Brookhaven National Laboratory

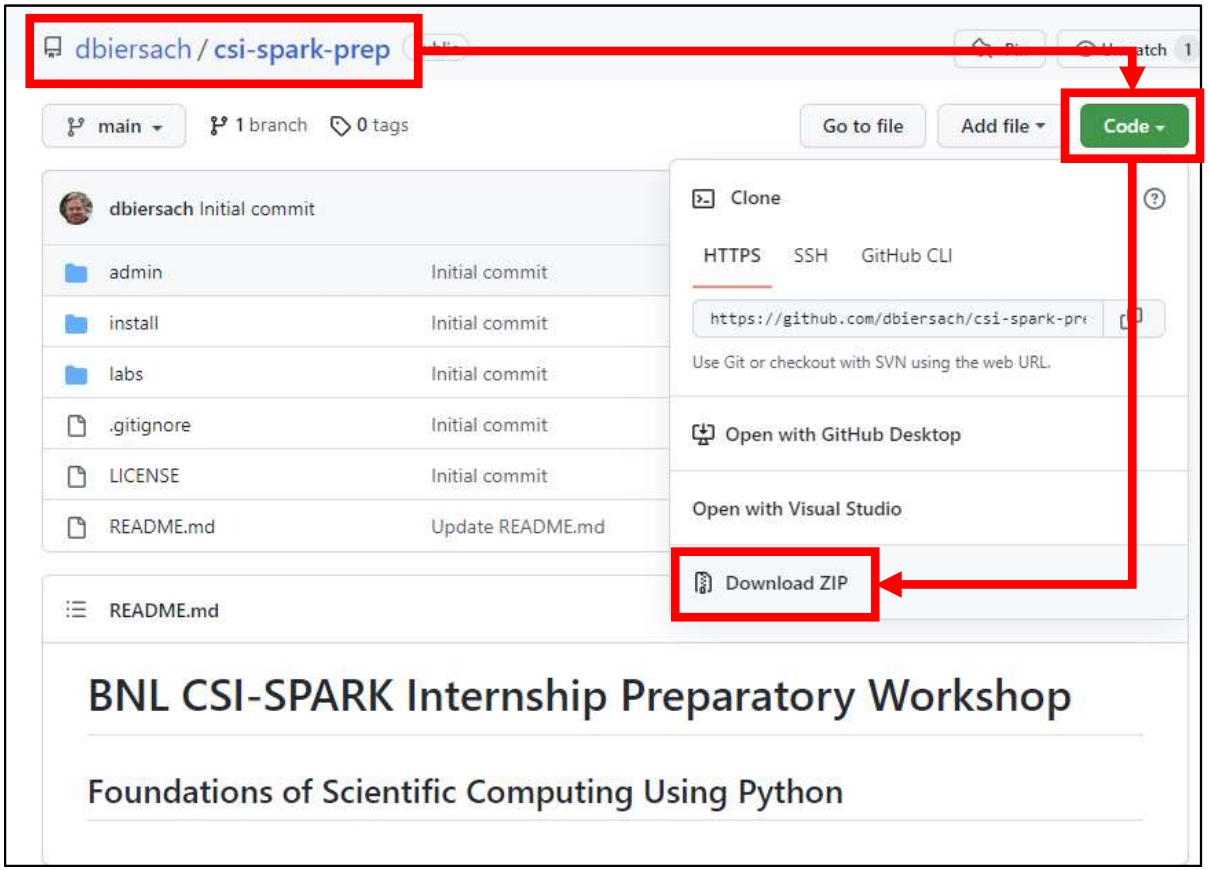
Foundations of Scientific Computing

Dave Biersach
dbiersach@bnl.gov

Session 01
Python
Fundamentals

Downloading the Courseware

The courseware can be downloaded as a large single **ZIP** file from https://github.com/dbiersach/csi-spark-prep



Welcome!

- My name is **Dave Biersach**
- I am a Senior Technology Architect at BNL
- I am a 1989 graduate of the United States Military Academy at West Point, and I served in the Persian Gulf War as a Combat Engineer
- I worked for DARPA developing satellite counter-reconnaissance search algorithms
- I have worked at **Microsoft** & **Pfizer**
- I have been married 30 years, have three adult children, and have mentored students for the past 15 years

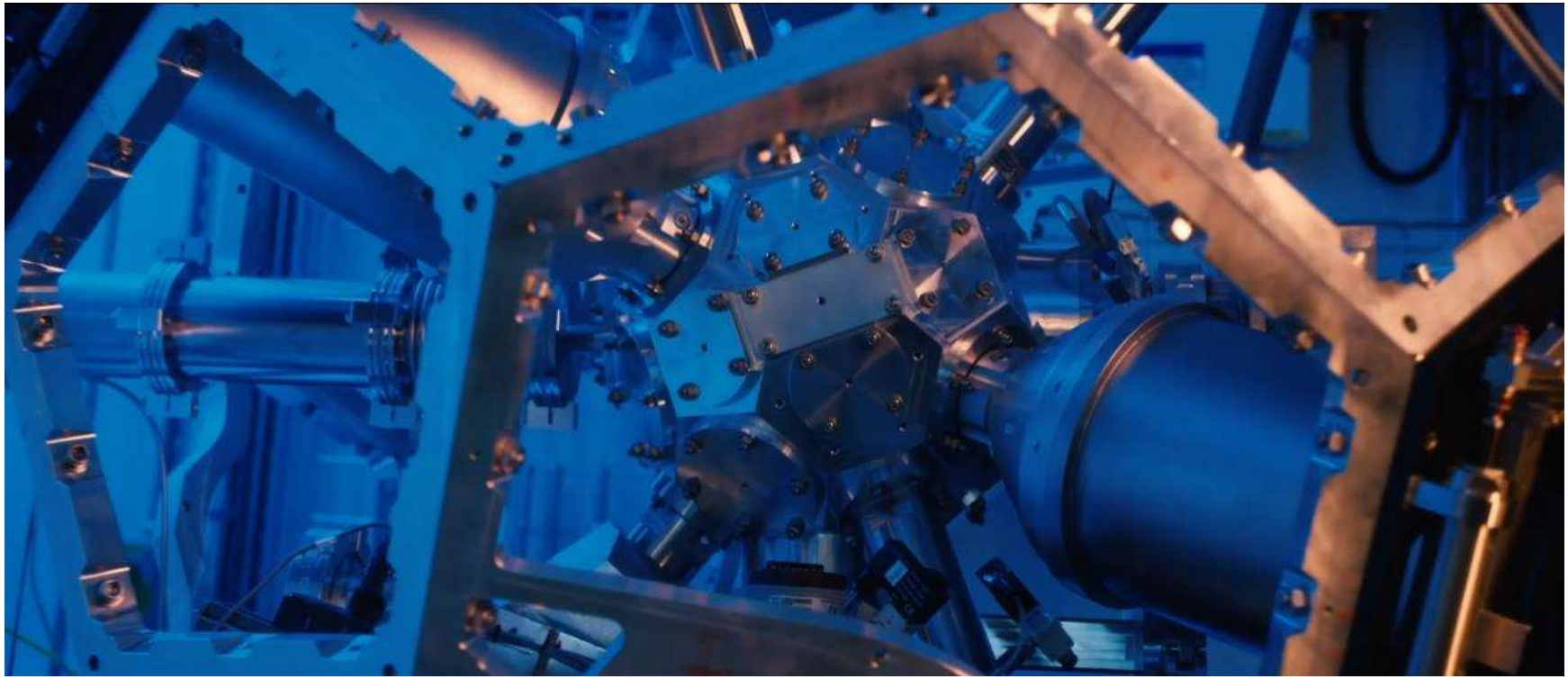
About Brookhaven National Laboratory

Who We Are

Brookhaven National Laboratory is a multipurpose research institution funded primarily by the U.S. Department of Energy's Office of Science. Located on the center of Long Island, New York, Brookhaven Lab brings world-class facilities and expertise to the most exciting and important questions in basic and applied science—from the birth of our universe to the sustainable energy technology of tomorrow.

We operate cutting-edge large-scale facilities for studies in physics, chemistry, biology, medicine, applied science, and a wide range of advanced technologies. The Laboratory's almost 3,000 scientists, engineers, and support staff are joined each year by more than 4,000 visiting researchers from around the world. Our award-winning history stretches back to 1947, and we continue to unravel mysteries from the nanoscale to the cosmic scale, and everything in between.

About Brookhaven National Laboratory



https://www.youtube.com/watch?v=csgXRPV0R3A

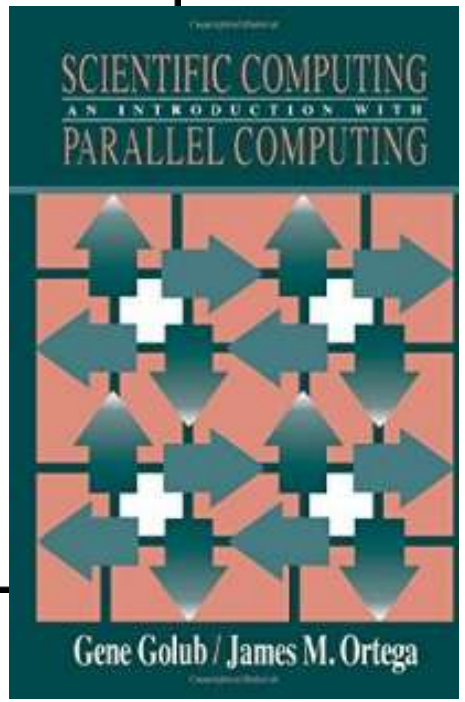
What is Scientific Computing?

Golub and Ortega: *"Scientific computing is the collection of tools, techniques and theories required to solve on a computer mathematical models of problems in science and engineering."*

Or a more narrow definition: *"Development and use of numerical methods and mathematical models to solve real-world problems efficiently on computers."*

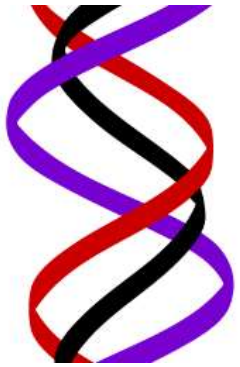
Interdisciplinary field requiring:

- knowledge about the underlying (physical) problem,
- ability to formulate a mathematical model,
- stable & accurate numerical schemes,
- efficient implementation on high performance computers.



What is Scientific Computing?

- Scientific computing problems **cannot be solved** using just a graphing calculator or a spreadsheet program
  - A computer should not be viewed as just another closed-form benchtop instrument with fixed functionality
  - SciComp does not require writing thousands of lines of code to answer problems – complete code usually fits on **one** slide!
- SciComp is **applied** computer science
  - The first name of CompSci is *computer*
  - The first name of SciComp is **science**
  - A **triple helix** of math, science, and computing



SciComp vs CompSci

Scientific Computing

- Probability and Statistics
- Simulation and Modelling
- Data Visualization
- Storing and Analyzing Very Large Datasets
- Parallel & Distributed Algorithms
- Speed and Accuracy Paramount
- Functional Languages
- Open-Ended Problems with Unknown Solutions

Computer Science

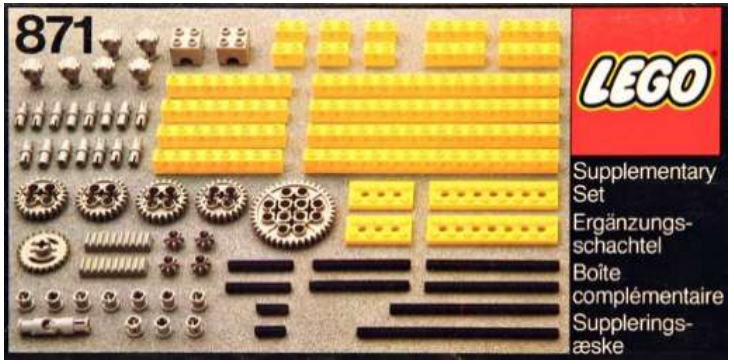
- General Data Structures
- Design Methodologies
- Procedural Languages
- Stand-Alone Programs
- Emphasis on Object-Orientation
- Simple Data Models
- Sequential Algorithms
- Less Graphics Intensive
- Directed Closed-Form Problems with Known Solutions

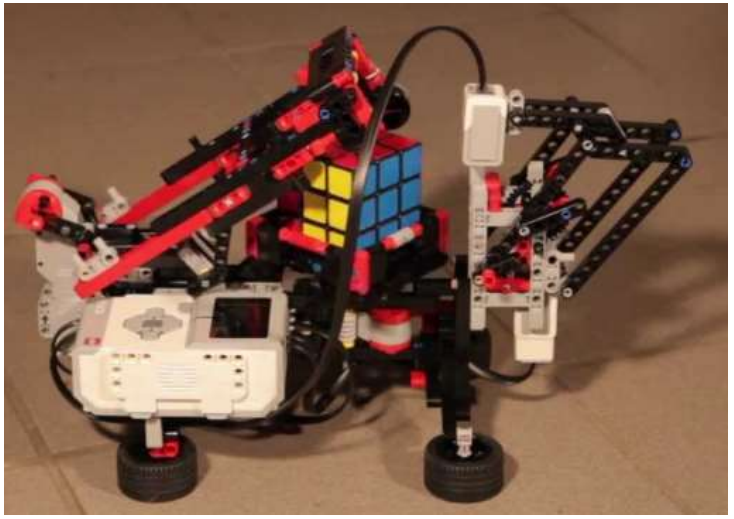
Scientific Computing with Python

- Python is quickly becoming one of the **most heavily used languages** in science projects
- Python runs on all major modern **operating systems** and is completely free and open-source (not vendor controlled)
- Python makes it easy for your code to directly integrate with a large spectrum of available 3rd party software
- Python code runs **consistently** on different platforms and scales well from small IoT devices to large server clusters
- Python benefits from a very active and growing user community that continues to enhance the language

SciComp = **Just Enough** CompSci

- Data **types**: int, float, bool, string
- Data **structures**: lists, arrays, dictionaries, classes, modules
- **Functions**: def, return
- **Statements**: if, for, while, break
- **Patterns**: vectorization, divide & conquer, map-reduce
- **Algorithms**: GCD, mean/variance, sorting, searching
- **Modules**: numpy, matplotlib, numba, scipy, sympy, scikit-learn

