in Anaconda

Running Jupyter Notebook



Running Jupyter Notebook



Running Jupyter Notebook

```
Jupyter Notebook  
[I 16:23:29.531 NotebookApp] Writing notebook server cookie secret to C:\Users\dyhong\AppData\Roaming\notebook_cookie_secret  
[I 16:23:30.405 NotebookApp] JupyterLab extension loaded from C:\Users\dyhong\Anaconda3\lib\site-packages  
[I 16:23:30.405 NotebookApp] JupyterLab application directory is C:\Users\dyhong\Anaconda3\share\jupyterlab  
[I 16:23:30.408 NotebookApp] Serving notebooks from local directory: C:\Users\dyhong  
[I 16:23:30.408 NotebookApp] The Jupyter Notebook is running at:
```

Home

localhost:8888/tree

Quit

After a while, the Jupyter Notebook will be run as shown above.

The base directory of the environment is shown here.

Files

Select items to perform actions on them.

Upload

0

▼

/

Name ▼

Last Modified

□

3D Objects

한 달 전

Creating a Notebook Document

The image shows the JupyterLab web interface. At the top left is the Jupyter logo. To its right are 'Quit' and 'Logout' buttons. Below these are tabs for 'Files', 'Running', and 'Clusters'. A message says 'Select items to perform actions on them.' Below this is a file browser showing a directory structure with folders like '3D Objects', 'Anaconda3', 'Contacts', and 'Desktop'. On the right side, there are 'Upload' and 'New' buttons. A red box highlights the 'New' button, with a blue arrow pointing to it from a box containing the number '1.'. A dropdown menu is open from the 'New' button, showing options: 'Notebook:', 'Python 3' (highlighted with a red box and a blue arrow from a box containing the number '2.'), and 'Other:' followed by 'Text File', 'Folder', and 'Terminal'.

jupyter

Quit Logout

Files Running Clusters

Select items to perform actions on them.

Upload New

Notebook:

Python 3

Other:

