

Python for Astronomers

An Introduction to Scientific Computing

Imad Pasha
Christopher Agostino

Copyright © 2014 Imad Pasha & Christopher Agostino

DEPARTMENT OF ASTRONOMY, UNIVERSITY OF CALIFORNIA, BERKELEY

Second Edition, January, 2017



Contents

1	Essential Unix Skills	7
1.1	What is UNIX, and why is it Important?	7
1.2	The Interface	7
1.3	Using a Terminal	8
1.4	UNIX Commands	8
1.4.1	Changing Directories	9
1.4.2	Viewing Files and Directories	10
1.4.3	Making Directories	11
1.4.4	Deleting Files and Directories	11
1.4.5	Moving/Copying Files and Directories	12
1.4.6	The Wildcard	12
1.5	SSH and Servers	13
1.5.1	Logging Into a Server	13
1.6	Setting up Aliases and Tab Complete	14
1.6.1	Tab Complete	14
1.6.2	Aliasing	14
2	Basic Python	17
2.1	Data-types	18
2.2	Basic Math	18
2.3	Variables	19
2.4	String Concatenation	21

2.5	Array, String, and List Indexing	22
2.5.1	Two Dimensional Slicing	23
2.6	Modifying Lists and Arrays	24
3	Libraries and Basic Script Writing	27
3.1	Installing Libraries	27
3.2	Importing Libraries	28
3.3	Writing Basic Programs	28
3.3.1	Writing Functions	29
3.4	Working with Arrays	30
3.4.1	Creating a Numpy Array	30
3.4.2	Basic Array Manipulation	31
4	Conditionals and Loops	33
4.1	Conditional Statements	33
4.1.1	Combining Conditionals	34
4.2	Loops	35
4.2.1	While-Loops	35
4.2.2	For-Loops	36
4.2.3	Nested For-Loops	37
4.3	Loading and Writing Files	37
5	Plotting	39
5.1	Basic Plotting	39
5.2	Subplotting	41
5.3	Plotting 2D Images	41
5.4	Loading Astronomical Fits Files	43
5.4.1	The Header	43
5.4.2	The Image	44
6	Classes and Object Oriented Programming	45
6.1	Defining Classes	45
6.2	Subclasses	46
7	LaTeX	49
7.1	Creating a Document	50
7.1.1	The Preamble	50
7.1.2	The Main Document	50

7.1.3	Sections and Subsections	51
7.1.4	Inserting Equations	51
7.1.5	Inserting Figures	52
7.1.6	Lists and Bullets	52
7.1.7	Wrap Up	53
8	HTML Editing	55
8.1	Basic HTML	55
8.2	CSS and Styles	57
8.2.1	Using divs	59
8.2.2	Tables and Lists	59
9	Glossary	61



1. Essential Unix Skills

1.1 What is UNIX, and why is it Important?

UNIX is an *operating system* which was developed at Bell Labs research center and was designed to be a multi-user system which could multitask more efficiently than previous systems. Many modern operating systems such as Mac OSX, Ubuntu, GNU, and many others are considered to be UNIX-like in the way they are designed—specifically in terms of their filesystems. To be clear, Unix is an operating system that handles files (and where they are stored within folders/*directories*) the same way your own computer does. The principle user-end differences between UNIX and other operating systems are the interface by which one interacts with the file system, and the way filesystems can be stored on servers that are accessible from any computer in the network (we'll talk more about this later).

1.2 The Interface

It follows that an understanding of UNIX can be extremely useful as many of its basic tenets can be applied to a variety of scientific systems. Alternatively, most current operating systems on personal computers offer what is called a graphical user interface*. This is the type of system most people are accustomed to, in which the primary means of interacting with the filesystem is via a mouse which can click to open up windows of different folders, etc. While GUI systems are usually more intuitive by nature, they are also inefficient (for example, to move a file from one folder to another one has to open two windows separately, usually by clicking through multiple other folders). In comparison, command line interfaces have a steeper learning curve, because they require knowledge of *syntax*, or the phrases and commands that can be interpreted by the computer. However, the advantage of learning this syntax is that the command line is an extremely efficient way of navigating an operating system. As a counterpart to the previous example, a single phrase in the command line (“`mv filename newlocation`”) can accomplish what took many clicks and drags in a GUI.



Note: Because these GUI's on personal computers usually operate over UNIX, one can usually find ways to access a command line interface for these computers as well, it just isn't the

primary interface by which most people interact with the operating system.

Perhaps most importantly, the majority of supercomputers and telescopes are operated by systems which utilize Linux operating systems which makes it especially useful for astronomers to learn. For example, one can often only access these systems through the use of the terminal's "secure shell" service, or more simply, **SSH**, which provides a remote, secure login. In utilizing the command line, or **terminal** as it will be referred to from here on, one can accomplish many tasks by simply typing a few commands rather than having to make several mouse clicks. Some examples which will be explicitly outlined later include removing files, creating folders, opening programs, and searching through droves of files for a specific keyword.

1.3 Using a Terminal

Most systems will have keyboard shortcuts to expedite and simplify the opening of a terminal application. In Ubuntu, one can press Ctrl+Alt+t to open a new terminal window. In Mac OSX, perhaps the simplest way to open the terminal application is to press Command + Spacebar then to type terminal and press enter. The new window should open in your home directory, which is the directory of your own user account which holds your documents, downloads, music, and other personal files. Not all operating systems have keyboard shortcuts for this but if one wishes, one could customize his/her own keyboard shortcuts to allow for this quicker functionality.

Once the terminal has been opened, there are really only two areas to understand. The **prompt** is a line of letters and symbols that appears on the left of the terminal window. What this prompt actually reads is different for different systems, and in fact can be customized to say whatever you like. Many systems will be set up so that the prompt indicates to some degree the current **path**. Path refers to the description of where you are in the filesystem, beginning with its most basic, or root directory. For example, if you were in your documents folder, the path to your location might be /root/home/users/your_name/documents. Different systems also have different names for their root directory. In some cases your prompt may be the full path, but it is often a shortened version which only indicates the name of the current directory; such a prompt might look like this: "systemuserid:documents%" (Prompts almost always end in a % or \$ symbol). Ultimately the prompt doesn't affect what you can or can't type; at most it can be a handy way of seeing where you are in a file system. The other area of the terminal is the actual command line, where you type the commands to be interpreted by the computer. We will cover in depth all of the commands you need to know to navigate a UNIX system. (See fig. 1.1).

1.4 UNIX Commands

UNIX terminals have a vast number of available commands which one can use but many of them are outside the scope of this course. For the majority of this course, you will need only a few simple commands which will be outlined and explored in this chapter. A more complete list of commands with shorter explanations is available in the appendix, and on the website as the "UNIX Guide".

We begin with the commands by which one actually navigates from folder to folder within a UNIX system. In the course of this textbook, we will frequently want to indicate commands to be typed into the terminal. Our format for doing so will be to represent the prompt with "», and to usually indent commands on separate lines: for example, to exit a terminal, type

>exit

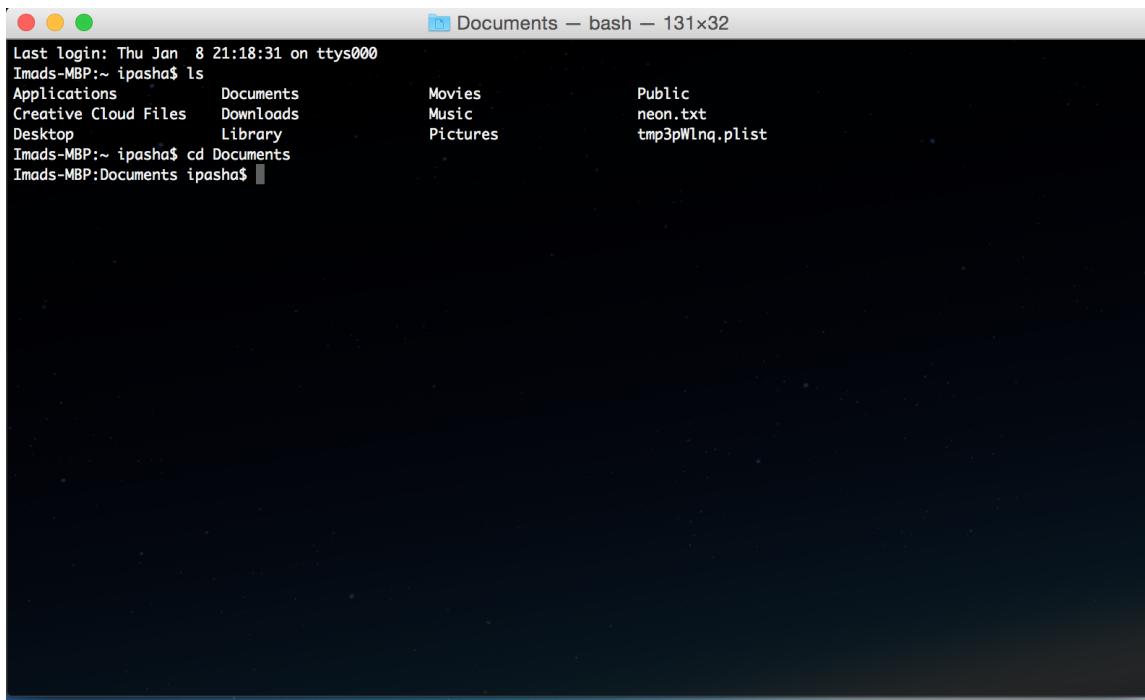


Figure 1.1: A typical terminal, with a prompt on the left hand side. An example command has been sent through, which displays the contents of a folder.

and the shell will close. Additionally, in this section of the text it will be useful to define a “typical” series of nested directories, so that as we practice navigation, we can use self consistent examples. For now lets use this typical tree:

```
root/physics/user/sally/documents/homework/python/week1/
```

1.4.1 Changing Directories

There is a single command by which one can navigate the entire UNIX directory tree of any system, and as there are several subtleties to it, we will discuss it in some detail. The command in question is “cd”. The syntax “cd” is interpreted by the computer to mean “change directories”. Clearly though, with just this command, it would be impossible for the computer to know where to change directories to. Because of this, the command cd takes what is called an *argument*. An argument is a part of the command necessary for it to function, but that is variable- the user can specify different values for the argument within a certain set of possibilities. In this example, the “cd” command takes as an argument a path location, for example in the command,

```
>> cd /root/physics/sally
```

/root/physics/sally serves as the argument to “cd”, it tells cd where to actually change directories to. The majority of commands in UNIX have arguments, although there are a few exceptions.

So the question is, how can we efficiently use the cd command to navigate between directories in UNIX? Luckily, the cd command has several built-in shortcuts that make navigating easier. However, we would like to point out that these shortcuts only work in certain situations, which will be described

below. On the other hand, there is one surefire argument for cd which will always work; unfortunately it is the most cumbersome. Below, we describe the various ways to phrase arguments for cd, the first being the sure fire method.

1. **The full path** : From any directory in a UNIX system, typing cd followed by a full path, starting with the root directory, will take you immediately to the specified location. This is possible because a full path is unique, and thus the computer knows exactly where you mean to go.
2. **A nested directory**: For example, if you are in the directory “/root/physics/sally/” and want to cd into the homework directory, you can simply type

»cd homework

This may seem confusing at first, because there is no “/” before homework. Essentially, the computer is interpreting your lack of a “/” to mean that the directory you are looking to cd into is within the one you are currently in. (It will complain if you give a directory as an argument that is not in the current one, if you use this syntax.)

3. **A secondary nested directory**: If you want to cd from a current location to two directories deeper in the nesting system, you can start with the syntax from (2), but continue it into a longer path, for example, if you were in “root/physics/sally/” and wanted to get into not just “homework”, but all the way into “python” :

»cd homework/python

and further nested directories can be strung onto the end, if desired.

4. **cd (space)** : defaults you to your username’s home directory. This will be different on different systems, but it is typically where your personal documents, downloads, pictures, etc., folders are stored.
5. **cd (space) .** : cd followed by a space and a period takes you to the current directory. In fact, a single period is always a shortcut for “current directory,” in lots of different commands. Using cd with just a period (as above) is pretty useless (it takes you nowhere). On the other hand:

»cd ..

is useful. This command will take you out one directory; i.e., if you were in the homework directory of the sample tree, “cd ..” would take you to the “sally” directory. This command can be strung together as well:

»cd ../../..

brings you out two directories, and so forth. The single dot is most useful in other functions, for example, when copying or moving files from remote directories to the directory you are in.

1.4.2 Viewing Files and Directories

You may notice that changing into a directory may change your prompt to reflect it, and/or typing “pwd” will “print working directory” and show you where you are. (As a side note, we type “print” though we aren’t printing things to a printer, because we are in essence printing the values from the memory to the console screen). But none of this actually tells you what is in the directory in question, unlike a GUI through which you actually see the files in a directory when you view it. No one’s memory is perfect, so there is of course a command for viewing the contents of a directory. Typing:

»ls

into the terminal will print a list of the files and folders in the current directory in which you are operating. ‘ls’ has many useful flags for various situations, which are listed in the appendix under the UNIX guide. Additionally, typing “»man(ls)” will bring up the manual for it (or any command you choose), right in the terminal. Note that ls is one of those few commands that does not take an explicit argument.

