



## Lecture 5 – Modules, Namespaces

<https://cw.fel.cvut.cz/wiki/courses/be5b33prg/start>

# Michal Reinštein

Czech Technical University in Prague,  
Faculty of Electrical Engineering, Dept. of Cybernetics,  
Center for Machine Perception

<http://cmp.felk.cvut.cz/~reinsmic/>  
[reinstein.michal@fel.cvut.cz](mailto:reinstein.michal@fel.cvut.cz)



# PLAGIARISM WARNING

[https://cw.fel.cvut.cz/wiki/help/common/plagiarism\\_cheating](https://cw.fel.cvut.cz/wiki/help/common/plagiarism_cheating)

05\_right  
EQD|PIA  
27.10.2017:

3

0

1

3

1

3

3

3

+++

3

3

3

0

3

3

3

+++

3

3

3

3

3

3

3

---

## Rules

These are the general rules for most courses using this CourseWare and UploadSystem. If you have any doubts, ask a lecturer.

- Code or text (from hereafter, only code) for each assignment has to be done independently. In programming, it is normal to (re)use someone else's code, however, for study purposes this is undesirable. <sup>1)</sup> If you are to use someone else's code, it will be specifically allowed in the assignment.
- Turned in work are checked for plagiarism. Any code, which is either a complete copy or which after simple mechanical syntax modification is identical to code of someone else, is a plagiarism.
- In case you turn in a plagiarism, you still have to turn in a original solution in original time, however, you receive zero points.
- In case you turn in copied code repeatedly<sup>2)</sup>, you are automatically classified with an F.
- Lecturer does not guarantee he will be able to differentiate between original work and plagiarism.
- Student who does not agree with his evaluation or classification can proceed according to [ČVUT Study and Examination process](#).

Table of Contents	
·Rules	
·What is a plagiarism	
·Practical notes	

## What is a plagiarism

**Plagiarism** is a text, code, illustration, method or idea taken from another's work without citing your sources.

Everyone who uses someone work or results of another, has to indicate he did so and reference the original author. Otherwise it is plagiarism. For study reasons, using other's work is also forbidden in some courses, even if it was properly labeled.

## Practical notes

In most assignments, you will primarily have to avoid using someone else's code. If you find a code (i.e. from a colleague or on the internet), which solves your assignment, try to wait for a day or two, before starting your own work, to limit the volume of copied work.

Students, we strongly appeal to your professional honor. Copying and Cheating are serious offences against academical ethics. If you are at your wit's end, ask lecturers for help. They will gladly provide you with the directions to the solution.

<sup>1)</sup> If you study someone else's code, try to wait for at least 24 hours, before you start writing your own implementation.

<sup>2)</sup> This means even across your courses, not only in the same course

source [https://cw.fel.cvut.cz/wiki/help/common/plagiarism\\_cheating](https://cw.fel.cvut.cz/wiki/help/common/plagiarism_cheating)



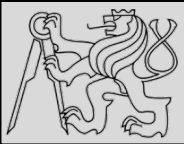
- The methods and variables are created on **stack memory**
- The objects and instances are created on **heap memory**
- New **stack frame** is created on invocation of a function / method
- Stack frames are destroyed as soon as the function / method returns
- Mechanism to clean up the dead objects is **Garbage collector**
- Everything in Python is **object**
- Python is **dynamically typed** language



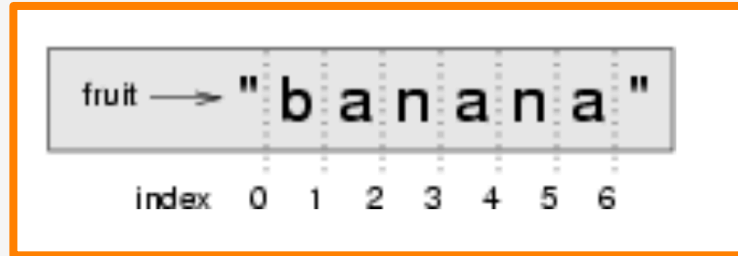
```
>>> my_string = "TEST"
>>> my_string[2] = "X"
Traceback (most recent call last):
  File "<interactive input>", line 1, in <module>
TypeError: 'str' object does not support item assignment
```

```
>>> my_list = ["T", "E", "S", "T"]
>>> my_list[2] = "X"
>>> my_list
['T', 'E', 'X', 'T']
```

