

Playful Python

Ken Youens-Clark

Version 1.0, 2020-1

Table of Contents

Introduction	1
Why Write Python?	2
Who Am I?	3
Who Are You?	4
Why Did I Write This Book?	5
A Note about the lingo	6
Why Not Notebooks?	7
Code examples, the REPL	8
Code formatting and linting	9
Organization	10
Using test-driven development	11
Suggestions for writing	12
Setting up your environment	13
Command line on Mac	13
Command line on Windows	13
Getting the code	14
How to write the world's best "Hello, World!"	15
Comment lines	15
The shebang	16
Making a program executable	16
String diagrams	17
Adding a parameter	17
Writing and testing functions	19
Adding a <code>main</code> function	21
Why is testing important?	22
Adding type hints	24
Verification with <code>mypy</code>	25
Manually validating and documenting arguments	26
Documenting and validating arguments using <code>argparse</code>	29
Checking style and errors	31
Making the argument optional	34
Starting a new program with <code>new.py</code>	35
Types of arguments	39
Using <code>argparse</code>	39
Manually checking arguments	42
Examples using <code>argparse</code>	44
Two different positional arguments	44
Two of the same positional arguments	47

One or more of the same positional arguments	48
Restricting values using <code>choices</code>	50
File arguments	51
Automatic help	53
Summary	55
Chapter 2: The Crow's Nest: Working with strings	56
Getting started	58
How to use the tests	58
Creating programs with <code>new.py</code>	58
Defining your arguments	59
Concatenating strings	61
Variable types	63
Getting just part of a string	64
Finding help in the REPL	65
String methods	66
String comparisons	67
Conditional branching	69
String formatting	70
Solution	72
Discussion	74
Defining the arguments with <code>get_args</code>	74
The <code>main</code> thing	75
Classifying the first character of a word	75
Printing the results	76
Running the test suite	77
Passing Tests	77
Summary	80
Going Further	81
Chapter 3: Going on a picnic: Working with lists	82
Starting the program	84
Writing <code>picnic.py</code>	86
Stuff about lists	88
Adding one element to a list	88
Adding many elements to a list	89
Indexing lists	92
Slicing lists	93
Finding elements in a list	94
Removing elements from a list	95
Sorting and reversing a list	97
Lists are mutable	99
Joining a <code>list</code>	100

Conditional branching with if/elif/else	102
Solution	103
Discussion	105
Defining the arguments	105
Assigning and sorting the items	105
Formatting the items	106
Printing the items	107
Testing	107
Summary	108
Going Further	109
Chapter 4: Jump the Five: Working with dictionaries	110
Dictionaries	112
Creating a dictionary	112
Accessing dictionary values	115
Other dictionary methods	116
Writing <code>jump.py</code>	119
Solution	122
Discussion	124
Defining the arguments	124
Using a <code>dict</code> for encoding	124
Method 1: Using a <code>for</code> loop to <code>print</code> each character	124
Method 2: Using a <code>for</code> loop to build a new string	127
Method 3: Using a <code>for</code> loop to build a new <code>list</code>	128
Method 4: List comprehension	128
Method 5: <code>str.translate</code>	129
(Not) using <code>str.replace</code>	130
Summary	131
Going Further	132
Chapter 5: Howler: Working with files and STDOUT	133
Reading files	135
Writing files	140
Writing <code>howler.py</code>	142
Solution	145
Discussion	147
Defining the arguments	148
Reading input from a file or the command line	148
Choosing the output file handle	149
Printing the output	149
Review	150
Going Further	151
Chapter 6: Words Count: Reading files/STDIN, iterating lists, formatting strings	152

Defining file inputs	154
Iterating lists	155
What you're counting	156
Formatting your results	157
Writing <code>wc</code>	158
Solution	159
Discussion	161
Defining the arguments	161
Reading a file using a <code>for</code> loop	161
A low-memory version	162
Review	164
Going Further	165
Chapter 7: Gashlycrumb: Looking items up in a dictionary	166
Writing <code>gashlycrumb.py</code>	168
Solution	172
Discussion	174
Handling the arguments	174
Reading the input file	174
Looping with <code>for</code> versus a list comprehensions	175
Dictionary lookups	176
<code>dict</code> vs <code>list</code> of <code>tuple</code>	176
Review	180
Going Further	181
Chapter 8: Apples and Bananas: Find and replace	182
Altering strings	186
Using <code>str.replace</code>	187
Using <code>str.translate</code>	188
Other ways to mutate strings	188
Solution	190
Discussion	192
Defining the parameters	192
Eight ways to replace the vowels	194
Method 1: Iterate every character	194
Method 2: Using <code>str.replace</code>	195
Method 3: Using <code>str.translate</code>	196
Method 4: List comprehension	199
Method 5: List comprehension with function	202
Method 6: The <code>map</code> function	205
Method 7: Using <code>map</code> with a defined function	208
Method 8: Using regular expressions	208
Refactoring with tests	211

Review	212
Going Further	213
Writing <code>abuse.py</code>	215
Validating arguments	216
Importing and seeding the <code>random</code> module	218
Defining the adjectives and nouns	219
Taking random samples and choices	219
Formatting the output	220
Solution	222
Discussion	225
Defining the arguments	225
Using <code>parser.error</code>	225
Program exit values and <code>STDERR</code>	226
Controlling randomness with <code>random.seed</code>	227
Iterating <code>for</code> loops with <code>range</code>	227
Constructing the insults	228
Review	230
Going Further	231
Chapter 10: Telephone: Randomly mutating strings	232
More briefing	233
Writing <code>telephone.py</code>	234
Calculating the number of mutations	236
The mutation space	237
Selecting the characters to mutate	238
Non-deterministic selection	238
Randomly sampling characters	238
Mutating a string	241
Time to write	243
Solution	244
Discussion	247
Defining the arguments	247
Mutating a string	247
Using a <code>list</code> instead of a <code>str</code>	248
Review	250
Going Further	251
Chapter 11: Bottles of Beer Song: Writing and testing functions	252
Writing <code>bottles.py</code>	253
Counting down	255
Writing a function	257
Writing a test for <code>verse</code>	258
Using the <code>verse</code> function	262

Solution	263
Discussion	266
Defining the arguments	266
Counting down	266
Test-Driven Development	266
The <code>verse</code> function	267
Iterating through the verses	268
1500 other solutions	271
Review	272
Going Further	273
Chapter 12: Ransom: Randomly captilizing text	274
Writing <code>ransom.py</code>	277
Mutating the text	277
Flipping a coin	278
Creating a new string	279
Solution	280
Discussion	283
Iterating through elements in a sequence	283
Writing a function to choose the letter	285
Another way to write <code>list.append</code>	285
Using a <code>str</code> instead of a <code>list</code>	286
Using a list comprehension	286
Using a <code>map</code> function	286
Comparing methods	288
Review	290
Going Further	291
Chapter 13: Twelve Days of Christmas: Algorithm design	292
Writing <code>twelve_days.py</code>	294
Counting	295
Creating the ordinal value	296
Making the verses	299
Using the <code>verse</code> function	302
Printing	303
Time to write	303
Solution	304
Discussion	307
Defining the arguments	307
Defining the output handle	307
Making one verse	307
Generating the verses	311
Printing the verses	312

Review	314
Going Further	315
Writing <code>rhymer.py</code>	319
Breaking a word	319
Using regular expressions	320
Using capture groups	323
Truthiness	326
Creating the output	327
Solution	329
Discussion	332
Stemming a word	332
Writing a regular expression	334
Testing and using the <code>stemmer</code> function	335
Creating rhyming strings	337
Writing <code>stemmer</code> without regular expressions	337
Review	339
Going Further	340
Chapter 15: The Kentucky Friar: More regular expressions	341
Writing <code>friar.py</code>	344
Splitting text using regular expressions	344
Shorthand classes	345
Negated shorthand classes	347
Using <code>re.split</code> with a captured regex	348
Writing the <code>fry</code> function	349
Using the <code>fry</code> function	354
Solution	356
Discussion	358
Breaking text into lines	358
Writing the <code>fry</code> function manually	359
Writing the <code>fry</code> function with regular expressions	361
Review	362
Going Further	363
Writing <code>scrambler.py</code>	365
Breaking the text into lines and words	365
Scrambling a word	367
Scrambling all the words	370
Solution	371
Discussion	374
Processing the text	374
Scrambling a word	376
Review	378

Going Further	379
Solution	381
Discussion	383
Defining the arguments.....	383
Reading the input file	383
Using regular expressions to find the pointy bits	383
Getting <code>input</code> from the user	385
Using <code>re.sub</code> to replace the values	386
Going Further	387
Gematria (Numeric encoding of text using ASCII values)	388
Solution	391
Discussion	393
Encoding one word	393
Finding the words	395
Putting it all together	396
Going Further	397
Mommy's Little Crossword Helper (Finding patterns of text in files)	398
Solution	400
Discussion	402
Regular Expressions	402
Manual Matching	404
Going Further	409
Workout Of the Day (Parsing CSV file, creating text table output)	410
Solution	413
Discussion	415
Defining the arguments	415
Reading a CSV file	415
Choosing the exercises	416
Formatting text tables with <code>tabulate</code>	418
Going Further	419
Password Strength	420
l33t-ify	422
Solution	423
Discussion	426
Review	427
Going Further	428

Introduction

"Codes are a puzzle. A game, just like any other game." - Alan Turing

I believe you can learn serious things through playing. This is a book of programming exercises that explores programming ideas by coding puzzles and games. Each chapter includes a description of a program you should write with examples of how the program should work. Most importantly, each program includes tests so that you know if your program is working well enough.

When you are done with this books you be able to:

- Write command-line Python programs
- Process a variety of command-line arguments, options, and flags
- Write and run tests for your programs and functions
- Manipulate of Python data structures including strings, lists, tuples, sets, dictionaries
- Use higher-order functions like `map` and `filter`
- Write and use regular expressions
- Read, parse, and write various text formats
- Use and control randomness
- Create and use graphs, kmers, Markov chains, Hamming distance, the Soundex algorithm, and more

At the core, I want you to learn just how much you can do with Python's basic data types and structures like strings, numbers, lists, tuples, dictionaries, and sets, as well as how to write and test small, focused functions. At no point will I demonstrate or encourage you to write object-oriented code. We will use some modules like `random` (to get random values) and `itertools` (to do some complex list operations), but otherwise I stick to the core Python language and functions.

I also want you to think all the time about how you can test your code. How can you be sure that a function does what you think it does? I want you to learn to write small, testable functions and stick them together to make larger, testable programs.

The ideas we explore in each exercise are not specific to Python. In fact, I think when you ought to write every program in every other language you know and compare what parts of a different language make it easier or harder to write your programs.

Why Write Python?

Python is an excellent, general-purpose programming language. You can use it to write command-line programs, web servers, adventure games, and business applications. There are Python modules to help you wrangle complex scientific data, explore machine learning algorithms, and generate publication-ready graphics.

Many college-level computer science programs have moved away from languages like C and Java to Python as their introductory language because Python is a relatively easy language to learn. Often Python codes reads like English—the statement `for item in cart` is a way to visit each `item` in the collection called `cart`. While languages like C, Java, Perl, and PHP use semi-colons (`;`) to mark the ends of statements, Python uses a newline (when you hit the `Enter` key). These same languages also tend to group statements into "blocks" by enclosing them in curly brackets (`{}`), but Python groups statements by how many spaces they are indented (usually four). Many people feel that Python's syntax and lack of excessive punctuation makes it more readable than other languages.

Who Am I?

My name is Ken Youens-Clark. I work at the University of Arizona as a Sr. Scientific Programmer in the lab of Dr. Bonnie Hurwitz. We are a bioinformatics lab that works mostly in metagenomics. That means we study biology almost entirely from the perspective of sequence data, not microscopes and petri dishes, and we spend most of our time trying to figure out which microscopic organisms (bacteria, archaea, fungi, viruses) are present in different environments (like your gut or a wound or in the soil or ocean) and what they are doing.

I began college as a music major (Jazz Studies playing the drumset) at the University of North Texas in 1990. I changed my major a few times and eventually ended up with a BA in English literature in 1995. I didn't really have a plan for my career, but I did like computers. I started tinkering with databases and HTML at first job out of college, building the company's mailing list and first website. I was definitely hooked! After that, I managed to turn a temporary technical writing gig into a full-time coding position where I learned VisualBasic on Windows 3.1. Through the next few years, I worked in several languages and companies before landing in a bioinformatics group at Cold Spring Harbor Laboratory in 2001 led by Lincoln Stein, an early advocate for open-source software and data in biology. In 2015 I moved to Tucson, AZ, to join Dr. Hurwitz, and finished my MS in Biosystems Engineering in 2019.

I am, perhaps, a reluctant Python programmer. The largest portion of my coding career—something like 18 years from 1998 to 2016 or so—I wrote almost exclusively in Perl. This is because Perl ruled the early days of the Internet and the world of genomics and bioinformatics. Since then, I've been studying Python, JavaScript, Haskell, Rust, Elm, Prolog, Ruby, Lisp, Julia, and anything else that seems interesting, but Python has clearly taken the lead in scientific computing partly due to a relatively easy syntax that closely mimics English.

When I'm not coding, I like playing music, riding bikes, cooking, reading, and being with my wife and children.

Who Are You?

I think my ideal reader is someone who's been trying to learn to code well but isn't quite sure how to level up. Perhaps you are someone who's been playing with Python or some other language that has a similar syntax like Java(SCRIPT) or Perl. Or maybe you've cut your teeth on something like Haskell or Scheme and you're wondering how does it go in Python Land? You are looking for interesting challenges with enough structure to help you know when you're moving in the right direction. Maybe you're interested in "test-driven development" and want to know how to write and use tests? Maybe you've heard of computer science ideas like Markov chains or n-grams or graphs and wonder how they could be useful?

This is a book that will try to teach you formal, structured methods to write Python well. This book introduces techniques I use in my own daily activity, from how I start writing a new program to how I verify that it actually works properly. This book shows you why you'd want to learn about regular expressions and algorithm design and statistics and random events. This book uses simple puzzles and games you already understand and asks you to teach the rules to Python.

This is not an ideal book for the absolute beginning programmer. While I cover some basics of the Python language, I quickly move into programs that assume a familiarity with coding in general. I especially assume you can use the command-line, which might be a leap if you have come from the world of graphical user interfaces like Windows and Apple. I would say this is probably your third or fourth book at least on Python or programming generally.

Why Did I Write This Book?

"The only way to learn a new programming language is by writing programs in it." - Dennis Ritchie

Over the years, I've had many opportunities to help people learn programming, and I always find it rewarding. The structure of this book comes from my own experience in the classroom where I think formal specifications and tests can be useful aids in learning how to break a program into smaller problems that need to be solved to create the whole program.

The biggest barrier to entry I've found when I'm learning a new language is that small concepts of the language are usually presented outside of any useful context. Yes, we all love to write "HELLO, WORLD!", but after that, I usually struggle to write a complete program that will accept some arguments and do something *useful*. For instance, how can I get the name of a file from the user, then read the file and do something with the contents, all the while handling the various errors that arise like not getting an argument from the user, the argument isn't actually a file, the file isn't readable, etc.? This book teaches you exactly how to write Python that accepts and validates arguments, presents usage messages, handles errors, and runs to completion.

More than anything, I think you need to practice. It's like the old joke: "What's the way to Carnegie Hall? Practice, practice, practice." These coding challenges are short enough that you could probably finish each in a few hours to days. This is more material than I could work through in a semester-long university-level class, so I imagine the whole book would take you several months, perhaps even a year or more. I hope you will solve the problems, then think about them and come back later to see if you can solve them differently, maybe using a more advanced technique or making them run faster. When you're done writing them in Python, I'd suggest you try to write them all again in some other language like Rust, Haskell, or Racket. Push yourself to compare what you know across languages to explore how they are the same and how they differ!

