

Python crash course

Basic python programming

with Jongbin Jung

All material publicly available [here \(https://github.com/5harad/css\)](https://github.com/5harad/css)

2. python programming - 1 of 3

Lines of python code can be saved to a plain text file, conventionally appended with a .py extension, which can be read by the interpreter to run a python program. A common setup for python development is to have

1. a text editor of choice
2. a command line/terminal open to run the .py file
3. (optionally) a interpreter for testing small pieces of code

Having your own setup can be great if you have a favorite text editor and love pushing commands around different windows. Some text editors have pretty good support for python development, too.

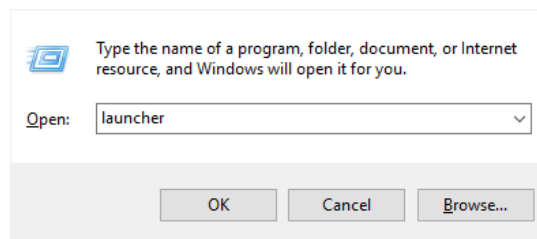
Another option is to use an IDE (Integrated Development Environment), specifically catered to your python development needs. PyCharm (<https://www.jetbrains.com/pycharm/>) is a pretty one, and will be familiar if you've worked with IDEA, Android Studio, WebStorm, PhpStorm, etc. (They're all based on the same platform.) Today, we'll be working with spyder because

1. it's free
2. it comes included with Anaconda
3. it loads faster (compared to PyCharm)

Launching spyder

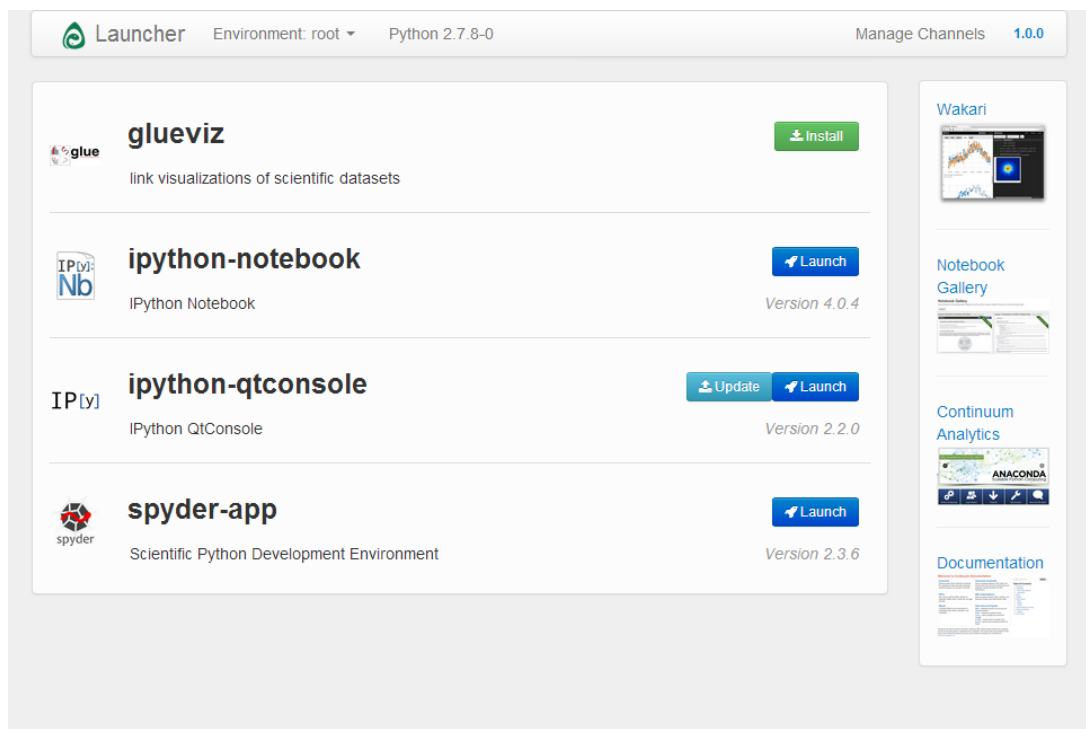
Spyder is best launched from Anaconda's launcher.