1.4.3 Making Directories

Now that we know how to view the contents in our directories, it becomes important for us to know how to create and delete files and directories as well. In order to create a directory, use the command,

»**mkdir desired_name**

which you may notice is a shortened version of the phrase “make directory”. For instance, if one wants to create a directory in which to store their specific information regarding this course, one would type into the terminal ‘mkdir python_decal’. If we use ‘ls’ we will see that this new directory is included in the contents listed. It is imperative to note that you should avoid using spaces when using the mkdir command as placing a space between two or more words will just end up creating directories named after each individual word. If you really must have a space in your directory, you can type a backslash before the desired space as this tells UNIX to ignore the space. On shared-network file systems, having an organized system of directories that make logical sense is very helpful, both for keeping yourself organized and for allowing you to direct others to specific files and folders more easily.

1.4.4 Deleting Files and Directories

Now that we know how to create directories, move between them, and look at the files inside, the next step is to learn how to delete things.

 NOTE: UNIX is not like a Windows or Mac where files are sent to a trash bin. When you hit delete, things are gone forever.

That being said, removing files is a relatively easy task in UNIX. If you are in the directory where the file to be deleted is stored, simply type

»**rm filename**

to delete it. (Note: ug astro servers have a setting turned on such that it will ask you if you are sure you want to delete. This is not native to the rm command, and on another system, do not expect rm to query you). Removing directories requires the use of a *flag*, or *option*. A flag/option is a modifier added after a command, before the argument, that changes exactly how the command is carried out. for example, typing “ls -a” will list “all” files, whereas “ls” alone usually ignores hidden files and directories. In this case, we need to make use of the “recursive” option:

»**rm -r directory_name**

which will go into a directory, delete the files within, and then delete the directory itself. The example here illustrates the syntax for using flags/options in general (with a dash preceding the flag).

There is a secondary way of deleting directories, which may be easier to remember:

»**rmdir directory_name**

which will also delete the directory in question.

1.4.5 Moving/Copying Files and Directories

The last major skill needed for operating in UNIX file systems is moving and copying files and directories from one place to another. Moving is done using the “move” command (who would’ve thought?):

```
>>mv filename new_location
```

(this assumes you are in the directory with the file to be moved. Depending on where you are moving the file to, the new_location could be as simple as “..” or as complex as a full pathname to another directory tree.).

Mv also gives you the option of changing the name of a file as you move it, for example:

```
>>mv file_name new_name new_location
```

would move “file_name” to “new_location”, changing its name to “new_name” along the way. Interestingly, because of this functionality, mv serves as the “rename” command as well. To rename a file, “move” it to a new name without specifying a new location to send it. If you want to copy a file instead of moving it, use:

```
>>cp filename new_location
```

which will create a copy and put it in “new_location”. Cp also has the ability to rename files in transit, by the same syntax as mv.

1.4.6 The Wildcard

One extremely useful thing to know about UNIX is the ability to use wild cards. Denoted by a “*” symbol, wildcards can stand for any character, or any number of characters. The strategic use of wildcards can save you a lot of time when working with large numbers of files. A few examples should make clear how wildcards are used:

1. **Deleting many files:** Say for example you wanted to delete all files in a certain directory that were of the type .doc (or docx for all you millenials). If you entered

```
>>rm *.doc
```

The wildcard would feed rm every file with any combination of characters that ended in .doc for deletion. In a similar vein, if you have a group of research files that all started with “simulation_run1” (where an example filename might be simulation_run10004.dat, simulation_run10005.dat, etc)

```
>> rm simulation_run*
```

would delete all of those files, as rm doesn’t care what comes after the “n” in run anymore.

2. **Copying files:** This is somewhat of a trivial expansion, but it is useful to note that more often than not you are going to be copying and moving large numbers of files rather than deleting them (archiving data for later is safer than losing it). It becomes clear now why many research processes that output many files have a very regular system for naming: it allows for the easy extraction of subsets or all files within UNIX systems. Wildcards also work within names, for example:

```
>>cp simulation*.dat newlocation
```

would copy all files starting with “simulation” and ending with “.dat” to a new location. This

can be handy if your software also outputs files with the same prefix but different file endings, and you only want the .dat files.

1.5 SSH and Servers

An extremely important aspect of working with the command line is ssh-ing into servers to work. A **server** is a computer or system of computers that store files and contain programs that are accessed and run remotely. Almost any computer can be converted into a server, though generally speaking servers are set up on computers with a lot of memory and free space. Astronomers use servers frequently because they allow for the storage of large (we are talking multiple terabytes) datasets. Additionally, it allows us to log in and work on our research from any computer with an internet connection, without needing all the data and programs installed on our personal machines. Finally, with multiple users on the same server, it becomes easy to share data, code, and any other file with collaborators, instead of having to email or otherwise transfer things to their computers.

1.5.1 Logging Into a Server

Usually, the only way to interact with a server is to log in via **SSH**. SSH is a terminal command standing for "secure shell host". When you run a command like

```
$ssh username@servername
```

in the terminal, your computer reaches out to the server and establishes a connection (assuming you have an account on the server). To give a concrete example, say you have an account under the name "sjohnson" on a server called "pepper" on UC Berkeley's astro network (it is typical for user accounts on servers to be first initial-last name, but it is up to the admin of the server how this works). You would type

```
$ssh sjohnson@pepper.astro.berkeley.edu
```

to log in to the server. The first time you try to SSH to a new server, you will be asked whether to trust the RSA key and add it to your trusted list, (just hit "y" and enter). The server will then ask you for a password. The admin for the server will have made one for you when they created your account; once you log in you can generally change this to something of your choosing using something like the "passwd" command (but this varies by system). Note that when you are typing in your passwords, nothing will appear on the screen- that's normal, just type the password and hit enter.

Now that you are in the server, everything works just like you are in a terminal on your own computer. You can ls, cd, and otherwise work with the files and programs installed on the computer you are ssh'd into. One extra step that's worth mentioning is that if you want to open programs with display windows (for example, ds9 which we will cover later), you will need to use the "-X" flag; that is,

```
$ssh -X sjohnson@pepper.astro.berkeley.edu
```

This will allow the windows to open on your computer. For this to work, you will need something called X11 forwarding. On a Mac, this involves installing something called "XQuartz" (easily googled), and on a PC it involves installing something called "Xming" and "Putty" which have X11 options. There is a guide to getting SSH working included with this bundle which covers this.

1.6 Setting up Aliases and Tab Complete

A lot of the commands we've covered so far are short and succinct- the point of the command line is to increase efficiency. Things like "ls" take very little time to type and can easily be used to navigate a file system. On the other hand, certain commands (particularly ones with long arguments, like the SSH commands above) are a pain to write out every time. There are two main ways of decreasing the amount of time you spend typing unnecessary information: *tab complete* and *aliases*.

1.6.1 Tab Complete

Tab complete is a feature of the terminal that allows you to quickly finish commands or filenames as you type them assuming they are unique. For example, let's say I have three files in a directory, "testrun1234876545635624.dat", "testrun49232450238472034.dat", and "testrun95432859234502598.dat". It would be extremely annoying to type these all out in something like a copy command. Notice, however, that after the word testrun, all three have a different character, which allows them to be differentiated by only the word testrun and the first number (i.e., if I asked you to give me "testrun9" there is only one you could choose). To quickly copy testrun9... we would simply

```
>> cp testrun9<tab>
```

and when we pressed tab, it would automatically complete the rest of the filename, letting us move on to typing in the new location. In fact, tab complete works on typing in locations as well- if you are typing in a long path name, you can tab complete each directory name as you type it, as soon as it's the only one with those letters/numbers in its name. Tab complete can also be useful when you haven't yet reached the unique part of a name- use it at any point while typing to see what options you have (everything starting with what you have already typed will pop up on screen, to remind you).

1.6.2 Aliasing

Certain commands cannot be tab completed. For example, if I start typing "ssh sjohnson..." I can't tab complete because Unix has no clue where I'm going with this command. But let's say there's a server we log into all the time- it would be very annoying to have to type the full ssh command every time. This is where aliases come in. An alias is a command you make up, stored in a special file on your computer (read by your terminal) that allows you to make your own shortcuts. The file that stores your aliases depends on whether your terminal is a bash, csh, tcsh, etc., but we will use bash here for reference (the format for the other terminal types is slightly different but easy to look up). Mac computers default to bash.

In your home directory (the one you are taken to by using "cd" with no arguments), there is a file called .bash_profile (it won't show up with "ls" unless you use "ls -a"). If you open it using your preferred text editor (see the included "vim guide"), you can add aliases (among other things) to be read by your terminal.

To add an alias, simply leave a line

```
alias 'mycommand'='realcommand'
```

For example, If I wanted to alias my long ssh call, I could do something like

```
alias 'pepper'='ssh -X sjohnson@pepper.astro.berkeley.edu'
```

Once the file is saved, open a new terminal, or source your bash profile by typing

```
source .bash_profile
```

in the terminal and you will find that typing in "pepper" runs the ssh command and asks you for your password to the server. This is also very useful for directory shortcuts; for example if you have a research directory buried in your filesystem somewhere, you could set an alias 'research'='cd researchfullpath' to make it easier to get to your research directory.

2. Basic Python

Python is a high level programming language which is easily readable and relatively simple in comparison to many other languages such as Java, Fortran, or C++. It has built-in support for object-oriented programming and functional programming as well. Ultimately, python has several key properties that has resulted in a trend towards its use in scientific computing, in astronomy and other fields alike. Those reasons are widely varied and often subtle, but we will try to cover some of the main ones below, and throughout this text.

- Python is non-compiled. With most older languages, (and those that require very high levels of speed maximization), the code you have written has to be *compiled* by a special program known as a compiler before it can be run. Programs that are compiled can be more efficient in execution, but removing that aspect allows python to be run line-by-line by the python interpreter, which is conceptually easier and often practically easier as well.
- Simplicity. Python is one of the simplest languages to work with, syntactically and organizationally. You can spend more time getting the work done rather than figuring out how on earth to code it.
- Variable declarations. As an extention of the above, many programming languages require that all variables (which will be covered soon) are defined explicitly at the top of a program (and what datatype they are). Python allows you to change these on the fly.

All of the above have made python one of the primary languages used by scientists to achieve their goals. As scientists, we are not really interested in the most subtle complexities of computer science, nor are we (usually) writing massive, conglomerate codes which require high speed and extensive abilities (C++ has over 775 pages of language legalese). Because of this, scientists have developed extensive libraries and tools to add on to python to make it even better for research, resulting in something of a positive feedback loop that is quickly leaving python the main choice for astronomers and physicists to code in. Of course, there are exceptions; when writing a specific type of program (like a hydrodynamics simulation), using a fast, compiled language like fortran is the way to go. But nowadays, people are writing fortran codes with python wrappers, so the end user can interact with the code entirely from python.

 For much of this course, we will be referring to the ipython interpreter rather than the regular python interpreter, as it has several added features which are useful to us.

2.1 Data-types

Python, like most programming languages, divides up all the possible “things” you can play around with into what are called ***data types***. These are the fundamental building blocks of your code, a property that every object/element/variable in your code will have, and which will determine the rules by which python operates on them. Some of these divisions seem obvious: clearly a word like ‘cat’ is a fundamentally different data type than a list of numbers [1,2,3,4,5]. Other divisions seem more arbitrary at first glance: for example python makes the distinction between integers (the counting numbers), and floats (numbers with decimals). It does so because of the way computer processors store information in bits, but it leads to the interesting (and important) characteristic that “42” and “42.” are different in python, and take up different amounts of computer memory. Essentially, what data-type a variable is determines what rules python applies to it. Some basic data types are listed and defined below, and you will learn more about them as we use them:

1. **Integers:** the counting numbers. Ex: 1,2,3,4,5,...
2. **FLOATS:** Decimal numbers. Ex: 1., 2.345, 6.3, 999.99999, ...
3. **STRINGS:** An iterable data-type most commonly used to hold words/phrases or path locations. Denoted by single or double quotes. Ex: ‘cat’ , ‘/home/ipasha’ , “1530”, ...
4. **LISTS:** Stored lists of any combination of data types, denoted with brackets. Ex: [1,2,’star’,’fish’] or [1, 2, [3, 4, 5], ‘star’] (notice that you can have lists within lists)
5. **ARRAYS:** Like lists, but can only contain one datatype at a time, and have different operations. Defined in numpy, not native python.
6. **TUPLE:** Also like a list, but immutable (un-changable). Somewhat like a read-only list. These are defined with parentheses. Ex: tuple1 = (‘hi’, 1, 4, ‘bye’)
7. **Dictionaries:** A collection of pairs, where one is a “key” and the other is a “value”. One can access the “value” attached to a key by indexing the dictionary by key:
»dictionary_name[‘key’] (more on this later).
8. **Boolean:** A data type with only two possible values: True, or False. They are used in conditional statements.