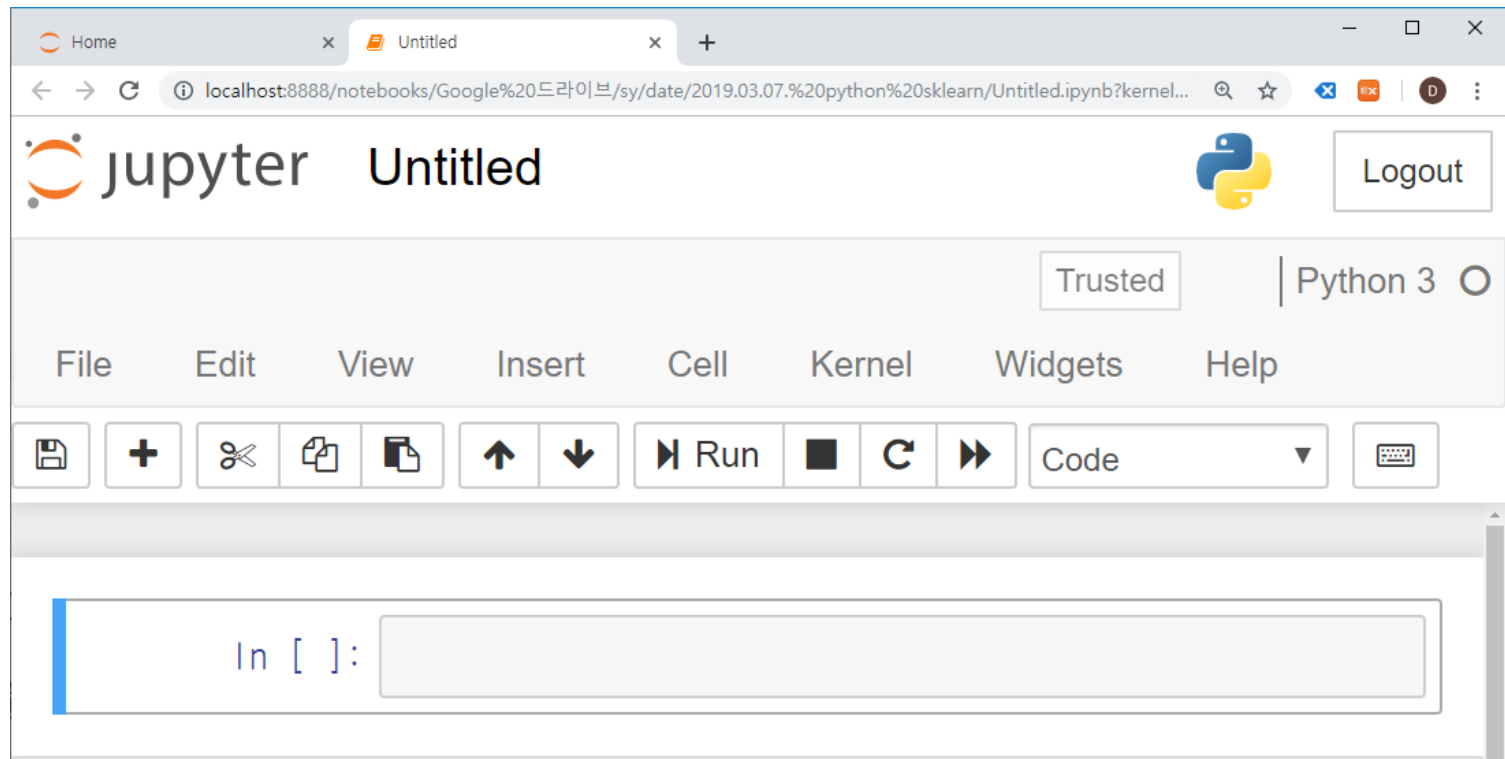
Text File

Folder

Terminal

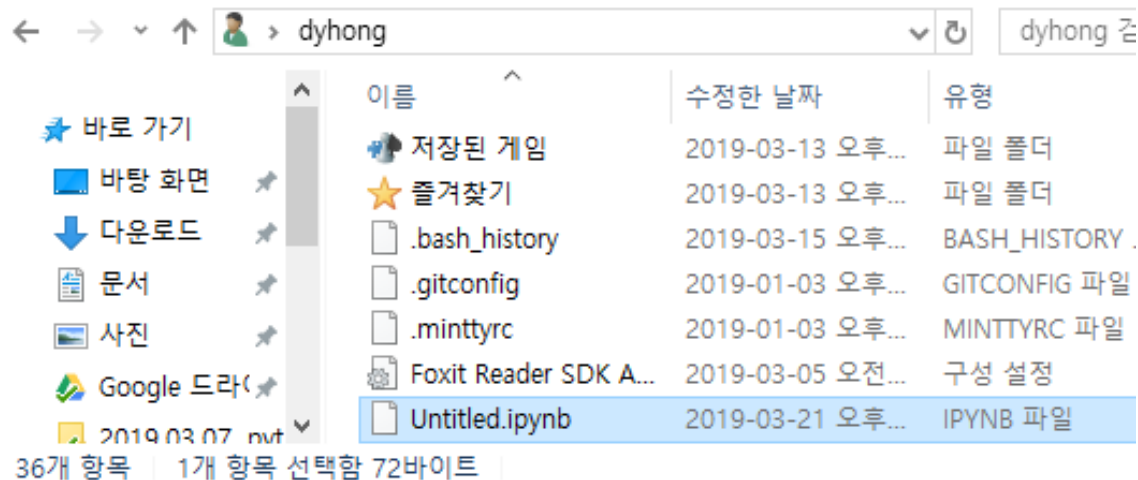
Creating a Notebook Document

- A Notebook document is created



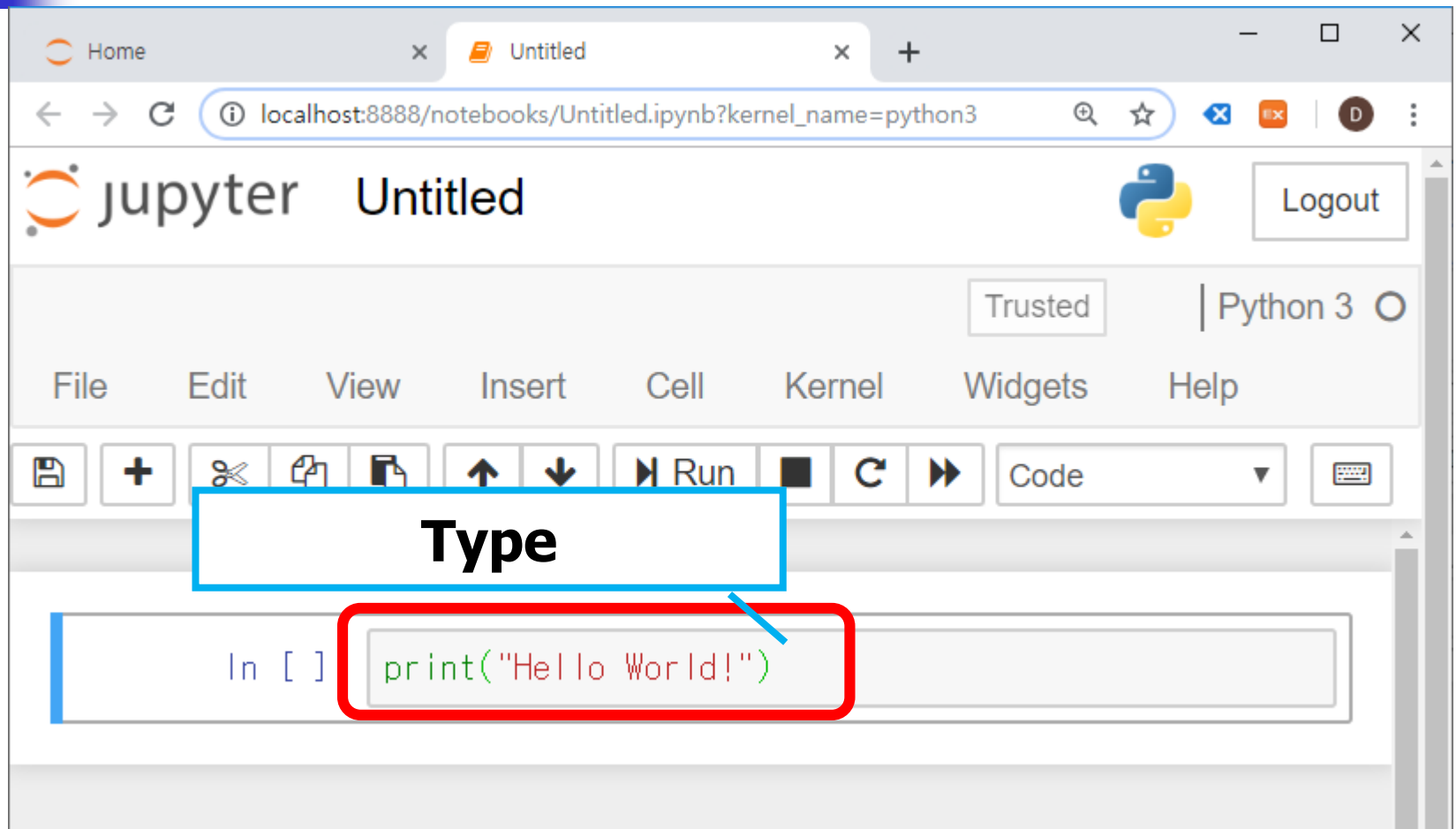
Creating a Notebook Document

- You can see the file (Untitled.ipynb) is created in the base directory
 - In this example, the base directory path is C:\Users\dyhong





Hello World!



The screenshot shows a web browser window with a Jupyter Notebook interface. The browser's address bar displays the URL `localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3`. The Jupyter interface includes a header with the Jupyter logo, the text "jupyter Untitled", a Python logo, and a "Logout" button. Below the header is a toolbar with icons for file operations (save, new, copy, paste, undo, redo), execution (run, interrupt), and cell management (code, raw, markdown). The main area contains a code cell with the prompt `In []` followed by the code `print("Hello World!")`. A blue box labeled "Type" points to the code input area, and a red box highlights the code itself.

Home x Untitled x +

localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3

jupyter Untitled Python Logout

Trusted Python 3

File Edit View Insert Cell Kernel Widgets Help

Run Code

Type

In [] `print("Hello World!")`

Hello World!

The screenshot shows the Jupyter Notebook interface in a web browser. The browser's address bar displays `localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3`. The Jupyter logo and the word "Untitled" are visible in the top left. A "Logout" button is in the top right. Below the header is a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", "Widgets", and "Help". The "Cell" menu is open, showing options: "Run Cells", "Run Cells and Select Below", "Run Cells and Insert Below", "Run All", "Run All Above", and "Run All Below". The "Run Cells" option is highlighted with a red box. A blue box with the number "1." points to the "Cell" menu, and another blue box with the text "2. Run the code" points to the "Run Cells" option. In the background, a code cell is visible with the text `In []: print("Hello`.

1.