- launch the launcher
 - **Windows**
 - `Win` + `R`, type `launcher` and hit `OK` (or `Enter`)

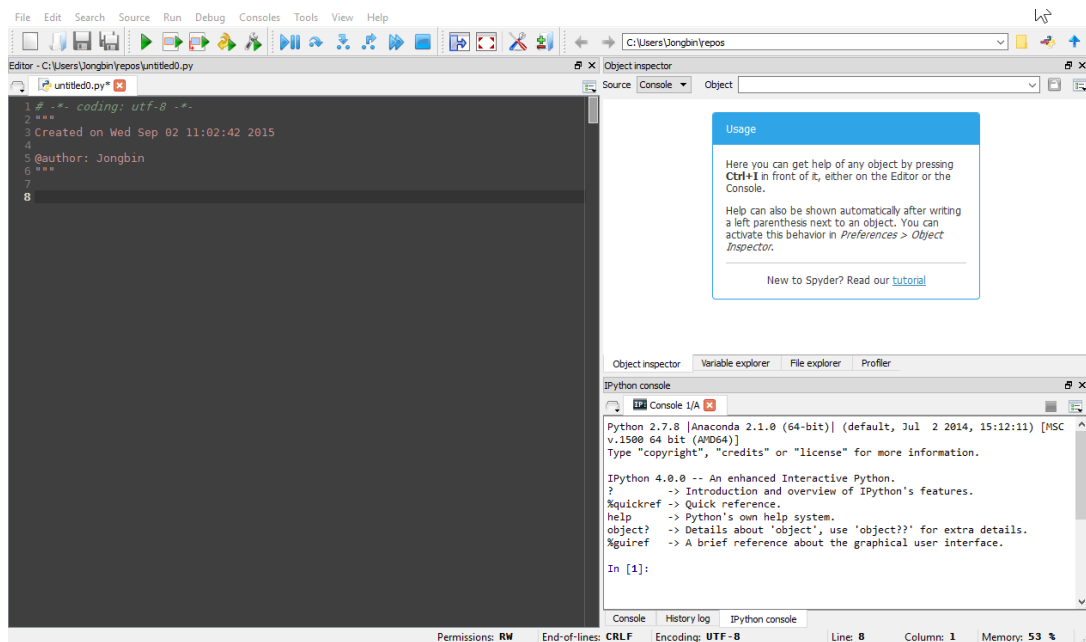


- **OS X / *nix**
 - Open a terminal
 - type `launcher` and hit `Enter`

```
✓ jongbinjung@JB-MacBook-Pro[~]
$ launcher
```



- hit **Launch** for spyder-app



By default, spyder has a text editor on the left pane, an interactive console (to which you can also send selected commands from the text editor pane with **Shift + Enter**) and object/variable/file browsers on the right side.

Feel free to explore and get used to the spyder environment before we move on.

(One feature I find particularly useful is the **Ctrl + i** shortcut, which displays documentation for the object at my current cursor, whether in the editor or console.)

if statements

An example should suffice

```
In [1]: x = int(input('Give me a BIG number: '))
        if x < 0:
            print('You\'re joking, right?')
        elif x < 1e3:
            print('Try harder ... ')
        else:
            print('Nice.')
```

Give me a BIG number: 10000
Nice.

Some notes on the above code:

- the `input()` function (as you've now seen), prompts the user for an input
- the `int()` (tries to) convert string values to integers (`raw_input()` will always return the user's input as a string)
- `elif` is short for `else, if`, and there can be none or more than one `elif` sequences
- the `else` clause is optional

One more thing that's implicit but **extremely** important: **Indents**.

- python, unlike many other languages out there, doesn't use curly brackets `{}`
- instead, blocks of grouped code are identified by the level of indents (this is something to get used to, if you've never seen it before)
- word of caution: NEVER USE `Tab` (don't worry, `spyder` changes all your `Tab`s to four spaces by default, which is the [PEP 8 spec for indentation \(https://www.python.org/dev/peps/pep-0008/#indentation\)](https://www.python.org/dev/peps/pep-0008/#indentation) in python)

for statements

The `for` statement in python iterates over the items of any sequence (e.g., lists and even strings!), in the order that they appear in the sequence.

```
In [2]: names = ['Jamie', 'Cersei', 'Jon', 'Sansa']

        for name in names:
            print(name, 'has', len(name), 'characters and starts with a', name[0])
```

Jamie has 5 characters and starts with a J
Cersei has 6 characters and starts with a C
Jon has 3 characters and starts with a J
Sansa has 5 characters and starts with a S

The example above introduces a few new concepts:

