

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try [etherpad.wikimedia.org](https://etherpad.wikimedia.org)).

Users are expected to follow our code of conduct: [https://docs.carpentries.org/topic\\_folders/policies/code-of-conduct.html](https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html)

All content is publicly available under the Creative Commons Attribution License:

<https://creativecommons.org/licenses/by/4.0/>

## Session 6

### ### Links for Noteable +GitRepo

<https://git.ecdf.ed.ac.uk/data-driven-chemistry/ddc-session6-student>

### ### Copy-paste code

### ### Questions and Answers

Q: what does `inplace = true` do?

A: Without this statement, pandas will create a copy of the data with the variables (e.g. NaN) replaced with your chosen value. Using `inplace = true` overwrites the original data frame.

Q: what does `.sort_values()` do ?

A: Sorts the rows of the DataFrame by their values (similar to sorting a list, but keeps rows of data together like a spreadsheet)

### ### Feedback

This last session has been very challenging and hard to understand, I hope the assignment will be a bit easier! The tasks were a bit unclear sometimes, which resulted in us being a bit confused in the breakout rooms (I only managed to do task 1 - even with the help of demonstrators, we were unable to do task 2-5 in time/ understanding the solutions was hard as well, they were explained very rushed). This was not helped by the fact the linear regression was not done by everyone during A levels (or at least to this extent - and we did not have a maths course covering it either) so it was a bit of multitasking to read up on linear regression whilst trying to code. I guess doing the session a bit slower would have been helpful, in any case I hope that having access to the solutions will clear most everything.

Regarding the assignment: it is a bit annoying to always have to look up the name of the function you are supposed to define in the tests/ guess the name. Perhaps for the coming years, it would be easier to state the function name the tests are going to assess with in the task, similar to the name of other variables. This

worked well for assignment 1-5.

---

## Session 5

### ### Links for noteable +GitRepo

<https://git.ecdf.ed.ac.uk/data-driven-chemistry/analysis-of-molecular-geometries>

### ### Copy-paste code

Solution to Exercise 3:

```
atmass = {'C' : 12, 'N' : 14, 'O' : 16, 'H' : 1}
mass = 0
#print(coordinates)
for coord in coordinates:
    print(coord.split()[0])
    mass += atmass[coord.split()[0]]
print(mass)
```

### ### Questions and Answers

Q: `view.addPropertyLabels('index',",")` adds 0,1,2,3,4... but what order is it going by?

A: They are indexed according to their order in the xyz file (i.e. C is the first atom encountered)

Q: I can't manage to do the `.split()` function

A: `.split()` works on strings (and requires a character to perform the splitting around, by default all whitespace). Add more detail if you want a more helpful answer!

Q. In the solution for exercises 3, I'm not sure what this function is doing :

```
mass += atmass[coord.split()[0]]
```

A: `+=` is a way to increment a value in-place (i.e. `a += 4` will add 4 to variable `a`). In this case, `coord.split()[0]` returns an element symbol (e.g. 'C') which we are then using to extract the mass value from `atmass` (e.g. `atmass['C']` will return a value of ~12). This way, within our loop we can handle multiple elements.

Q: Is there an advantage to using `math.sqrt` over `**0.5`?

A: They should return the same result. `math.sqrt` is much faster, however, so is probably better for calculating lots of sqrts. See e.g. <https://stackoverflow.com/questions/327002/which-is-faster-in-python-x-5-or-math-sqrtx>

`math.sqrt` will also be a bit clearer for other people (or future you) using this code after you have written it.

Q. What should the bond length be for EX.4?

A: The C=O distance is ~1.212 Ang

### ### Feedback

---

## SESSION 4

### ### Links for notable +GitRepo:

<https://git.ecdf.ed.ac.uk/data-driven-chemistry/numericalanalysis>

### ### Copy-Paste code:

### ### Questions and Answers:

QWhat does .ndim do?

A ndim returns the number of dimensions (axes) of the array (in this case the array was one dimensional  
sdfgying the text file, how do I do this?

A: Open it in a plain-text editor (e.g. Notepad, Notepad++, Emacs....). Not Word!

QWhat are unique elements?

A: The different values that appear in the array; i.e. [1,1,2,3,5,7,1,5] would have unique values [1,2,3,5,7]

Q what does .linspace do?

A: Generates an array of values between start and end, but returns a certain number of values (as opposed to .arange, where you define the spacing between the values, but don't care how many).

Q. How do I import using !pip? I get a "permission denied" error for the !conda in the notebook

A: I think you're using an older version of the notebook; !conda will fail to install mendeleev on Noteable at the moment. Replace the !conda install... with !pip install mendeleev==0.6.0 and it should work (but can lead to problems if you try to install other packages).