- Lists are **mutable** (we can change their elements)
- Strings are **immutable** (we cannot change their elements)
- Use **slicing principles** (indexes in between characters / items)



```
>>> s = "Pirates of the Caribbean"
>>> print(s[0:7])
Pirates
>>> print(s[11:14])
the
>>> print(s[15:24])
Caribbean
>>> friends = ["Joe", "Zoe", "Brad", "Angelina", "Zuki", "Thandi", "Paris"]
>>> print(friends[2:4])
['Brad', 'Angelina']
```



- A **substring** of a string is obtained by taking a **slice**
- Slice a list to refer to some **sublist** of the items in the list
- The operator **[n:m]** returns the part of the string from the *n*'th character to the *m*'th character, **including the first but excluding the last** (*indices pointing between the characters*)
- Slice operator [n:m] **copies** out the part of the paper between the **n** and **m** positions
- Result of **[n:m]** will be of **length (m-n)**



## Strings

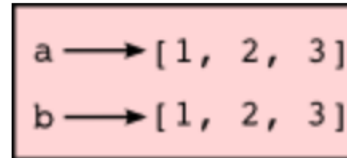
```
1 a = "banana"  
2 b = "banana"
```



```
>>> a is b  
True
```

## Lists

```
>>> a = [1, 2, 3]  
>>> b = [1, 2, 3]  
>>> a == b  
True  
>>> a is b  
False
```



- Variables **a** and **b** refer to string object with letters "banana"
- Use **is** operator or **id** function to find out the **reference**
- Strings are **immutable**, *Python optimizes resources by making two names that refer to the same string value refer to the same object*
- Not the case of lists: **a** and **b** have the same value (content) but do not refer to the same object



# RECAP: LISTS – ALIASING, CLONING



```
>>> a = [1, 2, 3]
>>> b = a
>>> a is b
True
```

```
>>> b[0] = 5
>>> a
[5, 2, 3]
```

```
>>> a = [1, 2, 3]
>>> b = a[:]
>>> b
[1, 2, 3]
```

```
>>> b[0] = 5
>>> a
[1, 2, 3]
```

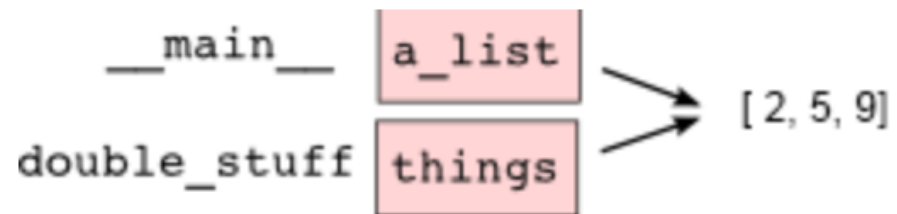
- If we assign one variable to another, both variables refer to the same object
- The **same list has two different names** we say that it is **aliased** (changes made with one alias affect the other)
- RECOMMENDATION: avoid aliasing when you are working with mutable objects
- If need to modify a list and keep a copy of the original use the **slice operator** (taking any slice of a creates a new list)



```
1 def double_stuff(a_list):  
2     """ Overwrite each element in a_list with double its value. """  
3     for (idx, val) in enumerate(a_list):  
4         a_list[idx] = 2 * val
```

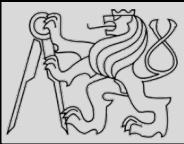
```
1 things = [2, 5, 9]  
2 double_stuff(things)  
3 print(things)
```

[4, 10, 18]

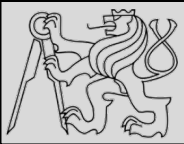


- Passing a **list as an argument** passes a **reference** to the list, **not a copy or clone** of the list
- So parameter passing creates an **alias**