- the variable `name` is defined along with the declaration of the `for` statement. It doesn't need to exist beforehand
- the `print` function can take multiple arguments of different types, (try to) change them to a string, and insert a space between each item (separated by commas)
- it's good practice to use plurals for collections (`names` for the list) and singulars for individual items (`name` for each name)

You can also loop over a string, one character at a time.

```
In [3]: vowels = ['a', 'e', 'i', 'o', 'u'] # make a list of vowels
        for name in names:
            vowel_count = 0 # initialize the vowel count
            for char in name:
                if char in vowels:
                    vowel_count += 1
            print(name, 'has', vowel_count, 'vowel(s)')
```

```
Jamie has 3 vowel(s)
Cersei has 3 vowel(s)
Jon has 1 vowel(s)
Sansa has 2 vowel(s)
```

You can use the built-in `range()` function to do a more 'classic' `for` loop over a sequence of numbers.

```
In [4]: for i in range(10): print(i)
```

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

`range(len)` generates the legal indices (starting from 0) for a sequence of length `len`. You can also use `range(start, stop[, step])` to specify the start, end, and (optionally) step to take.

(The `[, step]` notation in the function signature shows that the `step` argument is optional. It's useful to know such conventions when referring to the docs.)

```
In [5]: for i in range(4,8): print(i)
```

```
4
5
6
7
```

```
In [6]: for i in range(4,8,2): print(i)
```

```
4
6
```

```
In [7]: for i in range(20,4,-3): print(i)
```

```
20
17
14
11
8
5
```

You can combine `range()` with `len()` to iterate over the indices of a sequence.

```
In [8]: for i in range(len(names)):
        print('Name', i, 'is', names[i])
```

```
Name 0 is Jamie
Name 1 is Cersei
Name 2 is Jon
Name 3 is Sansa
```

But in such cases, the `enumerate()` function is usually more convenient.

```
In [9]: for i, name in enumerate(names):
        print('Name', i, 'is', name)
```

```
Name 0 is Jamie
Name 1 is Cersei
Name 2 is Jon
Name 3 is Sansa
```

As you might have guessed, the `enumerate()` function takes a sequence, and returns the (index, value) pairs for each item (the 'pairs' are actually called `tuple`s, but more on that later...), and you can assign items from a `tuple` to its own variable in the `for` statement.

```
In [10]: print(list(enumerate(names)))

[(0, 'Jamie'), (1, 'Cersei'), (2, 'Jon'), (3, 'Sansa')]
```

Occasionally, you might want to loop over two or more sequences at a time. You can pair the entries with the `zip()` function.

```
In [11]: title = 'Game of Thrones'
houses = ['Lannister', 'Lannister', 'Snow', 'Stark']
for char, house, name in zip(title, houses, names):
    print(char, '-', name, house)
```

```
G - Jamie Lannister
a - Cersei Lannister
m - Jon Snow
e - Sansa Stark
```

Note how `zip()` gracefully fits the iterator to the length of the shortest sequence, i.e., only the first four characters of the string 'Game of Thrones' were iterated.

break and continue statements

You can manage your loops in more detail using `break` and `continue` statements.

A `break` statement, as the name implies, will break you out of the smallest enclosing loop.

```
In [12]: for name, house in zip(names, houses):
        if house == 'Snow':
            break
        else:
            print(name, house)
```

```
Jamie Lannister
Cersei Lannister
```

A `continue` statement will simply skip over to the next item in the iterator, instead of breaking out of the loop.

```
In [13]: for name, house in zip(names, houses):
          if house == 'Snow':
              continue # compare to the previous example where we stopped the loop at S
          else:
              print(name, house)
```

```
Jamie Lannister
Cersei Lannister
Sansa Stark
```

Some more data structures

Before we move on, now might be a good time to cover a few more data structures.

dict (dictionary)

The most useful data structure in `python` (my very personal opinion)! Also known as *associative arrays* or *hash tables* in other languages, a `python` dictionary maps *hashable* values to *arbitrary* objects. Dictionaries can be created by placing a comma-separated list of `key:value` pairs within curly braces. Just remember that the `key` must be immutable (like a string).

```
In [14]: me = {'name': 'Jongbin', 'email': 'jongbin@stanford.edu'}
          print(me)

{'name': 'Jongbin', 'email': 'jongbin@stanford.edu'}
```