2. Run the code

File Edit View Insert Cell Kernel Widgets Help

Run Cells

Run Cells and Select Below

Run Cells and Insert Below

Run All

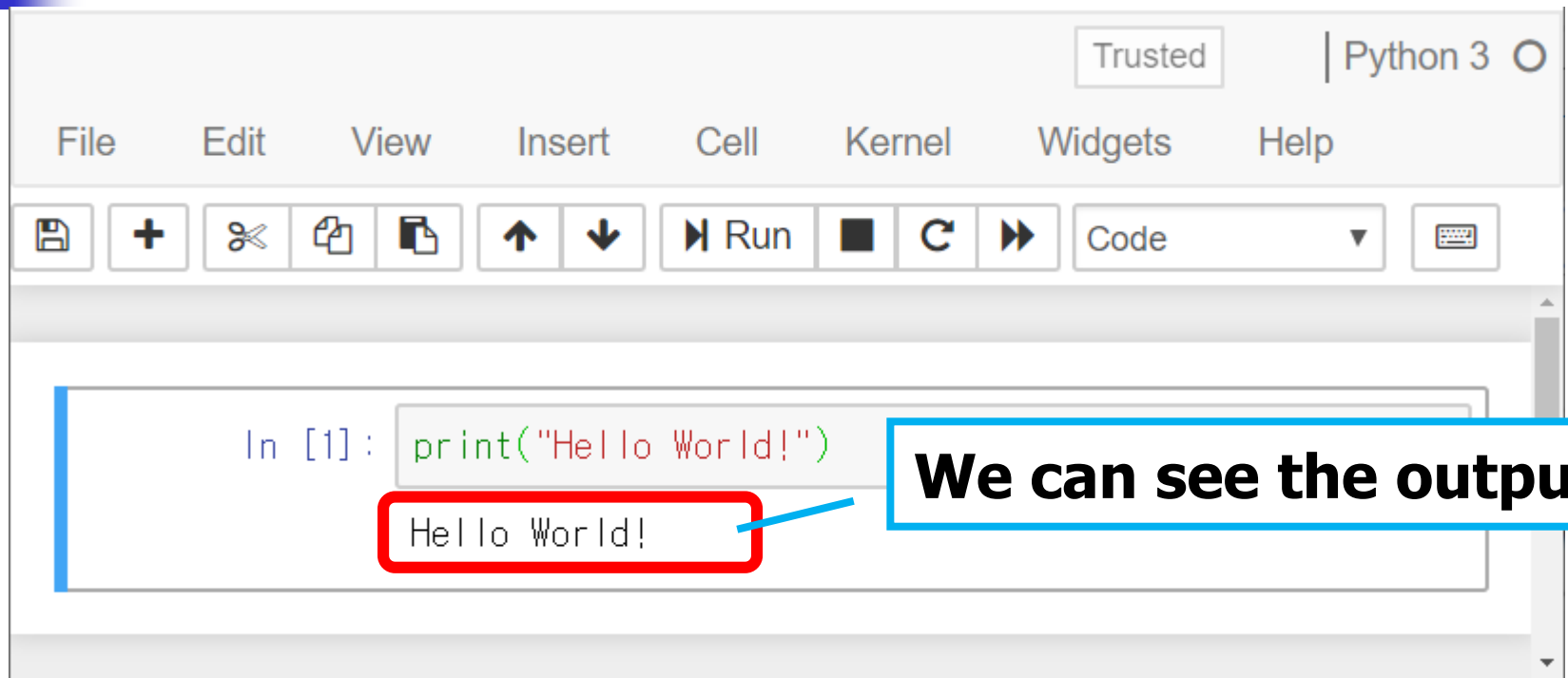
Run All Above

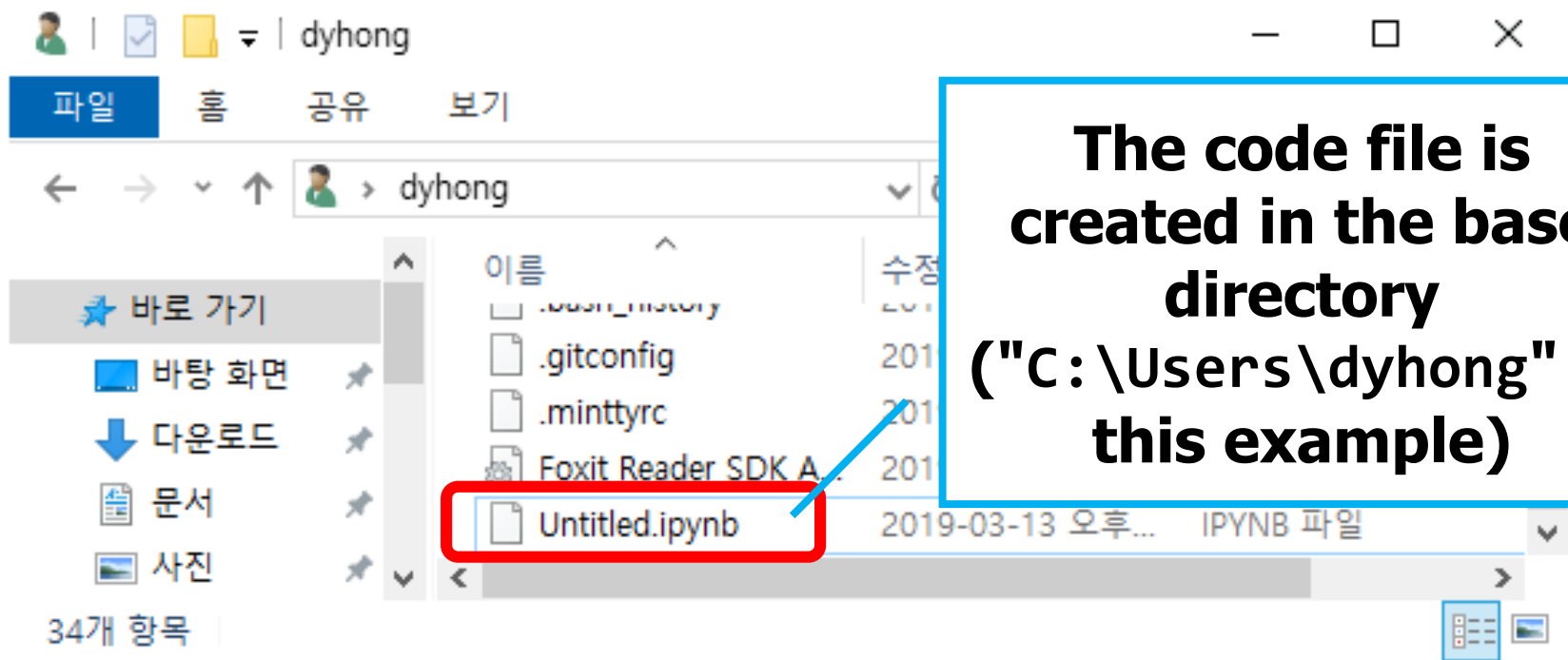
Run All Below

In []: `print("Hello`



Hello World!

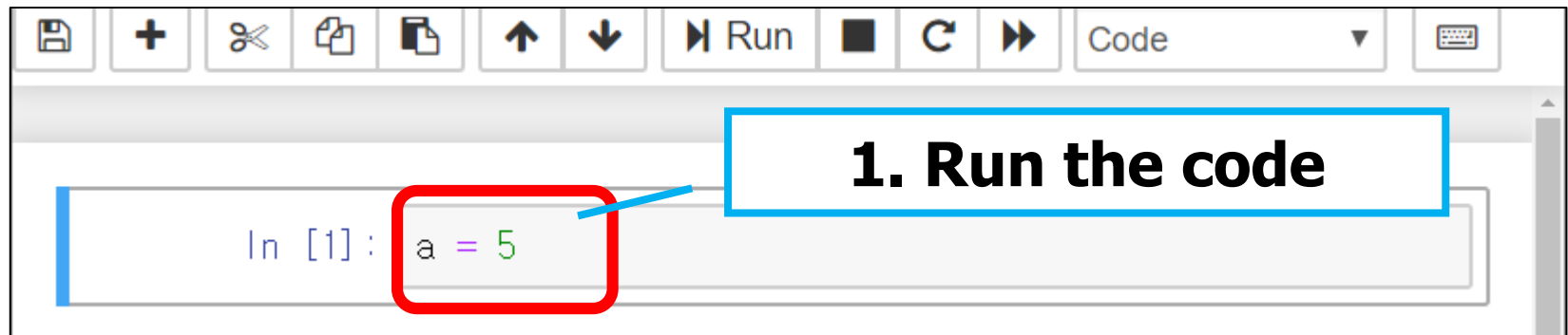




**The code file is
created in the base
directory
("C:\Users\dyhong" in
this example)**

Kernel

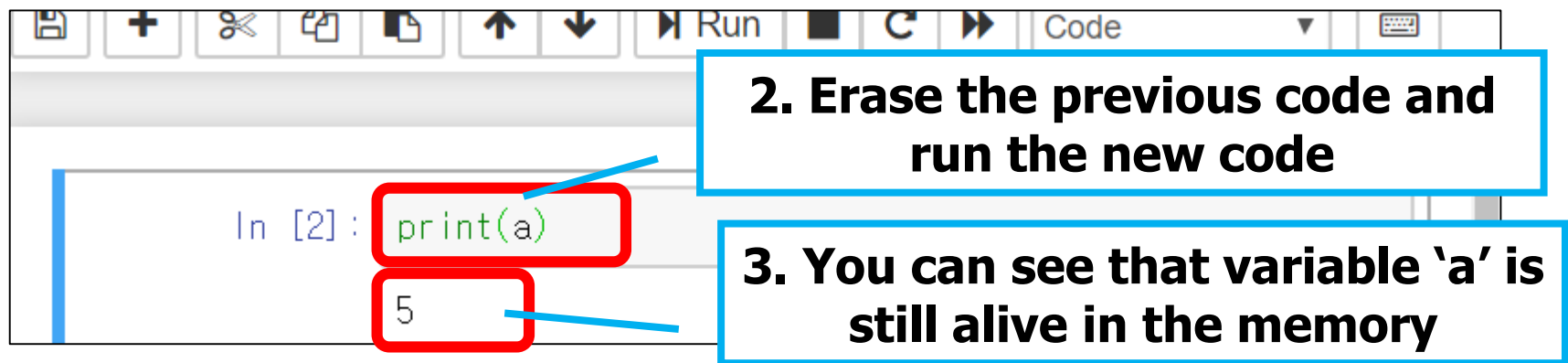
- Variables always remain in the memory unless you shut down the kernel in the Jupyter Notebook