2.2 Basic Math

Within the python interpreter (or indeed in any written code) you can perform simple to very complex mathematical operations. As a few introductory examples we can see how adding and subtracting works in ipython. From the terminal (assuming you have a distribution of python installed), simply type "ipython" to begin an ipython session, and you should see your prompt change to "[IN]:" with a line number. You are now in ipython. Try the following and see that they match up. "

```
[IN]: 3 + 5
[OUT]: 8
[IN]: 9-3
[OUT]: 6
```

We can also test out multiplication and division (denoted in python with * and /):

```
[IN]: 4*3
```

```
[OUT]: 12.0  
[IN]: 1/2  
[OUT]: 0
```

But wait! Clearly $1/2$ is not 0. (Unless we've jumped into some strange parallel reality). The reason we are getting 0 here is that python is performing integer division, meaning the answer has to be an integer. In this sort of situation, python simply rounds down to the nearest integer. The solution to this is to cast either the “1” or “2” (or both) as floats rather than integers. Only one is required to be a float because if one number in an operation (like addition, subtraction, multiplication, division, exponentiation, etc) is a float, it will convert all to floats and express the answer as a float. For your general knowledge, there is a function for converting integers to floats, and it looks like this:

```
[IN]: float(2)
```

However, there is a much faster way to create floats when you are entering a number manually, which is simply to add a decimal (period) to any number. Try it yourself: demonstrate that $1./2$ and $1/2.$ both output the proper answer. The place when the `float()` command comes in handy is when you have a variable (say, called “`x`”) in your code, and you don’t necessarily know what its value is, perhaps it is the sum of many calculations, but is just an intermediary holding value. If before the next stage of calculations you require it to be a certain data type, you can use this hard casting, like

```
[IN]: x = float(x)
```

or

```
[IN]: x = int(x)
```

(Which will convert it to an integer if it is not already).

The other basic math operation in python is exponentiation. In python this is denoted with a double asterisk (“`**`”). For example:

```
[IN]: 2**3  
[OUT]: 8
```

To perform more complicated math like `sin`, `cos`, `sqrt`, etc., requires the use of some additional packages, which is the primary focus of Chapter 3.

2.3 Variables

While using python as a calculator can be fun, the real power of programming comes in being able to store things (numbers, lists, etc) as *variables* and access them later. Declaring variables in python is easy; you simply type a desired variable name, an equal sign, and what you want it to be. For example:

```
[IN]: x = 5.0  
[IN]: y = 'cat'  
[IN]: Berkeley = 'no life ' + 'bad grades ' +'no sleep '
```

would set the variable `x` to the floating point number 5, set `y` to the string ‘cat’, and set `berkeley` to the concatenated string ‘no life bad grades no sleep’ (more on string concatenation in a bit). Notice that python doesn’t output anything when you declare a variable as it did when you entered a math operation. But rest assured, those values are stored in the computer. If you type:

```
[IN]: print x #print(x) in python 3.0+
```

[OUT]: 5.0

it will output the value attached to your variable. The print command is almost always how we check in to see what a variable's value is at a given point in a code, and it's an extremely useful way to begin debugging your code if something isn't working the way you think it should be.



Note, if you were wondering what the # symbol above indicated, it is what is called a comment. Comments in python allow you to write notes and guides to yourself and others in your code. On a given line in python, anything coming after a # symbol will be ignored when the code is run, allowing you to type anything you want there.

As a shortcut, in our version of python,

[IN]: y

[OUT]: 'cat'

simply typing a variable name alone will result in a printing of the variable value. Keep in mind this only works in the terminal. Later, when we start writing scripts, you can't just have a line "y", you will need to say "print y" for it to work.

Variables in python are mutable- that is, you can change them, within certain bounds. Most simply, if you consecutively typed:

[IN]: x = 5

[IN]: x = 3

then printed "x" you would find it is equal to 3. You can also use variables to change themselves:

[IN]: x = 5

[IN]: x = 2*x+3

In this case, the new value for x at the end of the line would be 2 times the value of x going in, plus 3. (in this case, 13). You can also add, subtract, and multiply variable, if they are of the right data type:

[IN] : x = 5.

[IN] : y = 6.

[IN] : z = x + y

[IN] : x = 2*z

[IN] : y = x/z

That is probably a bit confusing to follow, and illustrates why typically we avoid such oft redefining of variables, and instead come up with new variable names to store the various sums and products.



Note: these calculations worked because both 5 and 6 were floats, and it makes sense to perform math on them. If you attempted to multiply a string like 'bozo' with a number, python would throw an error (and rightfully so). However, there are times when the operators like +, -, *, and / can operate on non integers and non floats. (For example, as we saw before we can "add" (concatenate) strings together into a single string).

Additionally, an interesting property of the array data type is the following:

[IN]: Array = np.array([1,2,3,4,5]) #Dont worry about the np syntax right now

[IN]: x = 15.8

[IN]: z = x + Array

```
[IN]: print z
[OUT]: array([16.8, 17.8, 18.8, 19.8, 20.8])
```

Basically, if you add, subtract, multiply, or divide an array by a given value, each individual value in the array has that operation performed on it.

There is definitely subtlety in which data types can go together, and in what ways data types can be successfully changed (for example, the ‘int’ command we discussed can never convert ‘cat’ to an integer, and will throw an error). We hope to cover much of these intricacies in time, but much of it is common sense and experimentation.

2.4 String Concatenation

While the majority of data we work with is stored in arrays and lists rather than strings, there are still many useful tasks that can be accomplished with strings. As we know, strings are basically how you can work with words like ‘music’ in python. Additionally, strings are often an input type for arguments to functions, and are used to read in files. This is where we will focus. For example, to read in a txt file from your computer to python:

```
[IN]: file = np.loadtxt('/home/sally/file.txt')
```

will do the trick. (Don’t worry about the function np.loadtxt just yet, just notice that the format of the location of the file was within a string).

Now say that you want to a hundred data files, with names file0.txt, file1.txt,...,file99.txt. Typing the loadtxt command 100 times would be rather tedious. Instead we can use a for loop and string concatenation to simplify the task. There was an example of string concatenation earlier. In essence, it takes multiple strings and combines them into one:

```
[IN]: print 'cat' + 'dog'
[OUT]: 'catdog'
```

Notice that there is no space between cat and dog. Python shoves the two strings together exactly character by character. To create the string ‘cat dog’, we could either have defined the first string to be ‘cat ‘ (with a space after the t), or defined the dog string to begin with a space. Additionally, we could have concatenated in the space manually:

```
[IN]: print 'cat' + ' ' + 'dog' # One space between the second pair
of quotes
[OUT]: 'cat dog'
```

This allows us to do something like the following:

```
[IN]: prefix = '/home/sally/file'
[IN]: suffix = '.txt'
[IN]: for i in range(100):
[IN]: ... filename= prefix+ str(i) + suffix
```

Don’t worry about using for-loops yet, but notice that what we are doing is sequentially setting the variable filename to the ‘/home/sally/filei.txt’ where that “i” in there is 1, 2, 3, 4, 5, … , 100. This allows you to open multiple files and do things to them efficiently. Once we have learned how to use for-loops, if-statements, and while-loops, we will return to this example in more detail.

2.5 Array, String, and List Indexing

Generally speaking, as a scientist you will be working with large datasets which contain the above variable types (strings, floats, ints) within storage data-types like lists, dictionaries, and arrays. For lists and arrays, there is a uniform procedure for extracting information out of these container objects.

This procedure is known as slicing/indexing, and the method is as follows: given a list, array, or string (all 1-dimensional right now for simplicity), each entry is assigned an index. By convention (that you may reasonably find annoying), this index almost always starts with 0, rather than one. Below we have a sample list, with the indices for each entry listed below: i.e., ‘1’ is the 0th entry (or

```
[IN]: list_1 = [1 , 2 , 4 , 'cat' , 5 ]      #Spaces added for clarity
Index:      0     1     2     3     4
```

Figure 2.1: List elements and corresponding indices.

element) in the list, and 5 is the 4th. Lets say then that you wanted to extract the 0th entry from the list, to use for some other coding purpose. The way to slice a variable (of the proper data type) is by typing the variable name, attached on the right with closed brackets and an index number. For example, to extract the 0th element and set a variable for it:

```
[IN]: list1 = [1, 2, 4, 'cat', 5 ]
[IN]: x = list1[0]
[IN]: print x
[OUT]: 1
```

Notice that for a list, each entry is the “thing” between the commas, so typing

```
[IN]: print list1[3]
```

would print

```
[OUT]: 'cat'
```

as the string ‘cat’ is the third entry (if you start counting at 0).

Arrays can be sliced in precisely the same way as lists. Interestingly, strings can also be sliced. So if we had set var = list1[3], we could then type:

```
[IN]: print var[1]
```

and get

```
[OUT]: 'a'
```

Unfortunately, if you have a long integer like x = 1234456653453, you can’t slice through x the way you can with lists, arrays, and strings. What you can do is turn (or cast) x as a string:

```
[IN]: x = 123456789
[IN]: x = str(x)
```

Now that x is a string, you can happily index it:

```
[IN]: print x[0]
[OUT]: '1'
```

Normally if you try to convert a string like ‘cat’ to a float or int, python will hate you. But if you attempt to convert a string that only contains numbers, python can successfully make the conversion. So we can get the integer number of the 0th element of 123456789 like so:

```
[IN]: x = 123456789
```

```
[IN]: x = str(x)
[IN]: zeroth = int(x[0]) # or zeroth = float(x[0]) for the float
```

Sometimes we want more than a single value from a list/array/string. There is also a way to slice through multiple indices at once. The format is as follows. Take the previous example of the string '123456789'. Say we want the 0th, 1st, 2nd, and 3rd elements to be pulled, turned back into an integer, and set as the value of the variable H:

```
[IN]: H = int(x[0:4])
```

So basically, now instead of a single index in the brackets, we have a start index, a colon, and an end index. Also note, python will go up to, but not include the end index given. As a shortcut, if you are starting from the beginning, or slicing from some midpoint to the end, you can omit the 0 before the colon, or the final index after, i.e.,

```
[IN]: print x[0:4]
```

is equivalent to

```
[IN]: print x[:4]
```

and if you don't know how long an array is but want to index it from its nth element to the end, simply

```
[IN]: print x[n:]
```

You can also slice through an array backwards using what are known as **negative indices**, that is, and index of "-1" refers to the last element in an array/list/string, and "-2" the second to last, etc. An example of indexing from the last to 5th from last element might be

```
[IN]: print x[-1:-6]
```

2.5.1 Two Dimensional Slicing

Strings and lists now primarily excluded, often astronomical data (like images from telescopes) are stored in 2d arrays- essentially a large grid or matrix of numbers described by 2 indices, a row and a column. (If it helps, you can think of the arrays above as matrices with row length 5 and column height 1, so you only needed to index which column you were interested in). Lets cut to the chase with an example. Let's say "A" is a 2d array that looks like this:

```
print A
[[1 , 3, 4, 5, 6]
 [ 4, 5, 9, 3, 7]
 [ 9, 4, 6, 7, 1 ]]
```

Notice the way python is handling the list structure here; there are three one dimensional lists stacked within an extra set of brackets (like a list of lists). We slice it with two indices, row, then column.



Be careful, row then column translates into (y,x), which is the opposite of how we are usually taught to determine ordered pairs of coordinates.

To pull the 3 in the second row, we type:

```
>print A[1][3] # remembering that the "1" in the upper left corner
is in the 0th row and 0th column
```

Alternatively, you can use the comma syntax `A[1,3]` to equal effect. To pull the 6 in the first row:

```
>print A[0][4]
```

Try it out: what would be the way of slicing to pull the 4 in the last row? Using the same colon notation from above, how would you pull a whole row?

Given a 2D array, you may want to take a chunk of it, either end to end, or somewhere in the middle. The syntax for doing so is a combination of commas and colons. Remember that colons either separate a start and end index, or refer to a whole column if no start/end are specified. Lets say you have an image with 1000x1000 pixels, which you are viewing as a 2d array of 1000x1000 values. The following is a list of example slices, from which you can infer how to slice any section you'd like.