- A **module** is a file containing Python definitions and statements intended for use in other Python programs
- Modules are one of the main **abstraction layers** available and probably the most natural one
- Abstraction layers allow **separating code** into parts holding **related data** and **functionality**
- The most natural way to separate these two layers is to regroup **all interfacing functionality into one file**, and all **low-level operations in another file**
- Done with the **import** and **from... import** statements



## Where does Python look for imported modules?

- A list named **path** in module `sys` contains all the locations (and can be modified)
- The first item on the list is the directory with the program

## What happens when Python imports a module?

- Python searches the module among already imported modules in **sys.modules**
- If the module is not found in **sys.modules**, Python searches locations in **sys.path**, executes the module once it is found, and records that the module has already been imported
- Python creates names in local **namespace** for the imported module, or for all the variables, functions, etc. imported from the module



```
1 if __name__ == "__main__":  
2     print("This won't run if I'm imported.")
```

- Modules that can be both **run** and **imported**
- Special variable called **\_\_name\_\_** inside each module it contains:
  - the **name of the module** (of the *.py* file) when the module is imported
  - string **"\_\_main\_\_"** when the *.py* file is run as a program (script)
- Variable **\_\_name\_\_** is defined in both the calling namespace and in the namespace of the module



Very bad

```
[...]  
from modu import *  
[...]  
x = sqrt(4) # Is sqrt part of modu? A builtin? Defined above?
```

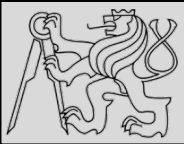
Better

```
from modu import sqrt  
[...]  
x = sqrt(4) # sqrt may be part of modu, if not redefined in between
```

Best

```
import modu  
[...]  
x = modu.sqrt(4) # sqrt is visibly part of modu's namespace
```

- To gain access to symbols defined in a module (i.e. in a different namespace) module has to be imported (3 ways)
- The statements **load a module**, **create a name** in the current namespace, and **bind the name to the loaded module**



```
1 import random
2
3 # Create a black box object that generates random numbers
4 rng = random.Random()
5
6 dice_throw = rng.randrange(1,7) # Return an int, one of 1,2,3,4,5,6
7 delay_in_seconds = rng.random() * 5.0
```

- To include **random decision-making process**
- To take **samples from probability distributions**
- To play a game of chance where the computer needs to throw some dice, pick a number, or flip a coin ...
- To shuffle a deck of playing cards randomly ...
- In **modelling and simulations**: weather models, environmental models, Monte Carlo method
- For encrypting banking sessions on the Internet



```
1 import random
2
3 # Create a black box object that generates random numbers
4 rng = random.Random()
5
6 dice_throw = rng.randrange(1,7) # Return an int, one of 1,2,3,4,5,6
7 delay_in_seconds = rng.random() * 5.0
```

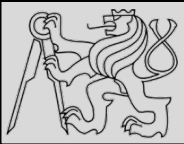
- The **random** method returns a *floating point number* in the interval  $[0.0, 1.0)$  — *the square bracket means “closed interval on the left” and the round parenthesis means “open interval on the right”* – 0.0 is possible, but all returned numbers will be *strictly less than 1.0*.
- It is usual to **scale the results** after calling this method to get them into an interval suitable for application.
- EXAMPLE: scaling to a number in the interval  $[0.0, 5.0)$  (*uniformly distributed numbers — numbers close to 0 are just as likely to occur as numbers close to 0.5 or close to 1.0*)



```
1 r_odd = rng.randrange(1, 100, 2)
```

- The **randrange** method generates an integer between its **lower** and **upper** argument
- The **randrange** method same semantics as range (so the lower bound is included, but the upper bound is excluded)
- All the values have an **equal probability** of occurring (i.e. the results are uniformly distributed).
- Randrange also takes an optional **step argument** (like range)
- EXAMPLE: We needed a random odd number less than 100

source <http://openbookproject.net/thinkcs/python/english3e/modules.html>



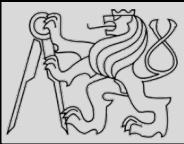
```
1 cards = list(range(52)) # Generate ints [0 .. 51]
2                               #     representing a pack of cards.
3 rng.shuffle(cards)      # Shuffle the pack
```