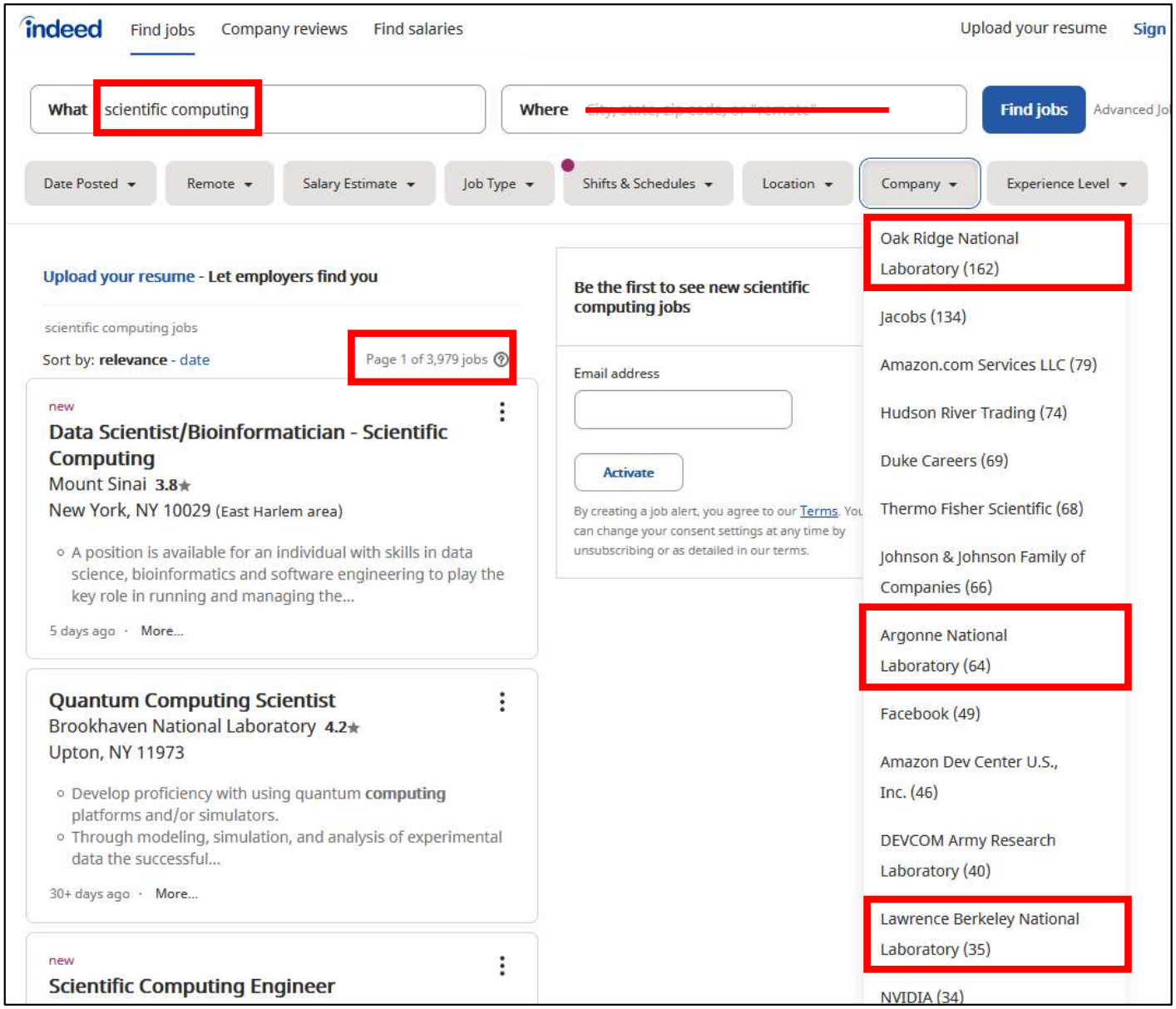




Useful Python Reference Sites

- https://docs.python.org/3/tutorial
- https://www.learnpython.org
- https://realpython.com
- https://www.w3schools.com/python
- https://www.fullstackpython.com

SciComp = The Pathway to Internships



Session **01** – Goals

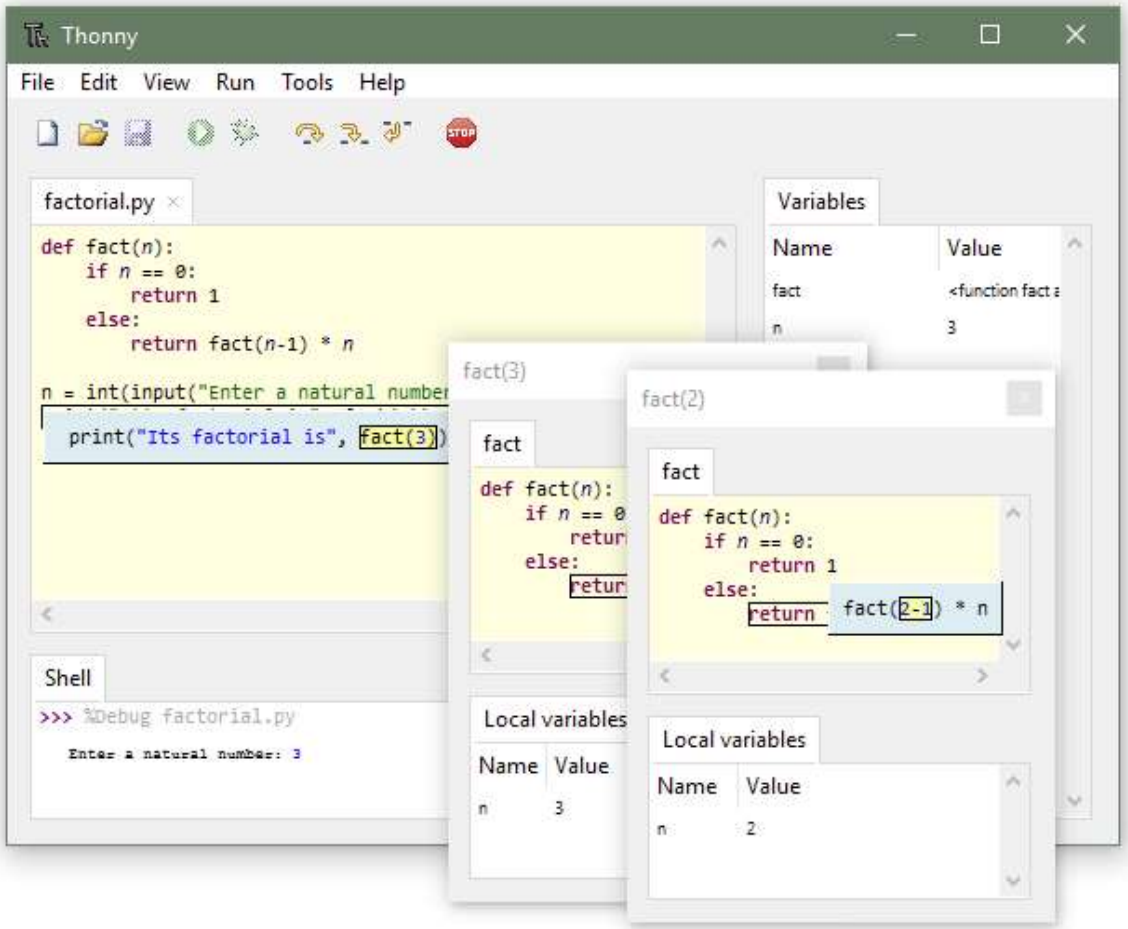
- Learn how to use the **Thonny** integrated development environment (IDE) and to install packages and plug-ins
- Review the rules for declaring **identifiers** in Python
- Declare **variables** and use built-in data **types**
- Display the values of variables in the output "Shell" window
- Review mathematical operators (PEMDAS) in Python
- Understand **statements** and **scopes**
- Create simple loops with the **for** statement
- Introduce **range()** as a "lazy" list definition

About the Thonny IDE

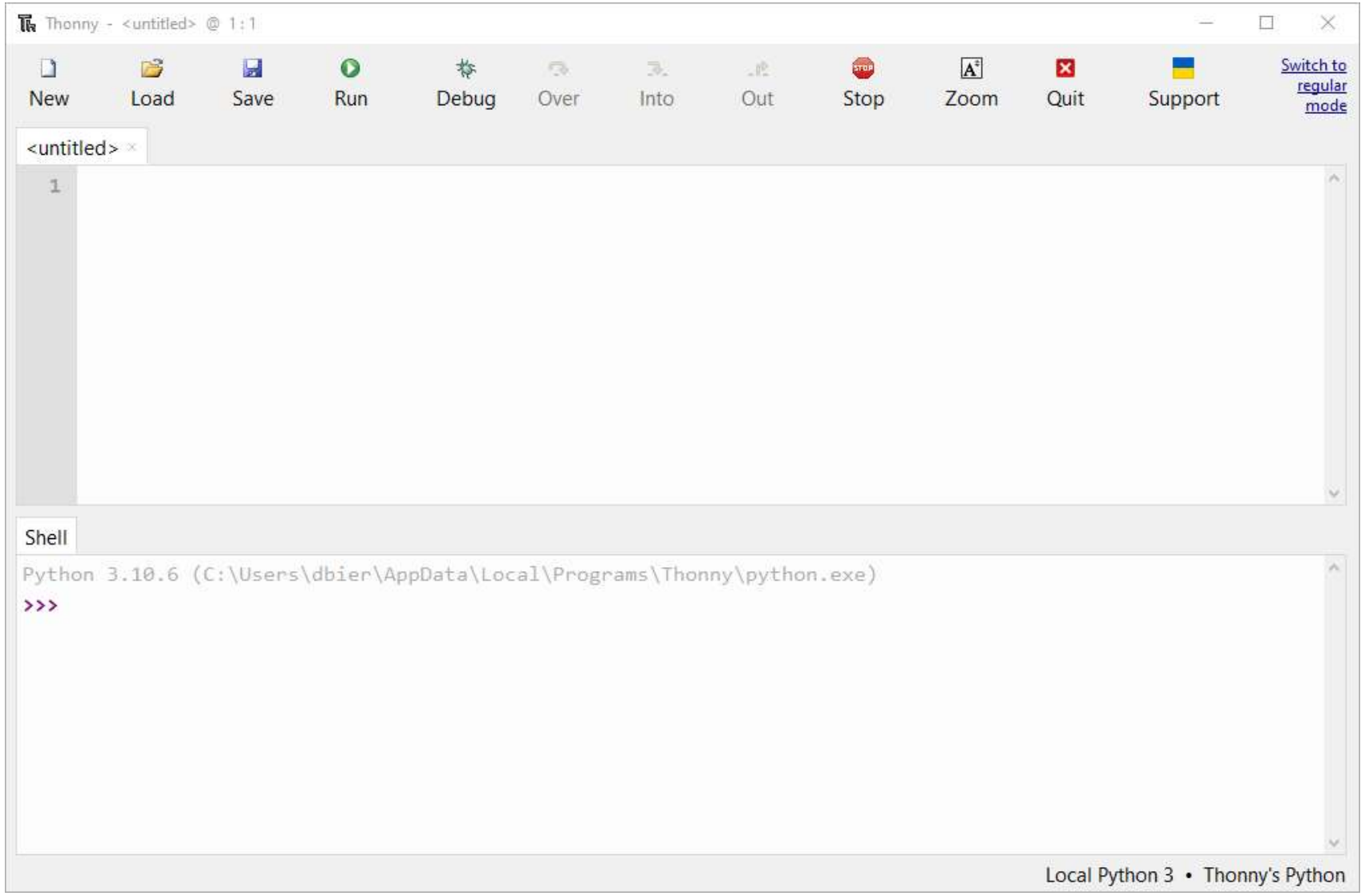
Thonny

Python IDE for beginners

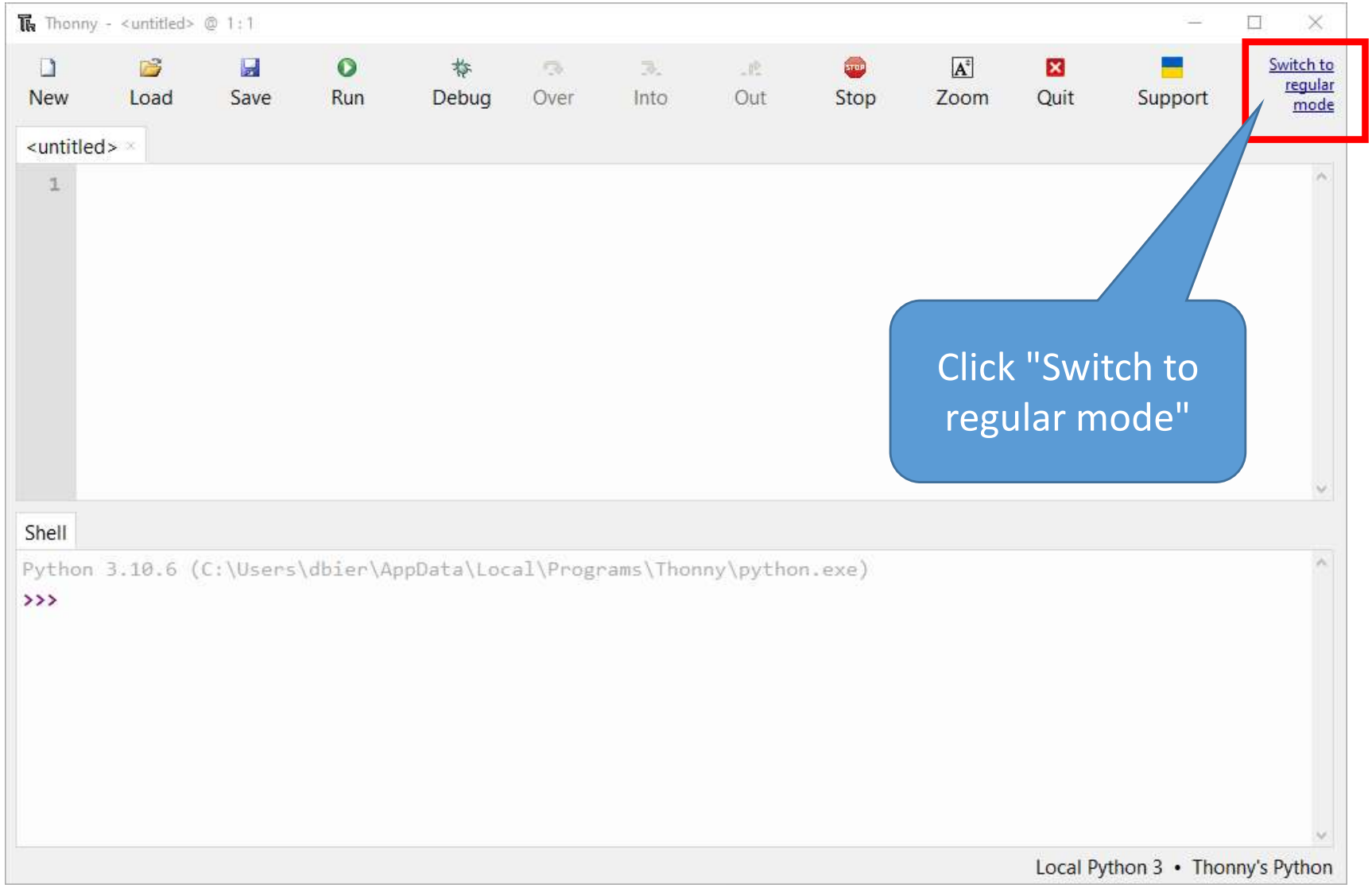
Download version 4.0.1 for Windows • Mac • Linux



