Python 3: Data Input and output

File handles / Data streams

The open function is used to open file handles. Good reference can be found at https://en.wikibooks.org/wiki/Python_Programming/Input_and_Output

Data streams could be from cmdline (eg STDIN)

```
$ cat file | python myscript.py
```

Can also be a file

```
filehandle = open(myfile,"r")
```

Getting data into your program

`open` - is for opening a file

There are two arguments, one is the filename, the other is how to open, reading, writing. For text data use "r" but if binary data use "rb".

```
file = "data1.dat"
fh = open(file,"r")
for line in fh:
    print(line)
```

Alternative structure for opening

This will throw a warning if the file cannot be opened

```
with open(myfile,"r") as fh:
    for line in fh:
        print(line)
```

Writing data

Write data or text to a file.

```
ofh = open("my_data.tab","w")
ofh.write("Species\tHabitat\tSize\n")
ofh.write("Crab\tBeach\tM\n")
ofh.write("Fish\tOcean\tS\n")
```

```
$ cat my_data.tab
Species    Habitat    Size
Crab    Beach    M
Fish    Ocean    S
```

Modules/Libraries/Packages

Modules are collections of code routines. We will talk more about functions/routines in next lecture. Can use these as tools.

- sys - System-specific parameters and functions
- urllib.request - URLs for opening web or network connections
- csv - Comma and Tab delimited data parsing

Reading in STDIN

Remember we can pass data into a program via STDIN if we use the '|' "pipes" in UNIX

```
import sys

counter = 0
for line in sys.stdin:
    counter += 1
print ("There are",counter, "lines")
```

Data streams don't have to be files

Can be a network connection (eg URL for web or FTP)

```
import urllib.request
orginfo = "https://biodataprog.github.io/programming-intro/data/random_exon
info = urllib.request.urlopen(orginfo)
for line in info:
    linestrip = line.decode('UTF-8').strip()
    print(linestrip)
```

CSV module

```
import csv

file2 = "test2.csv"
with open(file2) as csvfile:
    reader = csv.reader(csvfile,delimiter=",",quotechar='|')
    for row in reader:
        print("\t".join(row))

with open("outtest.csv","w") as csvfile:
    writer = csv.writer(csvfile,delimiter="\t")
    writer.writerow(["Name","Flavor","Color"])
    writer.writerow(["Apple","Sweet","Red"])
    writer.writerow(["Pretzel","Salty","Brown"])
```

command line arguments

- import sys
- get the command line input as a list from sys

```
$ python argparse.py arg1 arg2 arg3 this-is-one "this is one"
```

```
import sys
for n in range(len(sys.argv)):
    printf 'argv[%d] = %s' %(n, sys.argv[n])
```

```
argv[0] = argparse.py
argv[1] = arg1
argv[2] = arg2
argv[3] = arg3
argv[4] = this-is-one
argv[5] = this is one
```