This example shows how to **shuffle** a list.  
(shuffle cannot work directly with range object so list type converter first is necessary)

```
1 drng = random.Random(123) # Create generator with known starting state
```

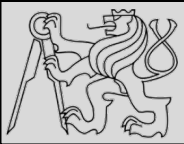
- Random number generators are based on a deterministic algorithm — **repeatable** and **predictable**
- Called **pseudo-random generators** (not genuinely random)
- Each time you ask for another random number, you'll get one based on the current **seed attribute**, and the **state of the seed**





```
1 drng = random.Random(123) # Create generator with known starting state
```

- Repeatability for **debugging** and for writing **unit tests** (*programs that do the same thing every time they are run*)
- Forcing the random number generator to be initialized with a **known seed** every time (*often this is only wanted during testing / back-testing*)
- Without this seed argument, the system probably uses something based on the OS time.



```
1 import random
2
3 def make_random_ints(num, lower_bound, upper_bound):
4     """
5     Generate a list containing num random ints between lower_bound
6     and upper_bound. upper_bound is an open bound.
7     """
8     rng = random.Random() # Create a random number generator
9     result = []
10    for i in range(num):
11        result.append(rng.randrange(lower_bound, upper_bound))
12    return result
```

```
>>> make_random_ints(5, 1, 13) # Pick 5 random month numbers
[8, 1, 8, 5, 6]
```

- EXAMPLE: generate a list containing  $n$  random ints between a lower and upper bound
- NOTE: that we got a **duplicates** in the result (often this is wanted, e.g. if we throw a die five times, we would expect some duplicates)



## How to take care of duplicates?

- If you wanted **5 distinct months**, then this algorithm is wrong
- In this case a good algorithm is to generate the list of possibilities, **shuffle** it, and **slice off** the number of elements you want:

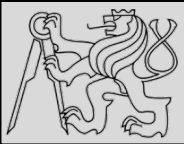
```
1 xs = list(range(1,13)) # Make list 1..12 (there are no duplicates)
2 rng = random.Random() # Make a random number generator
3 rng.shuffle(xs)       # Shuffle the list
4 result = xs[:5]      # Take the first five elements
```

source <http://openbookproject.net/thinkcs/python/english3e/modules.html>



- Allowing **duplicates** (usually described as pulling balls out of a bag with replacement)
- No **duplicates** (usually described as pulling balls out of the bag without replacement)
- Algorithm **“shuffle and slice”** is not ideal for case of choosing few elements from a very large domain

*(Suppose the need for five numbers between 1 and 10 million, without duplicates. Generating a list of ten million items, shuffling it, and then slicing off the first five would be a performance disaster.)*



```
1 import random
2
3 def make_random_ints_no_dups(num, lower_bound, upper_bound):
4     """
5     Generate a list containing num random ints between
6     lower_bound and upper_bound. upper_bound is an open bound.
7     The result list cannot contain duplicates.
8     """
9     result = []
10    rng = random.Random()
11    for i in range(num):
12        while True:
13            candidate = rng.randrange(lower_bound, upper_bound)
14            if candidate not in result:
15                break
16            result.append(candidate)
17    return result
18
19 xs = make_random_ints_no_dups(5, 1, 10000000)
20 print(xs)
```

This agreeably produces 5 random numbers, without duplicates:

```
[3344629, 1735163, 9433892, 1081511, 4923270]
```

**Choose wisely the algorithm based on the input data!**



- constant =  $\mathcal{O}(1)$
- logarithmic =  $\mathcal{O}(\log n)$
- linear =  $\mathcal{O}(n)$
- $n \log n$  =  $\mathcal{O}(n \log n)$
- quadratic =  $\mathcal{O}(n^2)$
- cubic =  $\mathcal{O}(n^3)$
- polynomial =  $\mathcal{O}(n^k)$
- exponential =  $\mathcal{O}(k^n)$
- factorial =  $\mathcal{O}(n!)$