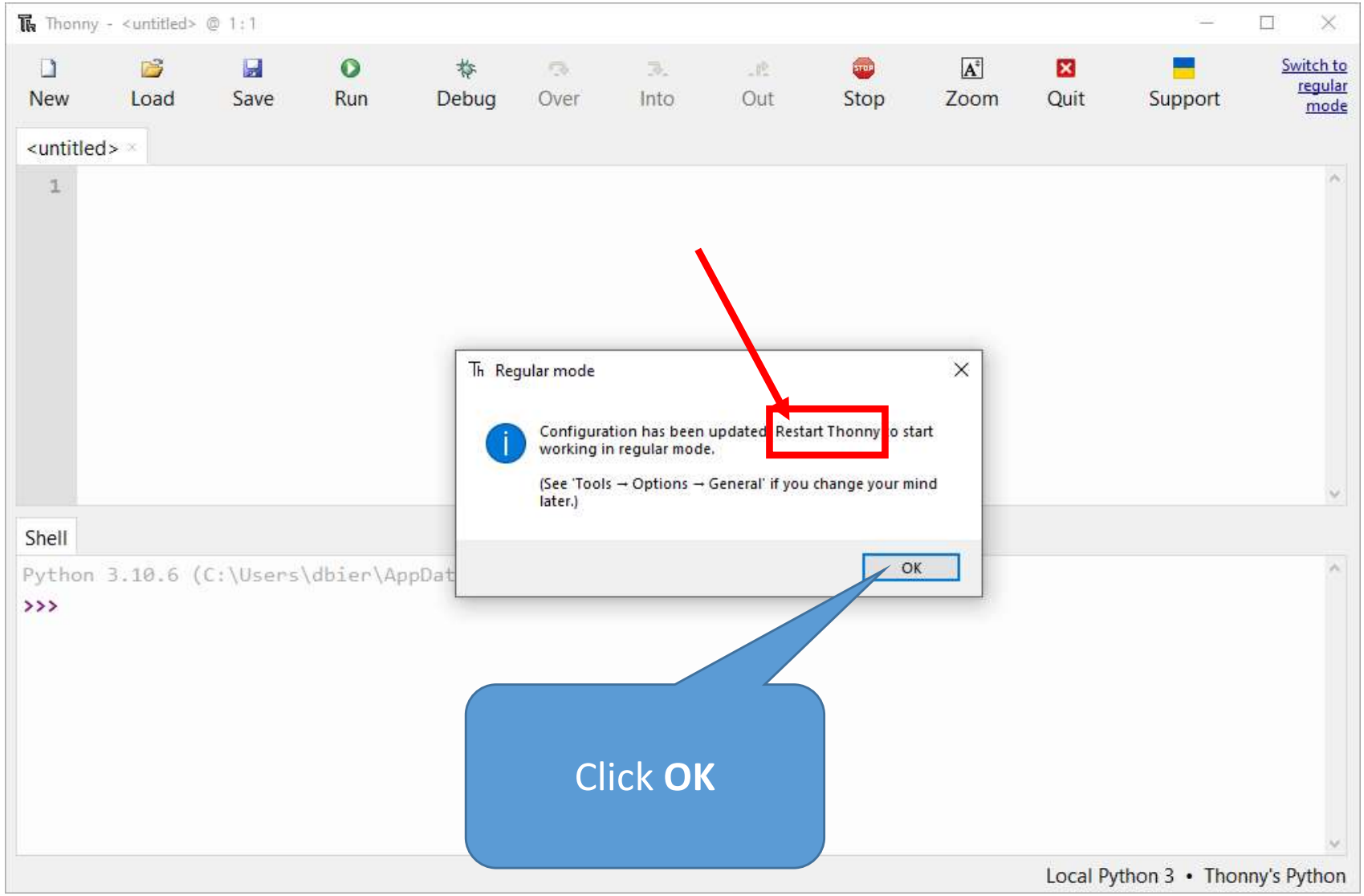
Run the Thonny IDE



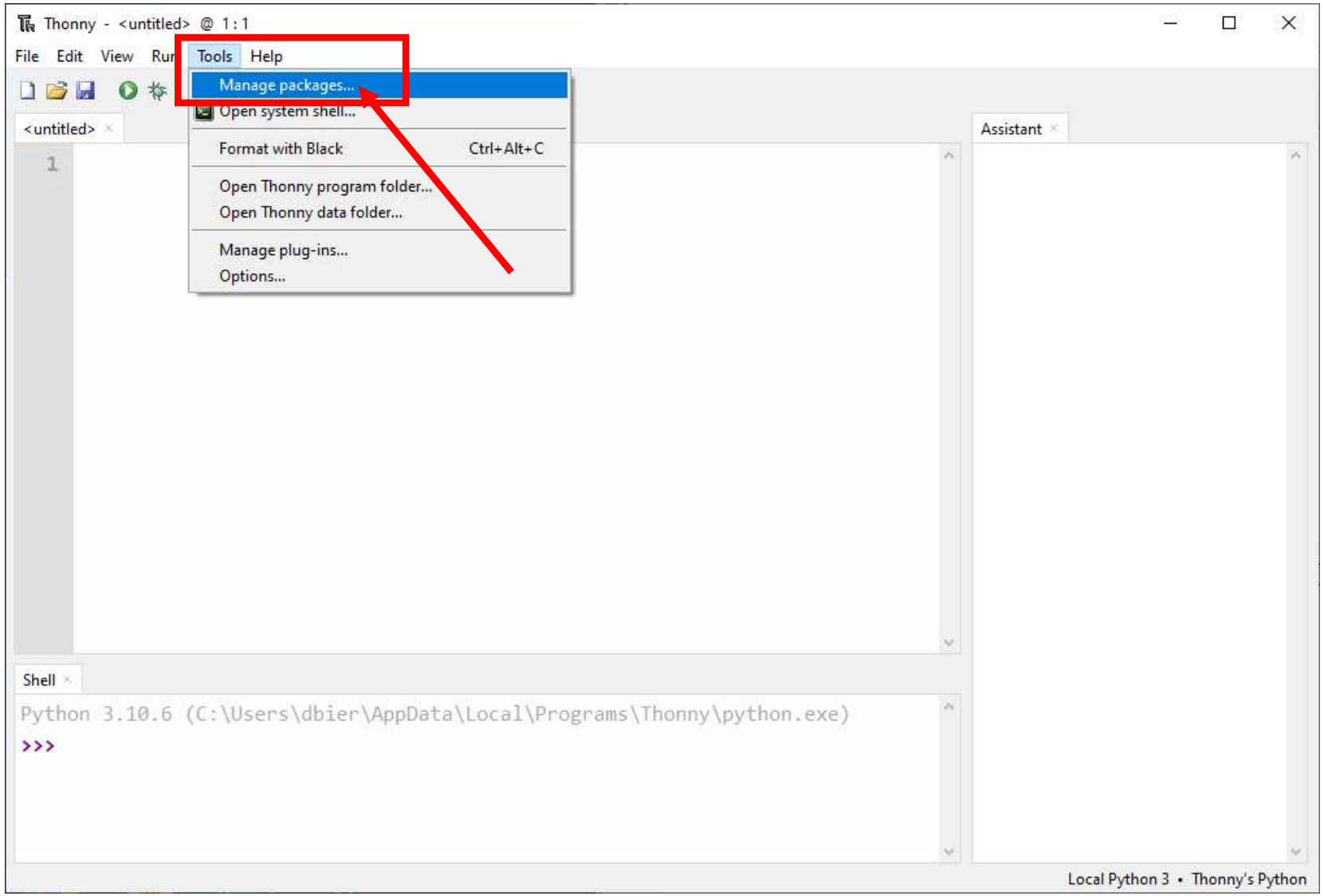
Switch the Thonny IDE into Regular Mode



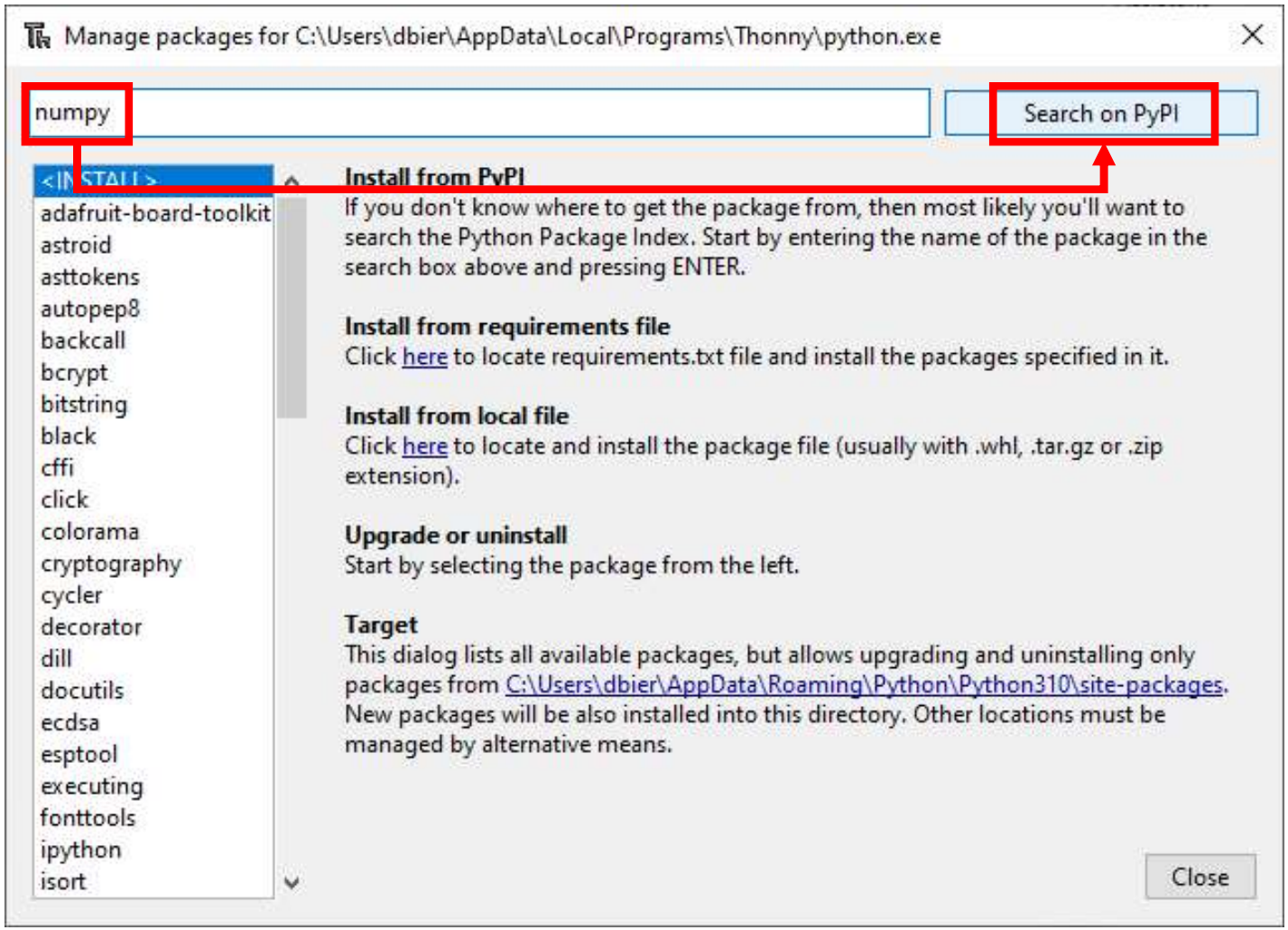
Restart the Thonny IDE



Installing Python Packages into Thonny

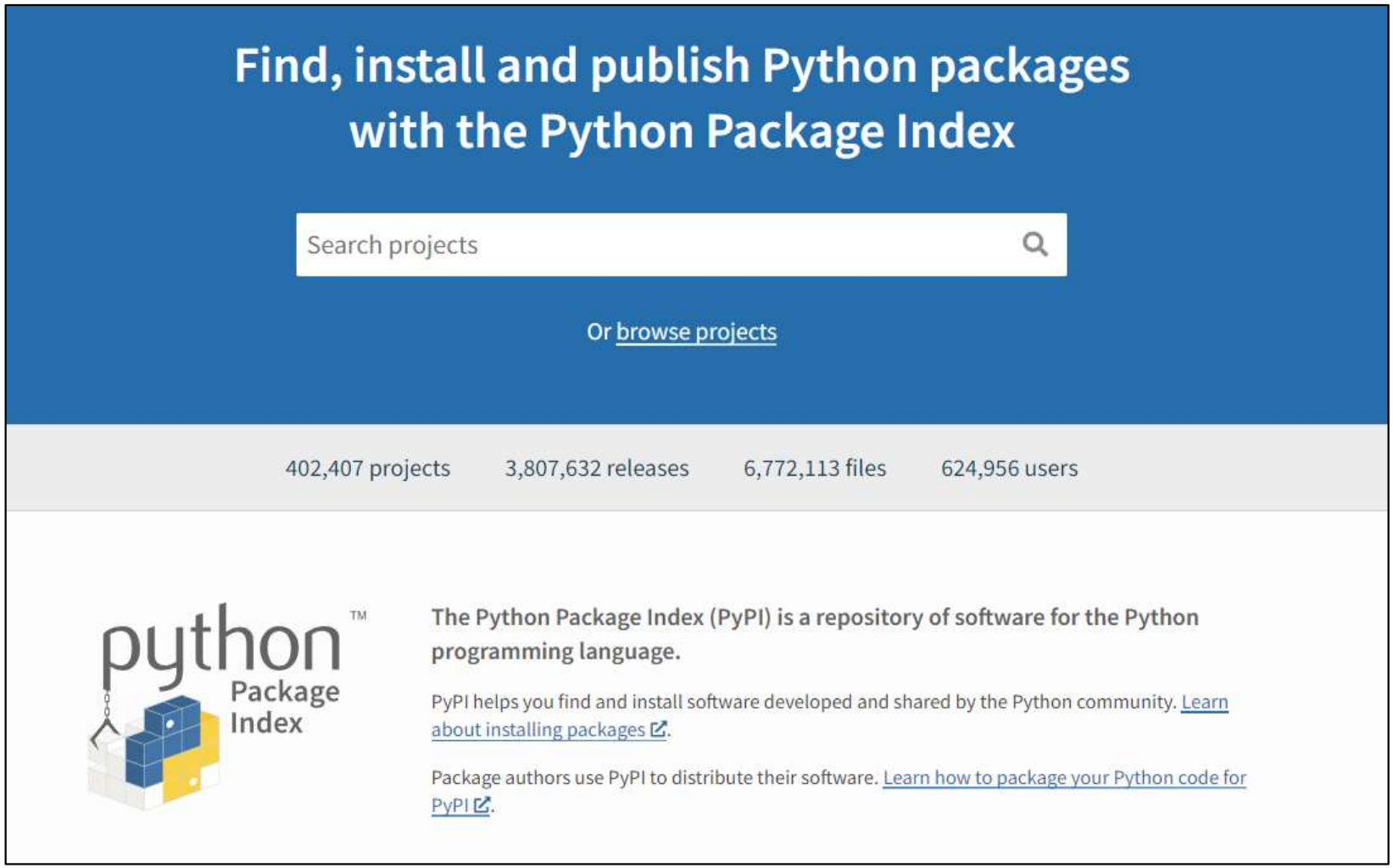


Find the Package Name on **PyPI**

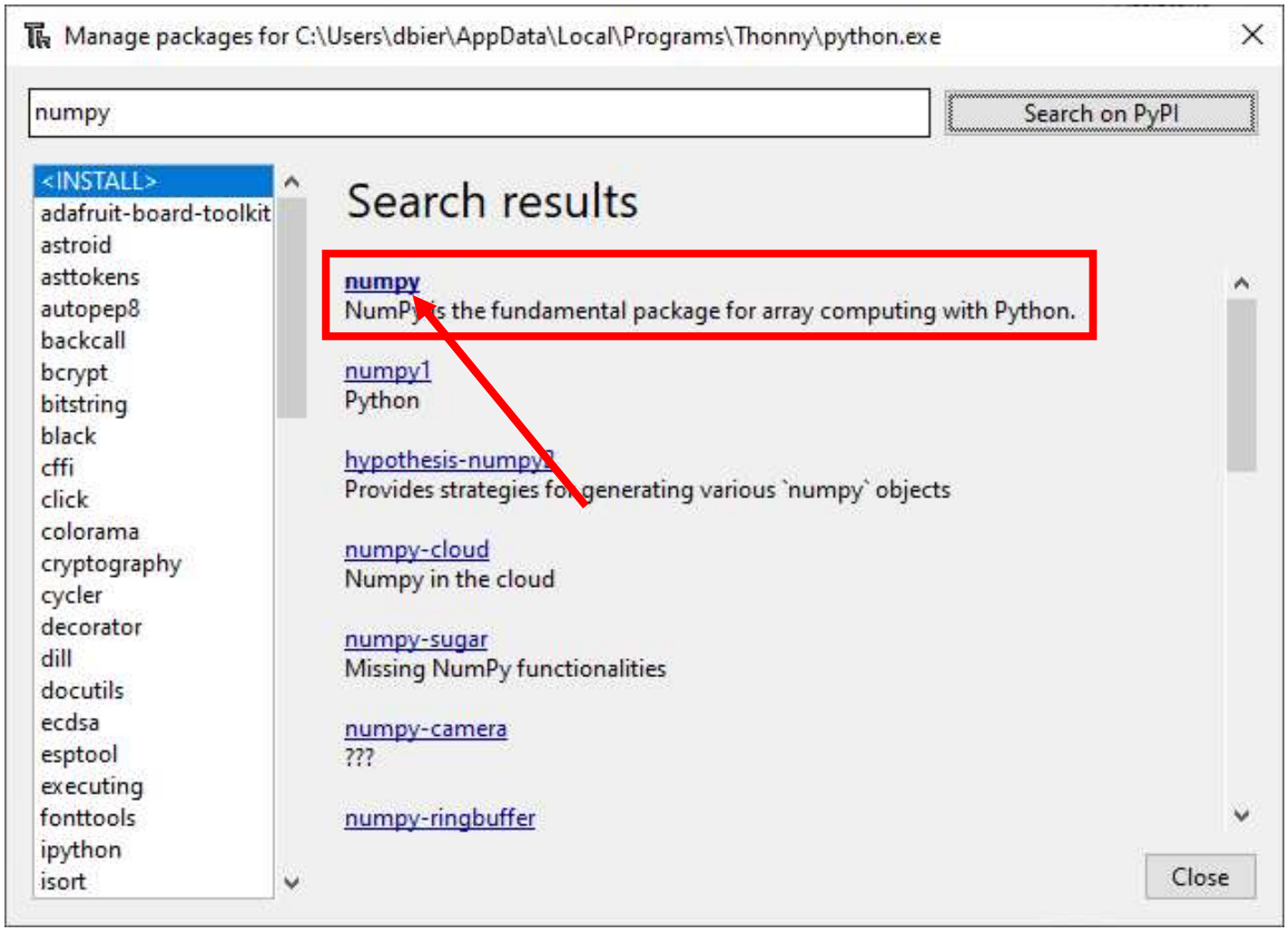


About **PyPI** (Python Package Index)

