

Learn to code for data analysis

Week 3 glossary

Concepts

A **conditional statement** is of the form

```
if condition1:
    statements1
elif condition2:
    statements2
...
else:
    statements
```

The computer evaluates the conditions from top to bottom and executes *only* the statements for the *first* condition that is true. If all conditions are false, it executes the `else` statements. If there is no `else` part nothing happens. The `elif` parts are optional too. Each block of statements must be indented, usually by four spaces.

A **constant** is a variable that is assigned only once, i.e. its initial value never changes. Constant names are conventionally written in uppercase, with underscores to separate multiple words.

A **function definition** is typically of the form

```
def functionName (argumentName1, argumentName2, ...):
    statements using arguments to compute the result
    return result
```

All statements in the body of the function must have the same indentation, usually four spaces. The statements use the arguments like normal variables. The execution of the function ends when a return statement is encountered.

A **join** is the merging of two tables on a common column. The resulting table has all columns of both tables (the common column isn't duplicated), and the rows are determined by the type of join. Rows in the two tables that have the same value in the common column become a joined row in the resulting table.

In a **logarithmic scale**, each major tick represents a value that is the multiplication by some constant (usually 10) of the value of the previous major tick.

A **method chain** is an expression

like `context.method1(args1).method2(args2).method3(args3)` where each method has and returns the same type of context, except possibly the last method, which can return any type of value.

The **p-value** is an indication of the significance of the result. Usually a p-value below 0.05 is taken to mean the result is statistically significant.

A **return statement** is of the form `return expression` and passes the value of the expression back to the code that called the function to which the return statement belongs.

The **Spearman rank correlation coefficient** of two series of values (e.g. two columns) is a number from -1 (perfect inverse correlation) to 1 (perfect direct correlation), with 0 meaning there is no rank correlation. Correlation doesn't imply causation. A rank correlation of 1 merely states that both values increase and decrease together, while a correlation of -1 states that if one value increases, the other decreases.

A **test** is some code that checks whether some other code works as expected, e.g. a boolean expression that compares the return value of a function call with the expected value.

Reserved Words

`def`, `elif`, `else`, `if` and `return` cannot be used as names.

Functions and methods

`col.apply(functionName)` returns a new column, obtained by applying the given one-argument function to each cell in column `col`.

`DataFrame(columns=listOfStrings, data=listOfLists)` returns a new dataframe, given the data as a list of rows, each row being a list of values in column order.

`download(indicator=string, country='all', start=number, end=number)` is a function in the `pandas.io.wb` module that downloads the World Bank data for the given indicator and all countries and country groups from the given start year to the given end year.

`merge(left=frame1, right=frame2, on=columnName, how=string)` returns a new dataframe, obtained by joining the two frames on the columns with the given common name. The `how` argument can be one of 'left', 'right', 'inner' and 'outer'.

`print()` is a Python function that takes one or more expressions and prints their values on the screen in a single line.

`frame.reset_index()` returns a new dataframe in which rows are labelled from 0 onwards.

`spearmanr()` is a function in the `scipy.stats` module that takes two columns and returns a pair of numbers: the Spearman rank correlation coefficient of the two series of values, and its p-value.