## How efficient and reliable is our code?

- One way to experiment is to time **how long** various operations take and what the **memory requirements** are (*related to algorithm complexity: <https://people.duke.edu/~ccc14/sta-663/AlgorithmicComplexity.html>* )
- The **time** module has a function **clock** that is recommended
- Whenever clock is called, it returns a **floating point number** representing *how many seconds have elapsed since your program started running (varies according to OS!)*



```
1 import time
2
3 def do_my_sum(xs):
4     sum = 0
5     for v in xs:
6         sum += v
7     return sum
8
9 sz = 10000000      # Lets have 10 million elements in the list
10 testdata = range(sz)
11
12 t0 = time.clock()
13 my_result = do_my_sum(testdata)
14 t1 = time.clock()
15 print("my_result      = {0} (time taken = {1:.4f} seconds)"
16       .format(my_result, t1-t0))
17
18 t2 = time.clock()
19 their_result = sum(testdata)
20 t3 = time.clock()
21 print("their_result = {0} (time taken = {1:.4f} seconds)"
22       .format(their_result, t3-t2))
```

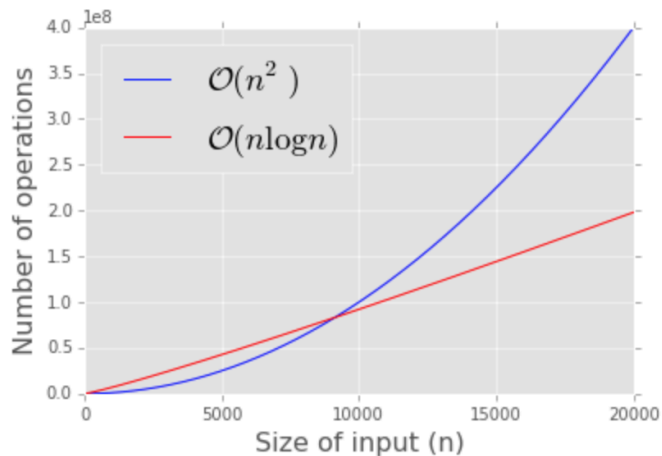
```
my_sum      = 49999995000000 (time taken = 1.5567 seconds)
their_sum   = 49999995000000 (time taken = 0.9897 seconds)
```

- EXAMPLE: Generating and summing up ten million elements in under a second
- Proprietary function runs **57% slower** than the **built-in** one.



## Comparing complexity of $\mathcal{O}(n^2)$ (e.g. bubble sort) and $\mathcal{O}(n \log n)$ (e.g. merge sort).

```
def f1(n, k):  
    return k*n*n  
  
def f2(n, k):  
    return k*n*np.log(n)  
  
n = np.arange(0, 20001)  
  
plt.plot(n, f1(n, 1), c='blue')  
plt.plot(n, f2(n, 1000), c='red')  
plt.xlabel('Size of input (n)', fontsize=16)  
plt.ylabel('Number of operations', fontsize=16)  
plt.legend([' $\mathcal{O}(n^2)$ ', ' $\mathcal{O}(n \log n)$ '], loc='best', fontsize=20);
```

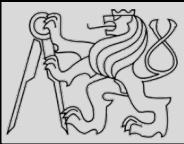


```
# Notice how much overhead Python objects have  
# A raw integer should be 64 bits or 8 bytes only  
  
print sys.getsizeof(1)  
print sys.getsizeof(1234567890123456789012345678901234567890)  
print sys.getsizeof(3.14)  
print sys.getsizeof(3j)  
print sys.getsizeof('a')  
print sys.getsizeof('hello world')
```

```
24  
44  
24  
32  
38  
48
```

source <https://people.duke.edu/~ccc14/sta-663/AlgorithmicComplexity.html>

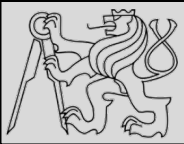




```
>>> import math
>>> math.pi           # Constant pi
3.141592653589793
>>> math.e           # Constant natural Log base
2.718281828459045
>>> math.sqrt(2.0)   # Square root function
1.4142135623730951
>>> math.radians(90) # Convert 90 degrees to radians
1.5707963267948966
>>> math.sin(math.radians(90)) # Find sin of 90 degrees
1.0
>>> math.asin(1.0) * 2 # Double the arcsin of 1.0 to get pi
3.141592653589793
```