A Note about the lingo

Often in programming books you will see "foobar" used in examples. The word has no real meaning, but its origin probably comes from the military acronym "FUBAR" (Fouled Up Beyond All Recognition). If I use "foobar" in an example, it's because I don't want to talk about any specific thing in the universe, just the idea of a string of characters. If I need a list of items, usually the first item will be "foo," the next will be "bar." After that, convention uses "baz" and "quux," again because they mean nothing at all. Don't get hung up on "foobar." It's just a shorthand placeholder for something that could be more interesting later.

We also tend to call errors code "bugs." This comes from the days of computing before the invention of transistors. Early machines used vacuum tubes, and the heat from the machines would attract literal bugs who could cause short circuits. The "operators" (the people running the machines) would have to hunt through the tubes to find moths and such, hence the term to "debug."

Why Not Notebooks?

Notebooks are great for an interactive and visual exploration of data, but the downsides:

- Stored as JSON not line-oriented text, so no good **diff** tools
- Not easily shared
- Too easy to run cells out of order
- Hard to test
- No way to pass in arguments

I believe you can better learn how to create testable, *reproducible* software by writing command-line programs that always run from beginning to end and have a test suite. It's difficult to achieve that with Notebooks, but I do encourage you to explore Notebooks on your own.

Code examples, the REPL

I always love when a language has a good REPL (read-evaluate-print-loop) tool. Python excels in this respect. For simplicity's sake, I show the standard REPL when you execute `python3` on the command-line. Unfortunately, you can't copy and paste my code examples directly into the standard REPL or into IDLE. I suggest instead that you use iPython (`ipython`) where the magical `%paste` mode handles the leading `>>>` prompts. Even better, though, I suggest you manually type all the code yourself as this builds muscle memory and forces you to interact with the syntax of the language.

Code formatting and linting

Every program included has been automatically formatted with `yapf` (Yet Another Python Formatter), a tool from Google that can be customized with a configuration file. I encourage you to adopt and regularly use a formatter (see also `black`) *after every modification to your program*. Sometimes I even set up the formatter to format every time I save my program.

I would also encourage you to look at code "linters" like `pylint` and `flake8` to find potential errors in your code that the Python interpreter itself will not complain about. The `mypy` tool will also be very helpful as we introduce type hints.

Organization

The exercises are roughly arranged from easier to harder. The first exercises deal mostly with strings and then move into lists and tuples. Eventually we move into more complicated data structures like dictionaries and sets. About 60% of my solutions are under 60 lines of code (LOC) and only one hits 100, so these are not meant to be terribly long programs. You should work at your own pace and incorporate frequent walks and naps as well as discussions with colleagues (or a rubber duck) about where you're getting stuck.

The shorter, easier examples may seem trivial to solve, but there is quite a bit that you learn along the way. For instance, have you ever used a test suite? Are you familiar with the idea of writing tests first, verifying that your code fails, then writing the code to fix the test, and then verifying that your code now works? That's "test-driven development," and it's a fundamental aspect of the exercises. Have you ever used a code generation tool like [new.py](#)? Have you ever used [argparse](#)? Working methodically through each exercise, you will see how the concepts build upon each other, so I'd encourage you to start at the beginning.

Using test-driven development

"Test-driven development" is described by Kent Beck in his 2002 book by that title as a method to create more reliable programs. The basic idea is that we write tests even before we write code. We run the tests and verify that our code fails. Then we write the code to make each test pass. We always run all the tests so that, as we fix new tests, we ensure we don't break tests that were passing before. When all the tests pass, we have at least some assurances that the code we've written conforms to some manner of specification.

Each program you are asked to write comes with tests that will tell you when the code is working acceptably. Once you have started your new program with `new.py`, you should open it in your favorite editor and change the example arguments in `get_args` to suit the needs of your app, then add your code to accomplish the task described in the README. You can run the test suite using the GNU `make` program with `make test` in the same directory as the `test.py` and `article.py` program. If your system does not have `make` or you just don't want to use it, type `pytest -xv test.py`.

Your goal is to pass all the tests. The tests are written in an order designed to guide you in how break the problem down. For instance, if the program alters some input text (like `howler.py` wants you to uppercase the input), the first test will see if you can handle input from the command line. The next test will then pass in a file and ask you to read that for the text. I would suggest you solve the tests in order. If you use `make test`, the command in the `Makefile` executes `pytest -xv test.py` where the `-x` flag will have `pytest` halt testing after it finds one that fails. There's no point in running every test when one fails, nor should you attempt to solve harder problems before solving the simpler cases.

Every exercise has a `test.py` file that contains tests so that you know if your program is working properly. Some also contain a `unit.py` that contains tests for specific functions that you are encouraged (but not required) to write. I suggest you read over the tests to see how they are written and consider how you might expand them or add your own tests.

The first test in every exercise is whether the expected program exists. The second test checks if the program returns something like a "usage" statement when run with `-h|--help`. If you use `new.py` to create your program, you should pass the first two tests. After that, your program will be exercised with various inputs and options. You'll fail a test, and fix it. Then run the tests again. You will probably fail the next test. Fix that test such that you don't break the previously passing one. Keep fixing each test until they all pass. Then you are done. It doesn't matter if you solved the problem the same way as in the `solution.py`. All that matters is that you solve it *on your own!*

Note that many of the tests use an input file that is a dictionary of English words. To be sure that you can reproduce my results, I include a copy of mine in `inputs/words.zip`.

Suggestions for writing

As a general rule, I try to write many very small functions and lots of tests. Every function should be as short as possible, but no shorter — generally 50 lines is my upper limit on function length. I have no lower bound, though. Even if a function is just one line of code, I like to give it a name and some tests to ensure it does what I think. This is often called "unit testing" in the world of software engineering, and it's a very good practice. As the challenges get harder, I start suggesting specific functions and tests you should write in your programs. If we know the smaller bits work individually, then we ought to have confidence that they will work in concert.

This is the basic idea of "compositionality" where we try to compose large systems from smaller ones, so I also write tests to check that my programs as a whole do what they should. This is called "integration testing" and is found in the `test.py` programs I've provided for you in each of these exercises where your programs are run with various options and checked that they produce the expected output.

So, in short, always start a program the same way. Always process the arguments the same way. Always have the same structure to your program (e.g., start with `main`). When you think you need a function in your program, write the tests first and then write the code that will pass the tests. Then integrate your function into the greater whole. Write a test suite outside your program that checks that it fails properly as well as works properly. Whenever you make a change to your program, run your test suite to see if you accidentally broke something that used to work!

Lastly, you should always use a source code manager. If you fork and clone the GitHub repo of the exercises, then you've made an important first step in learning how to keep track of your code. You don't have to use `git` or GitHub. Mercurial is fine, too. (It's also written in Python!) Find a tool that works for you, and commit to tracking your changes.

Setting up your environment

To write the programs and run the tests, you will need to install Python and have some sort of command-line access. I wrote and tested the programs with the Python version 3.7. I tend to use the Anaconda distribution Python because I find it relatively easy to install, but I've not used anything specific to Anaconda.

My personal computer is a Mac, but I also work extensively on Linux servers. I write all my code exclusively in a text editor call `vi` (well, `vim`, but you knew that). When I write and test code on my Mac, I typically open at least three terminal windows:

1. A window dedicated to writing code in `vi`.
2. A window dedicated to running the code and test suite.
3. A window with a Python REPL running so I can test and play with code ideas.

You may prefer to write code with an editor like Sublime Text, TextWrangler, or Textmate. Or you may prefer to use an integrated development environment (IDE) like Xcode, PyCharm, Spyder, or VSCode. Whatever works for you!

Command line on Mac

On a Mac, you can open the `Terminal` app and type `python3 --version` to see what version of Python you have installed. Use Unix commands like `cd` (change directory) to get to your source code directory. From there you can execute `make test` inside the exercise directories if you have the `make` tool installed. No worries if you don't! That just executes `pytest -xv test.py`, so you can type that directly to run the tests

Command line on Windows

I'm probably going to need some help to include setup instructions for Windows. In my class, I suggest students install Windows Subsystem for Linux in order to get a functioning command line. That seems to work, although I really don't understand the interaction between WSL and the regular file system.

Getting the code

1. Create an account on GitHub.com
2. Go to https://github.com/kyclark/playful_python
3. Fork the repository into your own.
4. Use Git to `clone` *your fork* of the repo to your local machine.
5. Inside your local Git repo, run `git remote add upstream https://github.com/kyclark/playful_python.git` to add my repo as a remote source.
6. Use `git pull upstream master` to get updates from my repo. = Chapter 1: How to write a Python program

In this chapter, I'm going to show how to write programs that are documented and tested using the following principles:

- Process and document command-line arguments using `argparse`
- Write and run tests for your code with `pytest`
- Use tools like `yapf` or `black` to format your code
- Annotate variables and functions with type hints and then use `mypy` to check correctness
- Use tools like `flake8` and `pylint` to find problems in your code

How to write the world's best "Hello, World!"

It's pretty common to write "Hello, World!" as your first program in any language, so let's start there. We're going to start as simply as possible and end with a program you'd be proud to submit for a code review.

Start off by creating a text file called `hello.py` and add this line:

```
print('Hello, World!')
```



We can run it with the command `python3 hello.py` to have Python version 3 execute the commands in the file called `hello.py`. (Note that my `python` is version 2, so I will be sure to run `python3`.) You should see this:

```
$ python3 hello.py  
Hello, World!
```

Comment lines



In Python, the `#` character and anything following it is ignored by Python. This is useful to add comments to your code or to temporarily disable lines of code when testing and debugging. It's always a good idea to document your programs with the purpose of the program and/or the author's name and email address. We can use a comment for that:

```
# Purpose: Say hello  
print('Hello, World!')
```

If you run it again, you should see the same output as before because the "Purpose" line is ignored. Note that any text to the left of the `#` is executed, so you can add a comment to the end of a line, if you like.

The shebang

We could write similar programs in other languages like bash, Ruby, and Perl, so it's common to document the language inside the program with a special comment character combination of `\#!`. The nickname for this is "shebang" (pronounced "shuh-bang"—I always think of the `\#` as "shuh" and the `!` as the "bang!"). As usual, Python will ignore this, but the OS will use it to decide which program to use to interpret the rest of the file.

Here is the shebang you should add:

```
#!/usr/bin/env python3
```

The `env` program will tell you about your "environment." If you run it like `env`, you should see lots of lines of output with lines like `USER=kyclark`. If you run it like `env python3`, it will search for and run the `python3` program. On my systems, I like to use the Anaconda Python distribution, and this is usually installed in my "home" directory. On my Mac laptop, this means that my `python3` is actually `/Users/kyclark/anaconda3/bin/python3`, but on another system like one of my Linux web servers will be an entirely different location. If I hardcoded the Mac location of `python3` in the shebang line, then my program won't work when I run it on Linux, so I always use the `env` program to find `python3`.

Now your program should look like this:

```
#!/usr/bin/env python3 ①
# Purpose: Say hello ②
print('Hello, World!') ③
```

- ① The shebang line telling the operating system to use the `/usr/bin/env` program to find `python3` to interpret this program.
- ② A comment line documenting the purpose of the program.
- ③ A Python command to print some text to the screen.

Making a program executable



Now I'd like to "run" the script. On a Unix-like system, we can make any file "executable" with the command `chmod` (*change mode*). Think of it like turning your program "on." Run this command to make `hello.py` executable — to *add* (+) the *executable* (x) bit:

```
$ chmod +x hello.py
```

Now you can run the program like so:

```
$ ./hello.py  
Hello, World!
```

That looks way cooler.

String diagrams



Throughout the book, I'll use "string diagrams" to visualize the inputs and outputs of the programs we'll write. If we create one for our program as it is, there are no inputs and the output is always "Hello, World!"

Adding a parameter

It's not terribly interesting for our program to always say "Hello, World!" Let's have it enthusiastically greet some name that we will pass as an *argument*. That is, we'd like our program to work like this:

```
$ ./hello.py Terra  
Hello, Terra!
```

The arguments to a program are available through the Python `sys` (system) module's `arg` (argument vector). We have to add `import sys` to our program to use it. Change your program so it looks like this:

```

1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import sys          ①
5
6 args = sys.argv[1:] ②
7 if args:            ③
8     name = args[0]   ④
9     print(f'Hello, {name}!') ⑤
10 else:              ⑥
11     print('Hello, World!') ⑦

```

① Allows us to use code from the `sys` module.

② `argv` is a list, slice from index 1.

③ If there are any arguments,

④ Then get the `name` from index 0.

⑤ Print `name` using an f-string,

⑥ Otherwise,

⑦ Print the default greeting.

Line 6 probably looks a bit funny, so let's talk about that. The argument vector `argv` is a record of how the program was run starting with the path to the program itself and followed by any arguments. If you run it like `./hello.py Terra`, then `argv` will look like this:

Python, like so many other programming languages, uses the *index 0* for the first element in a list. The element at `argv[0]` will always be the path to the program itself.

<code>command</code>	<code>\$./hello.py Terra</code>
<code>sys.argv</code>	<code>['./hello.py', 'Terra']</code>

`index`



So on line 6, we use `sys.argv[1:]` to say we

want a *slice* of that list starting from the index 1 and going to the end of the list and assign that to the variable called `args`. On line 7, we check if there is something in `args`—that is, there were some arguments to the program. If so, on line 8 we assign the variable `name` to the first element in `args`. If there are no `args`, we will print the default greeting on line 11. In chapter 3 (Picnic), we'll talk much more about lists and slices.



Now our program will say "Hello, World!" when there are no

arguments:

```
$ ./hello.py  
Hello, World!
```

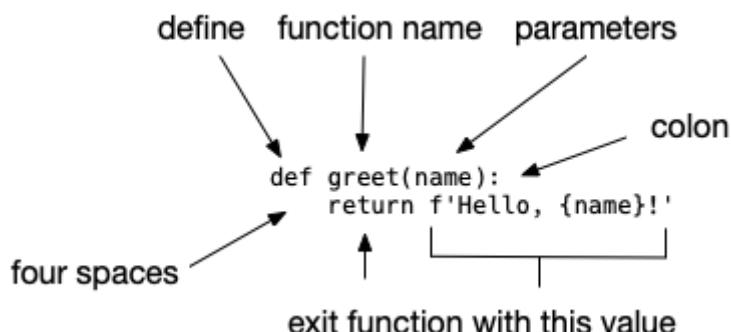
Or will extend warm salutations to whatever value is given as an argument:

```
$ ./hello.py Universe  
Hello, Universe!
```

Writing and testing functions

Our `hello.py` program now behaves differently depending on how it is run. How can we *test* our program to be sure it does the right thing? Let's find the part of our program that is variable and put that into a smaller unit of code that we can test. The unit will be a *function*, and the thing that changes is the greeting that we produce.

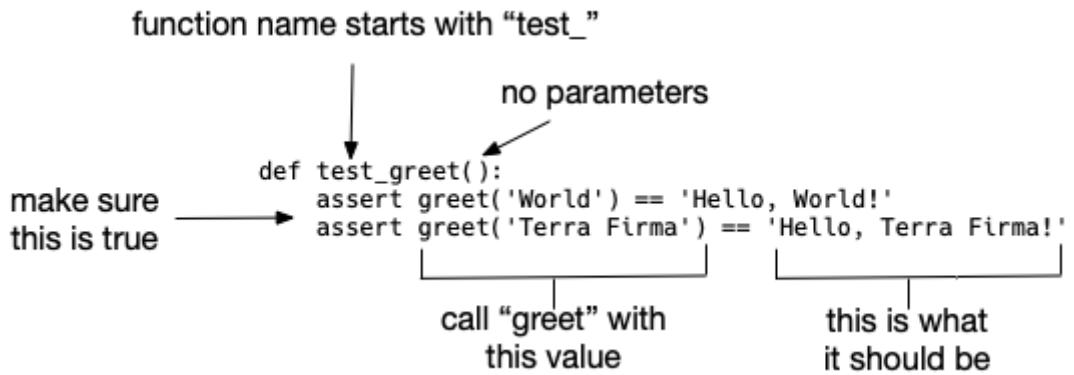
Create a function called `greet` with the `name` as a parameter.