The image shows a Jupyter Notebook interface with a toolbar at the top containing icons for saving, adding, deleting, copying, pasting, and running code. Below the toolbar is a code cell labeled 'In [1]:' containing the code 'a = 5'. The code 'a = 5' is highlighted with a red rectangular box. A blue arrow points from this box to a blue-bordered text box on the right that contains the text '1. Run the code'.

```
In [1]: a = 5
```

1. Run the code



The image shows the same Jupyter Notebook interface as above, but now with a second code cell labeled 'In [2]:' containing the code 'print(a)'. The code 'print(a)' is highlighted with a red rectangular box. A blue arrow points from this box to a blue-bordered text box on the right that contains the text '2. Erase the previous code and run the new code'. Below the code cell, the output '5' is displayed and highlighted with a red rectangular box. A blue arrow points from this box to a blue-bordered text box on the right that contains the text '3. You can see that variable 'a' is still alive in the memory'.

```
In [2]: print(a)
```

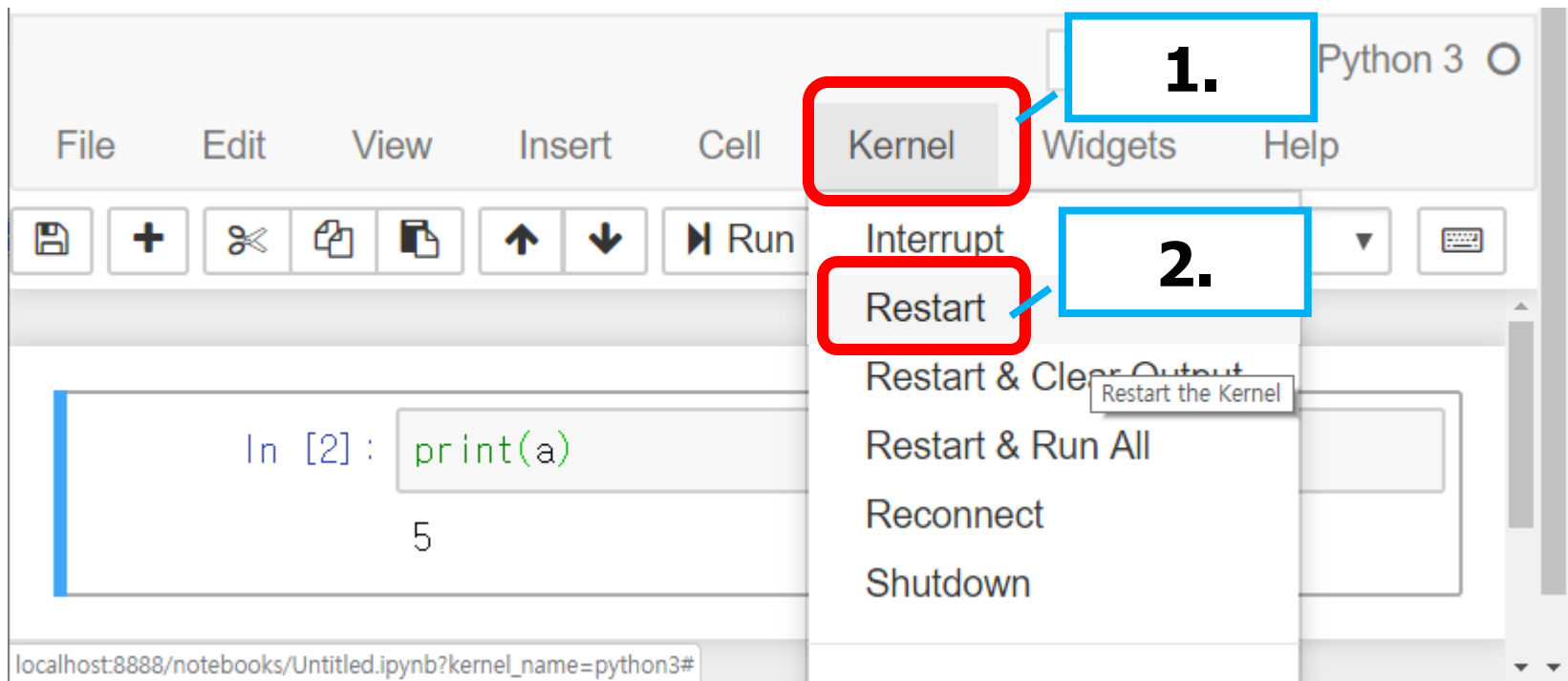
5

2. Erase the previous code and run the new code

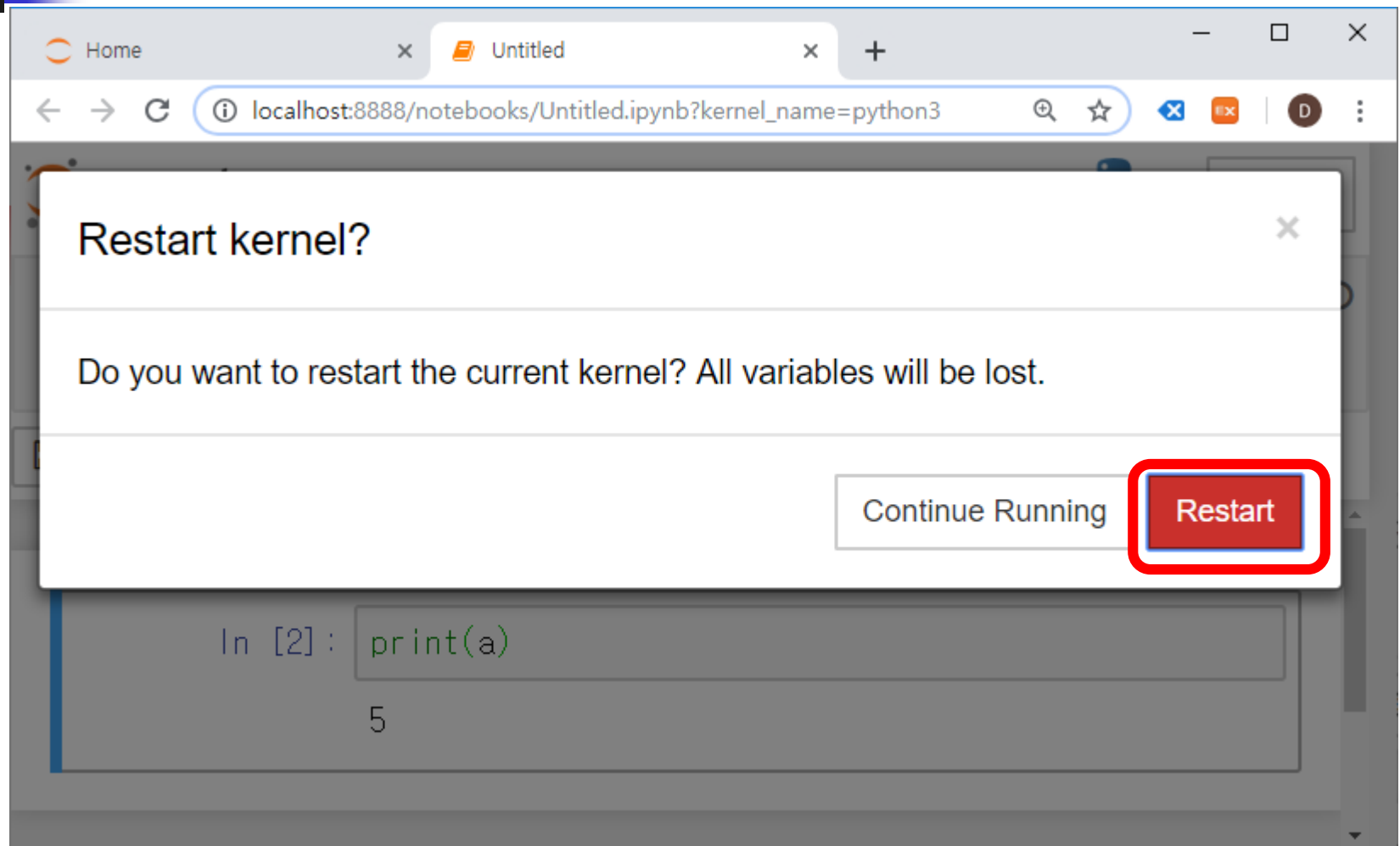
3. You can see that variable 'a' is still alive in the memory