Exercise 2.1 Slicing Images

1. »array[350:370,:]
 - takes the full rows 350-370 in the image (fig. 2.2)
2. »array[:,350:360]
 - takes the full columns 350-360 in the image (fig. 2.2)
3. »array[350:370, 350:360]
 - takes the box in fig. 2.2. (the region between/including rows 350-370 and cols 350-360)

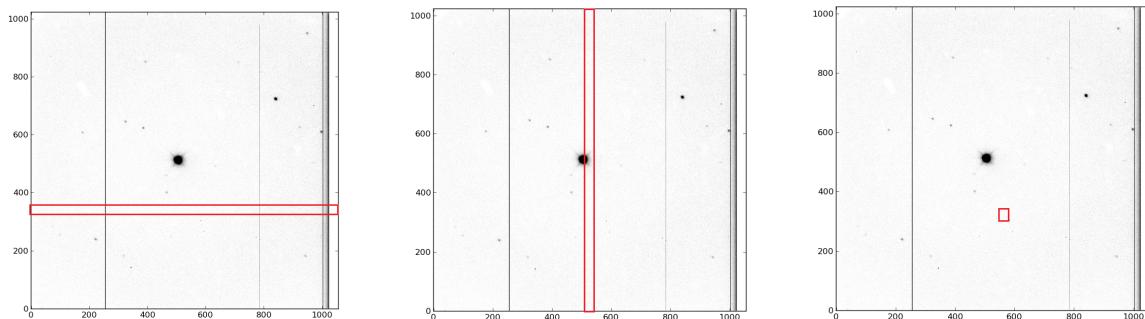


Figure 2.2: Left: Rows 350 to 370 pulled. Center: Columns 350 to 360 pulled. Right: Box of rows 350-370, cols 350-360.

2.6 Modifying Lists and Arrays

While we have shown how you can create a list of elements and how to extract and see specific values within them, we haven't talked about adding and removing, or changing, elements of lists and arrays. Say we have a list of integers as follows: `[1, 2, 3, 4, 5, 6, 7]`. The most simple way to change a value within the list is to set a new value equal to the slice of that list. For example:

```
[IN]: list1 = [1, 2, 3, 4, 5, 6, 7]
[IN]: list1[2] = 'hi'
```

When we print the list one now, we will see that the third element of the list (formerly the integer 3) will have been replaced:

```
[IN]: print list1  
[OUT]: [1,2,'hi',4,5,6,7]
```

Of course, now that there is a string in our list, we can't do things like sum(list1) and expect to get a proper value. Now let us see how to delete values out of a list. This will involve use of the 'del' command. If we continue using our list1 from above:

[IN]: del list1[-1]

will delete the last entry in the list. (Note, a -1 index means the last element in a list or array, and -2 references the second to last, etc). We could also have used forward indexing just fine.

 Be careful with this command. Remember that once you delete an entry, the indexes corresponding to all the remaining values get shifted. So you could run del[0] 3 times and it would keep deleting the “new” 0th entry in the list.

Now while the principles of what we've used applies equally well to arrays, the syntax of how everything is done will be somewhat different, due to the way numpy.array was created. We will discuss working with numpy arrays later on, after having formally introduced numpy and other scientific packages.

Before moving on we'd like to list 2 very basic image (2d array) manipulation commands that might come in handy. We will go through more in much more depth later.

Lets go back to our 1000x1000 entry 2D array. There are simple commands for if you want to flip the image vertically and horizontally. For a vertical flip (about the horizontal centerline):

[IN]: flip_vert_array = array[:: -1] # see fig. 5

(this is shorthand for array[:: -1, :] - it does the same thing but seeing it the second way makes the next command make sense). For a horizontal flip (about the vertical centerline):

[IN]: flip_hor_array = array[:, :: -1] # see fig. 6



3. Libraries and Basic Script Writing

We saw earlier that one can use the ipython interpreter to do basic math, and that there were various data-types that come “preinstalled” within python (like lists, strings, integers, etc). However, once a code requires more sophisticated analytical tools (especially for astronomical processes), it becomes apparent that the vanilla ipython functions are not sufficient. Luckily, there are hundreds of functions that have been written to accomplish these tasks, most of which are organized into what are called *libraries*. Most python distributions (like enthought and anaconda) come with a lot of these libraries included. There are 3 key libraries that we will be discussing in detail. When writing code for science, you will essentially always have to use these 3: numpy, matplotlib, and (sometimes) astropy. Numpy is an extremely versatile library of functions to do the things ipython can’t. For example, while you can create a polynomial yourself ($x^{**2} + 3*x + 1$), ipython provides no way to make sine and cosine functions. Matplotlib is a library with functions dedicated to plotting data and making graphs. Astropy is a library with functions for astronomical applications: we will be using it to import fits images (images taken by telescopes).

3.1 Installing Libraries

Most major distributions of python (like anaconda and enthought) come with important packages like numpy preinstalled. However, for most smaller packages, like astropy, or pyfits, or those for programs you are using written by other scientists, you will likely have to install them yourself. The easiest way, when available, is to use pip. If a library has been added to pip, (a package installer already on most windows, mac, and linux computers), then installing a new one is as simple as

```
>> pip install packagename
```

The easiest way to see if a package is in pip is to just try to pip install it, if it works then it is, if it says "not found" then you might have to do some google hunting to see how to install it- usually by downloading a folder and running a python script within called setup.py by the following:

```
>> python setup.py install
```

If, for example, you find you don't have pip already, you can download it from their website and install it via the above method. Then you can pip install other things.

3.2 Importing Libraries

Because these libraries are not automatically loaded up when python runs, we have to *import* them. The syntax variations for doing so are shown below:

```
[IN]: import numpy  
[IN]: import matplotlib.pyplot  
[IN]: import astropy.io.fits
```

Notice that there is a dot notation within some of the imports. This is associated with classes. These libraries are huge, and loading all of the functions in them is unnecessary if you know what you want. Since pretty much everything you need to plot is within the "pyplot" sub-library of matplotlib, we can just import that sub-library. Now that the functions are loaded, you can use them in your code. However, the syntax for using them is slightly different than that of normal python functions. Because python needs to know where the function you are calling is coming from, you have to first write the library, then the function, using the same dot notation as above. For example, a sin function might be:

```
[IN]: import numpy  
[IN]: x = numpy.arange(100)  
[IN]: y = numpy.sin(x)
```

Clearly, writing out numpy all over your code would take forever. Luckily, python allows us to import the libraries and name them whatever we want for the purposes of our code. 2 standard choices are

```
[IN]: import numpy as np  
and
```

```
[IN]: import numpy as n
```

There are multitudes of functions in these libraries, which can be intimidating. If you are trying to figure out what a numpy function is for a task you want to do, google and stack overflow are your friends. Additionally, we will include a list of commonly used numpy/matplotlib functions in the appendix of this textbook, and as a separate document available on the website.

3.3 Writing Basic Programs

Thus far, we have been working entirely in the ipython interpreter. While this is a quick and easy way to practice with python, it is unsuitable for the majority of things that you might want to accomplish in python. Say you wanted to program a basic video game (a good final project idea!). This would be essentially impossible to do in the command line, and you would only get to run it once (and would have to retype all the code to create the game every time you wanted to re do it).



Note: python cares about spacing when it involves tabbing at the beginning of a line, but doesn't care about spaces otherwise; i.e., `x=5` is the same as `x = 5` is the same as `x= 5` and `x =5`. We recommend using spaces between characters just to make your code easier to read.

A **program** is a self-contained list of commands that are stored in a file that can be read by python. Essentially, it is a text file, with each line being the exact syntax you would have typed into the

terminal. Python then opens up your program and runs it through the interpreter, line by line. For example, if this is what you did in interpreter before:

```
[IN]: import numpy as np  
[IN]: import matplotlib.pyplot as plt  
[IN]: x = np.arange(100)  
[IN]: y = x**2 + np.sin(3*x)  
[IN]: plt.plot(x,y)  
[IN]: plt.show()
```

then you could write a program in a text file that looked like this:

```
# START  
import numpy as np  
import matplotlib.pyplot as plt  
x = (100)  
y = x**2 + np.sin(3*x)  
plt.plot(x,y)  
plt.show()  
# STOP
```

You would type this up in any plaintext text editor (we will show you how to use vim and emacs, but feel free to use sublime or any coding editor of your choice), and save it as something like ‘simple_program.py’. Then to run it, simply open up the interpreter (in the same directory as the file) and type:

```
[IN]:run simple_program.py
```

and your plot will be output. You could’ve done the same thing in the interpreter, but now you can easily modify one element (like change the $3*x$ to a $4*x$) without having to change anything else, you can share the code with others, and you can write code more than a hundred lines without worrying about typing in all those lines every time!

3.3.1 Writing Functions

Remember how useful having those numpy functions was? Well, we will attempt to show that for a large complex code, having everything isolated in your own **functions** will make your life much easier. One of the main reasons why is debugging: if there is a problem in your code, as long as you know which functions are working properly (and you can test them individually), you can track down problems much more quickly.

When you define a function, you are defining a block of code that only gets run when someone calls your function, giving it the inputs you define. One thing to note is that a function can ONLY use variables that are input, or are defined within the function itself. So if you have a variable called ‘zeta’ in the main body of your code and you try to use that variable in a function without making it one of the inputs, python will throw an error. All of this seems pretty abstract, so let’s move to some concrete examples.

Exercise 3.1 A Function to Slice 2-D Arrays.

Start by remembering the syntax for 2d images: `img[start:end,start:end]` for rows, cols. Now, lets say you had a few hundred images that you wanted to slice in a similar way. Indexing each one individually would be tedious. We will now launch into defining the function. Try to follow, and we will meet up after to discuss and explain all the syntax along the way.

```
def image_slicer(image, rs,re,cs,ce):
    output = image[rs:re,cs:ce]
    return output
```

Several important notes: first, to define a function, you begin a line with “`def`”, a space, and then the name you want to give the function. You then put parenthesis right after the name, and define what inputs it should have, and then end the line with a colon (this is very important). Then, everything that’s part of the function is indented from the first line (one tab). (Also very important.) You can define functions within other functions and loops, and the tab rule still applies: if you define a new function, everything that’s part of it is tabbed in. Finally, there is a return statement at the end, which is what the function is outputting. Let’s look at the part of the function (`image, rs,re,cs,ce`). The names I give these quantities does not matter. They don’t have to match anything else in my code, they just delineate 4 different inputs. the only thing that matters is that those names (`image, rs, re, etc`) are the ones used consistently **WITHIN** the function. It is easy to mistake this and think that you have to input a variable named `rs` into this function: you don’t. You can input any variable, or just a number, so long as it satisfies the rules for what those values end up doing. To show this, here is an example of using this function, assuming I have 3 variables, `img1`, `img2`, and `img3`, which are 2d arrays:

```
img1_slice1 = image_slicer(img1,10,40,20,70)
img1_slice2 = image_slicer(img1,500,560,410,420)
img2_slice = image_slicer(img2,0,-1,0,-1)
```

What we have done here is set a variable named `img1_slice1` equal to the output of the function `image_slicer` (which returns the sliced image). That output is `img1[10:40,20:70]`. (So we can see that we will get errors if we put in pairs of numbers that are not ascending, or if we put in variables that aren’t integer numbers). Take a look at the last one. What slice of the image do you think it returns? ■

Once we have gone through the magic of conditional statements and for-loops, we will see that we can easily write a loop that requires only typing a call to our new function once, but will return slices for many images.

3.4 Working with Arrays

Earlier we discussed how you can initialize a list, add to it, replace values in it, etc. We will now repeat the discussion with the syntax for numpy arrays, given that we now know how to import numpy into our code.

3.4.1 Creating a Numpy Array

If you have numpy imported as `np`, then you can initialize a numpy array by setting your variable `= np.array([1,2,3])`. Basically, you make a list, surround it in parenthesis, and attach it to the call `np.array`.

You can also initialize empty arrays by typing `np.array([])`.

If you want to create an array of increasing counting numbers (1,2,3,4...), you can set a variable = `np.arange(start,stop,step)`. Start is 0 by default, and step is 1 by default.

If you instead want to make an array between two numbers where you specify some number of subdivisions, use `np.linspace(start,stop,divisions)`. For example, `np.linspace(0,1,100)` would make an array between 0 and 1 with counting by hundredths.

If you need an array of 0's of a certain length, you can use `np.zeros(length)`.

Note: You can convert any list of floats/ints into an array simply by setting it equal to `np.array(listname)`.

3.4.2 Basic Array Manipulation

Remembering back to lists, the syntax for appending was `listname.append(newvalue)`. For arrays, we call the specific function `np.append(arrayname, new value)`.

If you need to change a value in an array, the syntax is identical to before, simply set `arrayname[index]` equal to whatever you need.