The `def` is to *define* a function, and the function name follows after that. The function's parameters (the things that can change when it is run) go inside the parentheses. Here we have just one, and it's called `name`. The colon (`:`) lets Python know that a "block" of statements will follow, all of which need to be indented a consistent number of spaces (please indent with 4 spaces only, never tabs).

The `return` will leave the function, optionally returning any value that follows. If no value follows the `return`, then the special value `None` will be returned to the caller. Some languages will return the last value of the function, but not Python. You must explicitly `return` what you want. You may return multiple values by separating them with commas, and you are also allowed to have more than one `return` in a function.

Our new `greet` function does not `print` the greeting to the user, though. Its only job is to create the greeting so that we can write a test for it like so:



Here is what our whole program looks like now:

```

1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import sys
5
6 def greet(name):          ①
7     return f'Hello, {name}!'
8
9 def test_greet():          ②
10    assert greet('World') == 'Hello, World!'
11    assert greet('Terra Firma') == 'Hello, Terra Firma!'
12
13 args = sys.argv[1:]       ③
14 name = args[0] if args else 'World' ④
15 print(greet(name))        ⑤

```

- ① A function to create a greeting.
- ② A function to test the `greet` function.
- ③ The program starts here.
- ④ An `if` expression is a shorthand way to write `if/else` that fits on one line. The `name` will be set to `args[0]` if is something in `args`, otherwise use the value '`World`'. We'll talk more about these in chapter 2 (Crow's Nest) when we are dealing with *binary* (two) choices.
- ⑤ The `print` and `greet` functions are *two separate ideas*. I know that I can count on Python to reliably `print`, but I need to test my `greet` function to ensure it works properly.

I personally like to use the `pytest` module to run my tests, but you may prefer to use the `unittest` module, especially if you have a background in Java. `pytest` will execute every function that has a name starting with `test_`, so I call my function `test_greet`. This function will call the `greet` function two times, once with the argument '`World`' and once with the argument '`Terra Firma`' to `assert` that the value returned from the function is what I would expect.

To see the tests in action, run `pytest -xv hello.py`:

```
$ pytest -xv hello.py
=====
platform darwin -- Python 3.7.3, pytest-4.6.5, py-1.8.0, pluggy-0.12.0 --
/Users/kyclark/anaconda3/bin/python
cachedir: .pytest_cache
rootdir: /Users/kyclark/work/manning/playful_python/hello
plugins: openfiles-0.3.2, arraydiff-0.3, doctestplus-0.3.0, remotedata-0.3.1, cov-2.7.1
collected 1 item

hello.py::test_greet PASSED [100%]

===== 1 passed in 0.03 seconds =====
```



The `-v` flag is to have `pytest` run in the "verbose" mode, and the `-x` flag to `pytest` tells it to stop at the first failing test. It's common to combine short flags in this way (`-xv` or `-vx`, the order doesn't matter). Here the `-x` flag causes `pytest` to stop when it sees that `crowsnest.py` isn't there. There's no need to run any further tests if the program doesn't even exist. I include a `Makefile` with each exercise so that you can run `make test` to execute this for you. If you don't have `make` or don't want to use it, just manually run `pytest -xv test.py`.

Adding a `main` function

Our program is pretty good, but it's not quite idiomatic Python. Do you notice how we have two functions and then some other code just sort of hanging out (lines 13-15) flush left? It's good practice to put *all* your code inside functions, and it's very common to create a function called `main` where your program starts:

```
def main():
    args = sys.argv[1:]
    name = args[0] if args else 'World'
    print(greet(name))
```

We still have to tell Python to run this function, though, and the idiom for that in Python is to put these as the *last* two lines in your program:

```
if __name__ == '__main__':
    main()
```

Python reads your whole program from top to bottom. When it gets to these last lines, it will see if the special variable called `{dbl_}name{dbl_}` (one of the many special "double-under" or "dunder" variables) shows that we are in the "main" namespace. If you use `import` to bring your code into another piece of code, then it would *not* be in the "main" namespace. When you execute the code like a program, then the namespace will be "main" and so the `main` function will be run.

Here is our whole program now:

```
1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import sys
5
6 def greet(name):
7     return f'Hello, {name}!'
8
9 def test_greet():
10    assert greet('World') == 'Hello, World!'
11    assert greet('Terra Firma') == 'Hello, Terra Firma!'
12
13 def main():          ①
14     args = sys.argv[1:]
15     name = args[0] if args else 'World'
16     print(greet(name))
17
18 if __name__ == '__main__':  ②
19     main()                ③
```

① A new `main` entry function.

② Program starts here.

③ Call "main" function if in "main" namespace

Note that the call to run `main()` must occur in the file *after* the definition of the `main` function. Python will throw an exception if you try to call a function that hasn't yet been defined. Try moving these to the *top* of your program and running it again, and you'll see an error:

```
NameError: name 'main' is not defined
```

Why is testing important?



Would you fly on a plane or ride on an elevator if you knew it had never been tested? While lives may not depend so directly on the software you write, you still want to write code that is free from errors.

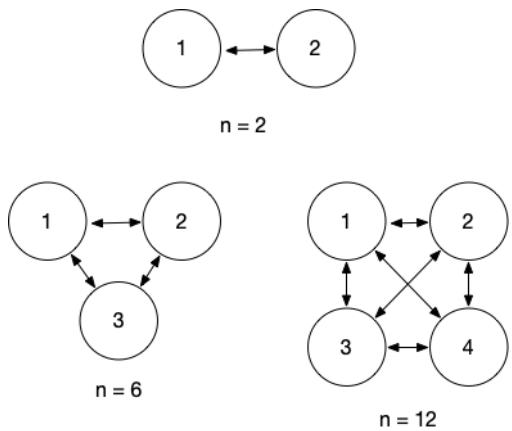


Figure 1. Complexity increases quickly!

As programs get longer, the tendency is that they get much harder to maintain. It's like adding a new team member. When there are just two people, there are only two ways to talk. Every new member increases the number of channels exponentially. Each new line of code or function you write interacts with all the other code in ways that become far too complicated for you to remember.

Tests help us check that our code still does what we think it does. Imagine we want to translate our `hello.py` to Spanish. We need to change the "Hello" to "Hola," but we misspell it "Halo":

```
#!/usr/bin/env python3
# Purpose: Say hello

import sys

def greet(name):
    return f'Halo, {name}!' ①

def test_greet():
    assert greet('World') == 'Hola, World!' ②
    assert greet('Terra Firma') == 'Hola, Terra Firma!' ③

def main():
    args = sys.argv[1:]
    name = args[0] if args else 'World'
    print(greet(name))

if __name__ == '__main__':
    main()
```

① "Hola" is misspelled.

② "Hola" is correct.

③ Here, too.

If you run `pytest` on this, you will see this failure:

```

=====
 FAILURES =====
-----
 test_greet -----
```

```

def test_greet():
>     assert greet('World') == 'Hola, World!'
E     AssertionError: assert 'Halo, World!' == 'Hola, World!'
E         - Halo, World!
E         ?
E         + Hola, World!
E         ? ++

hello07.py:10: AssertionError
=====
 1 failed in 0.07 seconds =====

```

I would encourage you to write many functions, each of which does a few things as possible. Each function should have a `test_`, either in the same source file (I usually place it immediately after the function it's testing) or in a separate file (which I usually call `unit.py` for "unit" tests).

Often I will even write my `test_` function *before* I write the function itself, and I will imagine how I would want the function to work under a variety of conditions. As a general rule, I usually will test a function with:

- No arguments
- One argument
- Several arguments
- Bad arguments

I will then test if the function works or fails as I expect it. For instance, what would you have `greet` do if given no arguments or given a *list* of arguments instead of one?

Adding type hints

Each variable in Python has a `type`. For instance, we execute `python3` on the command line to interact directly with the Python interpreter. You will see a `>>>` prompt that is waiting for you to type some Python code: Create a `name` variable set to "World":

```
>>> name = 'World'
```

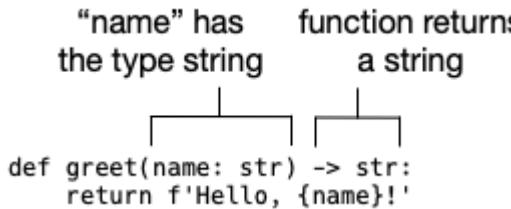
Now use the `type` command to see how Python represents this data:

```
>>> type(name)
<class 'str'>
```

Anything in quotes (single or double) is a "string," which in Python is represented by the class `str`. Python has many other types, for instance, the number `10` is an `int` or "integer":

```
>>> type(10)
<class 'int'>
```

We can add "type hints" to our code to ensure that we use the right types of data, like not trying to divide an `int` by a `str` which would cause an *exception* and crash our code.



Shown is how `greet` would look with type hints. You can read `name: str` as "name has the type string." Additionally, the `greet` function itself has been annotated with `-> str` to indicate that it will return a `str` value. If a function (like `main`) returns nothing, add the special `None` as the return type.

Here is what the entire program looks like with type hints. *Can you find the error?*

```
1  #!/usr/bin/env python3
2  # Purpose: Say hello
3
4  import sys
5
6  def greet(name: str) -> str: ①
7      return f'Hello, {name}!'
8
9  def test_greet() -> None: ②
10     assert greet('World') == 'Hello, World!'
11     assert greet('Terra Firma') == 'Hello, Terra Firma!'
12
13 def main() -> None: ③
14     args = sys.argv[1:]
15     print(greet(args))
16
17 if __name__ == '__main__':
18     main()
```

① The `name` argument is a `str`, and the function returns a `str`.

② The function returns `None`.

③ The function returns `None`.

Verification with `mypy`



Python itself completely ignores the type hints, but we can use the `mypy` program to find a problem in the code:

```
$ mypy hello.py
hello.py:15: error: Argument 1 to "greet" has incompatible type "List[str]"; expected
"str"
```

Can you see that we tried to pass `args` (which is a list of strings) to `greet` instead of a single `str` value? Here is the correct `main` function:

```
def main() -> None:
    args = sys.argv[1:]
    name = args[0] if args else 'World'
    print(greet(name))
```

Manually validating and documenting arguments

We're now going to change our program to require exactly one argument. If we don't get that, we need to print a "usage" statement that explains how to use the program:

```
$ ./hello.py
usage: hello.py NAME
```

And likewise if run with more than one argument:

```
$ ./hello.py Terra Firma
usage: hello.py NAME
```

When given one argument, it should work as expected:

```
$ ./hello.py "Terra Firma"
Hello, Terra Firma!
```

Here is the new version to make that work:

```
1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import os, sys
5
6 def greet(name: str) -> str:
7     return f'Hello, {name}!'
8
9 def test_greet() -> None:
10    assert greet('World') == 'Hello, World!'
11    assert greet('Terra Firma') == 'Hello, Terra Firma!'
12
13 def main() -> None:
14     args = sys.argv[1:]
15     if len(args) != 1:                                ①
16         prg_name = os.path.basename(sys.argv[0])      ②
17         print(f'usage: {prg_name} NAME')              ③
18         sys.exit(1)                                  ④
19     else:
20         print(greet(args[0]))
21
22 if __name__ == '__main__':
23     main()
```

① `args` must have 1 element.

② `sys.argv[0]` has program name. Use `basename` to remove the path.

③ Print "usage" statement.

④ Exit with error (not a 0 value).

On line 4, note that I can `import` more than one module by separating them with commas. Here I `import os` so that I can use the `os.path.basename` function to get the "basename" of `sys.argv[0]` which, you'll recall, is the path of the program itself. The "basename" of a path is the filename itself:

```
>>> import os
>>> os.path.basename('/path/to/hello.py')
'hello.py'
```

On line 15, I check if the `len` (length) of the `args` is *not* 1. If so, I'll print a "usage" statement on how to run the program. Note that I call `sys.exit(1)` to indicate a *non-zero exit value* because `0` (which is the default) indicates "zero errors." Any exit value that is not `0` indicates a failure.

Most Unix tools will also respond to `-h` or `--help` to show a "usage" statement. What happens when we try that?

```
$ ./hello.py --help
Hello, --help!
```

Well, that didn't work out. Just because we gave `--help` as an argument didn't mean that our program interpreted it correctly. Our program has been written to process *positional* arguments only, and the `--help` argument is a "flag," meaning a value that is `True` when present and `False` when absent. Here is a pretty ugly way to code this:

```
def main() -> None:
    args = sys.argv[1:]
    val = args[0] if args else ''
    if len(args) != 1 or val == '-h' or val == '--help':
        prg_name = os.path.basename(sys.argv[0])
        print(f'usage: {prg_name} NAME')
    else:
        print(greet(val))
```



Now we are setting a `val` ("value") variable equal to the first argument if one is present. Then we can check `len(args)` or if the `val` is equal to either the string '`-h`' or '`--help`'. I can verify that it works:

```
$ ./hello.py -h
usage: hello.py NAME
$ ./hello.py --help
usage: hello.py NAME
```

That big compound `if` test is pretty ugly. It's what I would call a "code smell"—it's trying to do too many things at once. We can write something much more elegant.

Documenting and validating arguments using argparse

The Python language has a standard module called `argparse` that will parse all the command line arguments, options, and flags. You have to invest a bit of time to learn it, but it will save you so much time in return. Here is a new version that will make our code much cleaner:

```
1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import argparse, os, sys
5
6 def greet(name: str) -> str:
7     return f'Hello, {name}!'
8
9 def test_greet() -> None:
10    assert greet('World') == 'Hello, World!'
11    assert greet('Terra Firma') == 'Hello, Terra Firma!'
12
13 def main() -> None:
14     parser = argparse.ArgumentParser(description='Say hello') ①
15     parser.add_argument('name', help='The name to greet') ②
16     args = parser.parse_args() ③
17     print(greet(args.name)) ④
18
19 if __name__ == '__main__':
20     main()
```

- ① Create argument parser.
- ② Add the `name` parameter.
- ③ Get the parsed `args`.
- ④ Print the greeting given the `args.name` value.



Now instead of directly dealing with `sys.argv`, we describe to `argparse` that we want a single argument called `name` and let it do the hard work of parsing and validating the arguments. This will be a *positional* argument because `name` does *not* start with a dash. We will set `args` equal to the result of our `parser` doing the work to `parse_args`.

Now if you run the program with no arguments, the "usage" statement will be generated by `argparse`. Though you can't see it directly, the exit value is also set to something other than `0` to indicate failure:

```
$ ./hello.py
usage: hello.py [-h] str
hello.py: error: the following arguments are required: str
```

And both the `-h` and `--help` flags will trigger a longer help document that looks like a Unix `man` page:

```
$ ./hello.py --help
usage: hello.py [-h] str

Say hello

positional arguments:
  str        The name to greet

optional arguments:
  -h, --help  show this help message and exit
```

As a matter of personal taste, I like to put all the `argparse` code into its own function that I always call `get_args`. For some of my programs, this can get quite long, and it makes the `main` function stay much shorter if this is separated. Besides, it is a functional unit in my mind, and so it belongs by itself. I always put `get_args` as the first function so that I can see it immediately when I read the source code. I usually put `main` right after it.