Kernel

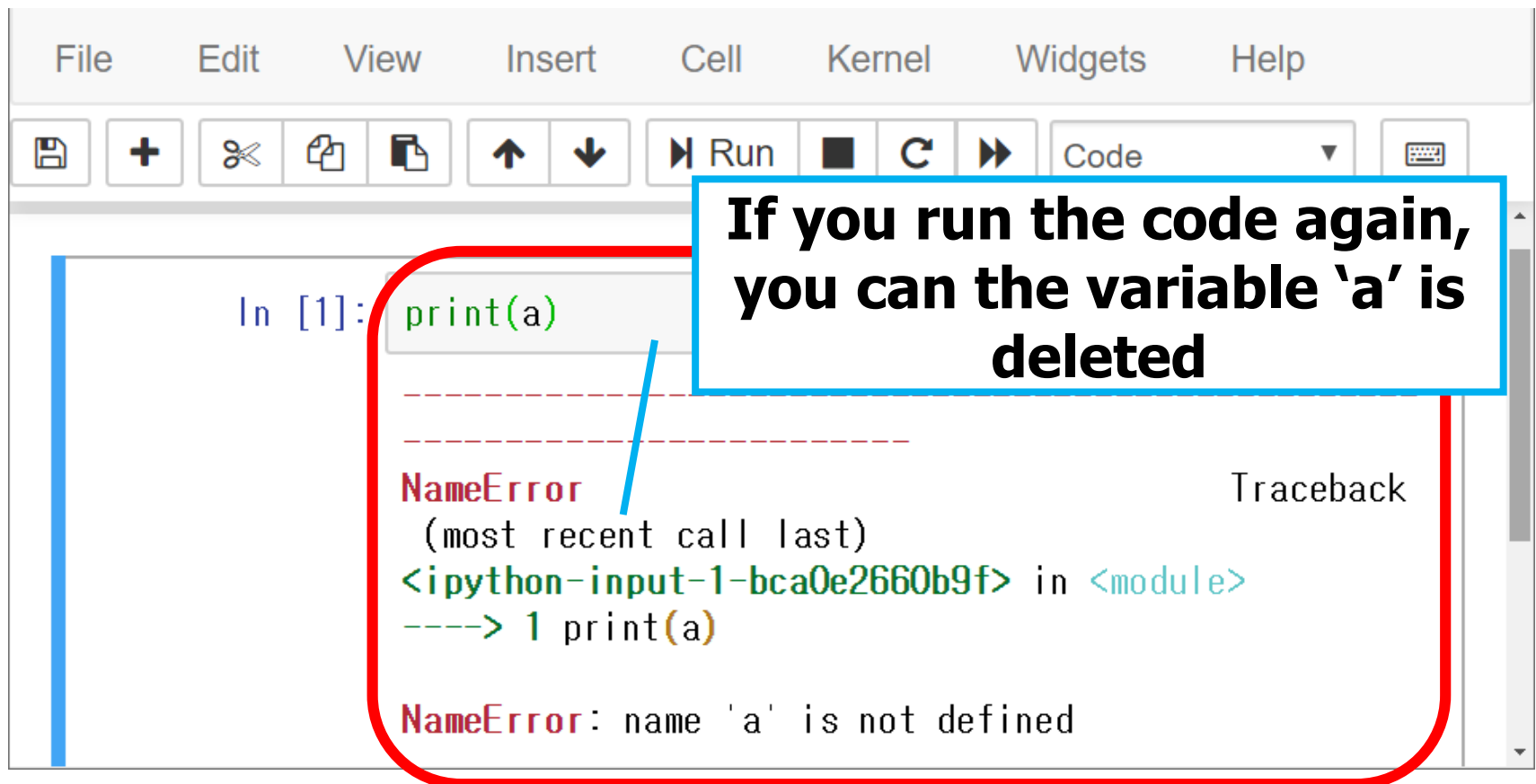
- If you want to start over a computation from scratch (e.g. variables are deleted, open files are closed, etc...), restart the kernel



Kernel



Kernel



The screenshot shows a Jupyter Notebook interface with a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for saving, adding cells, cutting, copying, pasting, and running code. The code cell contains the following text:

```
In [1]: print(a)
```

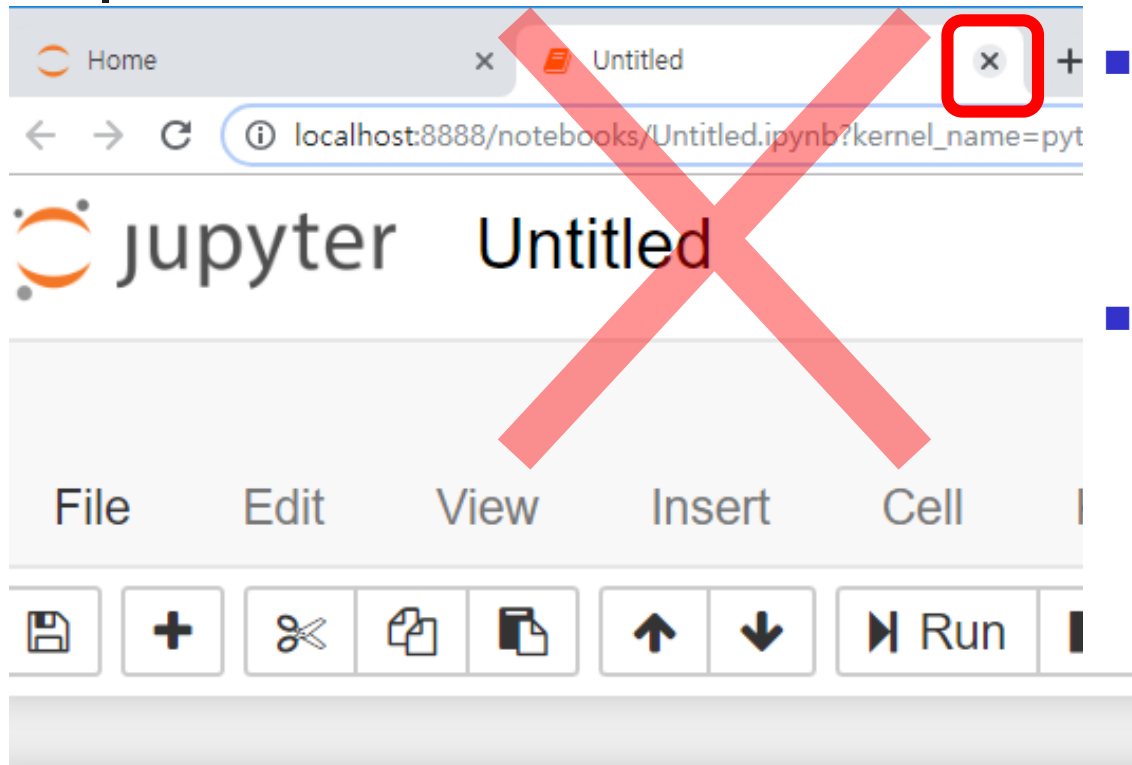
Below the code cell, a red-bordered box highlights the error message:

```
-----  
NameError                                Traceback  
(most recent call last)  
<ipython-input-1-bca0e2660b9f> in <module>  
----> 1 print(a)  
  
NameError: name 'a' is not defined
```