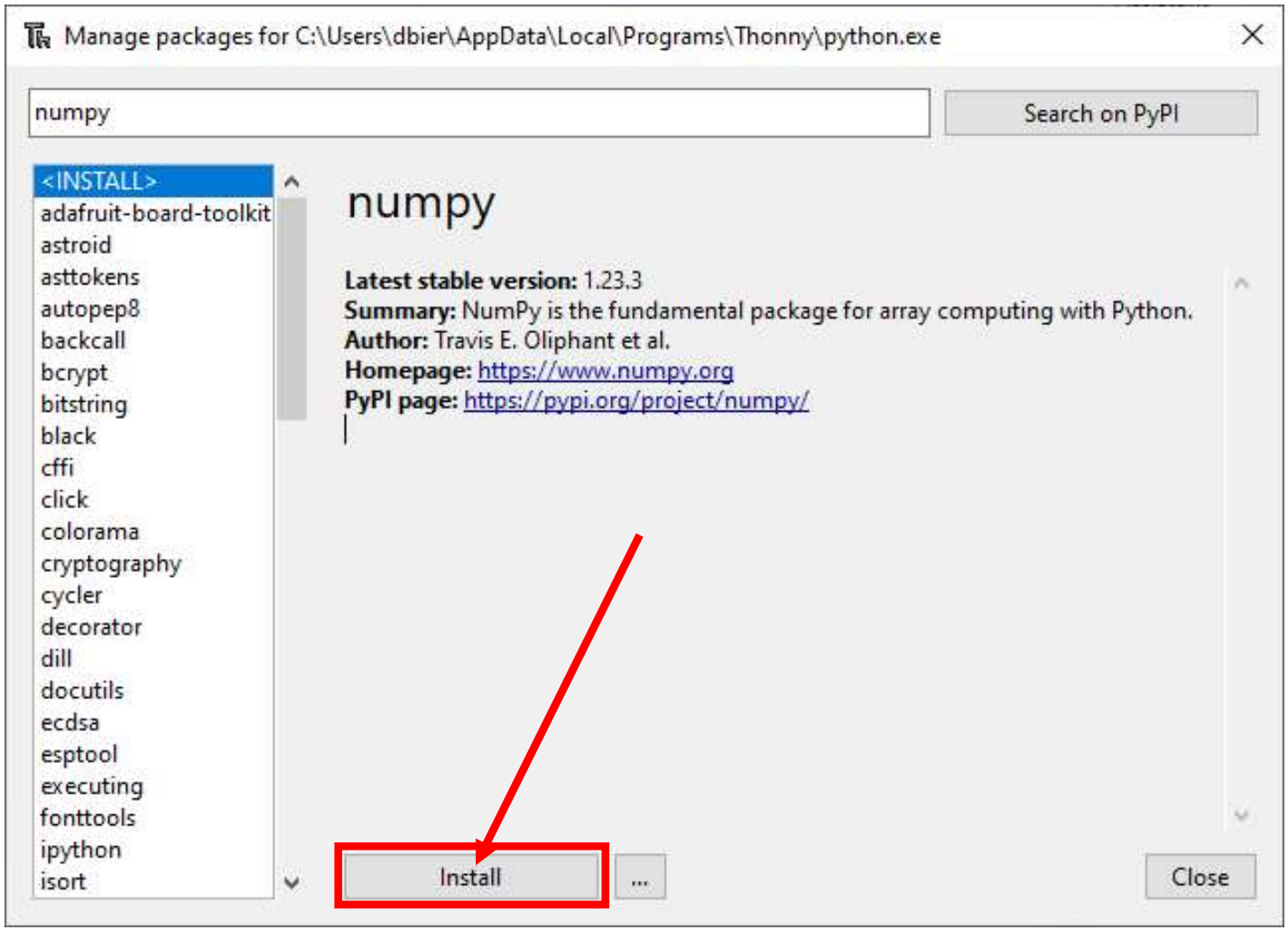
https://pypi.org



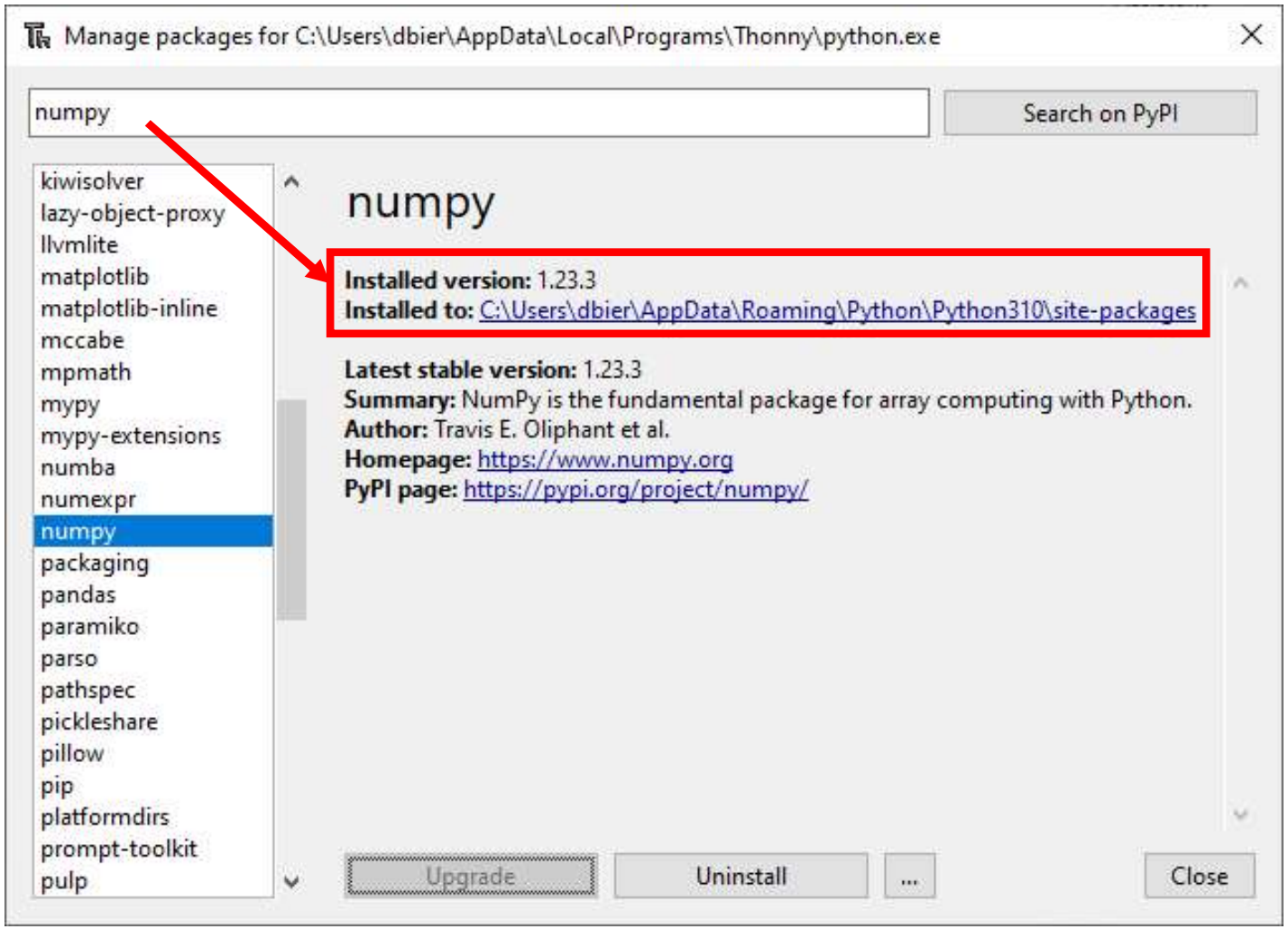
Select the Package Name



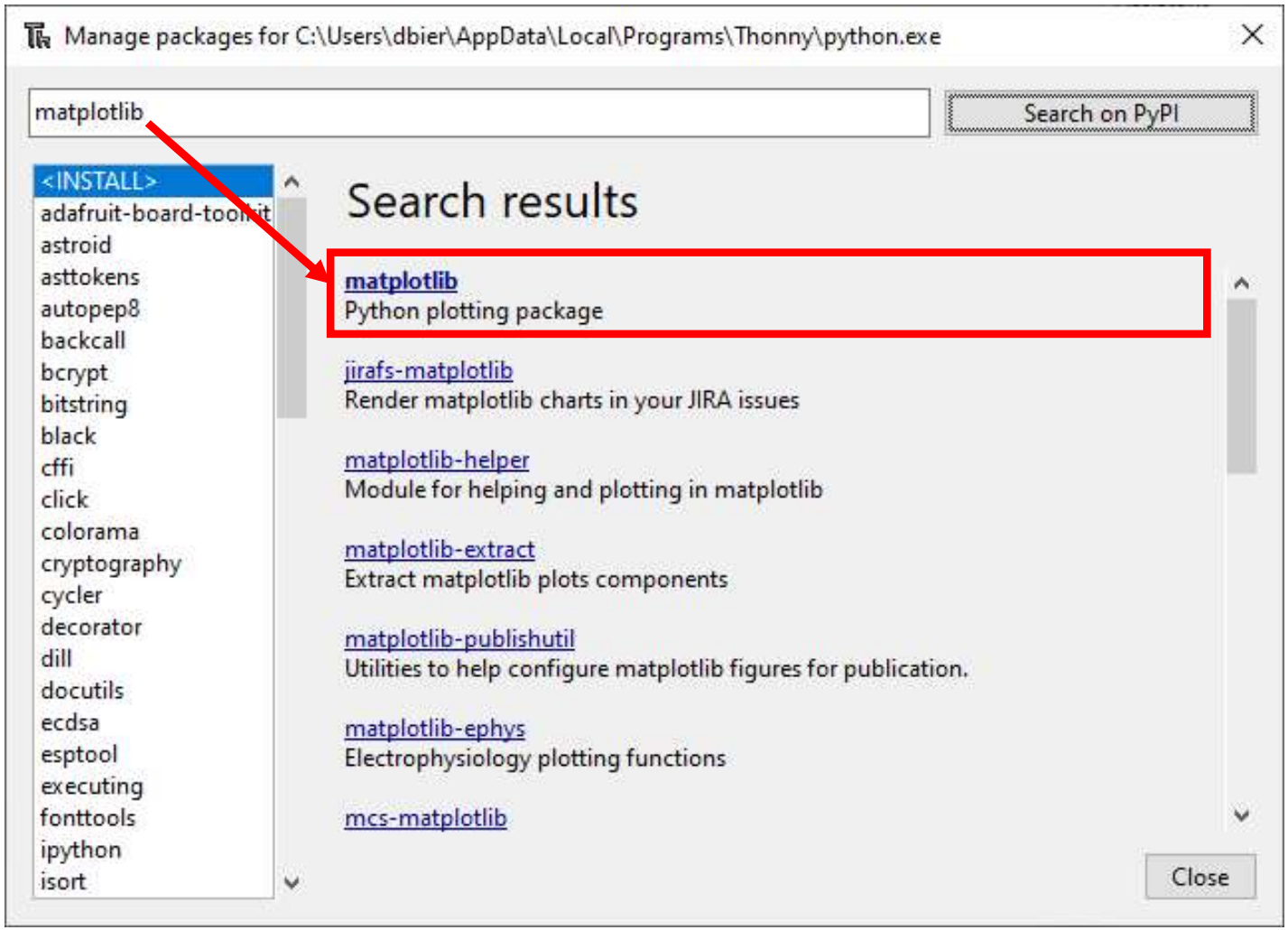
Install the Package



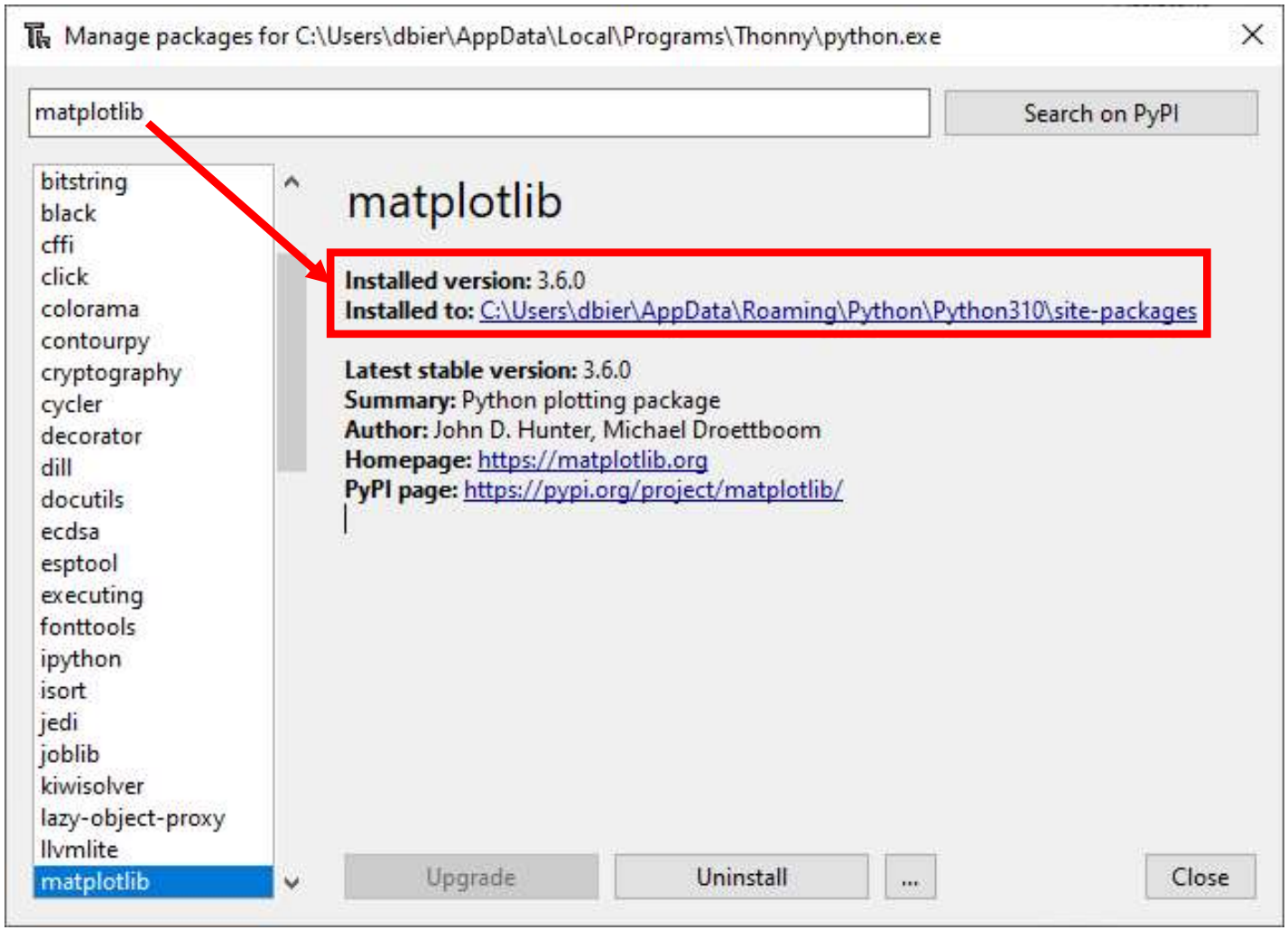