You can assign new keys to existing dictionaries.

```
In [15]: me['cel'] = '650-123-4567'
          print(me)

{'name': 'Jongbin', 'email': 'jongbin@stanford.edu', 'cel': '650-123-4567'}
```

Or delete existing `key:value` pairs with the `del` statement.

```
In [16]: del(me['email'])
          print(me)

{'name': 'Jongbin', 'cel': '650-123-4567'}
```

The `key` of a dictionary can't be a list (because lists are mutable), but the `value` sure can!

```
In [17]: me['siblings'] = ['Hanbyul', 'Hansol']
          print(me)

{'name': 'Jongbin', 'cel': '650-123-4567', 'siblings': ['Hanbyul', 'Hansol']}
```

Use the `keys()` method of dictionary objects to get a list of the keys used in the dictionary.

```
In [18]: me.keys()

Out[18]: dict_keys(['name', 'cel', 'siblings'])
```

And use the `in` keyword (compatible with all lists) to see if a certain key exists in the dictionary.

```
In [19]: 'name' in me.keys()
```

```
Out[19]: True
```

```
In [20]: 'email' in me.keys()
```

```
Out[20]: False
```

When the keys are simple strings, it is sometimes easier to specify pairs using the `dict` constructor.

```
In [21]: me = dict(name='Jongbin', email='jongbin@stanford.edu', siblings=['Hanbyul', 'Hansol'])
          print(me)
```

```
{'name': 'Jongbin', 'email': 'jongbin@stanford.edu', 'siblings': ['Hanbyul', 'Hansol']}
```

The `items()` method lets you loop over each `key:value` pair.

```
In [22]: for key, value in me.items():
          print(key, ': ', value)
```

```
name : Jongbin
email : jongbin@stanford.edu
siblings : ['Hanbyul', 'Hansol']
```

tuples

Tuples are pretty similar to lists, except for the fact that they are immutable. They consist of a number of values separated by commas (not necessarily, but often, enclosed in parentheses).

```
In [23]: description = 'male', 'dark hair'
          print(description)
```

```
('male', 'dark hair')
```

```
In [24]: description[0] # tuples are also sequences, and can be indexed
```

```
Out[24]: 'male'
```

```
In [25]: description[1:] # or sliced
```

```
Out[25]: ('dark hair',)
```

```
In [26]: description[0] = 'female' # but NOT changed, because they are immutable
```

```
-----
TypeError                                 Traceback (most recent call last)
<ipython-input-26-193ada17dc41> in <module>()
----> 1 description[0] = 'female' # but NOT changed, because they are immutable

TypeError: 'tuple' object does not support item assignment
```

While being immutable may seem like a minor difference from lists, the implications are quite big, and tuples are generally used for very different purposes compared to lists. For example, tuples can be used as the `key` for dictionaries (think sparse matrices).

```
In [27]: super_sparse_matrix = {(0, 0):1, (1000, 1000):1} # a 1000*1000 matrix with only two non-zero elements
print(super_sparse_matrix)

{(0, 0): 1, (1000, 1000): 1}
```

```
In [28]: word_matrix = {('apples', 'bananas'):1, ('apples', 'pears'):1} # a matrix indexed by words
print(word_matrix)

{('apples', 'bananas'): 1, ('apples', 'pears'): 1}
```

There are many more data structures commonly used in python, but lists, dictionaries, and tuples pretty much cover the basics (not to mention that these three constitute enough to fully represent the [JSON](http://json.org/) (<http://json.org/>) format in python)

List comprehension

List comprehension is python's way of creating lists (and also other data structures) in a concise manner. One way to create a list of squares would be:

```
In [29]: squares = [] # make an empty list
for x in range(10):
    squares.append(x**2)

print(squares)

[0, 1, 4, 9, 16, 25, 36, 49, 64, 81]
```

However, the more 'pythonic' way to do this, is to use list comprehension:

```
In [30]: [x**2 for x in range(10)]

Out[30]: [0, 1, 4, 9, 16, 25, 36, 49, 64, 81]
```

The command reads:

```
build a list out of the square of x (x**2), for the values of x in range(10)
```

List comprehension can be used to build a list of tuples too.

```
In [31]: [(x, y) for x in range(10) for y in range(10) if x*y == 21]

Out[31]: [(3, 7), (7, 3)]
```

This is equivalent to the nested for loop:


```
In [32]: twenty_one = []
        for x in range(10):
            for y in range(10):
                if x*y == 21:
                    twenty_one.append((x, y))

        print(twenty_one)

[(3, 7), (7, 3)]
```