A blue line points from the text box to the `print(a)` line in the code cell.

If you run the code again, you can the variable 'a' is deleted

Close a Notebook

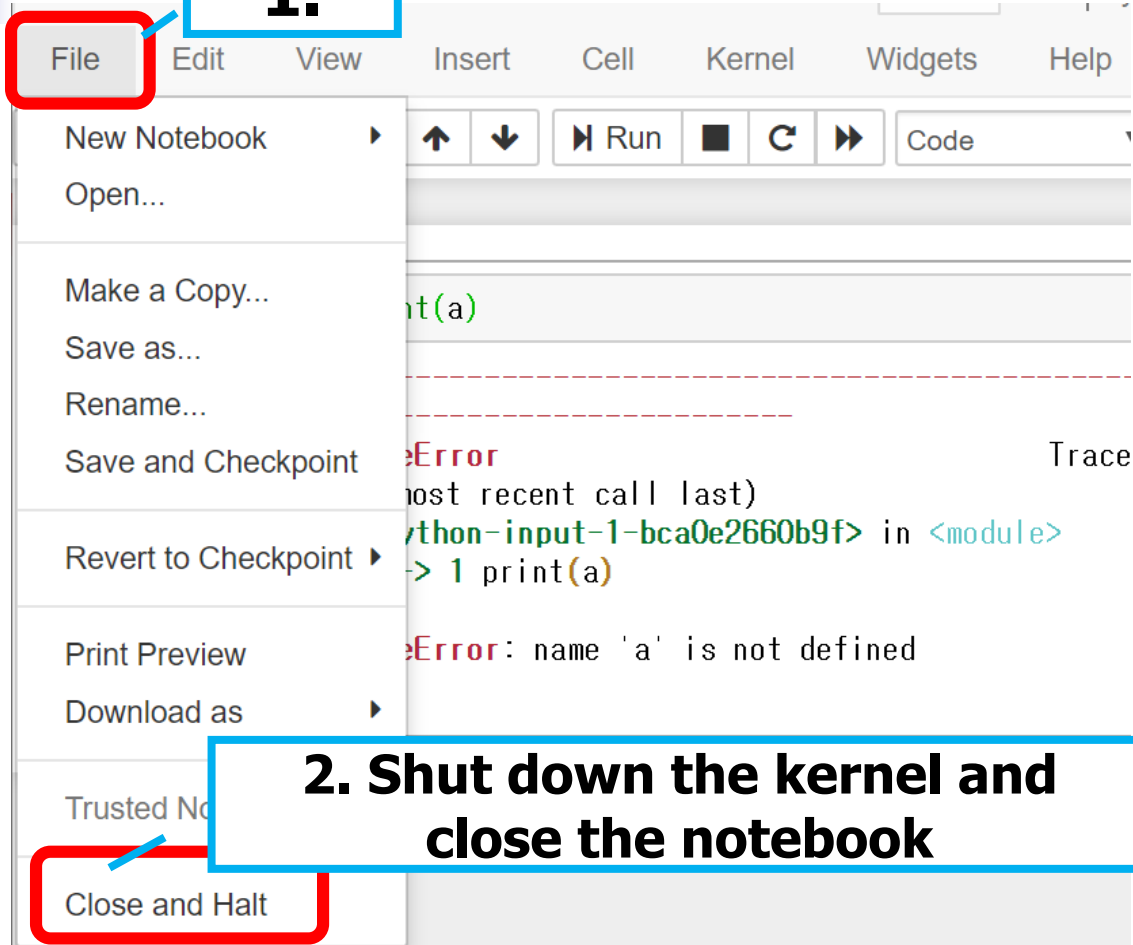


- Closing the notebook browser tab does not shut down the kernel
- Instead, the kernel keeps running until we explicitly shut down

```
In [1]: print(a)
```

Shutting Down

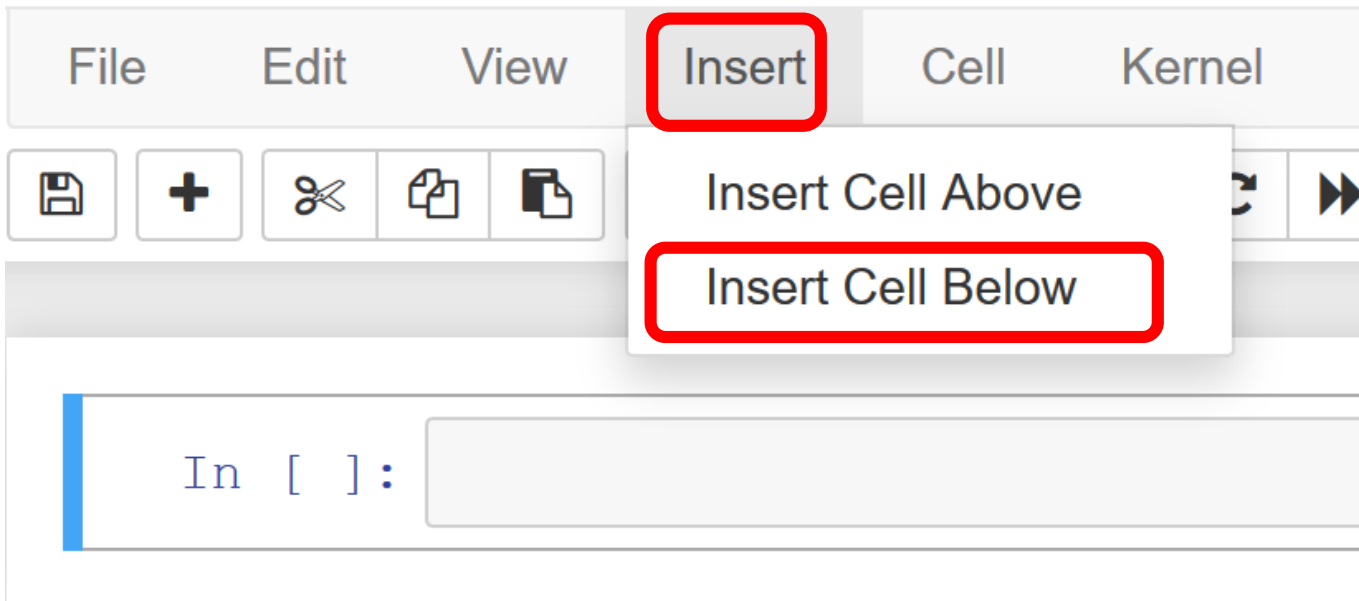
1.