Here is how the program looks now:

```

1 #!/usr/bin/env python3
2 # Purpose: Say hello
3
4 import argparse
5
6 def get_args() -> argparse.Namespace: ①
7     parser = argparse.ArgumentParser(description='Say hello')
8     parser.add_argument('name', metavar='str', help='The name to greet')
9     return parser.parse_args()
10
11 def main() -> None:
12     args = get_args() ②
13     print(greet(args.name))
14
15 def greet(name: str) -> str:
16     return f'Hello, {name}!'
17
18 def test_greet() -> None:
19     assert greet('World') == 'Hello, World!'
20     assert greet('Terra Firma') == 'Hello, Terra Firma!'
21
22 if __name__ == '__main__':
23     main()

```

① The `get_args` function dedicated to getting the command-line arguments

② Call `get_args` function to get parsed arguments.

Checking style and errors



We now have the bones of a pretty respectable program. I like to use the `flake8` and `pylint` tools to give me suggestions on how to improve my programs. This is a process called "linting," and the tools are called "linters." I find that `flake8` is unhappy with readability as it

tells me to put 2 lines after each function definition:

```
$ flake8 hello.py
hello.py:6:1: E302 expected 2 blank lines, found 1
hello.py:11:1: E302 expected 2 blank lines, found 1
hello.py:15:1: E302 expected 2 blank lines, found 1
hello.py:18:1: E302 expected 2 blank lines, found 1
hello.py:22:1: E305 expected 2 blank lines after class or function definition, found 1
```

The `pylint` program has other things to complain about, namely that my functions are missing documentation ("docstrings"):

```
$ pylint hello.py
*****
Module hello
hello.py:1:0: C0111: Missing module docstring (missing-docstring)
hello.py:6:0: C0111: Missing function docstring (missing-docstring)
hello.py:11:0: C0111: Missing function docstring (missing-docstring)
hello.py:15:0: C0111: Missing function docstring (missing-docstring)
hello.py:18:0: C0111: Missing function docstring (missing-docstring)

-----
Your code has been rated at 6.67/10 (previous run: 6.67/10, +0.00)
```

A docstring is a string that occurs just after the `def` of the function. It can be a single line of text enclosed in single or double quotes. It's also common to use several lines, and Python allows you to use triple-quotes for strings that have line breaks: You can assign them to a variable:

```
>>> multi_line = """
... I should have been a pair of ragged claws.
... Scuttling across the floors of silent seas.
... """
```

Or you can use them in place of the `#` for multi-line comments. For instance, I usually document my whole program by putting a docstring just after the shebang. Inside I put my name, email address, the purpose of the script, and the date.

To fix the formatting issues, I ran the whole program through the formatting program called `yapf` (Yet Another Python Formatter). Another popular formatter is `black`. It really doesn't matter which one you choose. Just choose one and use it religiously.

Here is a version that will silence all of our critics:

```

1 #!/usr/bin/env python3
2 """
3 Purpose: Say hello      ①
4 Author: Ken Youens-Clark
5 """
6
7 import argparse
8
9
10 # -----
11 def get_args() -> argparse.Namespace:
12     """Get command-line arguments"""\n    ②
13
14     parser = argparse.ArgumentParser(description='Say hello')
15     parser.add_argument('name', metavar='str', help='The name to greet')
16     return parser.parse_args()
17
18
19 # -----
20 def main() -> None:
21     """Start here"""
22
23     args = get_args()
24     print(greet(args.name))
25
26
27 # -----
28 def greet(name: str) -> str:
29     """Create a greeting"""
30
31     return f'Hello, {name}!'
32
33
34 # -----
35 def test_greet() -> None:
36     """Test greet"""
37
38     assert greet('World') == 'Hello, World!'
39     assert greet('Terra Firma') == 'Hello, Terra Firma!'
40
41
42 # -----
43 if __name__ == '__main__':
44     main()

```

① Triple-quoted, multi-line docstring for program/module.

② Triple-quotes can be used on a single line, too.

I also like to add the commented dashes to help me see the beginning of each function. I think it makes my code more readable, but this purely personal taste — you can omit these.

Making the argument optional

Let's return to making the argument to our program optional so that we can run with and without an argument. We'd like to run the program with no argument and have it default to using "World" for the `name` like so:

```
$ ./hello.py  
Hello, World!
```

If we make `name` an *optional* argument, it can no longer be a *positional* argument. Therefore, we have to use either `-n` or `--name` as the name for the optional argument:

```
$ ./hello.py --name Cleveland  
Hello, Cleveland!
```

The "short" name is a single dash and a single letter like `-n`, and the "long" name is two dashes followed by a longer name like `--name`. Here is what the usage looks like now:

```
$ ./hello.py -h  
usage: hello.py [-h] [-n str]  
  
Say hello  
  
optional arguments:  
-h, --help            show this help message and exit  
-n str, --name str    The name to greet (default: World)
```

And here is the code. Note the use of the `formatter_class` to have `argparse` show the default values for arguments in the "usage" output:

```
1 #!/usr/bin/env python3  
2 """  
3 Purpose: Say hello  
4 Author: Ken Youens-Clark  
5 """  
6  
7 import argparse  
8  
9  
10 # -----  
11 def get_args() -> argparse.Namespace:  
12     """Get command-line arguments"""  
13  
14     parser = argparse.ArgumentParser(  
15         description='Say hello',  
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter) ①
```

```

17
18     parser.add_argument('-n',          ②
19                     '--name',       ③
20                     default='World', ④
21                     metavar='str',   ⑤
22                     help='The name to greet')
23
24     return parser.parse_args()
25
26
27 # -----
28 def main() -> None:
29     """Start here"""
30
31     args = get_args()
32     print(greet(args.name))
33
34
35 # -----
36 def greet(name: str) -> str:
37     """Create a greeting"""
38
39     return f'Hello, {name}!'
40
41
42 # -----
43 def test_greet() -> None:
44     """Test greet"""
45
46     assert greet('World') == 'Hello, World!'
47     assert greet('Terra Firma') == 'Hello, Terra Firma!'
48
49
50 # -----
51 if __name__ == '__main__':
52     main()

```

① Show default values in "usage."

② Short option name

③ Long option name

④ The default value.

⑤ Hint to user of the data type.

Starting a new program with `new.py`

In my own practice, I almost never start writing a Python program from an empty page. I created a Python program called `new.py` that helps me start writing new Python programs. As most of my programs need to take parameters, I always use the `argparse` to interpret the command-line options.

I have put my `new.py` program into the `bin` ("binaries" even though these are just text files) directory of the GitHub repo, and I recommend you start every new program with this program.



A central tenet of this book is to create *documented* and *testable* programs. Anything that *can* change about a program should be passed as an *argument* to the program when it is run. For instance, if a program will read an input file for data, the name of that file should be an argument like `-f input1.txt`. Then the fact that the program takes an input file is now visible through the `--help` and we can pass in different files and test if the program processes the files correctly. It is *not* a requirement that you use `new.py` and `argparse`, however. As long as your programs process command-line arguments in the same way as `argparse` and always produce a usage on `-h|--help`, your programs should pass the test suites. The template that `new.py` provides is meant only to make it faster and more convenient to create new programs.

Here is how I would use `new.py` to create a new `hello.py` program:

```
$ new.py hello.py
Done, see new script "hello.py."
```

This is what will be produced:

```
1 #!/usr/bin/env python3
2 """
3 Author : Ken Youens-Clark <kyclark@gmail.com>
4 Date   : 2019-10-21
5 Purpose: Rock the Casbah
6 """
7
8 import argparse
9 import os
10 import sys
11
12
13 # -----
14 def get_args():
15     """Get command-line arguments"""
16
17     parser = argparse.ArgumentParser(
18         description='Rock the Casbah',
19         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
20
21     parser.add_argument('positional',
```

```

22                         metavar='str',
23                         help='A positional argument')
24
25     parser.add_argument('-a',
26                         '--arg',
27                         help='A named string argument',
28                         metavar='str',
29                         type=str,
30                         default='')
31
32     parser.add_argument('-i',
33                         '--int',
34                         help='A named integer argument',
35                         metavar='int',
36                         type=int,
37                         default=0)
38
39     parser.add_argument('-f',
40                         '--file',
41                         help='A readable file',
42                         metavar='FILE',
43                         type=argparse.FileType('r'),
44                         default=None)
45
46     parser.add_argument('-o',
47                         '--on',
48                         help='A boolean flag',
49                         action='store_true')
50
51     return parser.parse_args()
52
53
54 # -----
55 def main():
56     """Make a jazz noise here"""
57
58     args = get_args()
59     str_arg = args.arg
60     int_arg = args.int
61     file_arg = args.file
62     flag_arg = args.on
63     pos_arg = args.positional
64
65     print('str_arg = "{}"'.format(str_arg))
66     print('int_arg = "{}"'.format(int_arg))
67     print('file_arg = "{}"'.format(file_arg.name))
68     print('flag_arg = "{}"'.format(flag_arg))
69     print('positional = "{}"'.format(pos_arg))
70
71
72 # -----

```

```
73 if __name__ == '__main__':
74     main()
```

The arguments that this program will accept are:

1. A single positional argument of the type `str`. *Positional* means it is not preceded by a flag to name it but has meaning because of its position.
2. An automatic `-h` or `--help` flag that will cause `argparse` to print the usage.
3. A named string argument called either `-a` or `--arg`
4. A named integer argument called `-i` or `--int`
5. A named file argument called `-f` or `--file`
6. A boolean (off/on) flag called `-o` or `--on`

Each option here has both a "short" and "long" name. It is not a requirement to have both, but it is common and tends to make your program more readable.

After you use `new.py` to start your new program, you should open it with your editor and modify the argument names and types to suit the needs of your program. For instance, in the "Crow's Nest" chapter, you can delete everything but the positional argument which you should rename from '`positional`' to something like '`word`' (because the argument is going to be a word).

Note that you can control the `name` and `email` values that are used by `new.py` by creating a file called `.new.py` (note the leading dot!) in your home directory. Here is mine:

```
$ cat ~/.new.py
name=Ken Youens-Clark
email=kyclark@gmail.com
```

If you don't want to use `new.py`, then I have included a sample of the above program as `template/template.py` that you can copy. For instance, in the next chapter you should create the program `crowsnest/crowsnest.py`. (That is, from the root directory of the repository, go into the `crowsnest` directory and create a file called `crowsnest.py`.)

Either you can do this with `new.py`:

```
$ cd crowsnest
$ new.py crowsnest.py
```

Or the `cp` (copy) command:

```
$ cp template/template.py crowsnest/crowsnest.py
```

The main point is that I don't want you to have to start every program from scratch! I think it's much easier to start from a complete, working program and modify it.

Types of arguments

As you see above, command-line arguments can be classified as follows:

- **Positional arguments:** The order and number of the arguments is what determines their meaning. Some programs might expect, for instance, a file name as the first argument and an output directory as the second. Typically positional arguments are always required. Making them optional is difficult—how would you write a program that accepts 2 or 3 arguments where the second and third ones are independent and optional? In all the versions of `hello.py` up until the last one, the argument (a name to greet) was positional.
- **Named options:** Standard Unix format allows for a "short" name like `-f` (one dash and a single character) or a "long" name like `--file` (two dashes and a string of characters) followed by some value like a file name or a number. Named options allow for arguments to be provided in any order, so their *position* is not relevant; hence they are the right choice when the user is not required to provide them (they are "options," after all). It's good to provide reasonable default values for options. We changed the required, position `name` argument of `hello.py` to the optional `--name`. Note that some languages like Java might define "long" names with a single dash like `-jar`.
- **Flags:** A "Boolean" value like "yes"/"no" or `True/False` is indicated by something that starts off looking like a named option but there is no value after the name, for example, `-d` or `--debug` to turn on debugging. Typically the presence of the flag indicates a `True` value for the argument; therefore, its absence would mean `False`, so `--debug` turns *on* debugging while its absence means it is off. If you run `ls -s`, the `-s` is the flag to show that you want the files sorted by size.

Using argparse

Let's dig a bit deeper into the `get_args` function which is defined like this:

```
def get_args():
    """Get command-line arguments"""


```

The `def` keyword defines a new function. The arguments to the function are listed in the parentheses. Even though the `get_args` function takes no arguments, the parentheses are still required. The triple-quoted line after the function `def` is the "docstring" which serves as a bit of documentation for the function. Docstrings are not required, but they are good style and `pylint` will complain if you leave them out.

Creating the parser

The following line creates a `parser` that will deal with the arguments from the command line. To "parse" here means to infer some meaning from the order and syntax of the bits of text provided as arguments:

```
parser = argparse.ArgumentParser(  
    description='Argparse Python script',  
    formatter_class=argparse.ArgumentDefaultsHelpFormatter) ③
```

- ① Call the `argparse.ArgumentParser` method to create a new `parser`.
- ② A short, one-line summary of your program's purpose.
- ③ The `formatter_class` argument tells `argparse` to show the default values in usage. See the documentation for other values you can use.

A positional parameter

The following line will create a new *positional* parameter:

```
parser.add_argument('positional',  
    metavar='str',  
    help='A positional argument') ③
```

- ① The lack of leading dashes makes this a positional parameter, not the name "positional."
- ② A hint to the user for the data type. By default, all arguments are strings.
- ③ A brief description of the parameter for the usage.

Remember that the parameter is not positional because the *name* is "positional." That's just there to remind you that it *is* a positional parameter. The `argparse` interprets the string '`positional`' as such because it is not preceded with any dashes like the following options.

An optional string parameter

The following line creates an *optional* parameter with a short name of `-a` and a long name of `--arg` that will be a `str` with a default value of `''` (the empty string). Note that you can leave off either the short or long name in your own programs, but it's good form to provide both. Most of the tests for the exercises will use both short and long option names.

```
parser.add_argument('-a',  
    '--arg',  
    help='A named string argument',  
    metavar='str',  
    type=str,  
    default='') ⑥
```

- ① The short name.
- ② The long name.
- ③ Brief description for the usage.
- ④ Type hint for usage.
- ⑤ The actual Python data type (note the lack of quotes around `str`).
- ⑥

⑥ The default value.

If you wanted to make this a required, named parameter, you would remove the `default` and add `required=True`.

An optional numeric parameter

The following line creates the option called `-i` or `--int` that accepts an `int` (integer) with a default value of `0`. If the user provides anything that cannot be interpreted as an integer, the `argparse` module will stop processing the arguments and will print an error message and a short usage statement:

```
parser.add_argument('-i',  
                   '--int',  
                   help='A named integer argument',  
                   metavar='int',  
                   type=int,  
                   default=0)
```

① The short name.

② The long name.

③ Brief description for usage.

④ Type hint for usage.

⑤ Python data type that the string must be converted to. You can also use `float` for a floating point value (a number with a fractional component like `3.14`).

⑥ The default value.

An optional file parameter

The following line creates an option called `-f` or `--file` that will only accept a valid, readable file. This argument alone is worth the price of admission as it will save you oodles of time validating the input from your user. Note that pretty much every exercise that has a file input will have tests that pass *invalid* file arguments to ensure that your program rejects them.

```
parser.add_argument('-f',  
                   '--file',  
                   help='A readable file',  
                   metavar='FILE',  
                   type=argparse.FileType('r'),  
                   default=None)
```

① The short name.

② The long name.

③ Brief usage.

④ Type suggestion.

⑤ Says that the argument must name a readable ('r') file.

⑥ Default value.