Verify the Package Installation



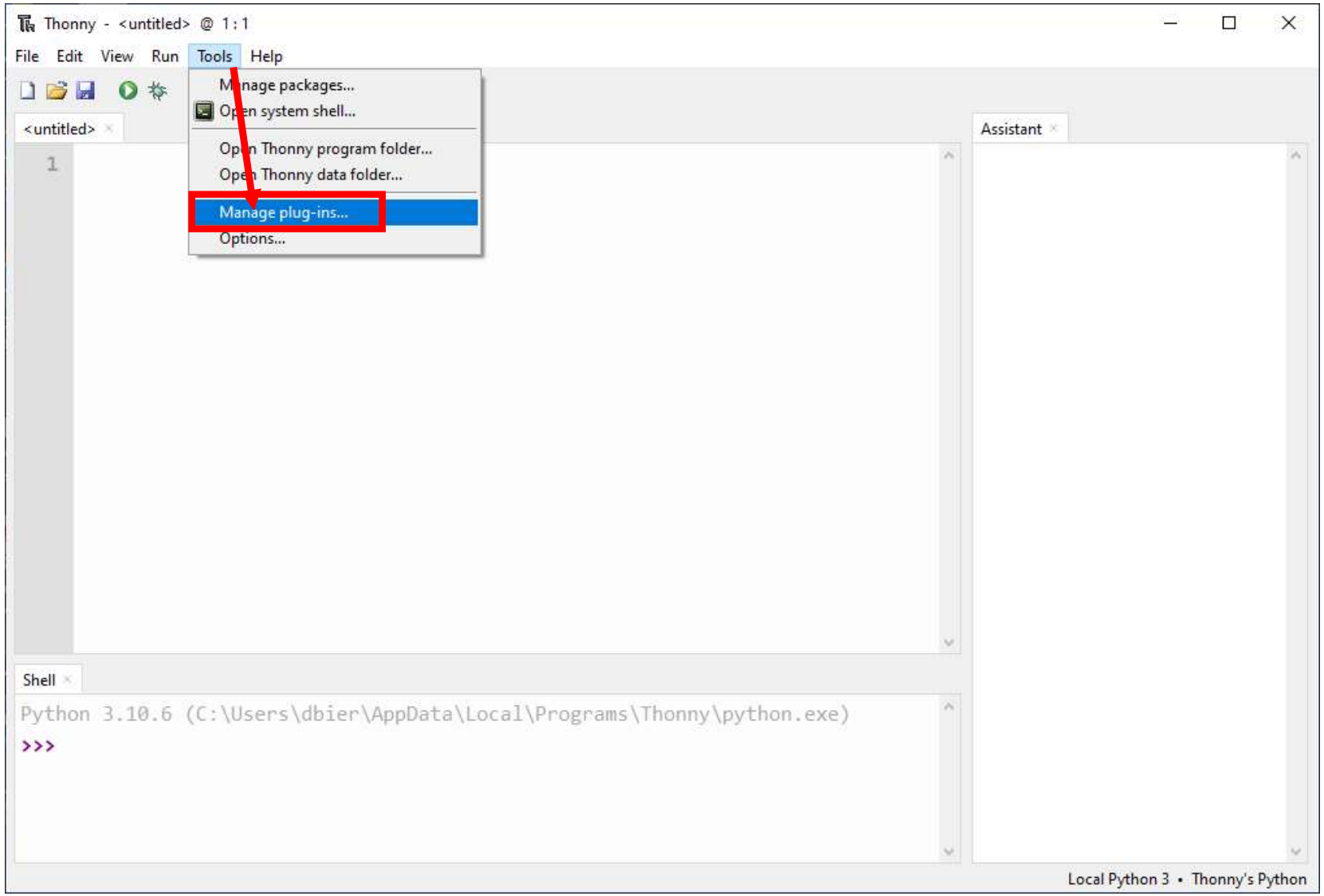
Install the **matplotlib** Package



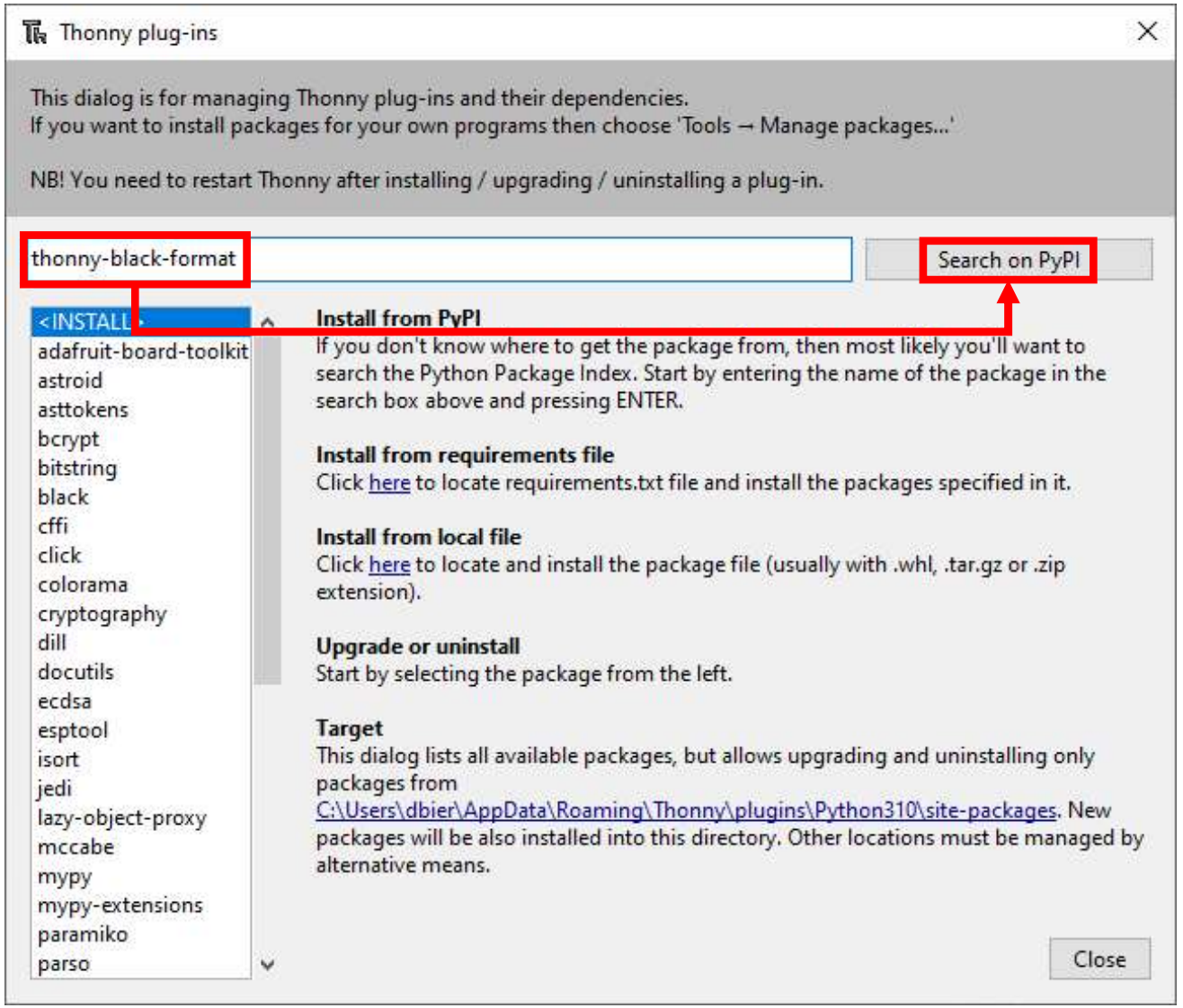
Verify the **matplotlib** Package Installation