2. Shut down the kernel and close the notebook

Cells

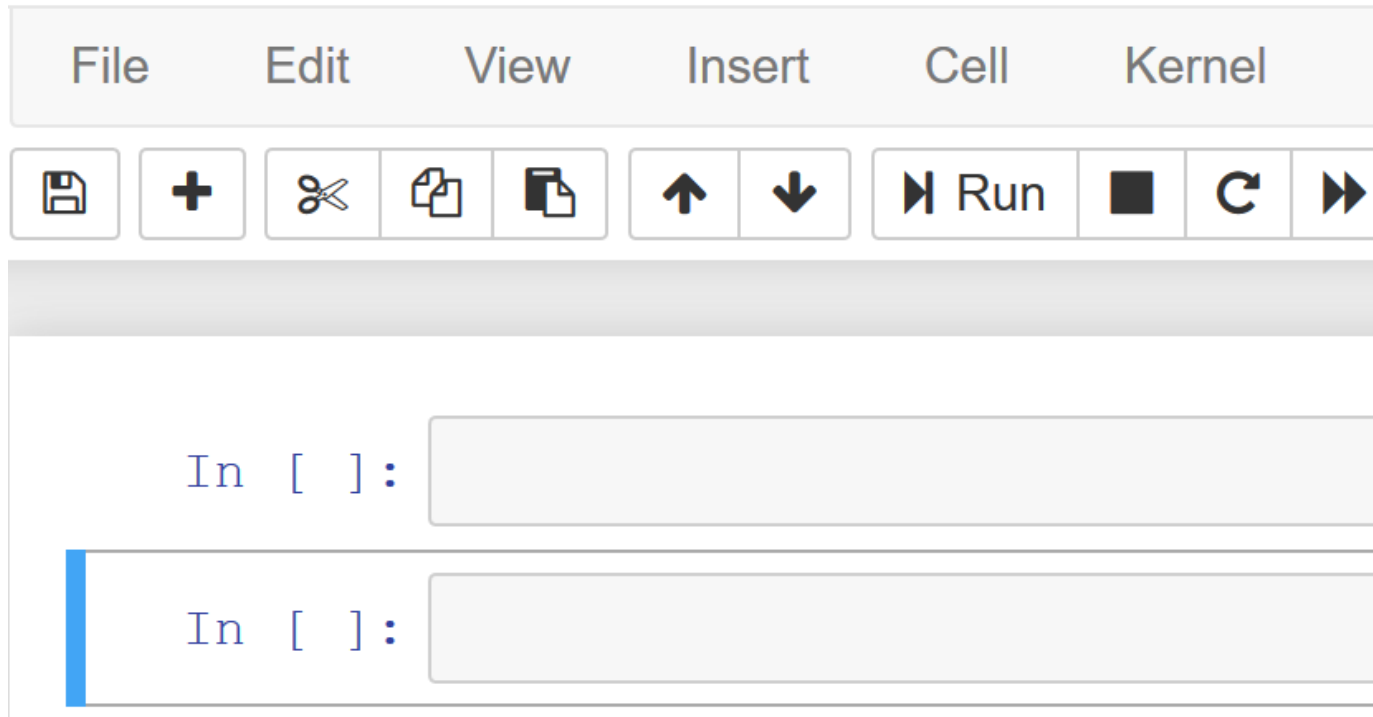
- By dividing the code into cells, you can run the code partially





Cells

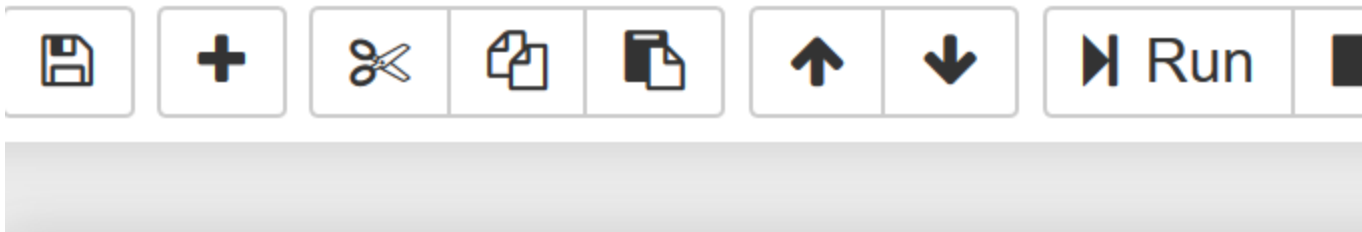
- A new cell is created





Cells

- Write down the code as follows

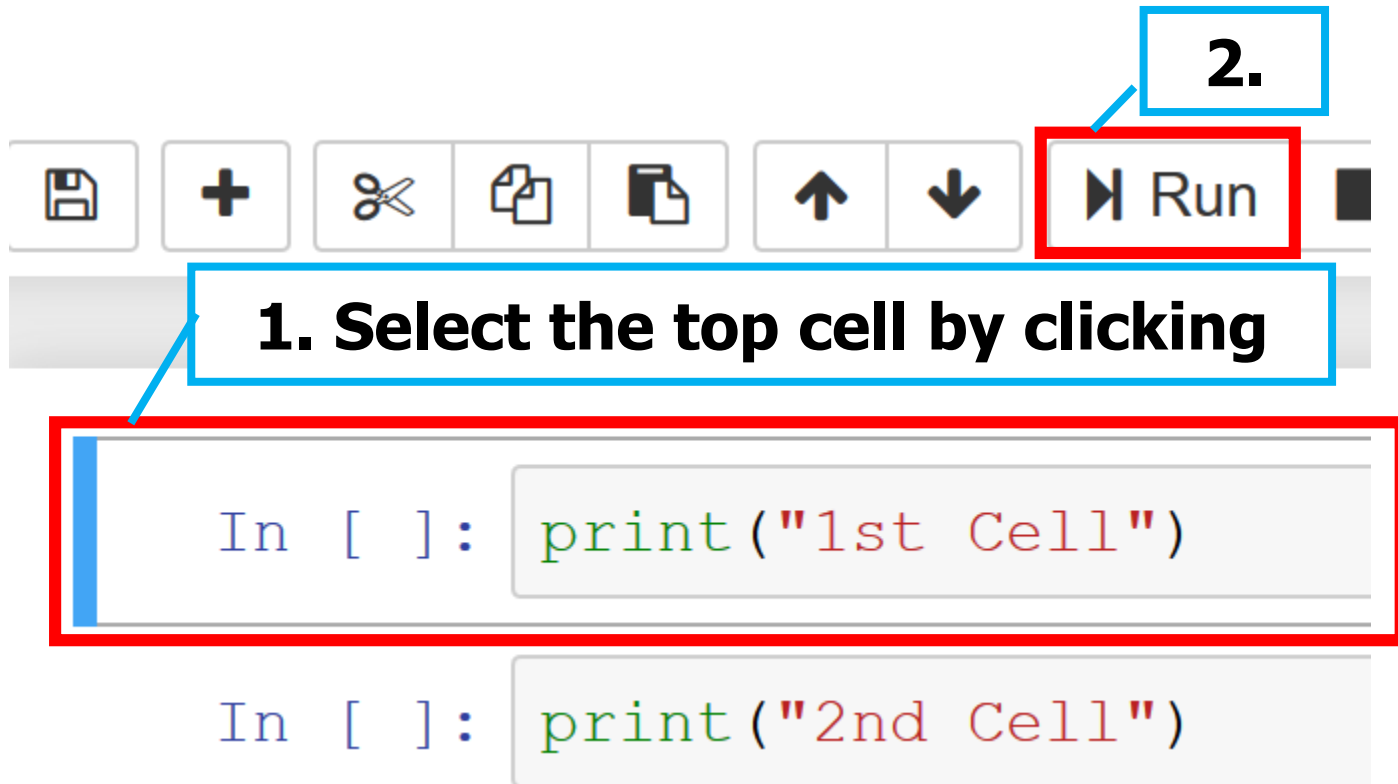


```
In [ ]: print("1st Cell")
```

```
In [ ]: print("2nd Cell")
```



Cells



2.