A flag

The flag option is slightly different in that it does not take a value like a string or integer. It's just the name part itself. Flags are either present or not. A common flag is `--debug` to turn *on* debugging statements or `--verbose` to print extra messages to the user. When `--debug` is not present, the default value is "off" (or `False`), which is why the `action` for this argument is `store_true`. It's not necessary to set a `default` value as it is automatically set to `False`.

```
parser.add_argument('-o',          ①
                    '--on',        ②
                    help='A boolean flag', ③
                    action='store_true') ④
```

① Short name.

② Long name.

③ Brief usage.

④ What to do when this is present. The default value is `False`, so `True` will be stored if present.

Returning from `get_args`

The final statement in `get_args` is to `return` the result of having the `parser` object parse the arguments. That is, the code that calls `get_args` will receive this value back:

```
return parser.parse_args()
```

This could fail because `argparse` finds that the user provided invalid arguments, for example, a string value when it expected a `float` or perhaps a misspelled filename. If the parsing succeeds, then we will have a way in our code to access all the values the user provided. Additionally, those values will be of the *types* that we indicated. That is, if we indicate that the `--int` argument should be an `int`, then when we ask for `args.int`, it will already be an `int`. If we define a file argument, we'll get an *open file handle*. That may not seem impressive now, but it's really enormously helpful.

Manually checking arguments

It's also possible to manually validate arguments before you `return` from `get_args`. For instance, we can define that `--int` should be an `int` but how can we require that it must be between 1 and 10? One fairly simple way to do this is to manually check the value. If there is a problem, you can use the `parser.error` function to halt execution of the program, print an error message along with the short usage, and then exit with an error:

```
args = parser.parse_args() ①
if not 1 <= args.int <= 10: ②
    parser.error(f"--int '{args.int}' must be between 1 and 10") ③

return args ④
```

- ① Parse the arguments.
- ② Check if the `args.int` value is *not* between 1 and 10.
- ③ Call `parser.error` with an error message. The entire program will stop, the error message and the brief usage will be shown to the user.
- ④ If we get here, then everything was OK, and the program will continue as normal.

Examples using argparse

A fair number of the tests can be satisfied by learning how to use `argparse` effectively to validate the arguments to your programs. I think of the command line as the boundary of your program, and you need to be judicious about what you let in. You should always expect and defend against every argument being wrong. Our `hello.py` program was an example of a single, positional argument and then a single, optional argument. Let's look at some more examples of how you can use `argparse`.

Two different positional arguments

Imagine you want two *different* positional arguments, like the `color` and `size` of an item to order. The color should be a `str`, and the size should be an `int` value. When you define them positionally, the order in which you declare them is the order in which the user must supply the arguments.

Here we define `color` first and then `size`:

```

1 #!/usr/bin/env python3
2 """Two positional arguments"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """get args"""
10
11     parser = argparse.ArgumentParser(
12         description='Two positional arguments',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('color',
16                         metavar='str',
17                         type=str,
18                         help='Color')
19
20     parser.add_argument('size',
21                         metavar='int',
22                         type=int,
23                         help='Size')
24
25     return parser.parse_args()
26
27
28 # -----
29 def main():
30     """main"""
31
32     args = get_args()
33     print('color =', args.color)
34     print('size =', args.size)
35
36
37 # -----
38 if __name__ == '__main__':
39     main()

```



Again, the user must provide exactly two positional arguments. No arguments triggers a short usage:

```
$ ./two_args.py  
usage: two_args.py [-h] str int  
two_args.py: error: the following arguments are required: str, int
```

Just one won't cut it. We are told that "size" is missing:

```
$ ./two_args.py blue  
usage: two_args.py [-h] str int  
two_args.py: error: the following arguments are required: int
```



If we give two arguments, the second of which can be interpreted as an `int`, all is well:

```
$ ./two_args.py blue 4  
color = blue  
size = 4
```

It's important to note that *all* the arguments coming from the command line are strings because pretty much everything in `bash` (or whatever your shell) is a string. The shell (here `bash`) doesn't require quotes around the `blue` or the `4`. To the shell, these are both strings, and they are passed to Python as strings. When we tell `argparse` that the second argument needs to be an `int`, then `argparse` will do the work to attempt the conversion of the string '`4`' to the integer `4`. If you provide `4.1`, that will be rejected, too:

```
$ ./two_args.py blue 4.1  
usage: two_args.py [-h] str int  
two_args.py: error: argument int: invalid int value: '4.1'
```

Positional arguments have the problem that the user is required to remember the correct order. In the case of switching a `str` and `int`, `argparse` will detect invalid values:

```
$ ./two_args.py 4 blue  
usage: two_args.py [-h] COLOR SIZE  
two_args.py: error: argument SIZE: invalid int value: 'blue'
```

Imagine, however, a case of two strings or two numbers which represent two *different* values like a car's make and model or a person's height and weight. How could you detect that the arguments are reversed? Generally speaking, I only ever create programs that take exactly one positional argument or one or more of *the same thing* like a list of files to process.

Two of the same positional arguments

If you were writing a program that adds two numbers, you could define them as two positional arguments, like `number1` and `number2`. Since they are the same kinds of arguments (two numbers that we will add), it might make more sense to use the `nargs` option to tell `argparse` that you want exactly two of some thing:

```
1 #!/usr/bin/env python3
2 """nargs=2"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """get args"""
10
11     parser = argparse.ArgumentParser(
12         description='nargs=2',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('numbers',
16                         metavar='INT',
17                         nargs=2,
18                         type=int,
19                         help='Two numbers')
20
21     return parser.parse_args()
22
23
24 # -----
25 def main():
26     """main"""
27
28     args = get_args()
29     n1, n2 = args.numbers
30     print('n1 =', n1)
31     print('n2 =', n2)
32
33
34 # -----
35 if __name__ == '__main__':
36     main()
```

The help indicates we want two numbers:

```
$ ./nargs2.py  
usage: nargs2.py [-h] INT INT  
nargs2.py: error: the following arguments are required: INT
```

On line 29, you see we can unpack the two `numbers` into `n1` and `n2` and use them:

```
$ ./nargs2.py 3 5  
n1 = 3  
n2 = 5
```



It's completely safe to unpack `numbers` in this way because we would never get to line 29 if the user hadn't provided exactly two arguments, both of which can be converted to `int` values. Also, notice that the `n1` and `n2` values were actually integers. If they had been strings, then our program would print `35` instead of `8` for the arguments `3` and `5` because the `+` operator in Python both adds numbers and concatenates strings!

```
>>> 3 + 5  
8  
>>> '3' + '5'  
'35'
```

One or more of the same positional arguments

You could expand your 2-number adder into one that sums as many numbers as you provide. When you want *one or more* of some argument, you can use `nargs='+'`:

```

1 #!/usr/bin/env python3
2 """nargs+"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """get args"""
10
11     parser = argparse.ArgumentParser(
12         description='nargs=+',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('numbers',
16                         metavar='INT',
17                         nargs='+',
18                         type=int,
19                         help='Numbers')
20
21     return parser.parse_args()
22
23
24 # -----
25 def main():
26     """main"""
27
28     args = get_args()
29     numbers = args.numbers
30
31     print('{} = {}'.format(' + '.join(map(str, numbers)), sum(numbers)))
32
33
34 # -----
35 if __name__ == '__main__':
36     main()

```

Note that this will mean `args.numbers` is always a `list`. Even if the user provides just one argument, `args.numbers` will be a `list` containing that one value:

```

$ ./nargs+.py 5
5 = 5
$ ./nargs+.py 1 2 3 4
1 + 2 + 3 + 4 = 10

```

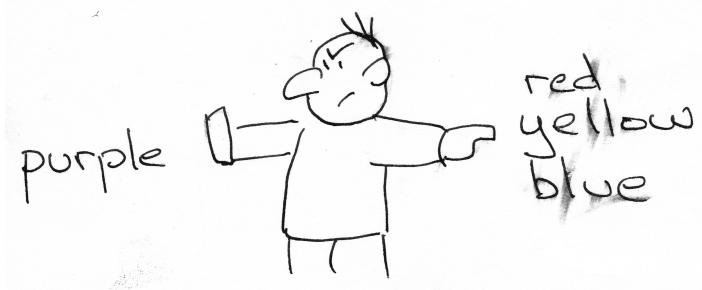
Restricting values using choices

Sometimes you want to limit the values of an argument. Maybe you offer shirts in only primary colors. You can pass in a `list` of valid values using the `choices` option. On line 18 in the following program, I restrict the `color` to one of "red," "yellow," or "blue."

```
1 #!/usr/bin/env python3
2 """Choices"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """get args"""
10
11     parser = argparse.ArgumentParser(
12         description='Choices',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('color',
16                         metavar='str',
17                         help='Color',
18                         choices=['red', 'yellow', 'blue'])
19
20     return parser.parse_args()
21
22
23 # -----
24 def main():
25     """main"""
26
27     args = get_args()
28     print('color =', args.color)
29
30
31 # -----
32 if __name__ == '__main__':
33     main()
```

Any value not present in the list will be rejected and the user will be shown the valid choices. Again, no value is rejected:

```
$ ./choices.py
usage: choices.py [-h] str
choices.py: error: the following arguments are required: str
```



If we provide "purple," it will be rejected because it is not in `choices` we defined. The error message that `argparse` produces tells the user the problem ("invalid choice") and even lists the acceptable colors!

```
$ ./choices.py purple
usage: choices.py [-h] str
choices.py: error: argument str: invalid choice: 'purple' (choose from 'red',
'yellow', 'blue')
```

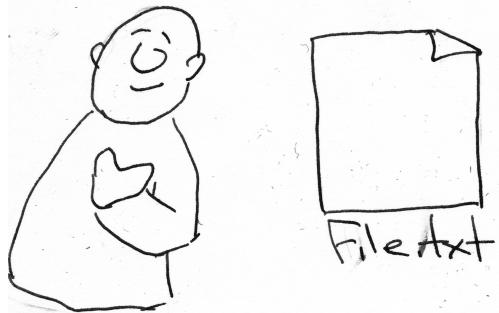
That's really quite a bit of error checking and feedback that you never have to write. The best code is code you don't write!

File arguments

So far we've seen that we can define that an argument should be of a `type` like `str` (which is the default), `int`, or `float`. There are many exercises that require a file as input, and you can use the `type` of `argparse.FileType('r')` to indicate that an argument must be a *file* which is *readable* (the '`r`' part).

Here is an example showing an implementation in Python of the command `cat -n` where `cat` will *concatenate* files and the `-n` says to *number* the lines of output (which is the command I use to create the following numbered line view of the program — how meta):

```
1 #!/usr/bin/env python3
2 """Python version of `cat -n`"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """Get command-line arguments"""
10
11     parser = argparse.ArgumentParser(
12         description='Python version of `cat -n`',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('file',
16                         metavar='FILE',
17                         type=argparse.FileType('r'),
18                         help='Input file')
19
20     return parser.parse_args()
21
22
23 # -----
24 def main():
25     """Make a jazz noise here"""
26
27     args = get_args()
28
29     for i, line in enumerate(args.file, start=1):
30         print(f'{i:6} {line}', end='')
31
32
33 # -----
34 if __name__ == '__main__':
35     main()
```



When I define an argument as `type=int`, I get back an actual `int` value. Here, I define the `file` argument as a file type, and so I receive an *open file handle*. If I had defined the `file` argument as a string, I would have to manually check if it were a file and then use `open` to get a file handle:

```
file = args.file          ①
if not os.path.isfile(file): ②
    print(f'{file} is not a file') ③
    sys.exit(1)                ④

fh = open(file)           ⑤
```

- ① Get whatever the user passed in for the `file`.
- ② Check if this is *not* a file.
- ③ Print an error message.
- ④ Exit the program with a non-zero value.
- ⑤ Proceed to `open` the `file`.

With the file type definition, you don't have to write any of this code. Plus we can directly read line-by-line with a `for` loop on line 29!

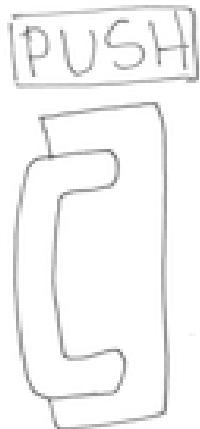
Automatic help

When you define a program's parameters using `argparse`, the `-h` and `--help` flags will be reserved for generating help documentation. You do not need to add these nor are you allowed to use these flags for other purposes.

I think of this documentation like a door to your program. Doors are how we get into buildings and cars and such. Have you ever come across a door that you can't figure out how to open? Or one that

requires a "PUSH" sign when clearly the handle is design to "pull"?

The book *The Design of Everyday Things* by Don Norman uses the term "affordances" to describe the interfaces that objects present to us which do or do not inherently describe how we should use



them.

The usage statement of your program is like the handle of the door. It should let me know exactly how to use it. When I encounter a program I've never used, I either run it with no arguments or with `-h` or `--help`. I *expect* to see some sort of usage statement. The only alternative would be to open the source code itself and study how to make the program run and how I can alter it, and this is a truly unacceptable way to write and distribute software!

When you start a new program with `new.py foo`, this is the help that will be generated:

```
$ ./foo.py -h
usage: foo.py [-h] [-a str] [-i int] [-f FILE] [-o] str

Argparse Python script

positional arguments:
  str                  A positional argument

optional arguments:
  -h, --help            show this help message and exit
  -a str, --arg str    A named string argument (default: )
  -i int, --int int    A named integer argument (default: 0)
  -f FILE, --file FILE A readable file (default: None)
  -o, --on              A boolean flag (default: False)
```

Without writing a single line of code, you have

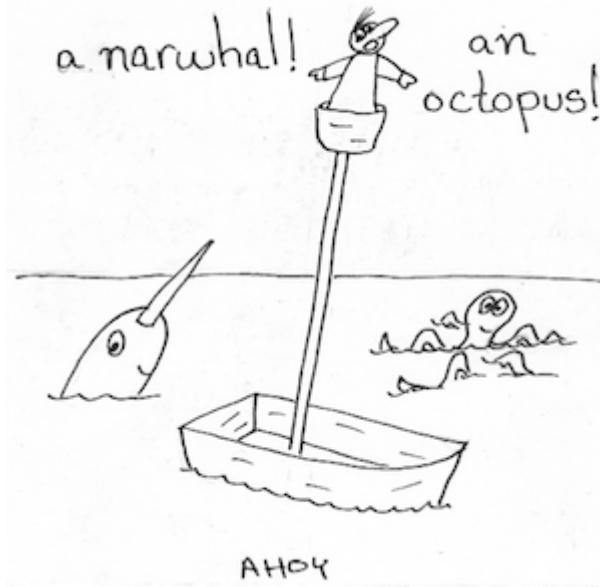
1. an executable Python program
2. that accepts command line arguments
3. and generates a standard and useful help message

This is the "handle" to your program.

Summary

- The `argparse` module will help you document and parse all the parameters to your program. You can validate the types and numbers of arguments which can be positional, optional, or flags. The usage will be automatically generated.
- You should write many small, limited functions and `test_` functions that you can run with `pytest`. It is often best to write the tests before you write the functions to imagine how they ought to work.
- You should run your tests after any change to your program to ensure that everything still works.
- You should automatically format your code with `yapf` and `black` (and probably other programs).
- You should add type hints to your code and use `mypy` to check for errors.
- You should "lint" your code with `pylint` and/or `flake8` to find both programmatic and style problems.
- You can use the `new.py` program to generate new Python programs that use `argparse`.

Chapter 2: The Crow's Nest: Working with strings



Avast, you corny-faced gollumpus! Ye are on the barrelman for this watch. D'ye ken what I mean, ye addle pated blunderbuss?! Ah, land lubber ye be! OK, then, you are the lookout in the crow's nest—the little bucket attached to the top of a mast of a sailing ship. Your job is to keep a lookout for interesting or dangerous things, like a ship to plunder or an iceberg to avoid. When you see something like a "narwhal," you are supposed to cry out, "Ahoy, Captain, **a narwhal** off the larboard bow!" If you see an octopus, you'll shout "Ahoy, Captain, **an octopus** off the larboard bow!" (We'll assume everything is "off the larboard bow" for this exercise. It's a great place for things to be.)