Just be aware that if the item of the list is a tuple, it must be parenthesized.

```
In [33]: [x, y for x in range(10) for y in range(10) if x*y == 21] # this won't work

File "<ipython-input-33-c0af109a84cc>", line 1
    [x, y for x in range(10) for y in range(10) if x*y == 21] # this won't work
        ^
SyntaxError: invalid syntax
```

Let's enhance our list of vowels from the previous exercises, by appending the uppercase letters as well.

```
In [34]: vowels = vowels + [V.upper() for V in vowels]
        print(vowels)

['a', 'e', 'i', 'o', 'u', 'A', 'E', 'I', 'O', 'U']
```

List comprehension can also be used to build dictionaries.

Try building a sparse matrix (represented by a dictionary), where the position is indexed by words x and y , and

$$(x,y) = \begin{cases} 1 & \text{if } y \text{ is longer than } x \\ 0 & \text{otherwise} \end{cases}$$

(But note that the example below is not always the best way to do this! Implementation should depend on your context - what do you want to do with the data/matrix?)

```
In [35]: words = ['anti', 'happy', 'evening', 'eagles', 'interior', 'zebra']
        {(x,y):1 for x in words for y in words if len(y) > len(x)}
```

```
Out[35]: {('anti', 'eagles'): 1,
          ('anti', 'evening'): 1,
          ('anti', 'happy'): 1,
          ('anti', 'interior'): 1,
          ('anti', 'zebra'): 1,
          ('eagles', 'evening'): 1,
          ('eagles', 'interior'): 1,
          ('evening', 'interior'): 1,
          ('happy', 'eagles'): 1,
          ('happy', 'evening'): 1,
          ('happy', 'interior'): 1,
          ('zebra', 'eagles'): 1,
          ('zebra', 'evening'): 1,
          ('zebra', 'interior'): 1}
```

Read/Write Files

Often, you will need to read some data into your `python` workspace, do something to/with said data, and then write the results to another file. We'll take a look at the most basic file read/write methods, which will get you started with your work, and look at some more advanced topics if we have more time.

File objects

Think of a `python` file object as a portal connecting your `python` workspace to a file on your hard drive. You can open a file object with the built-in `open(filename, mode)` function. The `filename` argument is a string specifying the file name, and the `mode` argument can be one of the following values, specifying whether you want to read from or write to the file:

- `'r'` : read
- `'w'` : write (overwrites any existing files with same filename)
- `'a'` : append (write additional to any existing data)

(you can also open files for both read/write with mode `'r+'` , but this best avoided if possible)

By default, files are opened in "text" mode (think: "Files that can be opened and read by a human in a text-editor.") Alternatively, you can open files in "binary mode" by appending a `b` to the `mode` argument (e.g., `wb` , `rb` , `ab`).

Remember that `open()` simply creates the 'portal', and you have to call additional methods on that file object to either read or write. Since reading can be a little more complicated, let's start with a simple write:

```
In [36]: f = open('example.txt', 'w')
         print(f)
```

```
<_io.TextIOWrapper name='example.txt' mode='w' encoding='UTF-8'>
```

Note that after creating the file object, the empty `filename` file (in the above example, `example.txt`) is created in your working directory. Now, let's actually write something to it:

```
In [37]: f.write('Something')
```

```
Out[37]: 9
```

You can only write strings to a file object:

```
In [38]: some_list = [1, 2, 3]
         f.write(some_list)
```

```
-----
TypeError                                 Traceback (most recent call last)
<ipython-input-38-70fc2ad98f1e> in <module>()
      1 some_list = [1, 2, 3]
----> 2 f.write(some_list)
```

```
TypeError: write() argument must be str, not list
```

To write anything other than a string, use the `str()` built-in function to convert it to a string first:

```
In [39]: f.write(str(some_list))
```

```
Out[39]: 9
```

You might notice that even though you've called `write()` a couple times, the actual file on your hard drive

doesn't necessarily get updated. That's because a file object's `write`s are kept in buffer. To complete all the `write`s and close the file object, call the `close()` method:

```
In [40]: f.close()
```

Note that using a closed file object will result in an error:

```
In [41]: f.write('...')
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-41-8be1c293604e> in <module>()
----> 1 f.write('...')

ValueError: I/O operation on closed file.
```

Reading from a URL

Reading data from a URL in `python` is pretty simple, using the `urllib.request` module. The `urllib.request` module lets you open URLs in `read` mode, as if they were file objects.