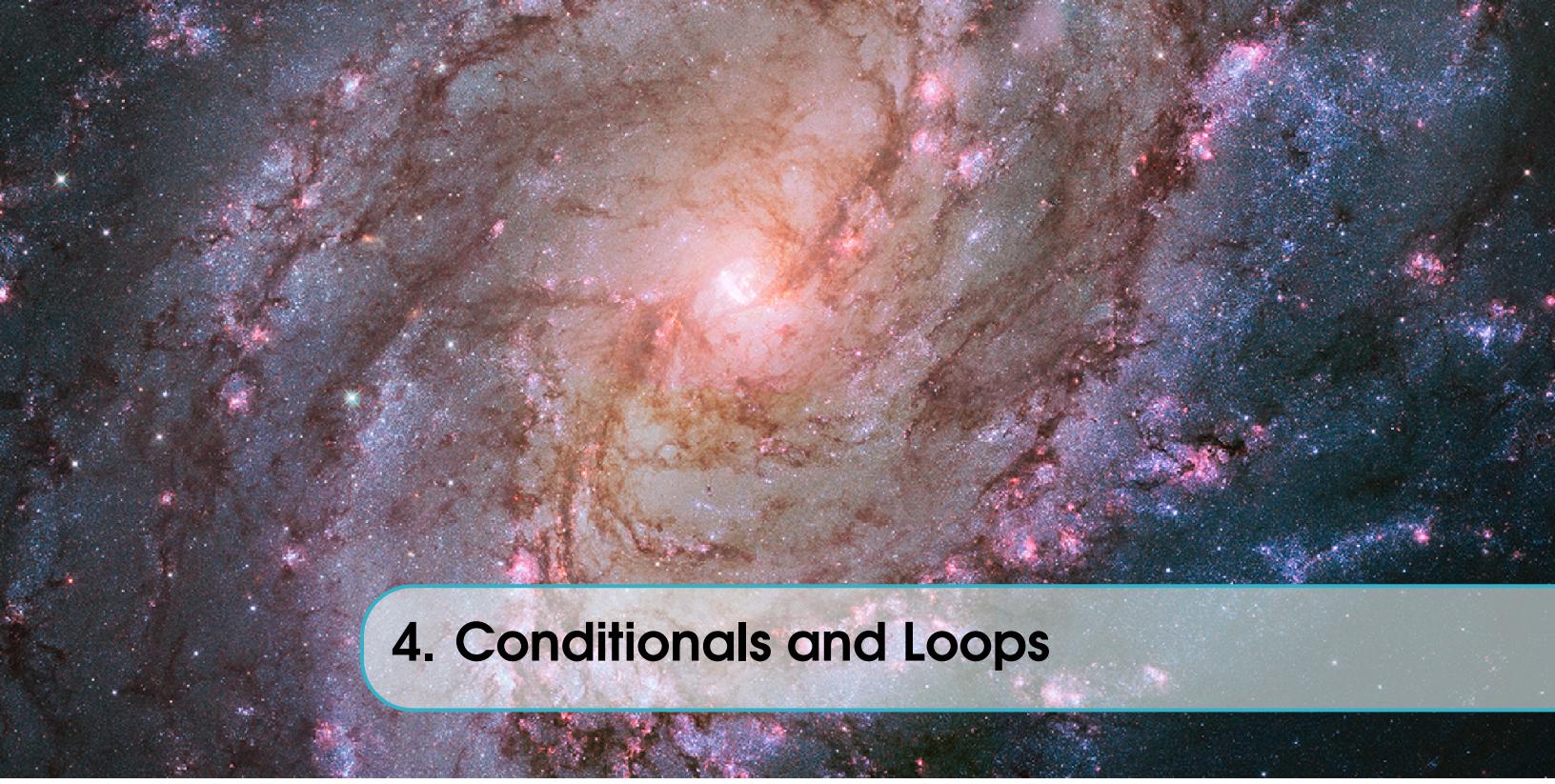
To delete values from an array, you can use `np.delete(arrayname, indices)`, where indices can be a single index or a range.

To insert values into an array, call `np.insert(arrayname, index, value)` and your value will be inserted at the index specified.

If you want to append one array onto the end of another (i.e., concatenate them), you can't use the '+' syntax used for strings, because you'll end up making a new array, the same size as the originals, with each new value being the sum of the two values in corresponding positions in the original arrays. Instead, we need to call `np.concatenate(arr1, arr2, ...)` to join them together.

Alternatively, if you have an array you need to split up, you can use `np.split(arr, indices)`. If you specify a single number, like 3, it will attempt to divide your array into 3 equal length arrays. If you provide a range of indices in order, it will know to split your array at those spots.

There is a ton more fiddly things you can do with arrays, particularly once you start working with 2 and 3 dimensional arrays. We will touch on that in Chapter 5, but primarily the `scipy` documentation and the web are good resources for learning about numpy array functions.



4. Conditionals and Loops

We saw in Chapter 3 how to create programs and run them in python. That powerful tool allows us to save text files containing coherent sets of python commands which python can run for us all at once. As of now, understanding how python interprets our simple programs is easy: it takes each line and enters it into the terminal. The real power of programming, however, lies in our ability to write programs that don't just contain a list of sequential commands. We can write code that repeats itself automatically, jumps around to different sections of the document, runs different functions depending on various inputs, and more.

4.1 Conditional Statements

As you might have guessed from the chapter title, we create programs like this by implementing various *conditional statements* and *loops*.

Definition 4.1.1 A **conditional statement** begins a defined, separated block of code which only executes (runs) if the conditional statement is evaluated by the interpreter to be “true”. Essentially, you are telling the computer “only run this block of code IF some condition is true.” The condition itself is determined by the programmer.

Let us start with some examples of conditional statements. The primary conditional you will use is “IF”. The syntax for creating an if-statement is as follows:

■ **Example 4.1 A Simple Conditional**

```
x = 5
y = 7
if 2*x**2 > y**2:
    print 'Wow, that's cool!'
```

■

Notice how the syntax is somewhat similar to defining a function. We start the line with the word “if”, which is a special word in python (and your text editor will probably color it differently) that

tells the interpreter to evaluate the truthiness of the rest of the line, up to the colon (again, the colon is important, don't forget it). In the case above, the if-statement would indeed print "Wow, that's cool!", because $2 * (5^2) = 50 > 49$. In this case of course, because x and y were simply defined to be numbers, the condition would always be true, and the print statement would always occur. But most of the time in your code, you have variables which are arrays, or parts of arrays, and the values have been changed in various steps of the code that you can't keep track of. Also note that, like for functions, all lines to be considered part of the conditional must be indented one tab.

To create a conditional with an "equals" condition, you have to use the strange syntax of the "==" double-equals, in the spot where you otherwise had > or <. The reason for the "double>equals" notation is that in python, a single '=' sign is reserved for setting the values of variables. As we will mention later, the "+=" notation means "set x = x+1". Some other conditional combinations are "not equal," given by "!=", greater than or equal to, ">=", and less than or equal to "<=".

Conditional	Symbol	Conditional	Symbol	Conditional	Symbol
Equals	==	Greater than	>	Less than	<
Not equals	!=	Greater than/equal to	>=	Less than/equal to	<=

Table 4.1: Symbols for various conditional statements

4.1.1 Combining Conditionals

We are not limited to one conditional per statement; we can combine as many as we need (within reason).

Exercise 4.1 Multiple Conditionals

```
x = raw_input('enter a number')
x = float(x)
y = 15
z = 20
if (x > y) and (x != z):
    print 'Nice!'
if (z > x) or (x != y):
    z = x+y+z
```

So here we have 2 if-statements, with the two possible combinations of conditionals, 'or' and 'and'. These statements can be combined indefinitely (for example, if ((a and b and c) and (d and f)) or (g +1>y) demonstrates how you can combine 'and' and 'or's' to suit your needs). ■

From now on, we will begin dropping new python commands and code into our examples, and will explain them either in comments in the code, or after the example. In this example, the command `raw_input('text')` prints 'text' to the screen and waits for the user to enter something. Whatever is entered is stored as a string in the variable x. (So above, if you said "enter a number" and a user entered a letter, the code wouldn't work).

So using the if-statement we have been able to set off blocks of code to be run only if some combination of conditionals is true. What happens otherwise? Typically we include an "else" statement following the if block, to determine all other cases.

■ Example 4.2 An Else Statement

```
x = raw_input('enter a number: ')
if int(x) ==5:
    print 'Wow, this was an unlikely coincidence.'
else:
    print 'Well, that's interesting.'
```

■

If your ‘else’ statement contains an if statement as well, you can use the “elif” command, which stands for else if. This saves you the trouble of an extra indent.

Exercise 4.2 Using Elif

```
if x < 0:
    print 'Negative'
else:
    if x==0:
        print 'Zero'
    else:
        print 'Positive'
```

Can be condensed to:

```
if x < 0 :
    print 'Negative'
elif x ==0:
    print 'Zero'
else:
    print 'Positive'
```

So now we know how to set up a “fork” in our code, to allow it to go in different directions based on various conditions. There is another type of block which instead continues to run the block over and over as long as some condition is met (to be clear, we refer to block as the indented section of code within various loops, conditions, functions, etc). This is known as a **while-loop**.

4.2 Loops

The two primary loops in Python are the while and for loops:

Definition 4.2.1 A **while-loop** is a set off block of code that will continue to run sequentially, over and over, so long as a certain condition is met.

Definition 4.2.2 A **for-loop** is a set off block of code that contains a temporary variable known as an iterator, and runs the block of code over and over for different specified values of that iterator.

4.2.1 While-Loops

Lets begin with a simple example of a while-loop.

■ Example 4.3 A while-loop

```
x = 100
while x > 5:
    print x
    x = x - 1
```

■

What's going on here? We initialize x to be some value. The next line of code read by the interpreter (remember it goes line by line) tells it that as long as x is greater than five, keep running the indented code over and over. The indented code in question prints x, then sets x = x-1. Eventually, after 95 times through the loop (and 95 prints), x would become $6-1 = 5$, which would no longer satisfy the while statement. The interpreter would then move on to the next line of code in the document. This brings up a very important point: you can see that if we had not included the "x = x-1" part of the code, x would *never* end up being 5 or less. Thus, your code would hang in this loop for all eternity. Luckily, if you find yourself in this situation, there is hope besides frantically shutting off the computer. Python interpreters have built in keyboard shortcuts to interrupt and stop your code from running. (In the lab computers this is ctrl+c). When using while loops, be sure you have included something within the loop that will eventually cause it to end. As a precaution, most programs that are more involved have special if statements within the while loop that will automatically break out of the while loop if, say, a certain threshold of time has passed. The rules for the conditionals themselves (the $x>5$ above) are the same as for if.

4.2.2 For-Loops

For-loops are one of the most powerful tools in Python. What they allow us to do is write a block of code that's like a template- it has the code we want to run, but without defining exactly "on what" the code acts. We then initialize a for-loop, picking a range of values, variables, etc., to plug into those designated spots in our block of code.

Exercise 4.3 A simple for-loop

```
arr = [1,2,3,4,5,6,7,8,9,10]
for i in arr:
    if i %2 ==0:
        print i
```

would print 2,4,6,8,10 (the even numbers). The % sign means "modulo," and the conditional would read "if i divided by two has a remainder of 0:". The letter i in this loop is a generalized iterator- when you type "for i in arr" you are telling the computer to run the block of code, replacing i in the block with the first second, third, etc. element in the array. (you could use any character/combination of characters for i, but i is standard practice (followed by j, and k if necessary). ■

The point of for-loops is that they are as generalizable as possible. In the above example, the array "arr" could be replaced with any variable that is an iterable data-type. You could say, "for i in range(15)" to have it plug the numbers 0 through 14 into your block of code, wherever a variable 'i' appeared. you could even iterate over a string, and it would plug in the elements of the string (as single character strings) into your block of code.

One common iteration practice is to iterate over an ascending list of numbers equal to the length

of a certain array. In this situation you could use "for i in range(len(array)-1)", where "array" is your array and len() is the command for returning the number of elements in an array, list, or string. The minus one is needed because the nth element of an array, list, or string is has an index of n-1.

4.2.3 Nested For-Loops

Just briefly, we'd like to mention that you can in fact nest multiple for-loops together, if you need to iterate over more than one value in your code. This often happens when dealing with two-dimensional arrays.

■ Example 4.4 Iterating a 2D Array

```
for i in range(len(x)-1):
    for j in range(len(y)-1):
        if arr[i,j]<1500.:
            arr[i,j]=0
```

In the above example, x would be a variable representing the x coordinates in the array, with a similar deal for y. This particular block of code would run through every combination of i, j to hit every spot on the 2D array, and if the value at any given point was below the 1500 threshold, it would just set that element to be 0.

4.3 Loading and Writing Files

At numerous points in this text we have alluded to the ability for loops to aid in the process of loading multiples files of data. Now that you know how to concatenate strings and generate for-loops, we can cover the file loading/writing process.

There are several ways of opening data files in python. Python itself has a built in mechanism for openin/writing files, and numpy also has support for file handling. To open a file in python's interface, we type:

```
»file1 = open('filename.txt','w')
```

where 'w' indicates we plan to write to the file. (We could instead use 'r' for read only, or 'a' for appending to a file that already contains data.

■ Example 4.5 Writing to a File

```
»file1 = open('file.txt','w')
»file1.write( 'this is a file')
»file1.write('this is not a drill')
»file1.close()
```

The close statement above tells python to close and save the file to the hard disk.

Exercise 4.4 Writing data to a file

You may have to analyze data in python but then export it to be analyzed more extensively by other programs. For example, you might have an array of planet distances and a second array with corresponding planet velocities that you wish to do some statistical analysis on with some other software. Likely you will want to save your data in a format that is easily usable in other programs. Thus, we can write it as such using a for loop. Assume we have already opened the

file in write mode and have predefined arrays of the same length.

```
for i in range(len(planet_dist)-1):
    file.write(planet_dist[i] + ' ' + planet_vel[i] + '\n')
```

where the `\n` is necessary for us to create a new line when writing the file so the data will be properly divided into their respective row and column. ■

Numpy also has a file input output framework that is often useful to use. The two we will discuss here are `np.loadtxt` and `np.genfromtxt`. These are useful tools because they have many specifiable options, and load your data straight into numpy arrays that you just love to work with!