From this point on, I will describe a coding challenge that you should write on your own. I will discuss key ideas you'll need to solve the problems as well as how to use the provided tests to help you know when your program is correct. You should have a copy of the Git repository locally (see the setup instructions). You should write your program in the chapter's directory, like this program should be written in the `crowsnest` directory where the tests for the program live.

In this chapter, we're going to start off working with strings. By the end, you will be able to:

- Create a program that accepts a positional argument and produces usage documentation
- Create a new output string depending on the inputs to the program
- Run a test suite

Your program should be called `crowsnest.py`. It will accept a single positional argument and will print the given argument inside the "Ahoy" bit along with the word "a" or "an" depending on whether the argument starts with a consonant or a vowel.

That is, if given "narwhal," it should do this:

```
$ ./crowsnest.py narwhal  
Ahoy, Captain, a narwhal off the larboard bow!
```

And if given "octopus":

```
$ ./crowsnest.py octopus  
Ahoy, Captain, an octopus off the larboard bow!
```

This means we're going to need to write a program that accepts some input on the command line, decides on the proper article ("a" or "an") for the input, and prints out a new string that puts those two values into the "Ahoy" phrase.

Getting started

You're probably ready to start writing the program! Well, hold on just a minute longer, ye duke of limbs. We need to discuss how we'll use the tests to know when our program is working and how we might get started programming.

How to use the tests

"The greatest teacher, failure is." — Yoda

In the code repository, I've included tests that will guide you in the writing of your program. Before you even write the first line of code, I'd like you to run `make test` or `pytest -xv test.py` so you can see how the first test fails. Be sure you are in the `crowsnest` directory for this!

Among all the output, you'll notice this line:

```
test.py::test_exists FAILED
```

[16%]

If you read more, you'll see lots of other output all trying to convince you that the expected file, `crowsnest.py` does not exist. Learning to read the test output is a skill in itself! It takes quite a bit of practice to learn to read test output, so try not to feel overwhelmed. In my terminal (iTerm on a Mac), the output from `pytest` shows colors and bold print to highlight key failures. The text in bold, red letters is usually where I start, but your terminal may behave differently.

Creating programs with `new.py`

In order to pass this test, we need to create a file called `crowsnest.py` inside the `crowsnest` directory where `test.py` is located. While it's perfectly fine to start writing from scratch, I suggest you use the `new.py` program to print some useful boilerplate code that you'll need in every exercise:

```
$ new.py crowsnest.py  
Done, see new script "crowsnest.py."
```

If you don't want to use `new.py`, you could copy the `template/template.py` program:

```
$ cp template/template.py crowsnest/crowsnest.py
```

At this point you should have the outline of a working `crowsnest.py` program that accepts command-line arguments. If you run your new `crowsnest.py` with no arguments, it will print a short usage statement like the following (notice how "usage" is the first word of the output):

```
$ ./crowsnest.py
usage: crowsnest.py [-h] [-a str] [-i int] [-f FILE] [-o] str
crowsnest.py: error: the following arguments are required: str
```

Those are not the correct parameters for our program, just the default examples given to you by `new.py`. You will need to modify them to suit this program.

If you run your tests again, you will pass the first *two* tests that check:

1. Does the program exist?
2. Does the program print something that looks like "usage."

And then you will fail the third test. There are more tests after this, but that's all you see if you run `pytest -xv test` because the `-x` flag tells `pytest` to stop at the first failing test. Note that we can combine the `-x` and `-v` flags into `-vx` and that the order doesn't matter, so `-vx` is fine, too:

test.py::test_exists PASSED	[16%]	①
test.py::test_usage PASSED	[33%]	②
test.py::test_consonant FAILED	[50%]	③

- ① This test checks if the file `crowsnest.py` exists.
- ② This test runs `crowsnest.py` to see if it produces a "usage" statement.
- ③ This test runs the program and passes it a word starting with a consonant to see if it produces the expected output. It has `FAILED`, and so all testing stops because of the `-x` flag.

Now we have a working program that accepts some arguments (but not the right ones). Next we need to make our program accept the "narwhal" or "octopus" value that needs to be announced, and we'll use command-line arguments to do that.

Defining your arguments

Here is a diagram sure to shiver your timbers showing the inputs (or *parameters*) and output of the program. We'll use these throughout the book to imagine how code and data work together. In this program, some "word" is the input, and a phrase incorporating that word with the correct article is the output.

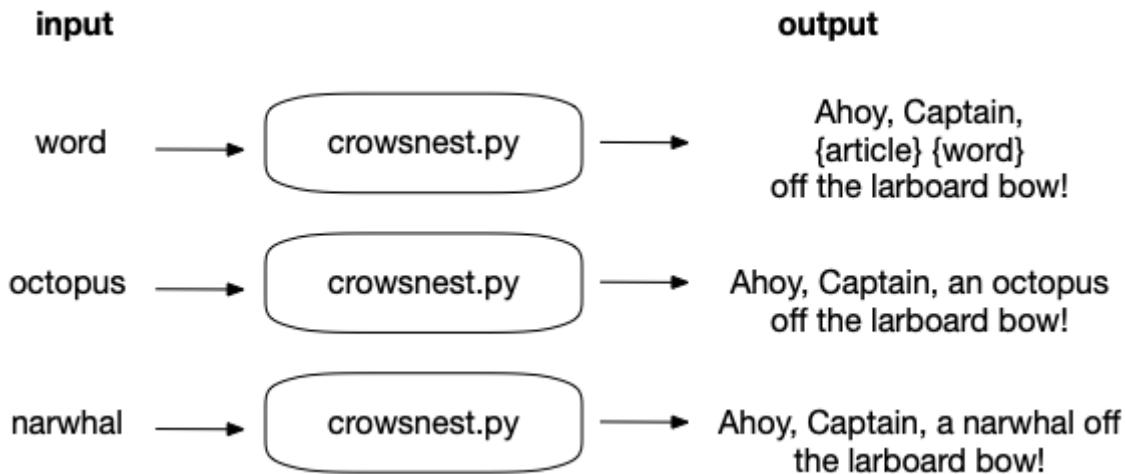


Figure 2. The input to the program is a word, and the output is that word plus its proper article (and some other stuff).

We need to modify the part of the program that gets the arguments—the aptly named `get_args` function. This function uses the `argparse` module to parse the command-line arguments. Refer to the `argparse` discussion in the introduction, particularly the section "A single, positional argument." The default `get_args` names the first argument '`positional`', and that's the only one you need. Remember that positional arguments are defined by their position and so don't have names that start with dashes. You can delete all the arguments except for the positional `word`.

Modify the `get_args` part of your program until it will print this usage:

```
$ ./crowsnest.py
usage: crowsnest.py [-h] str
crowsnest.py: error: the following arguments are required: str
```

Likewise, it should print a longer usage for the `-h` or `--help` flag:

```
$ ./crowsnest.py -h
usage: crowsnest.py [-h] str

Crow's Nest -- choose the correct article

positional arguments:
  str      A word ①

optional arguments:
  -h, --help show this help message and exit ②
```

① You need to define a `word` parameter. Notice that it is listed as a "positional" argument.

② The `-h` and `--help` flags are created automatically by `argparse`. You are not allowed to use these as options. They are used to create the documentation for your program.

When your program prints the correct usage, you can get the `word` argument inside the `main` function like so:

```
def main():
    args = get_args()
    word = args.word
```

Make your program print the given `word`:

```
$ ./crowsnest.py narwhal
narwhal
```

And now run your tests. You should still be passing two and failing the third. Let's read the test failure:

```
===== FAILURES =====
----- test_consonant -----
def test_consonant():
    """brigatine -> a brigatine"""

    for word in consonant_words:
        out = getoutput('{} {}'.format(prg, word))
>       assert out.strip() == template.format('a', word)
①
E           AssertionError: assert 'brigatine' == 'Ahoy, Captain, a brigatine off the
larboard bow!' ②
E               - brigatine
③
E               + Ahoy, Captain, a brigatine off the larboard bow!
④
```

- ① The line starting with `>` shows code that produced an error. The output of the program is compared to an expected string. Since it didn't match, the `assert` produces an exception.
- ② This line starts with `E` to indicate the "error."
- ③ The line starting with a `-` is what the test got when it ran with the argument '`brigatine`' — it got back just the word "brigatine."
- ④ The line starting with the `+` is what the test expected, "Ahoy, Captain, a brigatine off the larboard bow!"

So, we need to get the `word` into the "Ahoy" phrase. How can we do that?

Concatenating strings

Putting strings together is called "concatenating" or "joining" strings. To demonstrate, I'm going to enter some code directly into the Python interpreter. I want you to type along. No, really! Type everything you see, and try it for yourself.

Open a terminal and type `python3` or `ipython` to start a REPL, a "Read-Evaluate-Print-Loop" because Python will *read* each line of input, *evaluate* and *print* the results in a *loop*. (I pronounce this like "reh-pull" in a way that kind of, sort of rhymes with "pebble.") Here's what it looks like on my system:

```
$ python3
Python 3.7.3 (default, Mar 27 2019, 16:54:48)
[Clang 4.0.1 (tags/RELEASE_401/final)] :: Anaconda, Inc. on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

You may also like to use Python's IDLE (integrated development environment) program, or you can use Jupyter notebooks to interact with the language. I'll stick to the simple `python3` REPL for showing code examples. To exit the REPL, either type `quit()` or `CTRL-d` (the `Control` key plus the `d`).

The `>>>` is a prompt where you can type code. Let's start off by assigning the variable `word` to the value "narwhal." In the REPL, type `word = 'narwhal'<Enter>`:

```
>>> word = 'narwhal'
```

Note that you can put as many (or no) spaces around the `=` as you like, but convention and readability (and tools like `pylint` or `flake8` that help you find errors in your code) would ask you to use exactly one space on either side. If you type `word<Enter>`, Python will print the current value of `word`:

```
>>> word
'narwhal'
```

Now type `werd<Enter>`:

```
>>> werd
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'werd' is not defined
```



There is no `werd` variable because we haven't set `werd` to be anything. Using an undefined variable causes an *exception* that will crash your program. Python will happily create a `werd` for you when you assign it a value.

In Python, there are many ways we can concatenating strings. The `+` operator can be used to join strings together:

```
>>> 'Ahoy, Captain, a ' + word + ' off the larboard bow!'
'Ahoy, Captain, a narwhal off the larboard bow!'
```

If you change your program to `print` that instead of just the `word`, you should be able to four tests:

test.py::test_exists PASSED	[16%]
test.py::test_usage PASSED	[33%]
test.py::test_consonant PASSED	[50%]
test.py::test_consonant_upper PASSED	[66%]
test.py::test_vowel FAILED	[83%]

If we look closely at the failure, you'll see this:

```
E      - Ahoy, Captain, a aviso off the larboard bow!
E      + Ahoy, Captain, an aviso off the larboard bow!
E      ?           +
```

So we hard-coded the "a" before the `word`, but we really need to figure out whether to put "a" or "an" depending on whether the `word` starts with a vowel. How can we do that?

Variable types

Before we go much further, I need to take a small step back and point out that our `word` variable is a "string." Every variable in Python has a "type" that describes the kind of data they hold. Because we put the value for `word` in quotes ('narwhal'), the `word` holds a "string" which Python represents with a class called `str`. (A "class" is a collection of code and functions that we can use.)

The `type` function will tell us what kind of data Python thinks this is:

```
>>> type(word)
<class 'str'>
```

Whenever you put a value in single (' ') or double quotes (" "), Python will interpret it as a `str`:

```
>>> type("submarine")
<class 'str'>
```



If you forget the quotes, then Python will look for some variable or function by that name. If there is no variable or function by that name, it will cause an exception.

```
>>> word = narwhal
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'narwhal' is not defined
```

Exceptions are bad, and we will try to write code that avoids them or at least knows how to handle them gracefully.

Getting just part of a string

Back to our problem! We need to put either "a" or "an" in front of the `word` we're given based on whether the first character of `word` is a vowel or a consonant. In Python, we use square brackets and an *index* to get an individual character from a string. The index is the numeric position of an element in a sequence, and we must remember that indexing starts at `0`. You can use this with a

narwhal
0 1 2 3 4 5 6



variable:

```
>>> word[0]  
'n'
```

Or directly on a string:

```
>>> 'narwhal'[0]  
'n'
```

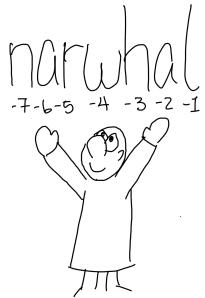
narwhal
0 1 2 3 4 5 6



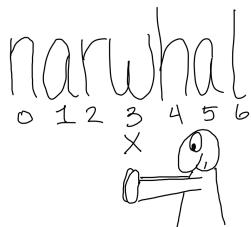
This means that the last index is *one less than the length*, which is often confusing. The length of "narwhal" is 7, but the last character is found at index `6`:

```
>>> word[6]  
'l'
```

You can also use negative numbers to count backwards from the end, so the last index is also `-1`:



```
>>> word[-1]  
'l'
```



You can use the "slice" notation `[start:stop]` to get a range of characters. Both `start` and `stop` are optional. The default value for `start` is `0` (the beginning of the string), and the `stop` value is *not inclusive*:

```
>>> word[:3]  
'nar'
```

And the default value for `stop` is the end of the string:

```
>>> word[3:]  
'whal'
```

In the next chapter, we'll see that this is the same syntax for slicing lists. A string is (sort of) a list of characters, so this isn't too strange.

Finding help in the REPL

The class `str` has a ton of functions we can use to handle strings, but what are they? A large part of programming is knowing how to ask questions and where to look for answers. A common refrain you may hear is "RTFM"—Read the Fine Manual. The Python community has created reams of documentation which are all available at <https://docs.python.org/3/>. You will need to refer to the documentation constantly to remind yourself how to use certain functions.



The docs for the string class are here:

<https://docs.python.org/3/library/string.html>

I prefer to read the docs directly inside the REPL by typing `help(str)`:

```
>>> help(str)
```

Inside the `help`, you move up and down in the text using the up and down cursor arrows on your keyboard. You can also press the `<Space>` bar or `CTRL-f` to jump forward to the next page, and `CTRL-b` to jump backward. You can search through the documentation by pressing `/` and then the text you want to find. If you press `n` (for "next") after a search, you will jump to the next place that string is found. To leave the help, press `q` (for "quit").

String methods



Now that we know `word` is a string (`str`), we have all these incredibly useful *methods* we can call on the variable. (A "method" is a function that belongs to a variable like `word`.) For instance, if I wanted to shout about the fact that we have a "narwhal," I could print it in UPPERCASE LETTERS. If I search through the help, I see there is a function called `upper`. Here is how to call it:

```
>>> word.upper()  
'NARWHAL'
```

You must include the parentheses `()` or else you're talking about the *function itself*:

```
>>> word.upper  
<built-in method upper of str object at 0x10559e500>
```

That will actually come in handy later when we use functions like `map` and `filter`, but for now we want Python to *execute* or *call* the `upper` function on the variable `word`, so we add the parens. Note that the function returns an uppercase version of the word but *does not change the value of word itself*:

```
>>> word  
'narwhal'
```

There is another `str` function with "upper" in the name called `isupper`. The name helps you know that this will return a True/False type answer. Let's try it:

```
>>> word.isupper()  
False
```

We can chain methods together like so:

```
>>> word.upper().isupper()  
True
```

That makes sense. If I convert the `word` to uppercase, then `isupper` is `True`.

I find it odd that the `str` class does not include a method to get the length of a string. For that, we



use a separate function called `len`, short for "length":

```
>>> len('narwhal')  
7
```

Are you typing all this into Python yourself? I recommend you do! Find other methods in the `str` help and try them out.

String comparisons

So now you know how to get the first letter of `word` by using `word[0]`. Let's assign it to the variable `char`:

```
>>> word = 'octopus'  
>>> char = word[0]  
>>> char  
'o'
```

Now we need to figure out if `char` is a vowel or a consonant. We'll say that letters "a," "e," "i," "o," and "u" make up our set of "vowels." You can use `==` to compare strings:

```
>>> char == 'a'  
False  
>>> char == 'o'  
True
```



Be careful to always use one equal sign (`=`) when *assigning a value* to a variable, like `word = 'narwhal'` and two equal signs (`==`, which, in my head, I say "equal-equal") when you *compare two values* like `word == 'narwhal'`. The first is a statement that changes the value of `word`, and the second is an *expression* that returns `True` or `False`.

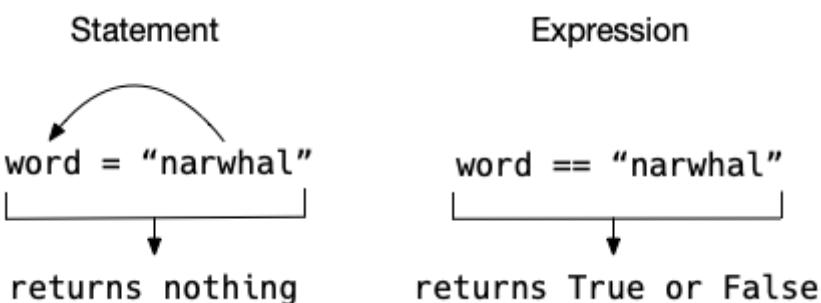


Figure 3. An expression returns a value. A statement does not.

We need to compare our `char` to *all* the vowels. You can use `and` and `or` in such comparisons and they will be combined according to standard Boolean algebra:

```
>>> char == 'a' or char == 'e' or char == 'i' or char == 'o' or char == 'u'  
True
```

What if the `word` is "Octopus" or "OCTOPUS"?