1. Select the top cell by clicking

```
In [ ]: print("1st Cell")
```

```
In [ ]: print("2nd Cell")
```


Cells

3. Run the next cell



1. You can see that only the first cell is executed

```
In [1]: print("1st Cell")
```

1st Cell

2. The next cell is automatically selected

```
In [ ]: print("2nd Cell")
```

- 
-
- The 2nd cell is also executed

```
In [1]: print("1st Cell")
```

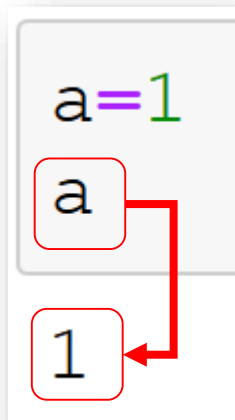
```
1st Cell
```

```
In [2]: print("2nd Cell")
```

```
2nd Cell
```



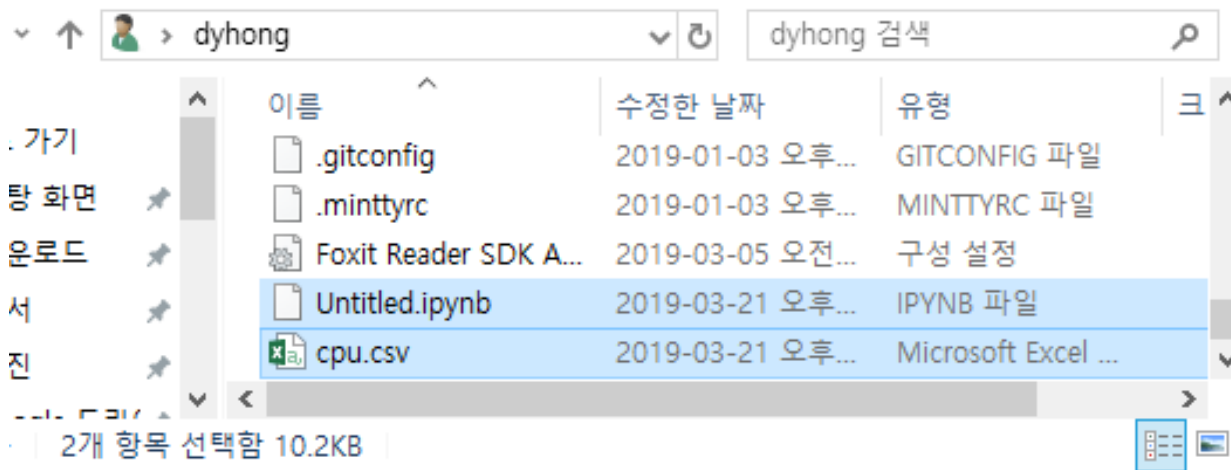
Cells



The value of the last line in the cell is automatically evaluated and printed

Load Data

- Download cpu.csv from <http://kdd.snu.ac.kr/python/>
- Save the csv file in the same directory as the source file (.ipynb)



Load Data

- *pandas*: A library that supports data analysis

Import the *pandas* library

and use alias 'pd' instead of 'pandas'

```
import pandas as pd
df = pd.read_csv("cpu.csv")
df[:5]
```

Read the csv file and store it in a pandas dataframe object

select the rows up to 5th row

cpu.csv							
MYCT	MMIN	MMAx	CACH	CHMIN	CHMAX	class	
125	256	6000	256	16	128	198	
29	8000	32000	32	8	32	269	
29	8000	32000	32	8	32	220	
29	8000	32000	32	8	32	172	
29	8000	16000	32	8	16	132	



Load Data

```
import pandas as pd
df = pd.read_csv("cpu.csv")
df[:5]
```

	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	class
0	125	256	6000	256	16	128	198
1	29	8000	32000	32	8	32	269
2	29	8000	32000	32	8	32	220
3	29	8000	32000	32	8	32	172
4	29	8000	16000	32	8	16	132



Load Data

- `df.values` returns an array form of the data

```
print(df.values)
```

```
[[ 125    256  6000 ...    16    128    198]
 [   29  8000 32000 ...     8     32    269]
 [   29  8000 32000 ...     8     32    220]
 ...
 [  125   2000  8000 ...     2     14     52]
 [  480    512  8000 ...     0      0     67]
 [  480   1000  4000 ...     0      0     45]]
```



Load Data

Select all the rows

Select all but
the last column

```
X = df.values[:, :-1]  
y = df.values[:, -1]
```

Select only the
last column

Load Data

```
In [3]: X = df.values[:, :-1]
        y = df.values[:, -1]
        print(X[:3])
        print(y[:3])
```

X[:3]

```
[[ 125   256  6000   256    16   128]
 [   29  8000 32000    32     8    32]
 [   29  8000 32000    32     8    32]]
```

Y[:3]

```
[198 269 220]
```

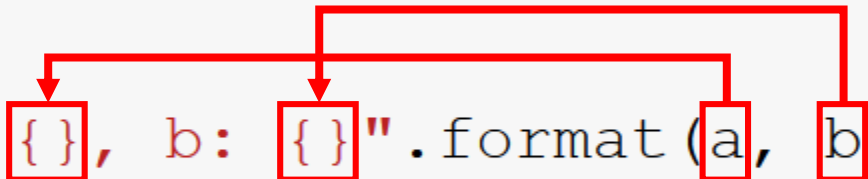
	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX	class
0	125	256	6000	256	16	128	198
1	29	8000	32000	32	8	32	269
2	29	8000	32000	32	8	32	220



Print

- You can control the formatting of output string by using `format()` function and `{}` mark in the string

```
a = 1
b = 2
print("a: {}, b: {}".format(a, b))
```



```
a: 1, b: 2
```



Shape

- You can check dimensions of array by using the *shape* attribute

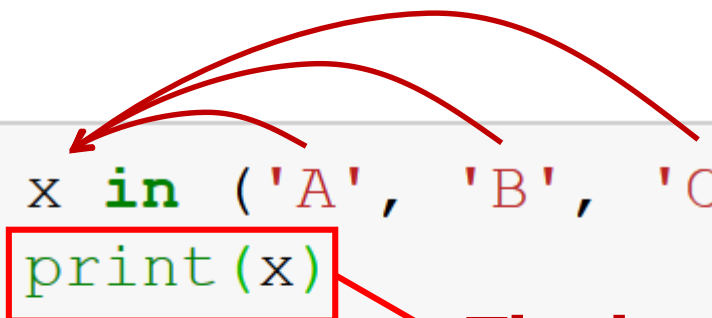
```
print("X.shape: {}, y.shape: {}".format(  
    X.shape, y.shape))
```

```
X.shape: (209, 6), y.shape: (209,)
```



For Loop

```
for x in ('A', 'B', 'C'):  
    print(x)
```



The body of the for loop

A
B
C



감사 합니다!