- The **math** module contains mathematical functions typically found on a calculator, including mathematical constants like pi and e
- Functions **radians** and **degrees** to convert angles
- Mathematical functions are **pure** and do not have any state

source <http://openbookproject.net/thinkcs/python/english3e/modules.html>



```
1 def remove_at(pos, seq):  
2     return seq[:pos] + seq[pos+1:]
```

```
>>> import seqtools  
>>> s = "A string!"  
>>> seqtools.remove_at(4, s)  
'A sting!'
```

- Create own **modules** – save script as a file with **.py extension**
- EXAMPLE: function `remove_at` in a script is saved as a file named `seqtools.py`
- The module must be first imported before use (**.py** is the file extension and is not included in the import statement)
- RECOMMENDATION: break up very large programs into manageable sized parts and keep related parts together



- Namespace is a **mapping** from names to objects
- Namespace is a **collection of identifiers** that belong to a module, function, or a class
- Namespace is **set of symbols** used to **organize objects** of various kinds so that we can refer to them by name
- Namespaces permit programmers to work on the same project without having **naming collisions** (allow name reuse)
- Often **hierarchically** structured
- Each name must be **unique in its namespace**
- Namespace is very general concept not limited to Python
- Each **module has its own namespace** – *we can use the same identifier name in multiple modules without causing an identification problem*

source <http://openbookproject.net/thinkcs/python/english3e/modules.html>

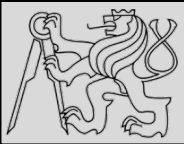


## How are namespaces defined in Python?

- Packages (*collections of related modules*)
- Modules (*.py files containing definitions of functions, classes, variables, etc.*)
- Classes, Functions ...

## What is the difference between programs and modules?

- Both are stored in **.py files**.
- **Programs** (scripts) are designed to be run (executed)
- **Modules** (libraries) are designed to be imported and used by other programs and other modules
- **Special case**: .py file is designed to be both a program and a module (*it can be executed as well as imported to provide functionality for other modules*)



```
1 # Module1.py
2
3 question = "What is the meaning of Life, the Universe, and Everything?"
4 answer = 42
```

```
1 # Module2.py
2
3 question = "What is your quest?"
4 answer = "To seek the holy grail."
```

```
1 import module1
2 import module2
3
4 print(module1.question)
5 print(module2.question)
6 print(module1.answer)
7 print(module2.answer)
```

```
What is the meaning of Life, the Universe, and Everything?
What is your quest?
42
To seek the holy grail.
```



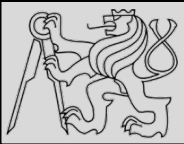
```
1 def f():
2     n = 7
3     print("printing n inside of f:", n)
4
5 def g():
6     n = 42
7     print("printing n inside of g:", n)
8
9 n = 11
10 print("printing n before calling f:", n)
11 f()
12 print("printing n after calling f:", n)
13 g()
14 print("printing n after calling g:", n)
```

```
printing n before calling f: 11
printing n inside of f: 7
printing n after calling f: 11
printing n inside of g: 42
printing n after calling g: 11
```

- Functions also have own namespaces
- Functions can read (**read-only**) variable in the outer scope
- EXAMPLE: the three **n**'s above do not collide since they are each in a **different namespace** — three names for three different variables

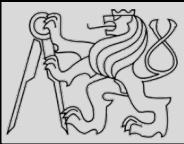


- Python has a convenient and simplifying one-to-one mapping: **one module per file** – giving rise to one namespace
- Python takes the **module name from the file name**, and this becomes the name of the namespace
- EXAMPLE: math.py is a **filename**, the module is called **math**, and its namespace is **math** (*in Python the concepts are more or less interchangeable*)
- In other languages (e.g. C#) one module can span multiple files, or one file to have multiple namespaces, or many files to all share the same namespace



- RECOMMENDATION: **keep the concepts distinct in your mind**
- Files and directories organize where code and data are stored
- Namespaces and modules are a **programming concepts**: help us organize how we want to group related functions and attributes.
- Namespaces are not about “where” to store things, and should not have to coincide with the file structures
- *If the file `math.py` is renamed, its module name needs to be changed, import statements need to be changed, and the code that refers to functions or attributes inside that namespace also needs to be changed accordingly*