Let's use `python`'s `urllib.request` to read Charles Dickens' "A Tale of Two Cities" from <https://goo.gl/fHIeOi> (<https://goo.gl/fHIeOi>)

(This is just for illustration. Note, there are other libraries that are usually more appropriate for reading/scraping web pages.)

```
In [42]: from urllib.request import urlopen # the import statement is used in python to im

link = urlopen('https://goo.gl/fHIeOi') # open the url
print(link)
text = link.read()
link.close() # just like file objects, url connections should be closed after you

<http.client.HTTPResponse object at 0x7f9c782c24a8>
```

The `text` variable now contains the entire text of "A Tale of Two Cities".

```
In [43]: print(text[0:20])
```

```
b'A Tale of Two Cities'
```

Notice the `b` in front of the quotes. This indicates that the data in our `text` object is saved as `bytes` not strings.

Now let's try writing `text` to a file.

```
In [44]: f = open('two_cities.txt', 'wb') # open file object in write (bytes) mode
f.write(text)
f.close()
```

Reading from file objects

And now, we have a file to practice reading from! We can create a file object just like we did for writing, but with the `'r'` mode specified:

```
In [45]: f = open('two_cities.txt', 'r') # open file object in read mode
```

A file object will iterate over the contents of the file it is connected to. For example, the `readline()` method will read the file, one line at a time. And consecutive calls to `readline()` will keep giving you the next line:

```
In [46]: print('first line:', f.readline()) # read the first line
print('second line:', f.readline()) # read the second line
```

first line: A Tale of Two Cities, by Charles Dickens

second line:

Since the file object essentially provides an iterator over each line of the file, you can loop over the file object line-by-line. This is memory efficient, fast, and leads to simple code:

```
In [47]: n = 1 # a simple counter to control the number of lines printed
for line in f:
    print(line)
    if n > 10:
        break
    n += 1
```

[A story of the French Revolution]

CONTENTS

Book the First--Recalled to Life

Just like when writing, don't forget to close files after you're done!

```
In [48]: f.close()
```

As your file I/O gets complex, opening and closing can become quite painful (e.g., what if an error occurs before you close the file object? what happens to the memory it's using?), and forgetting to close file objects is potentially dangerous. So, it's good practice to use the `with` and `as` keywords, which makes sure that the file is properly closed after operations are finished, even if an error occurs during operations:

```
In [49]: with open('two_cities.txt', 'r') as f:
        n = 1
        for line in f:
            print(line)
            if n > 10: break
            n += 1
```

A Tale of Two Cities, by Charles Dickens

[A story of the French Revolution]

CONTENTS

Handling exceptions

Sometimes, you will anticipate certain errors, and want your code to behave in a specific manner when such errors occur (e.g., if you're trying to parse tweets as JSON objects, and you know that some results will be junk that cause errors). Such errors that occur at runtime (distinguished from syntax errors) are known as exceptions. You can use `try - except` statements to catch certain exceptions.

One common example is assigning a value to an undefined dictionary key. Let's count the occurrence of each alphabet in a certain string:

```
In [50]: s = 'how many wood would a woodchuck chuck if a woodchuck would chuck wood'

d = {} # define an empty dictionary
for char in s:
    d[char] += 1 # increase the count of d[char] by 1 ... this will result in a K
print(d)
```

```
-----
KeyError                                Traceback (most recent call last)
<ipython-input-50-4a726879127a> in <module>()
      3 d = {} # define an empty dictionary
      4 for char in s:
----> 5     d[char] += 1 # increase the count of d[char] by 1 ... this will resu
lt in a KeyError
      6
      7 print(d)

KeyError: 'h'
```

As expected, the above code resulted in a `KeyError` (as we can see from the error message - or Traceback). Of course, we could avoid the error by checking if the key exists each time before increasing a count, but this would require an iteration through the entire list of keys every time we increase a value, and that could be potentially expensive if we have many keys. A more efficient way would be to capture the `KeyError` in a `try - except` statement. The `except` statement can specify what type of errors to handle, so we don't end up ignoring other meaningful errors:

```
In [51]: s = 'how many wood would a woodchuck chuck if a woodchuck would chuck wood'

d = {} # define an empty dictionary
for char in s:
    try: # try everything in this block, and goto the except block if an error occurs
        d[char] += 1 # increase the count of d[char] by 1 ... this will result in an error
    except KeyError: # execute the following block, only if the error is a KeyError
        d[char] = 1 # instead of incrementing, initiate a key:value pair in the dictionary
print(d)

{'h': 5, 'o': 11, 'w': 7, ' ': 12, 'm': 1, 'a': 3, 'n': 1, 'y': 1, 'd': 6, 'u': 6, 'l': 2, 'c': 8, 'k': 4, 'i': 1, 'f': 1}
```