■ Example 4.6 Loading files using `loadtxt`

```
data = np.loadtxt('filename.txt')
```

Lets say the file we loaded had three columns:times, positions, and velocities. These would all be stored in `data`, and could be singled out as such:

```
data = np.transpose(data)
times = data[0]
positions = data[1]
velocities = data[2]
```

■



Note: Because of the way columns/rows work in python, data in multiple columns are read in as individual data pairs. On the other hand, simply running an `np.transpose` on them sorts them to be 3 long separate arrays with all the times, all the positions, and all the velocities respectively.

Oftentimes data files have headers and footers- text that tells you what data is stored in the file. Of course, we don't want to try to read these into python as our data. For example, to skip a 5 line header and 3 line footer text, use

```
[IN]: data = np.genfromtxt('file.txt', skip_header=5, skip_footer=3)
```

This function is pretty versatile, and also has options for skipping columns, specifying data-types, etc.

Believe it or not, that's all there is to basic functional programming. By cleverly combining for loops, while loops, and conditional statements, we can do a lot of powerful analysis. While there is a lot more to python (for example, you can introduce classes and object-orientation (chapter 6), this is all you need to do the majority of scientific coding. What is missing in the above descriptions is the multitude of python and numpy functions you will need to use along the way. A list of useful functions is included in the appendix, and we will go over many functions in class.

5. Plotting

While we introduced the matplotlib library, and occasionally used plotting in examples, we'd like to go into more detail about plotting here, as being able to produce graphs and plots is not only important for use in scientific papers, etc., but also being able to quickly visualize data properly will save you a lot of time when working with large data sets.

5.1 Basic Plotting

Let's start with the basics. Say we have an independent variable (like time), and a measured variable (like position). This type of data could easily be read in from a 2 column text file and then plotted against each other.

Exercise 5.1 Plotting x vs. y

```
import matplotlib.pyplot as plt  
import numpy as np  
  
file_name = '/home/sally/data.txt'  
data = np.loadtxt(file_name)  
times = data[0]  
positions = data[1]  
# Now let's plot the data  
plt.plot(times,positions)  
plt.show()
```

■



Note that in python when plotting, the first argument is an array of x values and the second value is an array of y values, and the number of elements in the two arrays must match.

If you want to try this example, try creating an array of times using `np.arange(1,11)`, and position

values as an array you define manually: (using `position = np.array([1,2,6,34,56,57,...])`). Make sure that the number of positions you make is the same as the length of the times array. Try plotting as we did above. You'll notice that the default way python plots is by plotting the positions against the times and connecting them with blue lines. Now, as scientists, we know that raw data shouldn't be connected- what we were graphing was individual pairs of points. When using `plt.plot`, there are other optional settings you can specify. We will focus on color and linestyle. First, lets attempt to plot just discrete points, without a connecting line. Fig. () has a chart of how to specify colors and symbols within the plot command. If you choose a discrete symbol (like 'o' for circles or '+' for plusses), then python won't connect them automatically.

You can use a matplotlib shortcut to simultaneously choose a color and linestyle as follows:

```
»[IN]: plt.plot(times, positions, 'ro')
»[IN]: plt.show()
```

would plot the discrete points as red circles, while

```
»[IN]: plt.plot(times, positions, 'b+')
»[IN]: plt.show()
```

would plot the discrete points as blue plusses. You can also specify the size of the symbol by including the argument `ms=10` (play around with the number till you get the size you want).

If you are plotting multiple dependent variables against one axis (say, positions of multiple objects over the same time intervals), you'll want to create a legend to show which is which. to do so, use the optional command "label" within your `plt.plot` as follows:

```
»[IN]: plt.plot(times, positions_obj_1, 'k+', label='car one')
»[IN]: plt.plot(times, positions_obj_2, 'bo', label='car two')
»[IN]: plt.legend()
»[IN]: plt.show()
```

The labels defined in the plot functions will now show up in a legend. You can also comment out the `plt.legend` line- the labels will still exist but no legend will be shown. Legend has some optional inputs as well, primarily the one you need is `plt.legend(location=1)`, where 1 is a number 1-4 corresponding to the 4 corners of the plot. So if you find your legend covering up some of your data, try moving it to a new location.

One helpful plotting command to use is `plt.ion()`. This stands for interactive. It doesn't take any arguments. You may notice that generally when you use `plt.show()`, a plot pops up, and then your terminal stops accepting inputs until you close the plot. If you have a `plt.show` within the body of a large code, the rest of the code won't run until you manually close the figure. Using `plt.ion()` once, before all the plotting, will make it so that the plot opens, but the code continues running, and the terminal is still accessible. This allows you to make multiple plots in a row pop up as well. We recommend just putting `plt.ion()` right at the top of your programs next to the matplotlib import.

As you may have noticed above, if you type `plt.plot`(any arguments) multiple times, the graphs appear on the same plot (so you only want to do this when they share an axis). If you want to make two separate plots in a row , with different axes and such, just type `plt.figure()` in between the plots you are trying to make (so, for example, in between the first plot's `.show()` and the second plots first `.plot()`).

=====	=====
character	description
'-'	solid line style
--'	dashed line style
-.--'	dash-dot line style
--.'	dotted line style
'.'	point marker
',.'	pixel marker
'o'	circle marker
'v'	triangle_down marker
'^'	triangle_up marker
'<'	triangle_left marker
'>'	triangle_right marker
'1'	tri_down marker
'2'	tri_up marker
'3'	tri_left marker
'4'	tri_right marker
's'	square marker
'p'	pentagon marker
'*'	star marker
'h'	hexagon1 marker
'H'	hexagon2 marker
'+'	plus marker
'x'	x marker
'D'	diamond marker
'd'	thin_diamond marker
' '	vline marker
'_'	hline marker

Figure 5.1: How to specify different symbol types in matplotlib

5.2 Subplotting

5.3 Plotting 2D Images

Earlier we discussed 2-D arrays. Pyfits/the Astropy libraries have a way of displaying these as images. The easiest way to think about a 2-D array in terms of plotting is to pretend it is a black and white image. Each “pixel” is a value within the array. Some pixels might have low numbers (not bright), others higher (very bright). Matplotlib can generate an “image” based on this data- it simply assigns a color-table to follow the varying brightnesses, and displays the strength/intensity of each ‘pixel’ exactly as you would expect. (This is actually not a bad way to think, since images taken by telescopes are simply 2D CCD pixel arrays “counting” how many photons hit each pixel and returning a 2-D array with the totals).

Lets say we used astropy to read in a fits image, and turn it into a 2-D array (we cover how to do this in the next section). Now we have a two dimensional numpy array, with array[0,0] giving the number of photons in the top-left pixel, and so forth. To plot it, we would type:

character	color
'b'	blue
'g'	green
'r'	red
'c'	cyan
'm'	magenta
'y'	yellow
'k'	black
'w'	white

Figure 5.2: How to specify different symbol/line colors in matplotlib

```
»[IN]: plt.imshow(array, cmap='gray_r')
»[IN]: plt.show()
```



Note: for reasons that really don't matter, you don't have to call `plt.figure()` before using `imshow`, even if it is after plotting other things. It's a different kind of plot, and will show up in its own figure on its own.

In this example, we chose a `cmap` (color map) to be `gray_r`, which is essentially “reversed black and white”. Most of the time, when viewing images from telescopes, we want to use this setting, even though there are many wacky and colorful color-maps to choose from. The primary reason is that the astronomical image (unless taken with a specific filter), contains information only about raw numbers of light particles. So there is only one gradient- which is easily modeled as a transition between black and white. So why the reversal? Using the ‘gray’ `cmap` alone produces images that look quite a lot like the night sky anyhow. The answer is that when trying to pick out faint objects and stars, it is easier to see contrast between dark things on light backgrounds than the reverse. Furthermore, often times these figures end up in papers which are printed, and reversing the color-scheme saves on ink.

Plotting 2-D arrays, whether real images or other values, can sometimes be tricky. You are looking to get a certain level of contrast between light and dark, which maximally displays the information in the array (you don't want it to be washed out, or not visible). We encourage you to see the documentation for `plt.imshow()` to see how to select different scales (linear, quadratic, logarithmic, etc). This will also be covered in the image processing tutorial. An easy way to start pulling useful ranges within an images are the `vmin` and `vmax` commands. They are used to set the upper and lower range of the linear (by default) scale between black and white. Basically, if you set `vmin=50`, and `vmax=500`, it would create a linear scale from pure white to pure black between these two values; anything less than 50 is white, anything higher than 500 is just black. What this seemingly does is take away your ability to discern by a gradient a pixel of 500 and a pixel of 600. This is true. But when viewing astronomical images, often times there are several bright objects (like stars or galaxies), and a mostly black background (the sky). Since the difference between the sky brightness and object brightness is so huge, it doesn't make much sense to attempt to see the

“gradual” shift between them. Furthermore, in a simple plot like this, we only really care what is “not-sky” and what is “sky”, so we want a high level of contrast. If you need to know which pixels in a given star are brighter, and which are dimmer, comparatively, you would probably want to be more quantitative and write a piece of code to determine that for you.



A boring but important (sometimes) note: The convention amongst astronomers and scientists in general is the the “origin” of an image is in the lower left hand corner, (0,0), i.e., what we see is the first quadrant of a coordinate plane. Unfortunately, matplotlib has other ideas. When you use imshow, it displays like a matrix, the way arrays are defined, with (0,0) in the upper left corner. If you want to conform to convention and plot with (0,0) in the bottom left (which you should), you’ll want to use the command `origin='lower'` within your plt.imshow command. Unfortunately, doing THIS will end up flipping your image vertically. Sometimes, astronomical images come in upside down anyway (to the convention of north being up and east being left). Then `origin='lower'` actually solves your problem. But if your image was rightside up when plotted before, it is now upside down. Luckily we know how to fix this: just set your image = `image[::-1]` to flip the array, before plotting (and therefore flipping it).

5.4 Loading Astronomical Fits Files

Being able to manipulate data stored in images is one of the most important things you should try to take away from this course. In many physical science fields and especially in astronomy, images are taken using either microscopic or telescopic techniques and each pixel in the images corresponds to a specific intensity value. The meaning of these values is dependent on the actual instruments and the physical system you’re studying.

In astronomy, in particular, it is useful to work with FITS (flexible image transport system) files. FITS files are widely prevalent because of a feature they contain called a header. Headers often contain information about the image itself. For instance, they will often contain things like the declination, ascension, exposure time, as well as a description about the image itself.

FITS files are, like most things we will work with, not actually native to python itself. To use them, we will have to import a library called pyfits (or on the lab computers, astropy.io.fits); then we can start working with our images. Typically we import either of these as "pf". The syntax for opening a fits file is:

```
»hdu = pf.open(path)
```

where path is a string with the path location of the fits file, or, if your python file and fits file are in the same folder, then just a string with the filename is sufficient.

The reason we often refer to fits "files" rather than fits "images" is because a fits file actually contains more information in it than just an image. The two most important "sections" stored within a fits file are a header, and the image itself.

5.4.1 The Header

The header is a dictionary containing a lot of useful information about the images stored in the fits file: when they were taken, what exposure time was used, what type of filter was on the telescope, what the RA and DEC of the object viewed were, etc. Assuming we continue with hdu being the raw imported fits file, we can single out the header with dot notation:

```
»head = hdu[0].header
```

At this point, you could print "head" and see the entire header file. Alternatively, to see or pull

individual pieces of information from the header, you query it the way you would a dictionary, using a key.

■ **Example 5.1 Pulling from a header**

```
ra = head['RA']
dec = head['DEC']
time = head['EXPOSURETIME']
```

Note: the strings used in the dictionary call are not case sensitive. ■

5.4.2 The Image

To access the image itself, we call

```
»img = hdu[0].data
```

as the data attribute of the fits file contains the 2D array of the image, and the hdu[0] is due to the fact that sometimes fits files have multiple images stored within them.

Now, if you simply print img, you'll see a 2D array of data (likely with the whole center chopped out to save space). To see what that array looks like as an image, you can use plt.imshow(), as we discussed earlier in the chapter.

6. Classes and Object Oriented Programming

Object Oriented Programming is a relatively new system of organizing programs which is included in many higher level programming languages.

6.1 Defining Classes

Objects, or *Classes*, are useful for a variety of reasons; primarily we use them to easily organize groups of functions that act on a specific "object" that we define. There are two main components to a defined class: *attributes* and *methods*. Attributes are properties which are created when an *instance* of a class is initialized (when you create a new instance of a certain class in your code). Methods are functions which each instance of a class carries and can utilize. Accessing attributes and methods is done via dot notation, similarly to the way we utilize functions inherited from the libraries we imported. Essentially you can think of those libraries as large, complex classes.