- A **scope** is a textual region of a Python program where a namespace is directly accessible

## What types of scopes can be defined?

- **Local scope** refers to identifiers declared within a function (*these identifiers are kept in the namespace that belongs to the function, and each function has its own namespace*)
- **Global scope** refers to all the identifiers declared within the current module, or file
- **Built-in scope** refers to all the identifiers built into Python (*those like range and min that can be used without having to import anything*)



```
1 def range(n):  
2     return 123*n  
3  
4 print(range(10))
```

```
1 n = 10  
2 m = 3  
3 def f(n):  
4     m = 7  
5     return 2*n+m  
6  
7 print(f(5), n, m)
```

## What are the scope precedence rules?

- The same name can occur in more than one of these scopes, but the innermost, or local scope, will always take precedence over the global scope, and the global scope always gets used in preference to the built-in scope
- Names can be “**hidden**” from use if own variables or functions reuse those names
- EXAMPLE: variables **n** and **m** are created just for the duration of the execution of **f** since they are created in the local namespace of function **f** (precedence rules apply)



```
1 import math
2 x = math.sqrt(10)
```

```
1 from math import cos, sin, sqrt
2 x = sqrt(10)
```

```
1 def area(radius):
2     import math
3     return math.pi * radius * radius
4
5 x = math.sqrt(10)      # This gives an error
```

```
1 from math import *    # Import all the identifiers from math,
2                       # adding them to the current namespace.
3 x = sqrt(10)          # Use them without qualification.
```

```
1 >>> import math as m
2 >>> m.pi
3 3.141592653589793
```

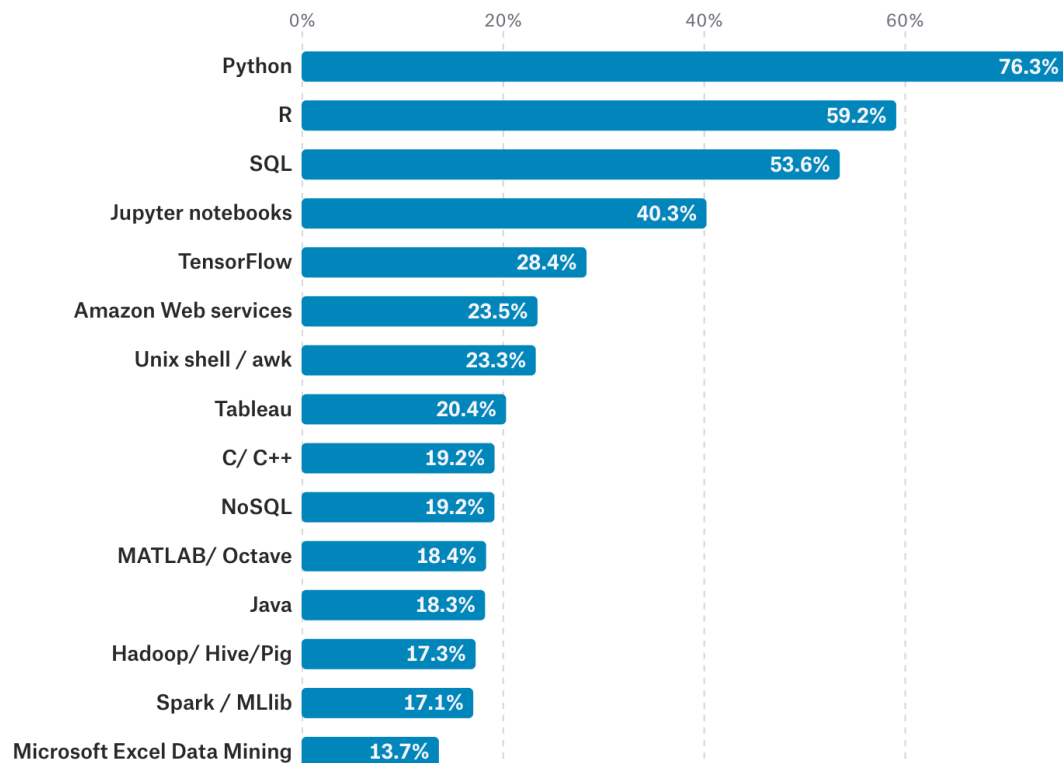
- Variables defined **inside a module** are called **attributes** of the module
- Attributes are accessed using the **dot** operator (.)
- When a dotted name is used it is often referred to it as a fully qualified name



## What tools are used at work?

Python was the most commonly used data analysis tool across employed data scientists overall, but more **Statisticians** are still loyal to R.