```
>>> word = 'OCTOPUS'  
>>> char = word[0]  
>>> char == 'a' or char == 'e' or char == 'i' or char == 'o' or char == 'u'  
False
```

Do we have to make 10 comparisons in order to check the uppercase versions, too? What if we were to lowercase the `word`?

```
>>> word = 'OCTOPUS'  
>>> char = word.lower()[0]  
>>> char == 'a' or char == 'e' or char == 'i' or char == 'o' or char == 'u'  
True
```

a e i o u



a



b

An easier way to determine if `char` is a vowel would be to use Python's `x in y` construct where we want to know if the value `x` is in the collection `y`. We can ask if the letter '`a`' is in the longer string '`aeiou`':

```
>>> 'a' in 'aeiou'  
True
```

But the letter '`b`' is not:

```
>>> 'b' in 'aeiou'  
False
```

Now use it to test on a lowercased `char` (which is '`o`'):

```
>>> char.lower() in 'aeiou'  
True
```

Conditional branching

Once you have figured out if the first letter is a vowel, you will need to select an article. We'll use a very simple rule that, if the word starts with a vowel, choose "an," otherwise choose "a." This misses exceptions like when the initial "h" in a word is silent. For instance, we say both "a hat" and "an honor". Nor will we consider when an initial vowel has a consonant sound as in "union" where the "u" sounds like a "y."

We can create a new variable called `article` that we will set to the empty string and then use an `if` / `else` statement to figure out what to put in it:

```
>>> article = ''          ①  
>>> if char in 'aeiou': ②  
...     article = 'an'   ③  
... else:  
...     article = 'a'    ⑤  
...  
...
```

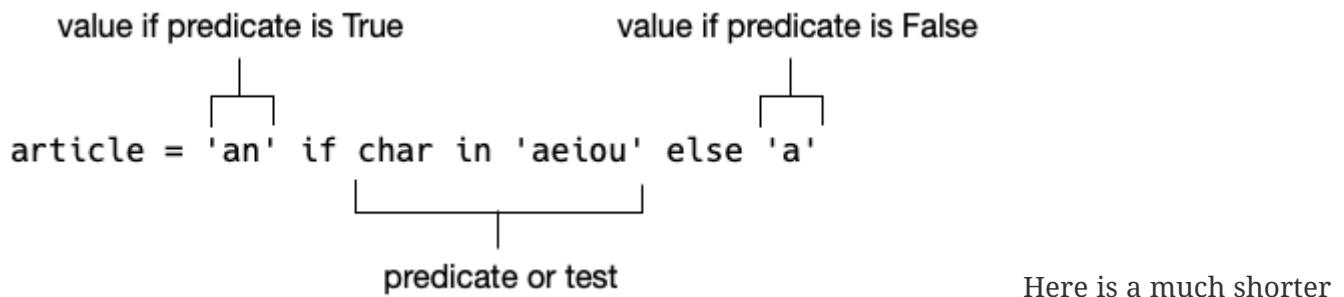
① Initialize `article` to the empty string.

② Check if `char` is a vowel.

③ If `char` is a vowel, set `article` to 'an'

④ Otherwise...

⑤ Set `article` to 'a'.



Here is a much shorter

way to write that with an `if` expression (expressions return values, statements do not). The `if` expression is written a little backwards. First comes the value if the test (or "predicate") is `True`, then the predicate, then the value if the predicate is `False`.

This way is also safer because the `if` expression is *required* to have the `else`. There's no chance that we could forget to handle both cases:

```
>>> article = 'an' if char in 'aeiou' else 'a'
```

And we can verify that we have the correct `article`:

```
>>> article  
'an'
```

String formatting

Now we have two variables, `article` and `word` that need to be incorporated into our "Ahoy!" phrase. We saw earlier that we can use the plus sign (`+`) to concatenate strings. Another method to create new strings from other strings is to use the `str.format` method. To do so, you create a string template with curly brackets `{}` that indicate placeholders for values. The values that will be substituted go as arguments to the `format`, and they are substituted in the same order that the `{}` appear:

```
'Ahoy, Captain, {} {} off the larboard bow!'.format(article, word)
```

Here it is in code:

```
>>> 'Ahoy, Captain, {} {} off the larboard bow!'.format(article, word)  
'Ahoy, Captain, an octopus off the larboard bow!'
```

Another method uses the special "f-string" where you can put the variables directly into the `{}`

brackets. It's a matter of taste which one you choose.

```
>>> f'Ahoy, Captain, {article} {word} off the larboard bow!'
'Ahoy, Captain, an octopus off the larboard bow!'
```

Python variables are very variable



A note that in some programming languages, you have to declare the variable's name and what *type* of data it will hold. If a variable is declared to be a number, then it can never hold a value of a different type like a string. This is called *static typing* because the type of the variable can never change. Python is a *dynamically typed* language because you do not have to declare a variable or what kind of data the variable will hold. You can change the value and type of data at any time. This could be either great or terrible news. As Hamlet says, "There is nothing either good or bad, but thinking makes it so."

Hints:

- Start your program with `new.py` and fill in the `get_args` with a single position argument called `word`.
- You can get the first character of the word by indexing it like a list, `word[0]`.
- Unless you want to check both upper- and lowercase letters, you can use either the `str.lower` or `str.upper` method to force the input to one case for checking if the first character is a vowel or consonant.
- There are fewer vowels (five, if you recall) than consonants, so it's probably easier to check if the first character is one of those.
- You can use the `x in y` syntax to see if the element `x` is `in` the collection `y` where "collection" here is a `list`.
- Use the the `str.format` or f-strings to insert the correct article for the given word into the longer phrase.
- Run `make test` (or `pytest -xv test.py`) after every change to your program to ensure your program compiles and is on the right track.

Now go write the program before you turn the page and study a solution! Look alive, you ill-tempered shabaroon!

Solution

```
1 #!/usr/bin/env python3
2 """Crow's Nest"""
3
4 import argparse
5
6
7 # -----
8 def get_args(): ①
9     """Get command-line arguments"""
10
11     parser = argparse.ArgumentParser(②
12         description="Crow's Nest -- choose the correct article", ③
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter) ④
14
15     parser.add_argument('word', metavar='str', help='A word') ⑤
16
17     return parser.parse_args() ⑥
18
19
20 # -----
21 def main(): ⑦
22     """Make a jazz noise here"""
23
24     args = get_args() ⑧
25     word = args.word ⑨
26     article = 'an' if word[0].lower() in 'aeiou' else 'a' ⑩
27
28     print(f'Ahoy, Captain, {article} {word} off the larboard bow!') ⑪
29
30
31 # -----
32 if __name__ == '__main__': ⑫
33     main() ⑬
```

- ① Defines the function `get_args` to handle the command-line arguments. I like put this first so I can see it right away when I'm reading the code.
- ② The `parser` will do the work of parsing the arguments.
- ③ The `description` shows in the usage to describe what the program does.
- ④ Show the default values for each parameter in the usage.
- ⑤ Define a positional argument called `word`.
- ⑥ The result of parsing the arguments will be returned to line 24.
- ⑦ Defines the `main` function where the program will start.
- ⑧ `args` contains the return value from the `get_args` function.

- ⑨ Put the `args.word` value from the arguments into the variable `word`.
- ⑩ Choose the correct `article` using an `if` expression to see if the lowercased, first character of `word` is or is not in the set of vowels.
- ⑪ Print the output string using an f-string to interpolate the `article` and `word` variables inside the string.
- ⑫ Check if we are in the "main" namespace, which means the program is *running*.
- ⑬ If so, call the `main()` function to make the program start.

Discussion

I'd like to stress that the preceding is *a* solution, not *the* solution. There are many ways to express the same idea in Python. As long as your code passes the test suite, it is correct.

That said, I created my program with `new.py` which automatically gives me two functions:

1. `get_args` where I define the arguments to the program
2. `main` where the program starts

Let's talk about these two functions.

(I imagine the above code could be annotated with big numbers that could be referenced by the following sections. E.g., a big "1" next to the `get_args` function and a big "2" next to `main`, etc. Or at least there could be some simple call-outs on the above code to generally explain what's happening, although I think the function names themselves are pretty explanatory.)

Defining the arguments with `get_args`

I prefer to put the `get_args` function first so that I can see right away what the program expects as input. You don't have to define this as a separate function. You could put all this code inside `main`, if you prefer. Eventually our programs are going to get longer, though, and I think it's nice to keep this as a separate idea. Every program I present will have a `get_args` function that will handle defining and validating the input.

Our program specifications (the "specs") say that the program should accept one positional argument. I changed the '`positional`' argument name to '`word`' because I'm expecting a single word:

```
parser.add_argument('word', metavar='str', help='Word')
```

I would really recommend you never leave the "positional" argument named '`positional`' because it is an entirely undescriptive term. Naming your variables *what they are* will make your code more readable. Since the program doesn't need any of the other options created by `new.py`, you can delete the rest of the `parser.add_argument` calls. The `get_args` function will `return` the result of parsing the command line arguments which I put into the variable `args`:

```
return parser.parse_args()
```

If `argparse` is not able to parse the arguments — for example, there are none — it will never `return` from `get_args` but will instead print the "usage" for the user and exit with an error code to let the operating system know that the program exited without success. (In the Unix world, an exit value of `0` means there were 0 errors. Anything other than `0` is considered an error.)

The main thing

Many programming languages will automatically start from the `main` function, so I always define a `main` function and start my programs there. This is not a requirement, just how I like to write programs. Every program I present will start with the `main` function which will first call `get_args` to get the program's inputs:

```
args = get_args()
```

I can now access the `word` by call `args.word`. Note the lack of parentheses. It's not `args.word()` because is not a function call. Think of `args.word` like a slot where the value of the "word" lives:

```
word = args.word
```

I like to work through my ideas using the REPL, so I'm going to pretend that `word` has been set to "octopus":

```
>>> word = 'octopus'
```

Classifying the first character of a word

To figure out whether the article I choose should be `a` or `an`, I need to look at the first character of the `word` which we can get like so. In the introduction, we used this:

```
>>> word[0]
'o'
```

I can check if the first character is `in` the string of vowels, both lower- and uppercase:

```
>>> word[0] in 'aeiouAEIOU'
True
```

I can make this shorter, however, if I use `word.lower` function so I'd only have to check the lowercase vowels:

```
>>> word[0].lower() in 'aeiou'
True
```

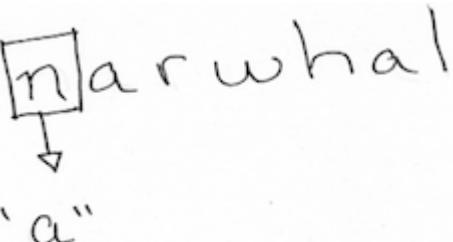
Remember that the `x in y` form is a way to ask if element `x` is in the collection `y`. You can use it for letters in a longer string (like the vowels):

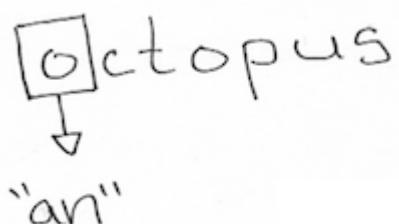
```
>>> 'a' in 'aeiou'  
True
```

Or for a string in list of other strings:

```
>>> 'tanker' in ['yatch', 'tanker', 'vessel']  
True
```

We can use membership in the "vowels" as a condition to choose "an," otherwise we choose "a":


"a"


"an"

As mentioned in the introduction, the **if** expression is the shortest and safest for a "binary" choice (where there are only two possibilities):

```
>>> article = 'an' if word[0].lower() in 'aeiou' else 'a'  
>>> article  
'an'
```

The safety comes from the fact that Python will not even run this program if you forget the **else**. We can change the **word** to "galleon" and check that it still works:

```
>>> word = 'galleon'  
>>> article = 'an' if word[0].lower() in 'aeiou' else 'a'  
>>> article  
'a'
```

Printing the results

Finally we need to **print** out the phrase with our **article** and **word**. As noted in the introduction, you can use **str.format**:

```
>>> article = 'a'  
>>> word = 'ketch'  
>>> print('Ahoy, Captain, {} {} off the larboard bow!'.format(article, word))  
Ahoy, Captain, a ketch off the larboard bow!
```

Python's f-strings will *interpolate* any code inside the `{}` placeholders, so variables get turned into their contents:

```
>>> print(f'Ahoy, Captain, {article} {word} off the larboard bow!')  
Ahoy, Captain, a ketch off the larboard bow!
```

However you chose to print out the article and word is fine as long as it passes the tests.

Running the test suite

"A computer is like a mischievous genie. It will give you exactly what you ask for, but not always what you want. - Joe Sondow"

Computers are a bit like bad genies. They will do exactly what you tell them but not necessarily what you *want*. In an episode of *The X-Files*, the character Mulder wishes for peace on Earth and a genie removes all humans but him.

Tests are what we can use to verify that our programs are doing what we actually want them to do. Tests they can never prove that our program is actually, completely free of errors, only that the bugs we imagined or found while writing the program no longer exist. Still, we write and run tests because they are really quite effective and much better than not doing so.

This is the idea behind "test-driven development":

- We can write tests *even before* we write the software.
- We run the tests to verify that our as-yet-unwritten software definitely fails to deliver on some task.
- Then we write the software to fulfill the request.
- Then we run the test to check that it now *does* work.
- We keep running all the tests to ensure that, when we add some new code, we do not break existing code.

Passing Tests

I would encourage you to look at the `test.py` program to see how it is testing your program. Eventually I'll recommend you write your own tests, but for now just see what's being expected of your code. I use the `pytest` module to write tests. There are other testing frameworks in Python, but I find `pytest` to be relatively easy to use. The `pytest` module will run any functions that begin with `test_` in the order they are found in the source code.

The first `test_` function is `test_exists` that uses the `assert` function to check if the `crowsnest.py` program exists. This is why your program must be named '`crowsnest.py`'. It must exist as this name so that we can run it and check the output:

```
prg = './crowsnest.py'

def test_exists():
    """exists"""

    assert os.path.isfile(prg)
```

The next is `test_usage` to check if the program will print something that looks like "usage" when run with `-h` and `--help` flags:

```
def test_usage():
    """usage"""

    for flag in ['-h', '--help']:
        rv, out = getstatusoutput('{} {}'.format(prg, flag))
        assert rv == 0
        assert out.lower().startswith('usage')
```

Inside `test.py`, there are two lists of words, one starting with consonants and one starting with vowels.