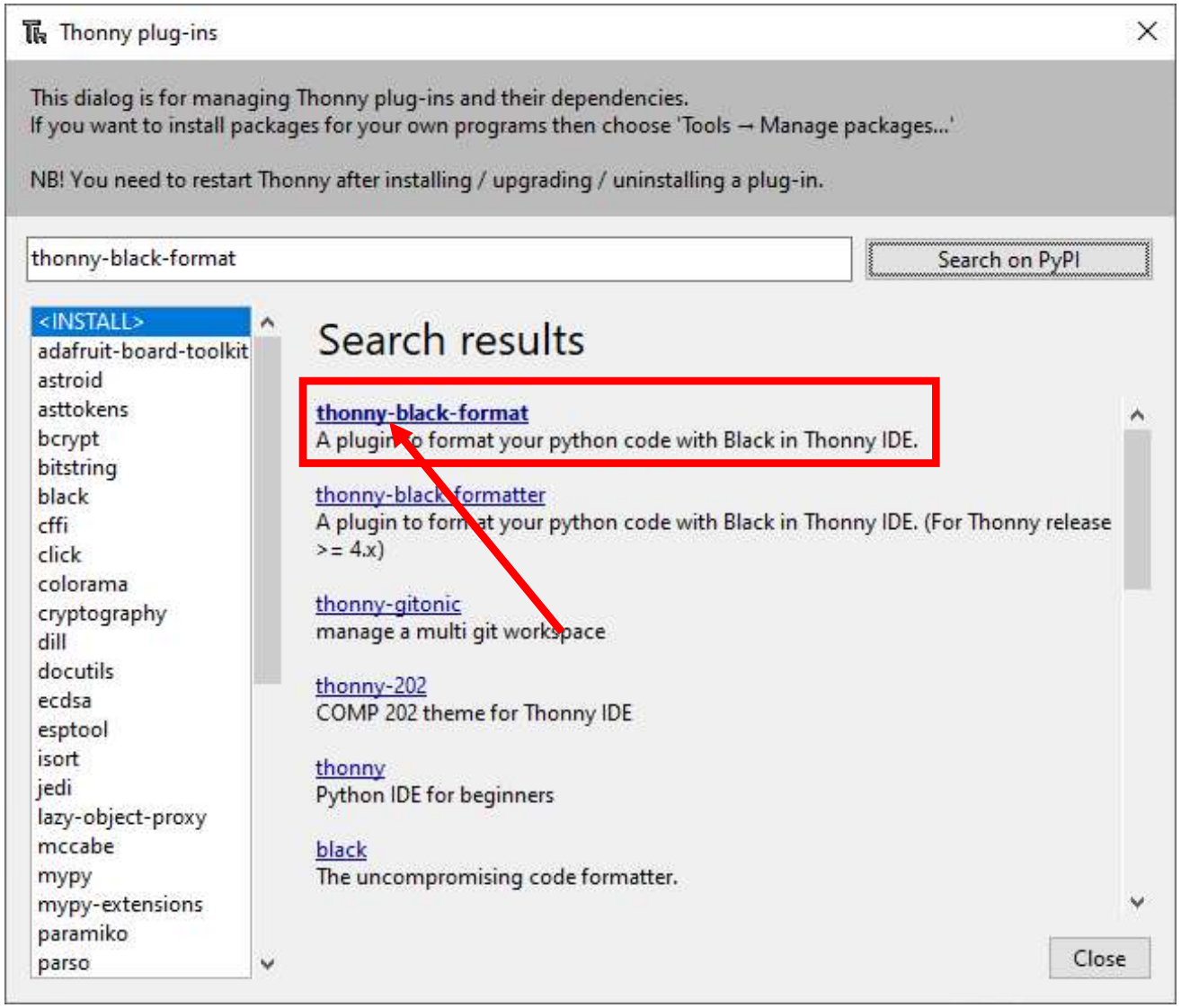
Installing Thonny **plug-ins**



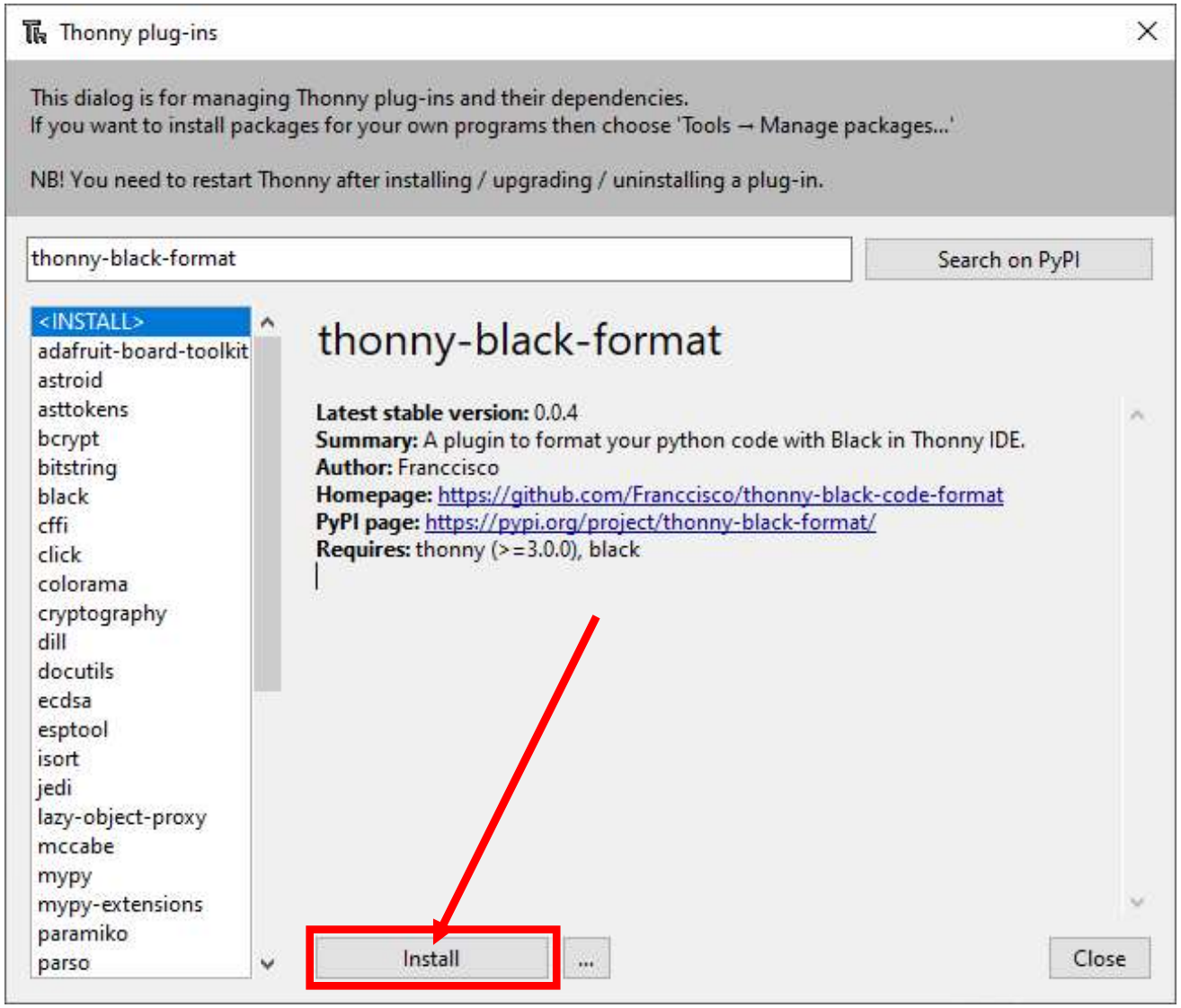
Find the plug-in name on **PyPI**



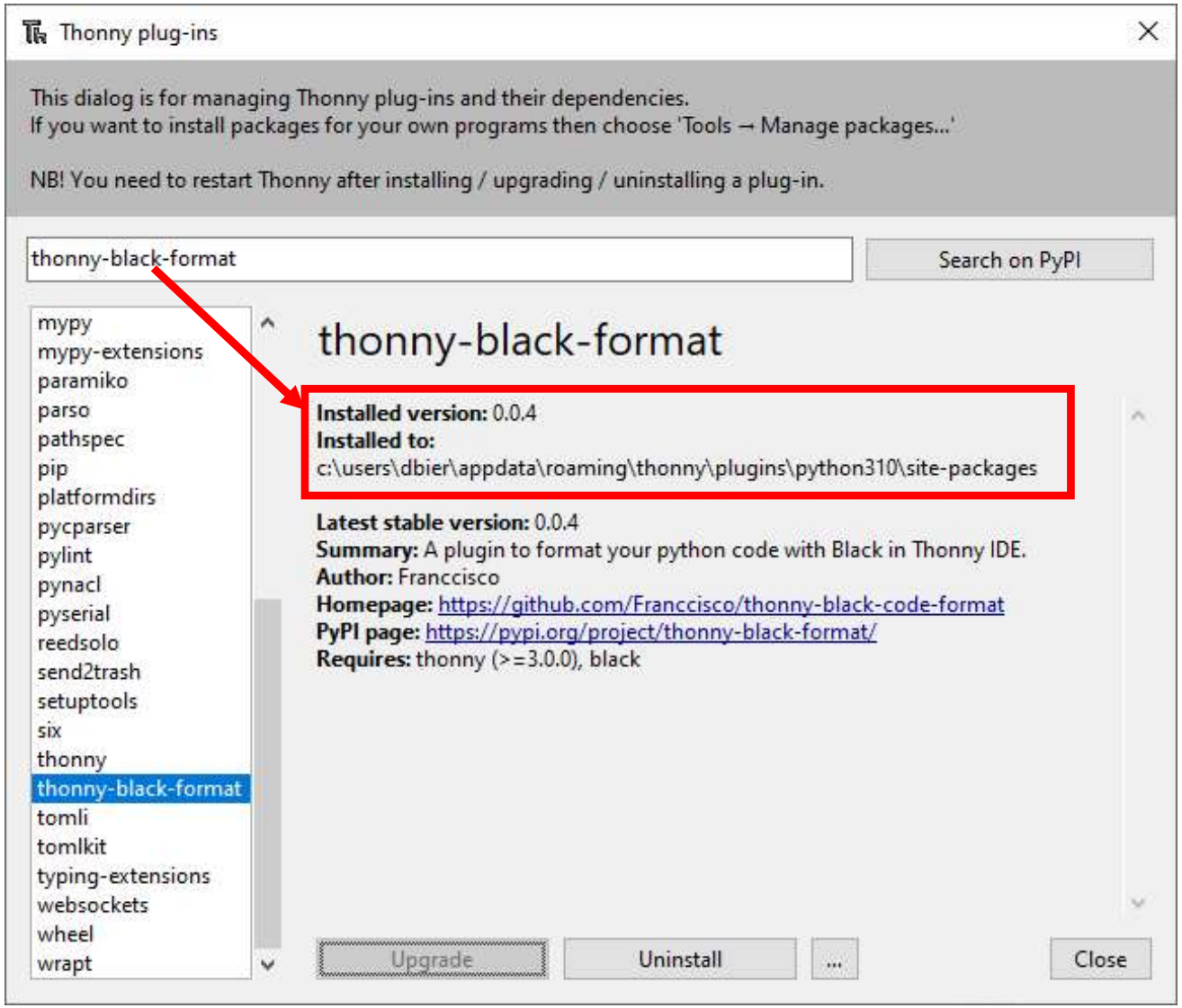
Select plug-in name



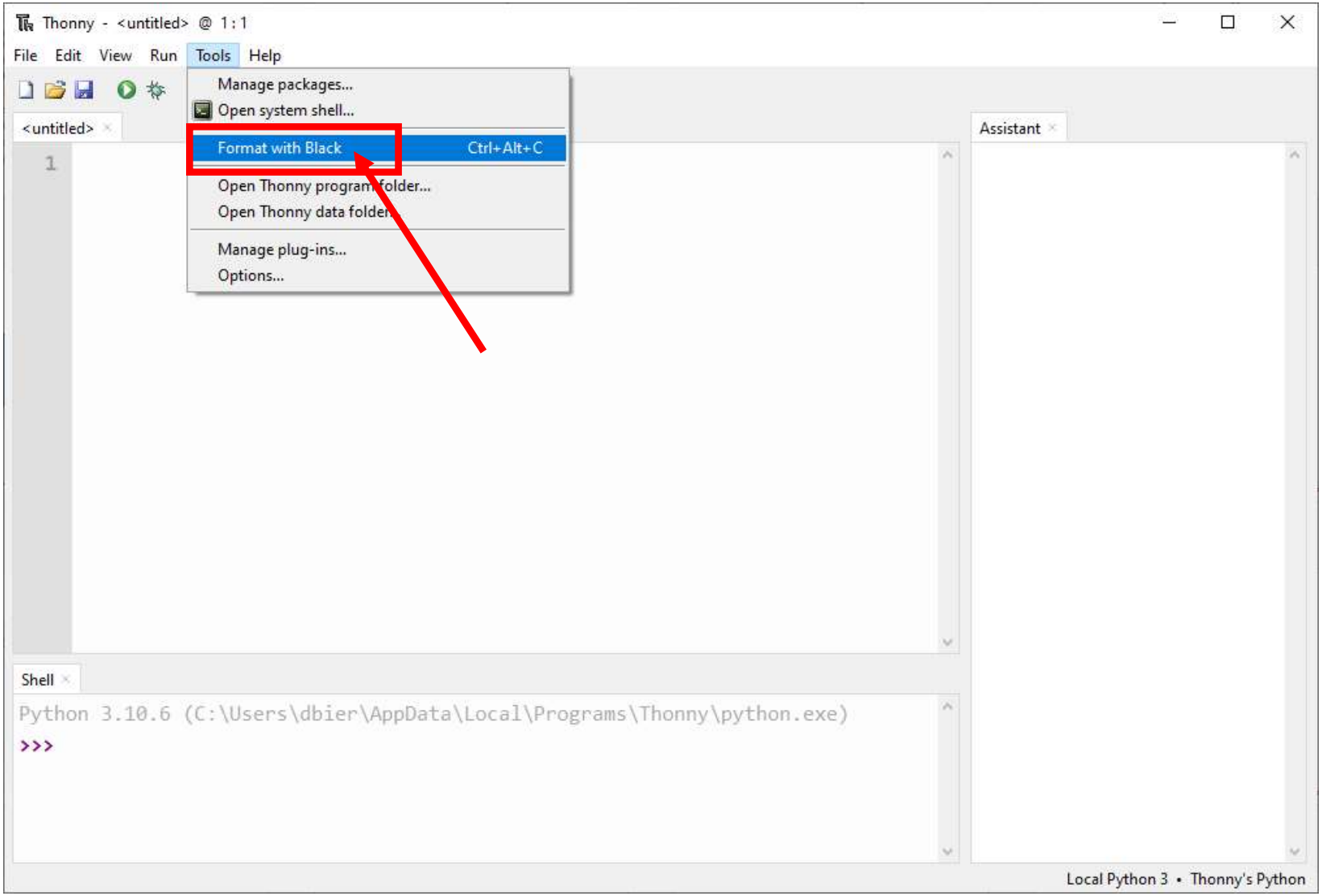
Install the plug-in



Verify the plug-in installation



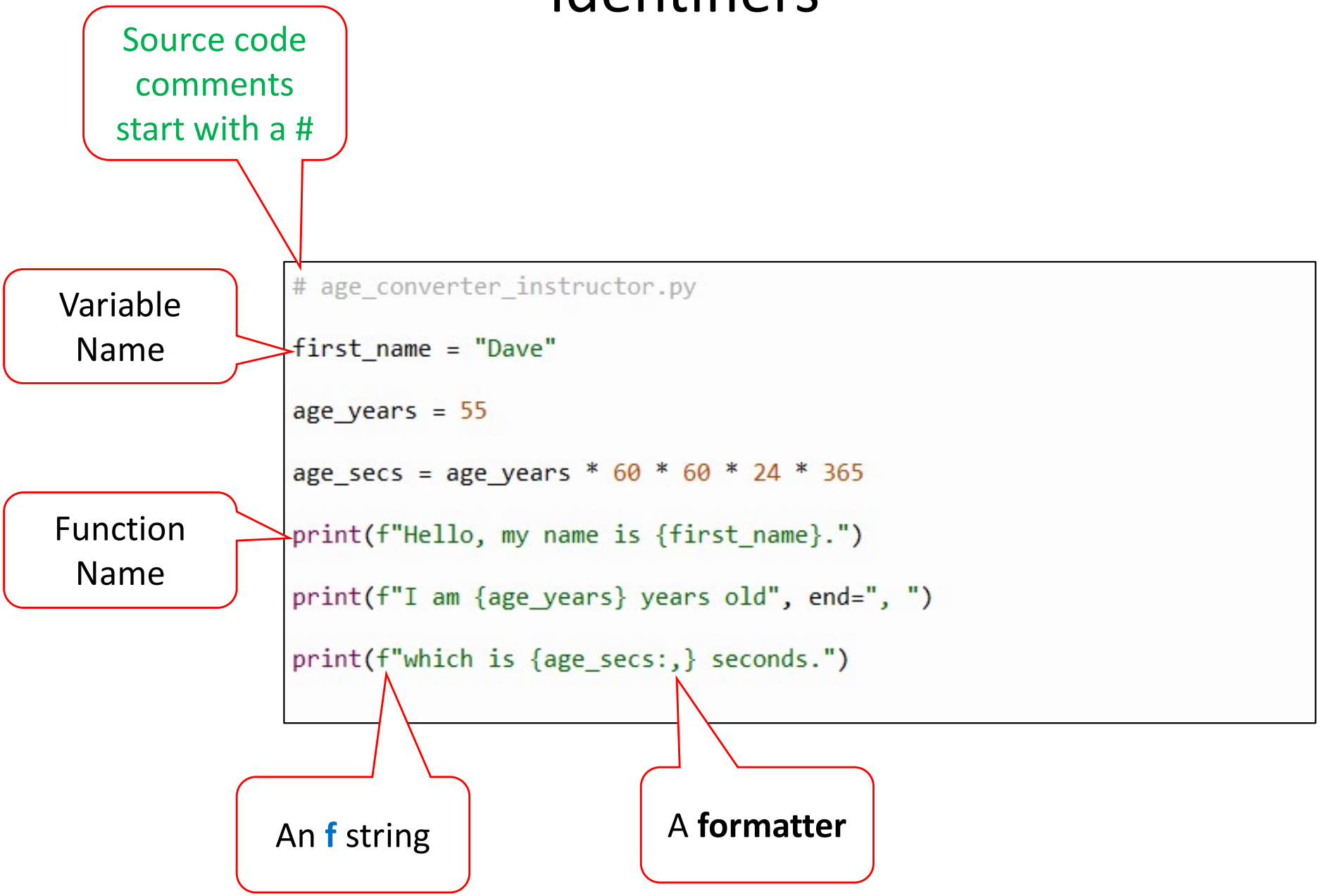
Restart the Thonny IDE to use a plug-in



Identifiers

- Identifiers are just **names** – everything in code has a name
  - Names must be < 64 chars in length
  - They can include upper- or lower-case letters & numbers
  - Identifiers must start with a letter and **cannot contain spaces!**
- Three types of identifier "casing"
  1. CamelCaseEachWord (first letter is Capitalized)
  2. camelCaseEachWord (first letter is not capitalized)
  3. all **lower_case** with underscores **(Snake case in Python!)** ✓
- Identifiers in Python are **case sensitive!!**
  - ***x*** *is not the same as* ***X***
  - Use ALLCAPS to define global constants (very rare)