### ### Feedback:

---

## PREVIOUS SESSIONS INFOS:

### Session 3:

#### ### LINKS

For notable import: <https://git.ecdf.ed.ac.uk/data-driven-chemistry/session-3-functions>

#### ### Copy and Past Code Session 3:

$$K = \exp\left(\frac{-\Delta G}{RT}\right)$$
$$x_+ = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$
$$x_- = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

#### ### Questions and Answers

Q: What does %matplotlib inline do?

A: This makes sure that any graphs we create appear within the Jupyter notebook (it is a Jupyter-specific 'magic' command)

<https://ipython.readthedocs.io/en/stable/interactive/magics.html>

Q: what the diff between the triple quotation mark and the hashtag for wirtting comments?

A: The triple quotation mark lets you have comments over more than one line. They are commonly used for documenting functions or modules. Hash tags are used within the code to give information on indivudual lines

Q:What happens if you define a function that already exists? does it overwrite it?

A: Yes. Do this with caution!

Q: Why would you use `print(f"Hello {name}")` rather than `print("Hello", name)`?

A: Both achieve the same result, but the f-string way gives more control over the formatting. For instance, imagine you wanted to print "Hello XXX" for a list of names, but wanted them all to appear right-aligned:

```
Hello    George
Hello    Catherine
Hello        Mo
```

This is very difficult to achieve with a simple print statement, but easy with f-strings, using some fancy formatting : `print(f"Hello {name:>12}")`

Q: what does \* mean beside the kernel?

A: The cell is still processing. There is also a chance that the kernal has crashed in that case restart the kernal using the options in the toolbar.

Q: What is the purpose of 'return'?

A: To literally 'return' a variable from a function so that you can do something useful with it such as assigning it to a variable, send it as input to another function etc. Without return, the value remains inside the function and cannot be used again. Print, in contrast, just writes the value to the notebook, but you can't access it again later (without copy and paste).

Q: Why does my \$ \$ (LaTeX) not work?

A: Make sure your cell is in Markdown mode

Q: What does the ".5f" mean in Task 3.5?

```
for peptide in ['CPHRALIAIT', 'NGQSVCGMSG', 'WPFYWRICNH', 'DLQVIDQMNW',
'CEWIMYVTDE']:
```

```
    print(f'Mass of {peptide} is {peptide_mass(peptide):.5f} Daltons')
```

A: This relates to the format of the output, and will print 5 characters following the "." - try it yourself with different numbers.

### Feedback

---

## PREVIOUS SESSIONS INFOS:

```
### CODE for copy paste
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline

print("Hello Data driven Chemistry!")
len("G") ## These are built in functions

def say_hello():
    print("Hello!")
```

To be executed into markdown cell :

$K = \exp\left(\frac{-\Delta G}{RT}\right)$

$x_{+} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$

$x_{-} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$

Below is the doc string for get\_roots

```
"""Computing the roots of a quadratic function
```

```
Parameters:
```

```
-----
```

```
a: float
```

```
parameter in front of x^2
```

```
b: float
```

```
parameter in front of x
```

```
c: float
```

```
third parameter in quadratic equation
```

```
Returns:
```

```
-----
```

```
x_plus: float
```

```
positive root
```

```
x_mins: float
```

```
negative root
```

""""

### ### Questions and Answers

Q: what does this line of code mean? `print("{n1} multiplied by {n2} is:`

`{multiplication}").format(n1=n1, n2=n2, multiplication=multiplication))`". Especially the `.format()` part?

A: The `.format()` part sets the values within the curly brackets. For example, if the string contains `"{toprint}"` and to replace this with the value of `n1`, you'd use `"{toprint}".format(toprint = n1)`.

Q: what does `f""` do and why do we use `{}`

A: Sometimes you want to create a string containing one or more variables, with a bit more formatting control than `print(var1, var2, var3...)`. f-strings are one solution; the 'f' tells python that any variable name within the string (contained in `{}`) should be replaced by the variable value.

Q: what does the `"""` do?

A: `"""` is a special (multi-line) string, which in this case is being used to document the function. You can add almost anything to a doc-string, but the most important thing is that it should tell someone what the function does, if they were to look at the help.

Q: What does it do when you write `return`?

A: `Return` literally 'returns' the value from the function so that you can use it later (e.g. save it to a variable).

Q: what is the difference between `return` and `print`?

`"print(str)"` just displays the string `str` whereas `return` is an integral part of a function (see above)

Q: What is `numpy`?

A: `NumPy` is a very commonly used package for numerical calculations with Python (we'll cover it more next session)

Q: What do dollar signs do?