Let's start with the basic syntax of classes. Before we can initialize an object of a certain class, have to define one first. Say we want to create a class called Planet which has certain attributes such as planet name, revolution period, and mass.

■ Example 6.1 Defining a Planet Class

```
class Planet(object):
    def __init__(self, planet_name, rev_period, mass):
        self.planet_name = planet_name
        self.rev_period = rev_period
        self.mass = mass
```

This looks a little funky, but its just the basic syntax for defining a class. The first function in a class is always the init function (which has a double underscore before and after the word init). The first

argument of init is always "self" (the word self isn't special, it could be any word, so long as it was consistent in the rest of the class. But self is the overwhelmingly common choice. The position of self as being the first argument is what makes it special). After this first argument, you list the other attributes that the user can define for a class object when they attempt to make one. Here, we let them specify a name, revolution period, and mass. Now, within the init function, we have to initialize "self" to be all of the things the user specifies. For example, by setting self.mass = mass (where mass comes from the user when they call the init function), then later someone can name a planet "mars" by typing mars = Planet('mars', 3,25), and call the mass by typing mars.mass. In short, when a person initializes a class object (like we just did in the previous sentence), the variable (in this case mars), stores all the information that "self" does within the init function.



Note: The reason we have the word 'object' in Example 6.1 is that we are actually creating a subclass of the superclass object. Don't worry too much about it, (we will get to subclasses in a moment), but just know that when defining a class in your code, you will probably use object in the class line.

Once we have created an init function, we can put other functions in our class as well. These are known as methods. For example, our Planet class could be updated to easily return the semimajor axis of the planet's orbit using Kepler's third law.

■ Example 6.2 Methods

```
class Planet(object):
    system_name = '2014-B178h'
    def __init__(self, planet_name, rev_period, mass):
        self.planet_name = planet_name
        self.rev_period = rev_period
        self.mass = mass

    def semimajoraxis(self):
        return (self.rev_period**(2./3.))
```

As you can probably infer, this method will be inherent to any instance of the Planet class you create and you can use it on any of them, just as any other method you include in the Planet class.

In addition to instance attributes (attributes set when an object is made) there can also be class attributes which are the same regardless of the instance. You simply set them within your class, outside the functions. In example 6.2, every planet you create will have the attribute of a system_name, and it will always be '2014-B178h'.

6.2 Subclasses

When discussing the class initialization, we mentioned that you use the word object within the parenthesis to pull methods and attributes from the object superclass. We usually don't call the classes we create subclasses though, since the object class is basically inherently necessary to working with classes. Thus, the classes in your code are the highest level ones you create. It is possible to make

subclasses of your own, which *inherit* all the attributes and methods of their parent class, while having some specialized methods and attributes of their own. For example, we can make a subclass of Planet called Dwarf:

■ **Example 6.3 Subclasses**

```
class Dwarf(Planet):
    def describe(self):
        return self.name + ':' + 'Mass-' + self.mass + 'Period-' + self.rev_period
```

■

Notice how instead of using object, we call Planet within the class call. Because Dwarf inherits everything from Planet (including the init function), we only have to worry about the special methods and attributes we want to apply only to the dwarf planets we define.

Classes and subclasses have lots of applications. For example, if you were designing a chess game, you might have a "board" class and a "piece" class, and then subclasses for the different kinds of pieces and the rules they follow, all while still having the common attributes of a piece.



7. LaTeX

Latex is a typesetting compiler used to produce high quality documents such as papers for publication, books, and more. For example, this textbook was written in Latex, as well as all the tutorial guides. Latex has somewhat of a learning curve, but once you learn it, it has distinct advantages over trying to write up documents in Word, or other normal processors. Here are some of those advantages:

- Latex has built in support for easily including complicated mathematical equations and expressions, as is often needed in a scientific paper.
- Figure/numbering handling. Latex lets you insert figures, equations, etc., all over your document and handles the numbering for you- if you insert a figure before fig. 3, all the figures in the document automatically renumber accordingly.
- Format and placement. Latex basically controls the format of your document for you- where figures go, how the text is laid out (based on certain styles you can specify). At first it's infuriating not to have complete control over every detail, but after a while you are thankful that it just does everything pretty nicely.
- Journal Styles: If you are submitting to an academic journal such as ApJ (the astrophysical journal), you can import that style into latex and it will automatically format your document to fit ApJ standards.
- Internal referencing. Latex has a built in reference system, so that you can label a figure as "star", and reference that figure throughout the document, and your text will say fig. 4 (where the 4 will change anytime the figure numbering does).
- Citations. There is a package known as bibtex that allows you to easily and short hand cite authors or papers throughout your own, and it automatically assembles a reference page at the end of your paper.

There are more advantages, based on the situation, but for now we will leave it at that. If you are planning on taking the Upper division lab, you will need to type up your reports in Latex. If not, it's still a valuable tool to know how to use.

7.1 Creating a Document

Latex is a compiled format, which means that you type in your text with a bunch of special commands to define things like new sections, equations, etc. Then, a Latex compiler turns all those commands into the appropriate formatting, and outputs a pdf of your document. To create a new latex document, you can type vim text.tex into terminal. Additionally, there is an awesome site called overleaf.com (previously writelatex.com), which lets you write and compile latex documents online (and store your stuff in the cloud for safety and for sharing with collaborators). In a moment, we will be going over the preamble and what needs to go there, but know that overleaf has plenty of templates with these finicky terms already in place, so you can focus on content.

7.1.1 The Preamble

The preamble to a latex document is a collection of commands that specify the overall formatting of the document. The first line in almost any latex file is

```
\documentclass[12pt]{article}
```

which determines the "class" of document and default fontsize you will be working with. There are many classes (which you can look up online), some typical ones include article, book, report, minimal (mostly for debugging), slides, memoir, and letter. You can also import specific classes to match a journal you want to submit to. After the document class is defined, you will want to import some packages (just like importing libraries in python) to make the environment more rich. These are the barebones ones you need to make sure you can include pictures and math:

```
\usepackage[margin=1in]{geometry}
\usepackage{amsmath,amsthm,amssymb}
\usepackage{graphicx}
```

When you are working you may need to download or install more packages to suit your needs. Overleaf has the majority of commonly used packages already imported online for you to access. Once you have all the packages and formatting you need done, you can move to a new line and start the document.

7.1.2 The Main Document

We start the document by having a line that reads

```
\begin{document}
```

"Document" here is what is known as an environment. Environments are set off from the rest of the latex document by "begin" and "end" statements (so your paper will end with an "end{document}" line). Other environments include equations, figures, etc. The rule of thumb is, if you have to begin something, you have to end it.

The first thing we want to do is create a nice title for our paper. We can do that by typing

```
\title {A witty title}
```

Most document classes come with titling code which also takes an author, which we can set by typing

```
\author{Author}
```

There are some other things you can put in a title, but this is the basics of it. Now, once those lines of code are down, you need a line

```
\maketitle
```

to actually have latex make the title when it compiles your document.

7.1.3 Sections and Subsections

Now you are ready to start dividing up your paper, book, assignment, etc., into sections and subsections. (These will all be numbered for you automatically). To start a new section, we type

```
\section{Section Name}
```

And then type our text in a block. If you need a new line, two backslashes in a row will do the trick. To create a subsection within a section, the syntax is identical to section creation, just with the word subsection instead. Keep in mind that this is not like an outline: you can't jump out of a subsection to the main section once a subsection has been defined. For scientific papers, doing so should really be unnecessary; sections will have introductions and then be divided into smaller subsections as needed.

7.1.4 Inserting Equations

There are two primary ways of inserting equations into latex: inline, and ownline. Inline refers to when you see a short equation or expression within a line of code. For example, few would think the equation $F = ma$ would require its own line. The way we wrote that equation just there was by surrounding everything we wanted in the equation by dollar signs. These set off the inline equation environment, which italicizes letters, allows for subscripts and superscripts, etc. To create an equation on its own line, we use could use something like the following:

■ Example 7.1 Defining an Equation

```
\begin{equation}
U = -\frac{Gm_1 m_2}{r}
\end{equation}
```

which would translate to:

$$U = -\frac{Gm_1 m_2}{r} \quad (7.1)$$

■

There are too many different subtle syntax variations for actually entering the symbols and math format needed, but the rule of thumb is basically to google it, and after a while the common ones become second nature.

7.1.5 Inserting Figures

If you have pictures (like .jpg, .png) that you've saved and want to include in your document, you can use the following template:

■ Example 7.2 Inserting Figures

```
\begin{figure}[htp]
\centering
\includegraphics[width=.55\textwidth]{img.png}
\caption{Enter a caption here}
\label {enter a label here for referencing}
\end{figure}
```

The [htp] tells latex where youd LIKE the figure to be, in this case we use the default of "here" (in its position in the latex document with reference the the text). Latex might decide it will go better at the top of the page, bottom, next page, etc. Usually you just let it do what it wants until the end, when all figures and text are entered, then you can start trying to push things around a little. a [h!] will attempt to force latex to put a figure "here". No promises though.

The centering line tells latex to center the image within the text or column of text it is in. the [width=] tag lets latex know how wide to scale the image; here we say .55 of the text width (you can choose any percentage, or just use in or cm if you want). The caption you specify appears under the image. The label is invisible (no one will see it), but it allows you to reference this figure in your text by a name you can remember, so that latex can replace that reference with the proper figure number later. ■

In the body of your text you may want to make reference to a figure you inserted by number, eg, (see fig. 1). But remember that latex numbers for you, and inserting another figure somewhere could throw off all your numbers if you do it manually. The solution is labels. If we specify a label tag on our figures and equations, we can then type see fig. \ref{labelname} and the ref labelname part will automatically show the figure number in question when we compile.

7.1.6 Lists and Bullets

In the tutorials we often had a list of commands, or bullet points, or both. Lists (both numbered and bulleted) are their own environments within latex.

You can begin a numbered list by having a line \begin{enumerate} (and of course a corresponding "end" command at the end. Inside, you can make individual items by typing \item (a space) and then the content of that number-bullet. It will automatically number each item in order. To make a bullet list, the syntax is identical, except you use "itemize" rather than "enumerate" in the begin and end statements.

You can also nest enumerates and itemizes together: if you have a full itemize chunk within an enumerate chunk, it will make an indented bullet between two of the numbers. You can look back at the tutorials for how this type of thing looks.

7.1.7 Wrap Up

That's about it for basic Latex! If you want to work with latex files from scratch, you'll need a latex compiler for your computer. The compiler installed on the lab computer is pdflatex, and typing pdflatex filename.tex into terminal will compile and create a pdf of your document (and or warn you of compile errors). If you are on your own computer, you will also have to download and install a lot of the packages for different formatting styles. There are some latex editors that will automatically try to download packages you don't have if you call them in your document.

Honestly, overleaf is a great resource if you need to write in latex. Not only does it store your files, images/figures, etc online (in case your computer crashes), but it has pretty much all the commonly used packaged "installed". Additionally, it has the cool feature of autocompiling a live preview in a pane next to where you are typing, so you can see quickly what your formatting looks like and if you've made a mistake. Learning latex is a lot of trial and error and google, but once you get a basic hang of it, and learn where to find the templates that handle a lot of the nasty junk for you, you can produce some really good looking documents.



8. HTML Editing

HTML is the language of the web. HTML stands for HyperText Markup Language and it allows you to build simple, functional web pages. This book will not delve deeply into the intricacies of web development but will give you a brief foray into the basics of websites. In addition to HTML, we will touch on the basics of Cascading Style Sheets, or CSS. CSS files, as their name suggests, allow you to manage the style of the page including the fonts, colors, text size, and various other factors.

8.1 Basic HTML

Just as you create a python file by typing ‘vim filename.py’, you create HTML files by typing ‘vim filename.html’ or by initializing a ‘.html’ file in whatever text editor you choose. Once you do so, it is pretty standard practice to start off your file by typing <!DOCTYPE HTML> This line indicates to your web browser which version of HTML to use. Once you have this, you include the <html> tags such as this:

```
<!DOCTYPE HTML>
<html>
</html>
```

As you have likely noticed, actual HTML code is placed in between brackets and the closing bracket always has a slash before the html tag. In HTML, anything placed between the two brackets is called a tag. Now that you’ve created your basic HTML, it’s time to get started with some actual content. In between the <html> tags, you will likely want to include the <head> tag. The <head> tag is where you will put information about the site title, author, and description and where you will include the code to tell the HTML files to incorporate CSS, Javascript, jQuery, or any other files you may need to include. For now, we will focus simply on the title tag. The title is what appears in the actual tab. Say we were to make a website about Jupiter and want to have our title be Jupiter.

■ **Example 8.1 Creating a Title**

```
<!DOCTYPE HTML>
<html>
  <head>
    <title>
      Jupiter
    </title>
  </head>
</html>
```

■



Note: The indenting we have demonstrated here is merely for aesthetics: ease of reading and editing. The browser itself doesn't care about indenting, only about start tags and end tags.