Identifiers



Variable Types

- **Variables** store data in memory to be used later
  - Variables can be called whatever you want
  - Pick variable names that mean something to a human
  - Use **snake_case** (all lower case, underscores to break words)
- Python supports many built-in **data types** for variables:
  - **int** = Stores integers only
  - **float** = Stores real numbers with 15 digits of precision
  - **bool** = Stores **True** or **False** (Boolean logic)
  - **str** = Stores zero or more letters & numbers
- Python mostly "infers" the type of a variable

Displaying Variables

- The **print**() function is used to display the value of variables
  - When running inside Thonny, the **print**() output will show up in the "Shell" terminal window at the bottom of screen
  - String *literals* must be enclosed in quotation marks (single or double)
- Python can substitute a variable's value into a **placeholder**
  - To make a **placeholder** (aka a *replacement field*) you enclose the variable name between curly braces **{}**
  - Substituting a variable's actual value into its *replacement field* is called *string interpolation*
- Placeholders can also contain **format specifiers**
  - You can specify number of digits to the right of the decimal, etc.
  - You can specify left/right/center justification, column width, etc.

print() and **f-strings**

```
for fahrenheit in range(-44, 217, 4):
    celsius = (fahrenheit - 32) * 5 / 9
    print(f"{fahrenheit:>6.2f} F = {celsius:>6.2f} C")
```

Placing a lowercase **f** before the first quote in a **print**() statement indicates you will use some **placeholders**

A placeholder contains the variable's name sandwiched between curly **braces** {}

A **colon** after the variable's name starts a **format specifier**

Some Common Format Specifiers

:<	Left aligns the result (within the available space)
:>	Right aligns the result (within the available space)
:^	Center aligns the result (within the available space)
:,	Use a comma as a thousand separator
:f	Fix point number format
:n	Number format
:%	Percentage format

Using format specifiers is optional but makes your output more professional

print() and **f-strings**

```
for fahrenheit in range(-44, 217, 4):
    celsius = (fahrenheit - 32) * 5 / 9
    print(f"{fahrenheit:>6.2f} F = {celsius:>6.2f} C")
```

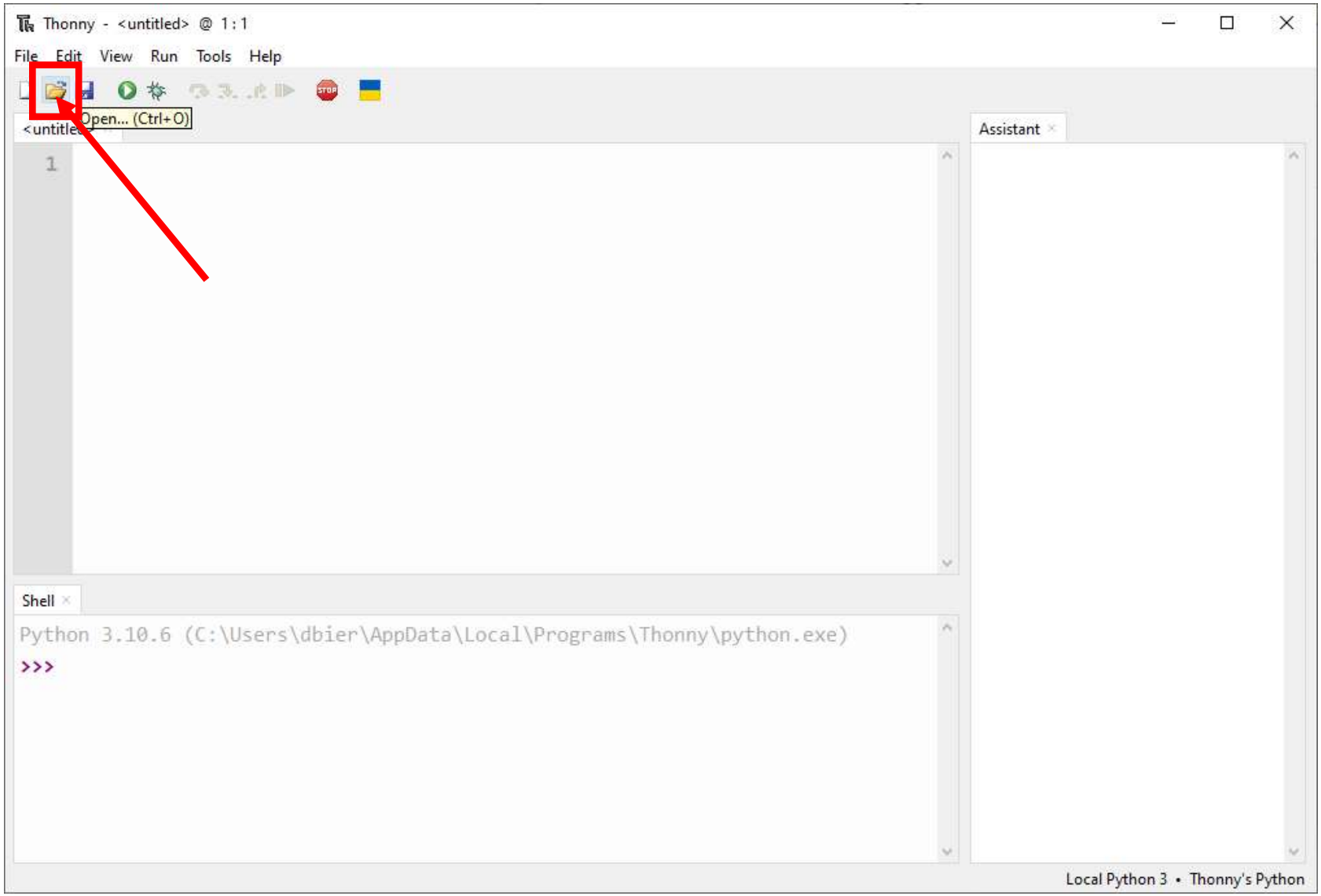
Print the current value of the variable **fahrenheit** right-justified (**>**) in a column six characters wide (**6**) and round to two digits (**.2f**) to the right of the decimal

Mathematical Operators

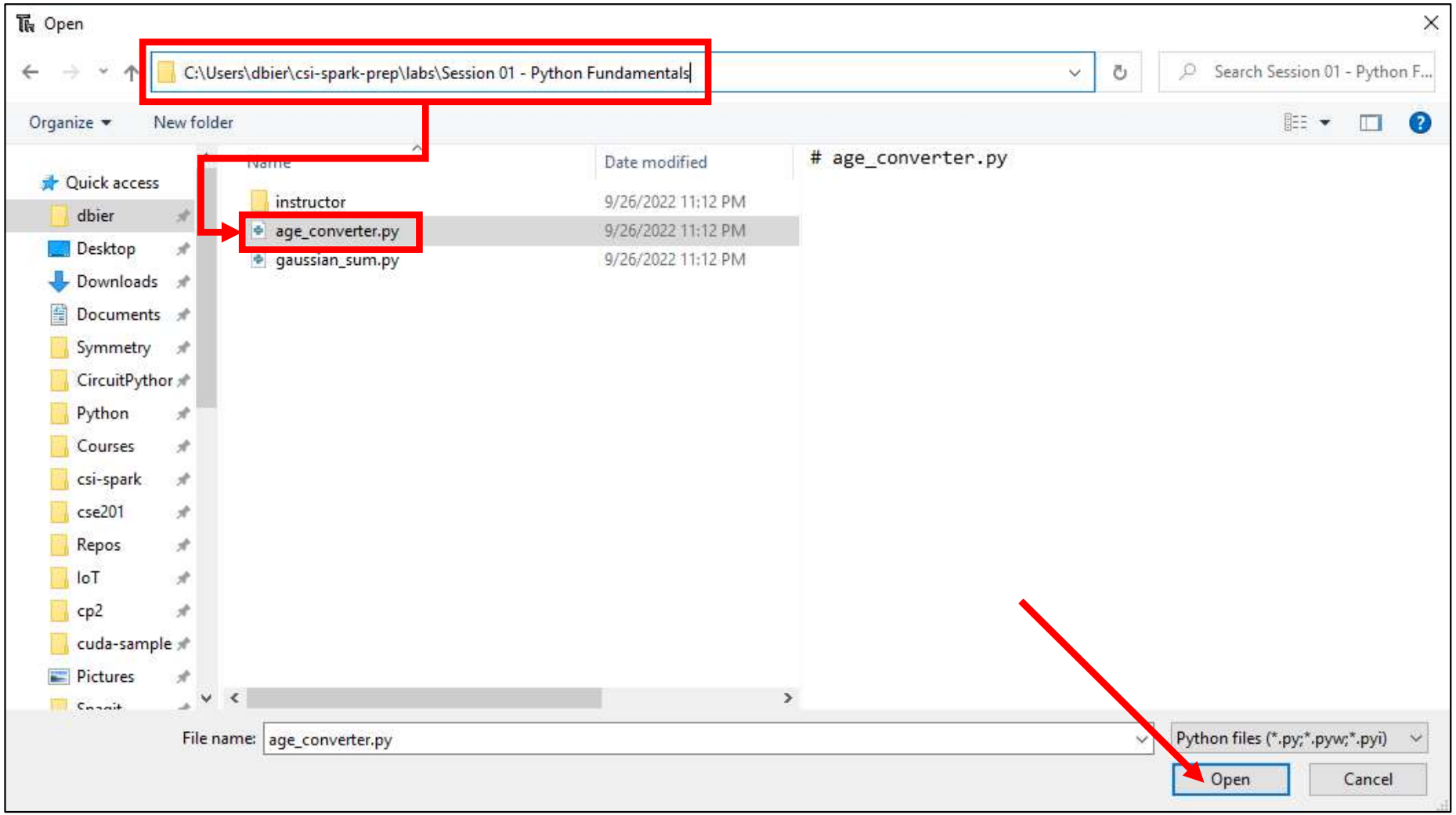
- Python operators obey normal **PEMDAS** precedence
  - Expressions are evaluated left to right in your source code
  - Use **=** to assign a value to a variable
  - Use ***** for multiplication and **/** for division operators
  - Use parenthesis to explicit specify the order of operations
  - The “greater than or equal to” operator is **>=**
  - The “less than or equal to” operator is **<=**

```
celsius = (fahrenheit - 32) * 5 / 9
```