A: In a markdown cell they render equations in LaTeX (math) mode. Surrounding an equation with single dollar signs will put it 'in-line' with the text, while double dollars (`$$y=mx$$`) will render it as a 'separate' equation (e.g. centred and on a new-line)

Q: What does `.format()` do?

A: `"some string".format` is another (older) way of putting variables in strings. `"{1} and {2}".format('a','b')` will produce the string `"a and b"`.

Q: What does `adjust` do?

Q. Are all the assessments marked by computer? Is there any incentive to write efficient code that's futureproof against larger tasks?

### ### FEEDBACK:

Can we have more time in the breakout room - i always cant manage to finish all the tasks before it ends

---

### ###LINKS

GitLab 1st session

<https://git.ecdf.ed.ac.uk/data-driven-chemistry/introduction>

##section of code to add:

```
print("Printing cumulative sum from 1-10:")
```

```
total = 0
```

```
for i in range(1, 11):
```

```
    total += i
```

```
    print( f'Sum of 1 to {i} is {total}')
```

```
print( 'Done printing numbers.' )
```

### ### Questions and Answers

Q. If we add our own code in separate cells in the assignment, will it mess it up, or are extra cells considered a form of playground? same with commenting, does the marker pay attention to it?

Q: How do I get the files?

Answer: On noteable - go to the +gitrepo button on the top right. and click.

copy and paste : <https://git.ecdf.ed.ac.uk/data-driven-chemistry/introduction>  
and hit enter!

Q: How do I open a new python notebook?

Answer: on the top right of noteable, click on the button "new".

Then click on python3.

This will open a new window.

Q: why do you need #print?

Answer: # signifies a comment, we did not want to use that line of code. so instead of deleting it, we jsut comment it out with #.

Q:what does str() do?

A: have a look at the manual online:) [https://www.w3schools.com/python/ref\\_func\\_str.asp](https://www.w3schools.com/python/ref_func_str.asp)

Q: how do you convert to a string?

answer: using the function str(my\_variable\_that\_isnt\_a\_string).

e.g. str(42) or str(4.5)

Q: `print(a[1])` works, why do I need to use `str()`?

if you declared variable `a` as a string [i.e. `a = '123'` ], you will not need to convert it onto the string.

if you declared variable `a` as an integer [i.e. `a = 123`, note no ' '], you will need to first convert it into the string [`s = str(a)`] and then ask for the character [1] [i.e. `s[1]`], or shorter skipping `s` , just `str(a)[1]` ]



Q: Why doesn't first + int(third) work? (third = "1.1")

A: 1.1 isn't an integer. Going from a string to float to integer does work (ie. task 4 part 1+2) but string to integer doesn't.

Q: What does len() mean?

A: len() is the function that gives the length of an argument. This works for string but not integers.

Q: What are tuples/lists/dictionaries?

A: - A list is a 'general' container for a collection of items, which can later be modified

- A tuple is like a 'fixed' list; it cannot be altered after creation

- A dictionary is for storing key:value pairs, but the order in which these are retrieved is not defined

Q. What does the f do?

### ### Feedback

the session was understandable, but very very long and harder to follow as the time went by

Session goes really quick/ confusing given that a lot of people would never have done this before.

Good explanation about everything, a lot of new terminology to take in, but I think everything will become clearer upon reading the documents in my own time. Thank you!

Got into complex topics too quickly, with too much time spent on talking and not enough individual/team working, which is the main draw of labs. Coding is interesting, but I found that the demonstrators couldn't get the real lesson across when we get such little time to work on the code ourselves. Practically a 3-hour lecture throwing us in the deep end. I just don't think Coding is conducive to long 3-hour sessions. Thank you for the effort you put in so far, I just think it just needs ironing out.

Good teaching and help on offer but moved through topics too quickly and vaguely especially for people who have never done this before.

Too much talking and not enough independent/peer working. Basically a 3 hour lecture where it was incredibly difficult to keep concentration by 2 hours in. Complex topics started too quickly - many of us haven't coded before. Was difficult to learn when demonstrators kept telling us to 'google it' instead of explaining how to do things.

Probably not enough independent study advice, for those who haven't done coding before. As the person above said, 'google it' isn't helpful unless they have to be taught to google, and have learnt that most programmers spend more time on stackoverflow than coding itself.

Too Fast - Can't keep up

Go through everything too fast and teaching as if we know what to do already. Can't access feedback for assignment 1 so have no idea if we are doing it correct which will have a knock on effect for assignment 2 :((

More time in breakout rooms as we rarely have enough time to finish all tasks.

The demonstrations are good, but I always end up in breakrooms with people who don't talk, so can't engage in peer learning. Would be nice if the sessions were organised in a way that stimulate more people to engage

The tasks have nothing to do with what it was taught and are extremely complicated