Now if we take this code and save it in an HTML file then open it with a browser, we would notice that the tab now reads “Jupiter” where it would have read the URL (or file path if it's locally hosted) before (note that this isn't titling the web page itself, only the tab bar). After the head tag, you can start writing the code which contains the actual content of the web page. This will be contained inside the body tags of the page. We will now go into some of the basic html tags, which are summarized in table 8.1.

Once you have created the enclosing body tags, we can get into the meat of the page. You'll probably want to include a title to indicate what the site is about. There are multiple title-types and they are designated with numbers `<h1>` through `<h6>`, where the increasing number corresponds to a decrease in the size of the text (h1 headers are the largest). These header tags are used typically to separate the page into subsections.

Typically, normal text is contained inside paragraph tags, designated `<p>`. Perhaps the most ubiquitous tag used in HTML code is the `<div>` tag. The div tag is used to enclose sections which will have the same styles. They are often most useful in conjunction with CSS classes and IDs. Often

html tag	description	example
<code><h1></code>	headers, size by number	<code><h1>A title </h1></code>
<code><a></code>	linking to files/web pages	<code>Link text</code>
<code><p></code>	denotes paragraphs	<code><p>paragraph text </p></code>
<code><div></code>	denotes chunks for given style	<code><div>content chunk </div></code>
<code></code>	insert an image	<code></code>

Table 8.1: Basic Html tags and their functions

we need to create hyperlinks which when clicked, go to a new web page. This is accomplished with the `<a>` tag. The `<a>` tag is not as straightforward as some of the other ones. It is best displayed through an example.

`Click here!`

The text in between the `<a>` tags will be the text displayed in link form on the actual web page and the address in the quotations will be where the user is redirected to. It should be noted that you can

also use path addresses to a specific file or folder in your directory instead of a URL (say, if you'd uploaded a pdf to your directory you wanted to share).

Images are an important part of many websites and are dealt with using the `` tag. Like the link tag, images point to a source file or address. In addition to the source attribute, images also have an alt attribute which displays some specified text in case the images are unable to load. Say for example we want to load a picture of Jupiter in our directory called "jupiter.png" we would write

```
<img src= "jupiter.png" alt="picture of Jupiter">
```

After specifying the alt, you can also specify a height and width in px or inches by the same syntax, (height="42", etc). If you want to be really fancy, you can make images into clickable links like this:

```
<a href= "http://en.wikipedia.org/wiki/Jupiter"> <img src= "jupiter.png"> </a>
```

Armed with this information, you can create a basic web page to present yourself to the world.

8.2 CSS and Styles

However, you have likely noted that this will all be plain text, which isn't necessarily a bad thing as it will get its point across but in this Web 2.0 World it's nice to give your site a little flash. This is where CSS comes in. As previously mentioned, CSS files contain the "styles" of the web page. First off, we need to know how to include CSS files in our HTML web pages. Inside our head tags, we use the `<link>` tag to link to the external files as such `<link rel="stylesheet" type="text/css" href="stylesheet.css">` This is, of course, assuming we have named our stylesheet to be simply "stylesheet.css". Let's put together all our code before into some basic website.

Exercise 8.1 Basic Web Page

```
<!DOCTYPE HTML>
<html>
  <head>
    <link rel="stylesheet" type="text/css" href="stylesheet.css">
    <title>
      Jupiter
    </title>
  </head>
  <body>
    <h1> Jupiter </h1>
    <p> Jupiter is the largest planet in our solar system </p>
    <p> Jupiter is a gas giant </p>
    <img src= "jupiter.png" alt="picture of Jupiter">
  </body>
</html>
```

Now say we wanted to make the body of the document have a blue background but we wanted `<h1>` and `<p>` tags to be two different colors and fonts. First off let's initialize our CSS file and then

we can begin looking at the syntax of CSS. In CSS, you include the HTML tag you wish to modify followed by a set of brackets. Inside these brackets are the actual stylistic attributes. Let's say we want to make the background color of our body be green. We would write in our file:

```
body {  
    background-color: green;  
}
```

where the semicolon tells the compiler that the attribute has received its assigned value. Now let's say we wanted to make the h1 tag some wild, exotic color, like orange and we wanted to center it, just for kicks. We would write that as such:

```
h1 {  
    color: orange;  
    text-align: center;  
}
```

You can just keep adding attributes as you wish and there are so many to choose from.

Now we want to specialize our <p> tags. However, we have two of them and we may want to give them separate styles. This is accomplished through the beauty of classes and IDs. IDs and classes work in essentially the same way but differ in one specific way that is crucial to success of your web page. IDs can only be used on one element of a web page, whereas classes can be used as many as you may like. For that reason, it is likely a better idea to use classes where possible as it will avoid most complications. These are implemented in your html code when you initialize a paragraph. We simply adapt the typical <p> marker to be:

<p class = "classname">
inside the one of the two <p> tags (in this case the first one). We then write some CSS:

```
p  
{  
    color: blue;  
    background-color:red;  
}  
  
.classname  
{  
    text-align: right;  
    color: red;  
    background-color:blue;  
}
```

and if we update our files and load the page, we will find that the paragraph we specified with a different class now appears different than any normal paragraph tagged lines (which default to the normal css description). The dot before the class name tells the browser to interpret the following thing as a class selector. For IDs, one would use a number sign before the identifier, but the rest would be the same.

8.2.1 Using divs

As things get more complex, it is often useful to enclose larger sections in div tags. There are a few reasons for this; it makes the code easier to read and edit, and more importantly, allows us to specialize our css further. For example, say we were to have a div section with a h1, h2, and a p section. However, all of those tags appear in other div sections and we do not want the same style in those sections. Thankfully, CSS has a simple way of implementing such a thing. As in the previous example, we specify the div we are interested in with a special class name, and then define the special formatting of that class in the CSS. The easiest way to learn about how these structuring elements work is to load up a simple website and look at the html and css code that went into it. One easy way to do this is to google "free css templates". There are numerous site with free html templates that come with a css sheet as well. Knowing what we've discussed above, you can use the structure in place to enter your own content, move things around, and so on. Viewing the page in a browser, you should try to follow along the html, see where the different styles are linked to, etc.

8.2.2 Tables and Lists

One thing we want to touch on before wrapping up is how to format a simple table or list, since they are effective ways of placing a lot of information.

Tables have several HTML tags associated with them. In order to initialize the table environment, one types `<table>...</table>` where the other information will go in between the two tags. First, you will need to create a row. This is done by writing the row tags `<tr>...</tr>`. Inside the row tags, you will include your column headers in between `<th>...</th>` tags. This tells the browser that these are table headers and it will automatically bold them and center them, which you could eliminate with CSS if you really wanted to. Once you have that, you can start creating your regular table rows using the `<td>...</td>` tags, which contain the information to wanted to show in the first place.

Exercise 8.2 Creating a table

```
<table>
  <tr>
    <th>Planet</th>
    <th>Distance from Sun (AU)</th>
  </tr>
  <tr>
    <td>Earth </td>
    <td>1 </td>
  </tr>
</table>
```

In HTML, there are essentially two types of lists: ordered and un-ordered lists. These are designated by the tags `` and `` respectively. Inside a set of `` and `` (same fo ``), you list individual items inside `` and `` tags. This is very similar to lists in Latex- with `enumerate` vs. `itemize` and the `\item` commands. Here's an example:

■ **Example 8.2** Say you had a to-do list of things you wanted to do in a specific order.

```
<ol>
    <li>Get groceries</li>
    <li>Cook Dinner</li>
    <li> Shower</li>
    <li> Cry self to sleep</li>
</ol>
```

whereas if you had just a list of things you might want for your birthday you may write

```
<ul>
    <li> Socks</li>
    <li> Underwear</li>
    <li>Ties</li>
    <li>Carrots</li>
</ul>
```

There is more to lists than simply in paragraph denotation. The added usefulness of lists is rooted in the power of CSS and other helping languages. For instance, most navigation menus are created using unordered lists.

Ultimately, HTML and CSS are not the most astronomically relevant languages to learn but their usefulness lies in their ability to convey information to those around you. Make your research look exciting and it will likely get more people behind it. What we do as scientists is research, and then communicate that research, so effective communication is essential to success in the field. Now, you should not expect yourself to be able to create intricate websites overnight and in actuality, there really is not even a need for it. There are already a plethora of splendid HTML templates out there which we are confident you can fully implement now. These are often much more complex than the stuff you learned in this textbook but you can break it down into the elements you learned here. Work smarter, not harder. Do not waste your own time reinventing the wheel. After all, that is why we program in the first place.



9. Glossary

Argument

An argument is essentially an input to a function. The term can be seen in the mathematics application, in which the sine function takes an argument, such as $(x-5)$, in the form $\sin(x-5)$. In UNIX, the commands like mkdir, rm, and vim serve as built-in functions, and the syntax for applying arguments is by typing the function, a space, and then the argument. Within python, functions, both built in and user defined, are referenced much like in the math example: via parenthesis attached to the function call. (Ex: np.sqrt(15), my_function(name, dob, gender)).

CLI

Command-line Interface (terminal). A means of interacting with a computer system via successive lines of text in the form of commands.

Command

A word, phrase, or instruction that can be understood and interpreted by a computer system which then executes the command in question.

Conditional Statement

A statement defining a certain condition, using operators like greater than, less than, equals to, their opposites, or some combination. These statements enclose a block of code that runs only if the conditional statement is evaluated to be true.

Data Type

Refers to the different types of objects in python to which python places certain rules. For example, integers and floats can be added, but not indexed.

Directory

Within a UNIX system, a directory is analogous to “folders” on a PC or Mac. It’s where all your files are stored.

Element

A single discrete object within an iterable set. For example, a single character in a string, or a single entry in a list or array, or an entry in a dictionary.

FITS

Flexible Image Transport Syste. A file format typical to astronomical images.

Flag

Also known as an option, a flag is a way to modify a UNIX command to alter the way it performs a task. Flags are entered between the command and the argument, with spaces in between both the command and the flag, and the flag and the argument. The typical syntax for a flag is a dash (-) followed by a letter or short combination thereof. (Ex: ls -l). To see what flags are available to use for any UNIX command, typing man(command) will have them listed.

For-loop

A for loop is a block of code that contains some iterative variable like "i" within it, with "i" cycling through different values defined in the initiation of the loop.

Function

A function is an operator that takes some inputs (or none) and, when called, performs some operations and outputs something (or multiple things). Functions can come from libraries, within python, or be user defined.

GUI

Graphical User Interface. A means of interacting with a computer system via graphical icons and visual indicators, through the direct manipulation of graphical elements (e.g., clicking, double clicking, and dragging)

Index

Given a list, array, or other sliceable data type (these are called iterable data types), every element is assigned an index based on its position from the leftmost element, starting with $i = 0$, and “reverse index” which starts on the rightmost element with -1 , and gets more negative as you move left.

Library

A large collection of functions that can be used in Python by importing the library. Calling functions from libraries usually requires the dot notation call of the library name (dot) function name. Libraries have defined names like numpy and matplotlib, but you can import them into your program as anything you like.

Loop

A block of code that is run multiple times, either due to iteration through some predefined range of values, or indefinitely so long as some condition is met.

Operating System

Software that manages the hardware and software resources for a computer. A vast majority of applications and programs require an OS be installed on a computer in order to function. Common OS's include Windows, MacOS, and Ubuntu.

Path

The description of your location within a file system, indicated by the names of successively nested directories, usually separated by slashes (e.g., home/documents/project/) Prompt: In essence, the computer is “prompting” you to issue a command. This string of letters could be anything, but is usually set to be some variance on the current path (meaning the prompt will change as the directory does).

Prompt

A line of text appearing on the left of the command line interface, usually containing some sort of path information. It can also be altered to read anything. You type in terminal next to be prompt.

Root Directory

The root directory is the directory within which all other folders/directories and files are contained. In essence, one could start from any directory in a file system and move up in directories until the root directory is reached.

SSH

Secure Shell Host. A way of logging into a server of networked computers via the terminal of any external computer.

Syntax

The specific set of words, phrases, and commands that can be successfully interpreted and understood by a computer. For example, a computer can understand that upon receiving the typed command “ls” it should display all files in the current directory. Had the user typed “list files” the computer would have thrown an error as this phrase is not in its syntax.

Terminal

Also called a shell, the terminal is where the prompt and command line interface are contained. It is the program through which one can issue commands directly to the computer’s OS through text commands.

UNIX

An operating system developed by Bell Labs upon which many other systems are currently built. UNIX is also the primary operating system on scientific machines like telescopes and supercomputers.

Variable

A variable is a user defined set of characters (could be a word, or number, or combination), which is assigned a value, array, etc. Variables allow us to implement large arrays and strings etc. in shorthand throughout our code.

While-loop

A while loop is a block of code that will run over and over so long as a conditional statement is still true after each run through of the block. This of course means that there must be something within the loop that will eventually force the condition to become false.