Company Size ▾ Industry ▾ Job Title ▾



7,955 responses

Only displaying the top 15 answers. There are 38 answers not shown.

source <https://www.kaggle.com/surveys/2017>

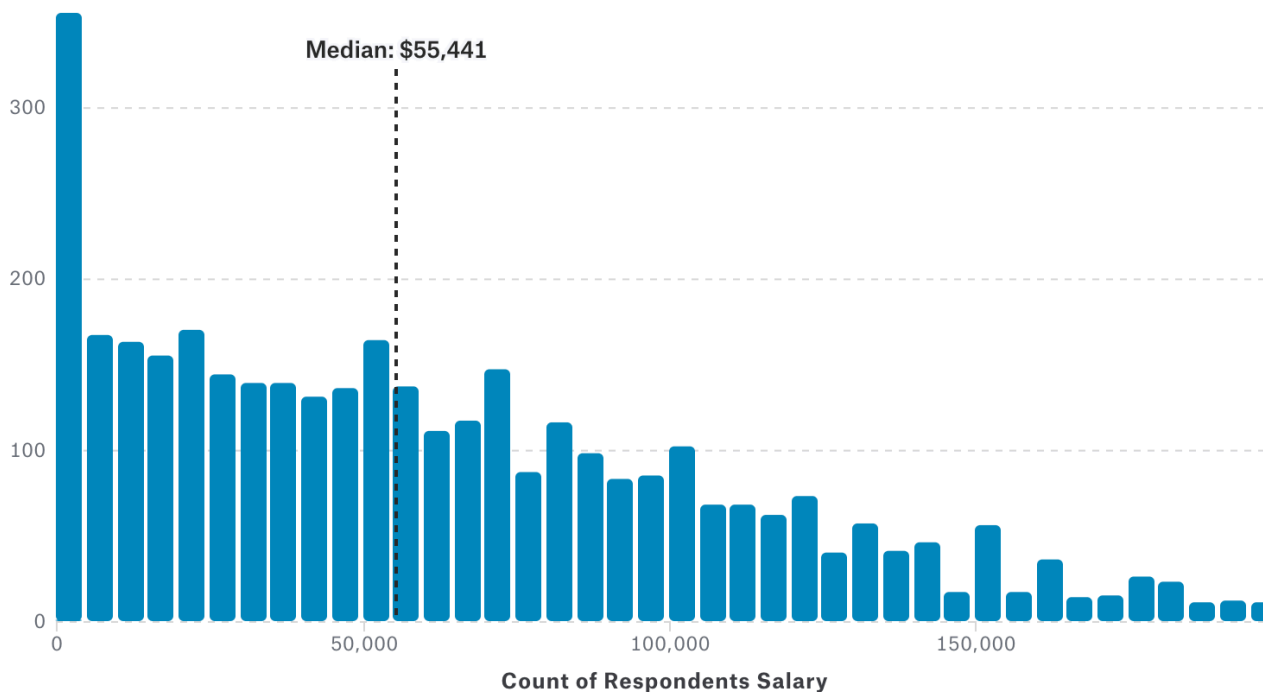


## What is your full-time annual salary?

Although “compensation and benefits” was ranked as slightly less important than “opportunities for professional development” in our survey, it’s still good to know what should be considered reasonable compensation. In the US,

**Machine Learning Engineers** bring home the most bacon (on average).

Country Experience Job Title **FILTER GENDER** All Female Male Other



3,771 responses

131 responses are not shown that exceed the maximum value shown, but they are included in the median calculation.

source <https://www.kaggle.com/surveys/2017>