While catching errors can be useful, always be careful of which errors you catch. Ignoring unexpected error messages can render your results useless!

Exercise 2.

1. Declare three lists:

```
names = ['Harry', 'Ron', 'Hermione']
verbs = ['likes', 'hates', 'eats']
objects = ['pie', 'owls', 'the snitch']
```

Use list comprehension to create a list of all (name, verb, object) combinations. (The final result should have length 27.)

2. With the list of combinations from the previous question, use a `for` loop to print sentences of what Harry eats. (Hint: You can use an `and` / `or` statements to check two or more conditions in a single `if` statement, e.g.,
`if condition_1 and condition_2:...`

The final output should be properly formatted as a sentence, e.g.,

```
...
Harry eats pie.
...
```

3. Create a `python` dictionary that counts the occurrence of every word, delimited by white spaces, in "A Tale of Two Cities".
4. Find words that occur between 500 and 700 times in "A Tale of Two Cities".

Extra: Catching non-standard exceptions

I mentioned a specific case of catching errors that arise when trying to parse junk JSON strings. Let's take a closer look at how to actually do that!

In [52]: **import** json

```
# Create some example JSON input
lines = ['{"name": "Jongbin", "msg": "Hello!"}',
        '{"name": "Jessica", "msg": "Hi to you, too!"}']

for line in lines:
    # the json string is loaded as a dictionary
    obj = json.loads(line)
    print("%s says %s" % (obj["name"], obj["msg"]))
```

Jongbin says Hello!

Jessica says Hi to you, too!

Now let's see what happens when we try to parse some broken JSON string

```
In [53]: # Create some example JSON input with one broken input
lines = [
    '{"name": "Jongbin", "msg": "Hello!"}',
    '\n', # This entry isn't really a JSON object
    '{"name": "Jessica", "msg": "Hi to you, too!"}',
]

for line in lines:
    # the json string is loaded as a dictionary
    obj = json.loads(line)
    print("%s says %s" % (obj["name"], obj["msg"]))
```

Jongbin says Hello!

```
-----
JSONDecodeError                                Traceback (most recent call last)
<ipython-input-53-d9ce257682ab> in <module>()
      8 for line in lines:
      9     # the json string is loaded as a dictionary
--> 10     obj = json.loads(line)
     11     print("%s says %s" % (obj["name"], obj["msg"]))

/usr/lib64/python3.6/json/__init__.py in loads(s, encoding, cls, object_hook, parse_float, parse_int, parse_constant, object_pairs_hook, **kw)
     352         parse_int is None and parse_float is None and
     353         parse_constant is None and object_pairs_hook is None and not
    kw):
--> 354         return _default_decoder.decode(s)
     355     if cls is None:
     356         cls = JSONDecoder

/usr/lib64/python3.6/json/decoder.py in decode(self, s, _w)
     337
     338     """
--> 339     obj, end = self.raw_decode(s, idx=_w(s, 0).end())
     340     end = _w(s, end).end()
     341     if end != len(s):

/usr/lib64/python3.6/json/decoder.py in raw_decode(self, s, idx)
     355         obj, end = self.scan_once(s, idx)
     356     except StopIteration as err:
--> 357         raise JSONDecodeError("Expecting value", s, err.value) from None
     358     return obj, end
```

JSONDecodeError: Expecting value: line 2 column 1 (char 1)

Notice the error message, and specifically, what exception is raised: `JSONDecodeError` from the `json` package.

Now we know what error we want to catch!


```
In [54]: for line in lines:
        try:
            # First, try to load the string as a json object
            obj = json.loads(line)
            print("%s says %s" % (obj["name"], obj["msg"]))
        except json.JSONDecodeError:
            # Silently ignore lines that trigger a JSONDecodeError
            pass
```

```
Jongbin says Hello!
Jessica says Hi to you, too!
```