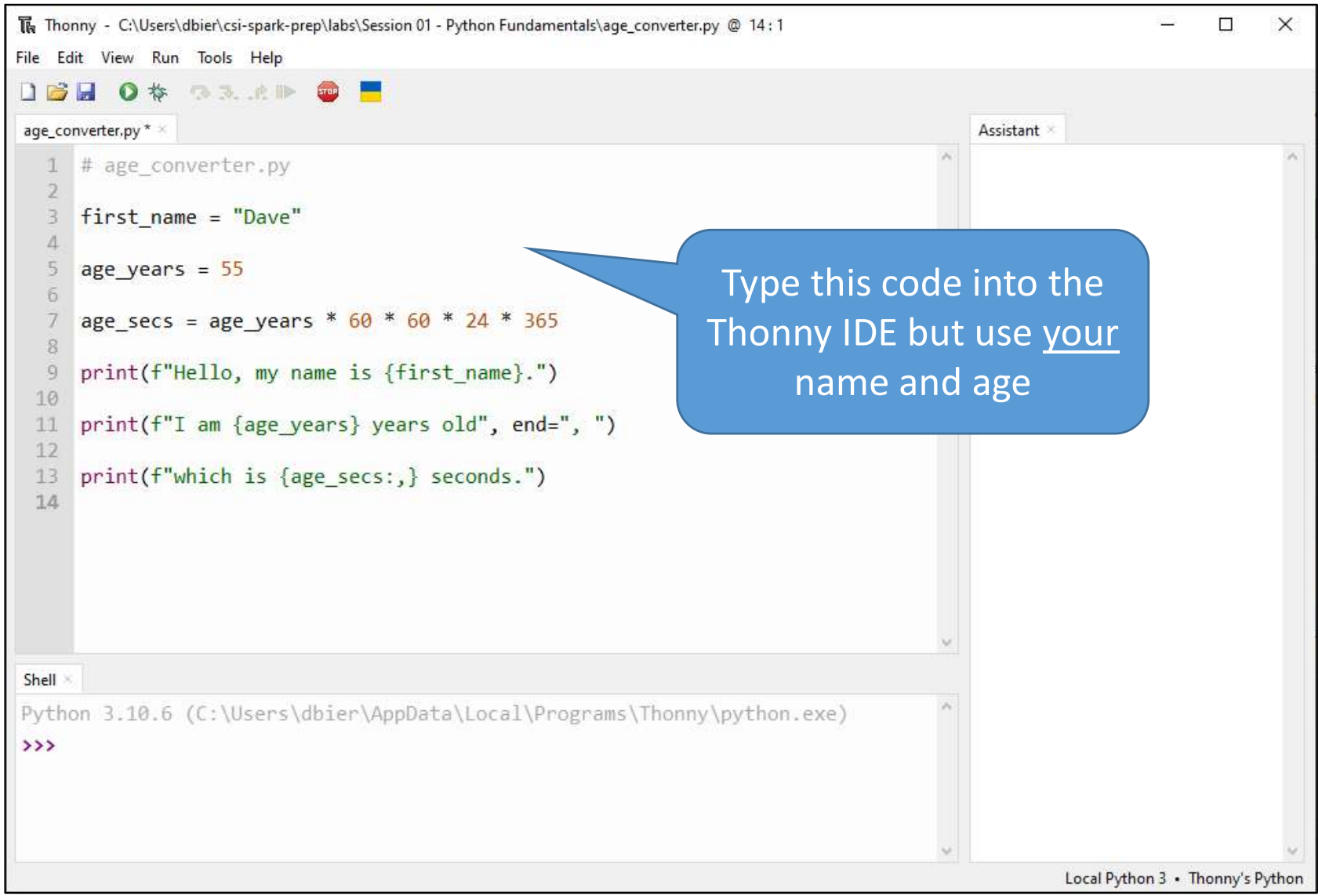
Open age_converter.py



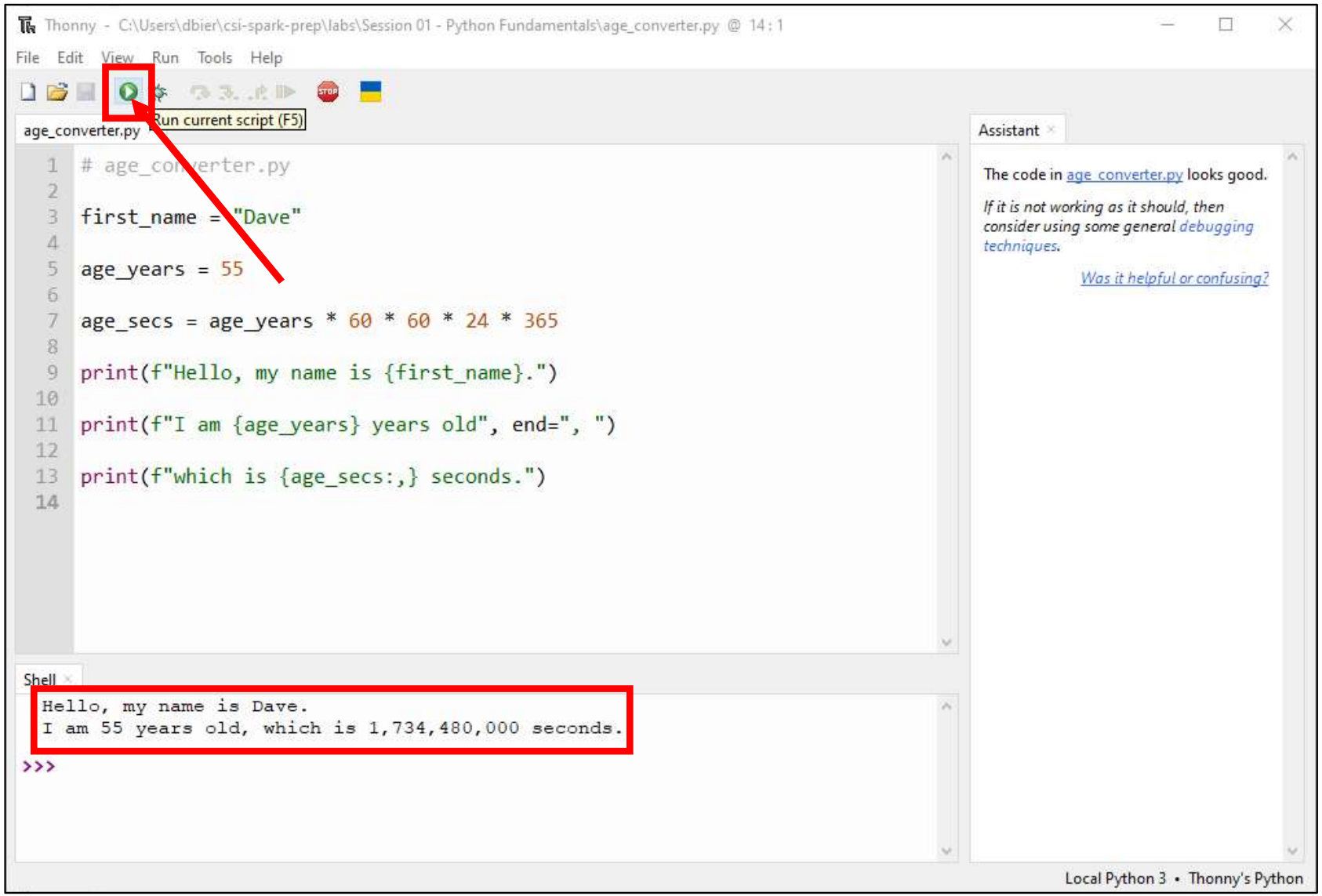
Open age_converter.py



Edit age_converter.py



Run age_converter.py



Statements & Scopes

- A **statement** does something (think: *sentence*)
  - A statement is either a declaration, a keyword, or a function
  - Statements are executed from top to bottom of a scope
- A **scope** contains one or more statements (think: *paragraph*)
  - A scope in Python begins with a **colon :**
  - Scopes are also denoted via **indentation**
  - All the statements in a scope **must** start at the same **column**
  - Scopes can be *nested* – each inner scope is further indented
  - Certain Python statements "introduce" (**require**) a new scope
  - In Python, white space is significant – indentation matters!

Statements and Scopes

```
# gaussian_sum.py

n = 1_000_000

sum = 0

for k in range(n + 1):
    sum = sum + k

print(f"Sum by looping = {sum:,}")

sum = n * (n + 1) // 2

print(f"Gaussian sum   = {sum:,}")
```

A **for** loop introduces a scope with a **colon :**

The statements in a **for** loop **scope** are all indented by the *same* amount

Identifiers *live* only within the scope in which they are defined
(except for globals & closures)

for loops

- A **for** loop executes all the statements within its scope for **each** item in the **list** passed *into* the **for**
- It is common to use the **range() function** to describe the list of numbers passed into the **for** statement
- The **range()** function takes three parameters: (**start**, **stop**, **step**)
  - The **stop** value is required but the **start** and **step** values are *optional*
  - The *default* value (if unspecified) for **start** is **0** and for **step** is **1**
  - The range is inclusive, *exclusive*: **[start, stop)**

```
for n in range(10):
    print(n, end=' ')
```

⇒

```
0 1 2 3 4 5 6 7 8 9
```

A Shortcut

Carl Friedrich Gauss
(1777 – 1855)

Sum the integers from 1 to 100

1
2
3
4
5
6
7
8
9
10

1	9
2	8
3	7
4	6
5	
10	

n = 10
4 matched rows that each sum to 10

1 row that is = 10 / 2 = 5

1 row that is = n = 10

$$n\left(\frac{n}{2}-1\right)+\frac{n}{2}+n \qquad = \frac{n*(n+1)}{2}$$

= 55

Another Shortcut

Sum of first **n** *natural* numbers:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2},$$

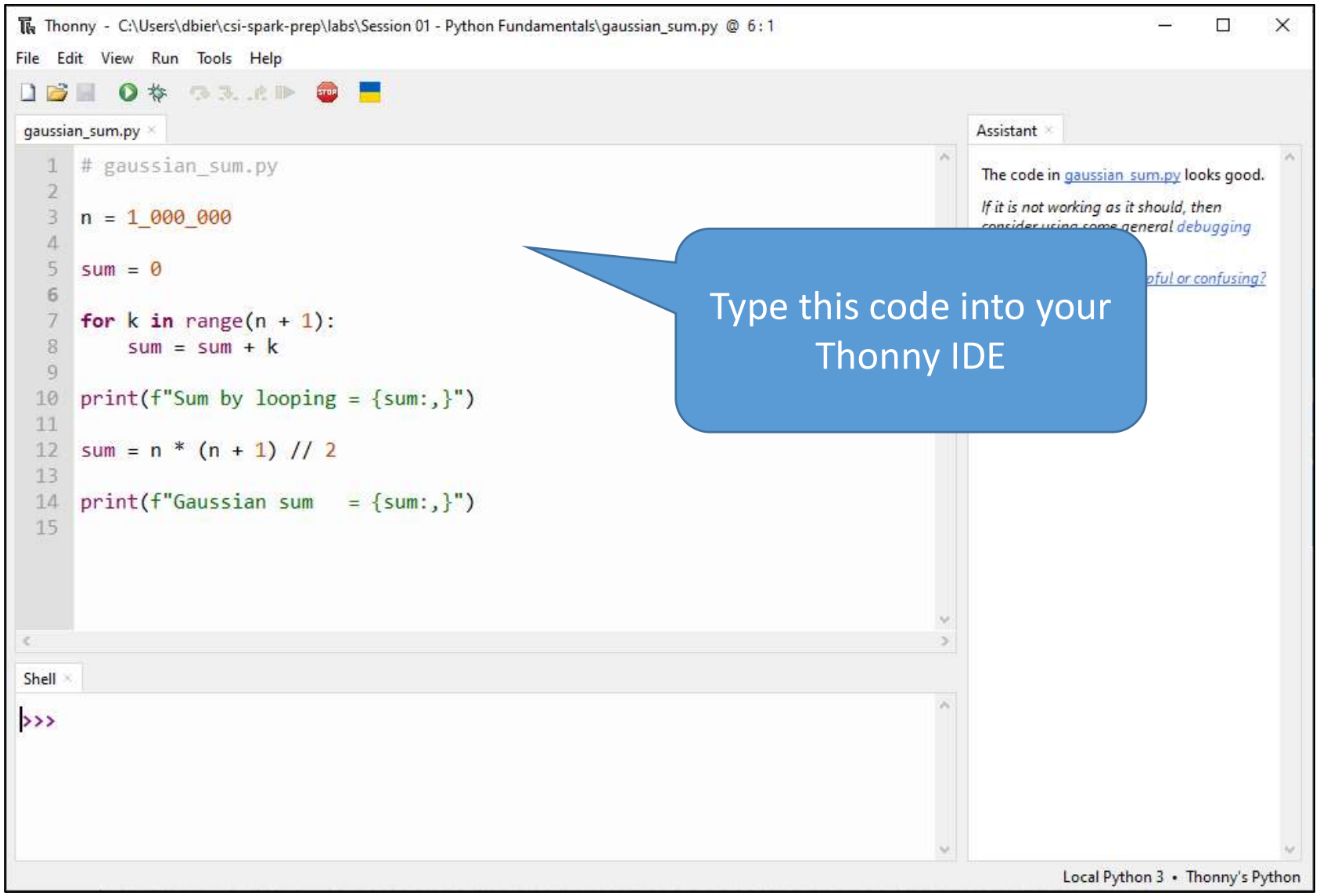
Sum of squares of first **n** *natural* numbers:

n	n^2	Sum
1	1	1
2	4	5
3	9	14
4	16	30
5	25	55
6	36	91
7	49	140
8	64	204
9	81	285
10	100	385

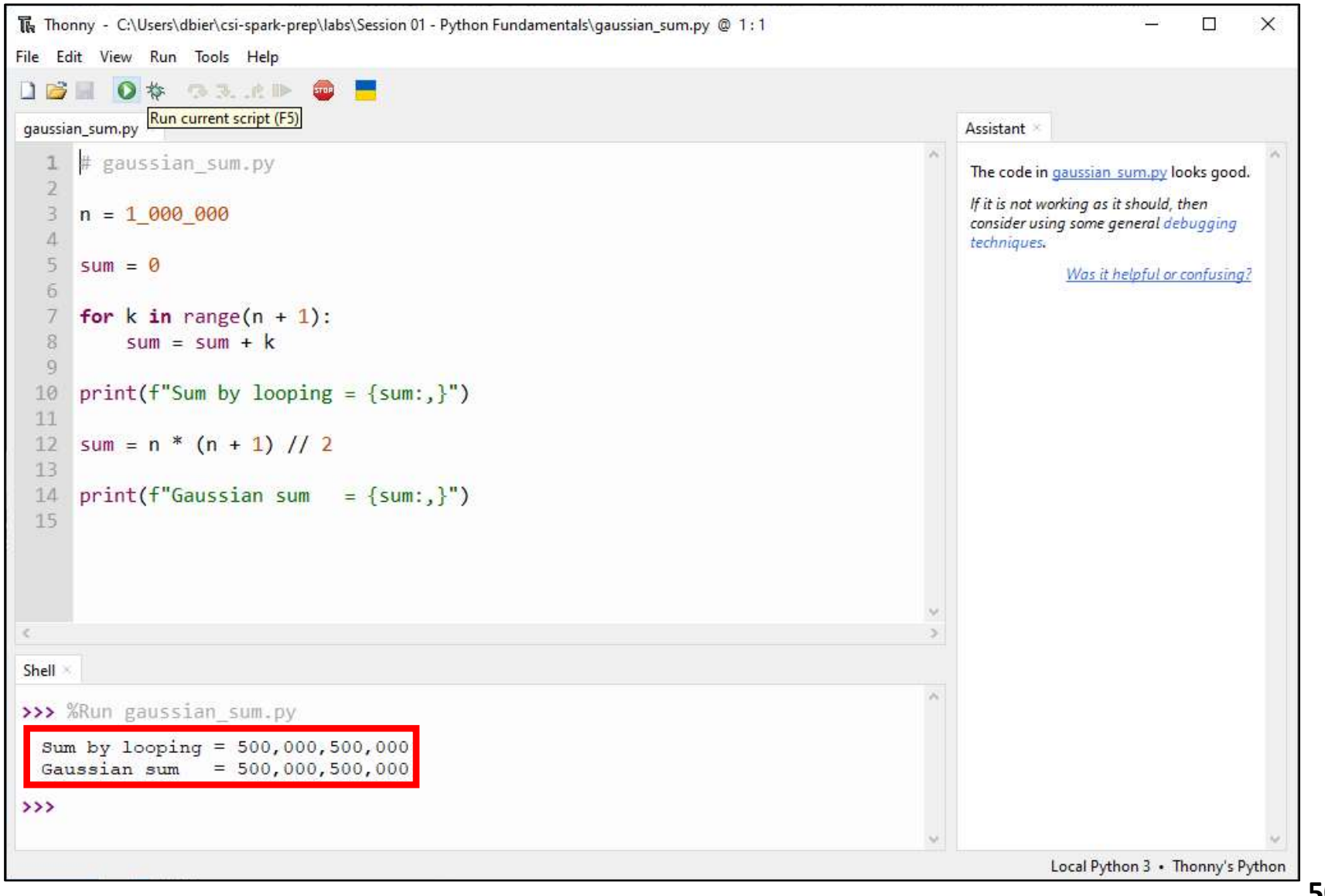
$$P_n = \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} = \frac{2n^3 + 3n^2 + n}{6}.$$

These are functional equations - we can now calculate the sums *immediately* without having to loop over every element!

Edit gaussian_sum.py



Run gaussian_sum.py



Session **01** – Now You Know…

- How to install packages and plug-ins into the Thonny IDE
- How to define variables in Python (**snake_case**)
- How to use **print**() to show variable values on screen
- How to use **f-strings** with **{}** placeholders (replacement fields) for variable names
- How to indicate **format specifiers** in order to make your code output more human readable and professional looking
- How to pass the lazy list generated by **range()** into the **for** statement to enumerate each value in that sequence

TASK 01-01

- Create a console (terminal/text mode) Python program called **celsius_to_fahrenheit.py** that converts a range of temperatures in the Celsius scale to the equivalent temperatures in the Fahrenheit scale
- The program should display all temperatures between $-44°$C and $106°$C inclusive in steps of $4°$C and the corresponding temperature in Fahrenheit
- Each C/F pair should be displayed on its own output line, with two digits to the right of the decimal for each temperature scale
- Verify the correctness of your program by checking your values for $-40°$C, $0°$C, and $100°$C