```
consonant_words = [
    'brigantine', 'clipper', 'dreadnought', 'frigate', 'galleon', 'haddock',
    'junk', 'ketch', 'longboat', 'mullet', 'narwhal', 'porpoise', 'quay',
    'regatta', 'submarine', 'tanker', 'vessel', 'whale', 'xebec', 'yatch',
    'zebrafish'
]
vowel_words = ['aviso', 'eel', 'iceberg', 'octopus', 'upbound']
```

There is also a string `template` for what the program should print:

```
template = 'Ahoy, Captain, {} {} off the larboard bow!'
```

The `test_consonant` test runs through each of the `consonant_words` and checks if the program puts an "a" in front of the word.

```

def test_consonant():
    """brigatine -> a brigatine"""

    for word in consonant_words:
        out = getoutput('{} {}'.format(prg, word))
        assert out.strip() == template.format('a', word)

```

The next function does the same thing but uses a capitalized version of the consonant word. The next two tests then use the `vowel_words`, checking both lower- and uppercase versions.

When all tests are passing, this is the output you should see:

```

$ make test
pytest -xv test.py
=====
platform darwin -- Python 3.7.3, pytest-5.0.1, py-1.8.0, pluggy-0.12.0 --
/Users/kyclark/anaconda3/bin/python3
cachedir: .pytest_cache
rootdir: /Users/kyclark/work/manning/playful_python/crowsnest
plugins: openfiles-0.3.2, arraydiff-0.3, doctestplus-0.3.0, remotedata-0.3.1, cov-2.7.1
collected 6 items

test.py::test_exists PASSED [ 16%]
test.py::test_usage PASSED [ 33%]
test.py::test_consonant PASSED [ 50%]
test.py::test_consonant_upper PASSED [ 66%]
test.py::test_vowel PASSED [ 83%]
test.py::test_vowel_upper PASSED [100%]

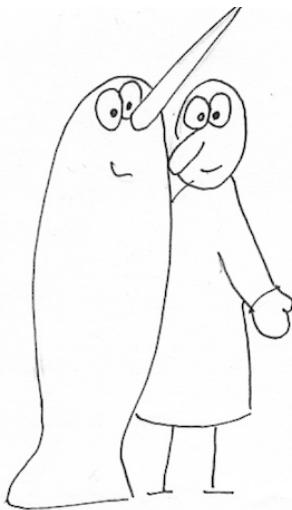
===== 6 passed in 2.28 seconds =====

```

Summary

- All Python's documentation is available on <https://docs.python.org/3/> and with the `help` command in the REPL.
- Variables in Python are dynamically typed according to whatever value you assign them and they come into existence when you assign a value to them.
- Strings have methods like `upper` and `isupper` that you can call to alter them or get information.
- You can get parts of a string by using square brackets and indexes like `[0]` for the first letter or `[-1]` for the last.
- You can concatenate strings with the `+` operator.
- The `str.format` method allows you to create a template with `{}` placeholders that get filled in with the arguments.
- F-strings like `f'{article} {word}'` allow variables and code to go directly inside the brackets.
- The `x in y` expression will report if the value `x` is present in the collection `y`.
- Statements like `if/else` do not return a value while expressions like `x if y else z` do return a value.
- Test-driven development is a way to ensure programs meet some minimum criteria of correctness. Every feature of a program should have tests, and writing and running test suites should be an integral part of writing programs.

Going Further



- Have your program match the case of the incoming word, e.g., "an octopus" and "An Octopus." Copy an existing `test_` function in the `test.py` to verify that your program works correctly while still passing all the other tests. Try writing the test first, then make your program pass the test. That's *test-driven development*!
- Accept a new parameter that changes "larboard" (the left side of the boat) to "starboard" (the right side.^[1]). You could either make an option called `--side` that defaults to "larboard," or you could make a `--starboard` flag that, if present, changes the side to "starboard."
- The provided tests only give you words that start with an actual alphabetic character. Expand your code to handle words that start with numbers or punctuation. Should your program reject these? Add more tests to ensure that your program does what you intend.

[1] "Starboard" has nothing to do with stars but with the "steering board" or a rudder which typically would be on the right-side of the boat for right-handed sailors!

Chapter 3: Going on a picnic: Working with lists

Writing code makes me hungry! Let's write a program to consider some tasty foods we'd like to eat. So far we've just handled *one* of something like a name to say "hello" to or a nautical-themed object to point out. In this program, we want to eat one or more foods which we will store in a **list**, a variable that can hold any number of items. We use lists all the time in life. Maybe it's your top-five favorite songs, your birthday wish-list, or a bucket list of the best types of buckets.

In this exercise, we're going on a picnic, and we want to print a list of items to bring. You will learn to:

- Write a program that accepts multiple positional arguments
- Use **if**, **elif**, and **else** to handle conditional branching
- Find and alter items in a list
- Sort and reverse lists
- Format a list into a new string



The items will be passed as positional arguments. When there is only one item, you'll just print that:

```
$ ./picnic.py salad  
You are bringing salad.
```



What? Who just brings salad on a picnic? When there are two items, you'll put "and" in between them:

```
$ ./picnic.py salad chips  
You are bringing salad and chips.
```



Hmm, chips. That's an improvement. When there are three or more items, you will separate the items with commas:

```
$ ./picnic.py salad chips cupcakes  
You are bringing salad, chips, and cupcakes.
```

There's one other twist. You will also need to accept a `--sorted` argument that will require you to sort the items before you print them, but we'll deal with that in a bit. So, your Python program must:

- Store one or more positional arguments in a `list`
- Count the number of arguments
- Possibly modify the list like maybe to sort the items
- Use the list to print a new a string that formats the arguments according to how many items there are.

How should we begin?

Starting the program

I will always recommend you start programming either by running `new.py` or by copying `template/template.py` to the program name. This time the program should be called `picnic.py`, and we need to create it in the `picnic` directory:

```
$ cd picnic
$ new.py picnic.py
Done, see new script "picnic.py."
```

Now run `make test` or `pytest -xv test.py`. You should pass the first two tests (program exists, program creates usage), and fail the third:

```
test.py::test_exists PASSED [ 14%]
test.py::test_usage PASSED [ 28%]
test.py::test_one FAILED [ 42%]
```

The rest of the output is complaining about the fact that the test expected "You are bringing chips" but got something else:

```
===== FAILURES =====
----- test_one -----
def test_one():
    """one item"""

    out = getoutput(f'{prg} chips') ①
>     assert out.strip() == 'You are bringing chips.' ②
E     assert 'str_arg = "...nal = "chips"' == 'You are bringing chips.'
E         + You are bringing chips. ③
E             - str_arg = "" ④
E             - int_arg = "0"
E             - file_arg = ""
E             - flag_arg = "False"
E             - positional = "chips"

test.py:31: AssertionError
===== 1 failed, 2 passed in 0.56 seconds =====
```

① The program is being run with the argument "chips."

② This is the line that is causing the error. The output is tested to see if it is equal (`==`) to the string "You are bringing chips."

③ The line starting with a `+` sign is show what was expected.

④ The lines starting with the `-` sign is showing what was returned by the program.

Let's run the program with the argument "chips" and see what it gets:

```
$ ./picnic.py chips
str_arg = ""
int_arg = "0"
file_arg = ""
flag_arg = "False"
positional = "chips"
```

Right, that's not correct at all! Remember, the template doesn't have the *correct* arguments, just some examples, so the first thing we need to do is to fix the `get_args` function. Here is what your program should print a usage statement if given *no arguments*:

```
$ ./picnic.py
usage: picnic.py [-h] [-s] str [str ...]
picnic.py: error: the following arguments are required: str
```

And here is the usage for the `-h` or `--help` flags:

```
$ ./picnic.py -h
usage: picnic.py [-h] [-s] str [str ...]

Picnic game

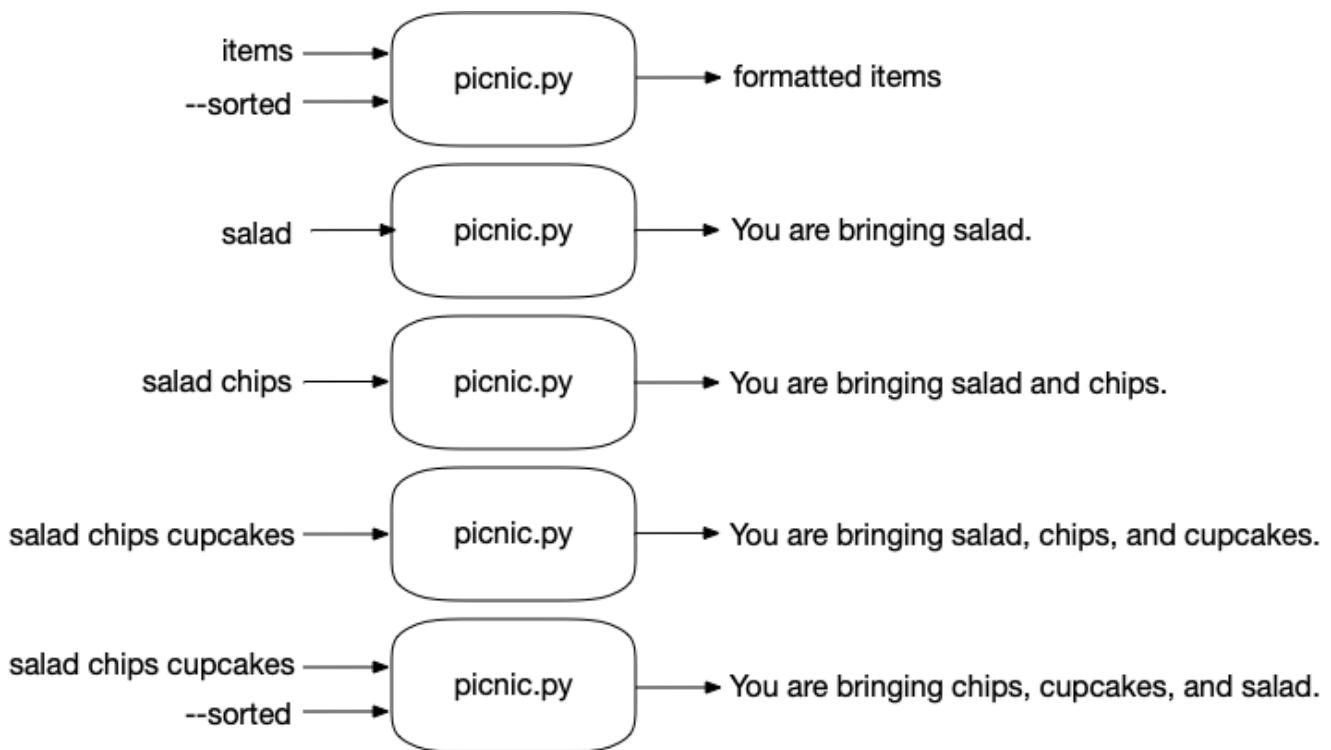
positional arguments:
  str           Item(s) to bring

optional arguments:
  -h, --help    show this help message and exit
  -s, --sorted  Sort the items (default: False)
```

We need a single positional argument and an optional flag called `--sorted`. Modify your `get_args` until it produces the above output. Note that there should be one or more of the `item` parameter, so you should define it with `nargs='+'`. Refer to the section "One or more of the same positional arguments" in chapter 1.

Writing `picnic.py`

Here is a tasty diagram of the inputs and outputs for the `picnic.py` program we'll write:



The program should accept one or more "positional" arguments as the items to bring on a picnic as well as a `-s` or `--sorted` flag to indicate whether or not to sort the items. The output will be "You are bringing" and then the list of items formatted according the following rules:

1. If one item, just state the item:

```
$ ./picnic.py chips  
You are bringing chips.
```

2. If two items, put "and" in between the items. Note that "potato chips" is just one string that happens to contain two words. If we leave out the quotes, then there would be three arguments to the program. Also, it doesn't matter here if we use single or double quotes:

```
$ ./picnic.py "potato chips" salad  
You are bringing potato chips and salad.
```

3. If three or more items, place a comma and space between each item and the word "and" before the final element. Don't forget the comma before the "and" (sometimes called the "Oxford comma") because your author was an English lit major and, while I may have finally stopped using two spaces after the end of a sentence, you can pry the Oxford comma from my cold, dead hands:

```
$ ./picnic.py "potato chips" salad soda cupcakes  
You are bringing potato chips, salad, soda, and cupcakes.
```

Be sure to sort if given the `-s` or `--sorted` flag:

```
$ ./picnic.py --sorted salad soda cupcakes  
You are bringing cupcakes, salad, and soda.
```

In order to figure out how many items we have, how to sort and slice them, and how to format the output string, we need to talk about the `list` type in Python in order to solve this problem.

Stuff about lists

We briefly touched on the idea of a `list` in the `hello.py` program when we looked at `sys.argv`, the "argument vector" for a program. If we run this:

```
$ ./picnic.py salad chips cupcakes
```

Then the arguments `salad chips cupcakes` would be available in `sys.argv` as the `list` of the strings

```
['salad', 'chips', 'cupcakes']
```

Note that they would be in the same order as they were provided on the command line. Lists always keep their order!

Let's go into the REPL and create a variable called `items` to hold some tasty foods we plan to bring on our picnic. I really want you to type these commands yourself, too, whether in the `python3` REPL or `jupyter` or a Jupyter Notebook. It's very important to interact in real time with the language!

To create a new, empty list, we can either use the `list()` function:

```
>>> items = list()
```

Or use empty square brackets:

```
>>> items = []
```

Check what Python says for the `type`. Yep, it's a `list`:

```
>>> type(items)
<class 'list'>
```

One of the first things we need to know is how many `items` we have for our picnic. Like a `str`, we can use `len` (length) to get the number of elements in `items`.

The length of an empty `list` is `0`:

```
>>> len(items)
0
```

Adding one element to a list

An empty list is not very useful. Let's see how we can add new items. We use `help(str)` in the last

chapter to read the documentation about the string *methods*, the functions that belong to every `str` in Python. Here I want you to use `help(list)` to find the `list` methods:

```
>>> help(list)
```

You'll see lots of "double-under" methods like `{dbl_}len{dbl_}`. Skip over those, and the first method we can find is `append`, which we can use to add items to the end of the list. If we evaluate our `items`, we see that the empty brackets tell us that it's empty:

```
>>> items  
[]
```

Let's add "sammiches" to the end:

```
>>> items.append('sammiches')
```

Nothing happened, so how do we know that worked? Let's check the length. It should be `1`:

```
>>> len(items)  
1
```

Hooray! That worked. In the spirit of testing, use the `assert` method to verify that the length is `1`. The fact that nothing happens is good. When an assertion fails, it triggers an exception that results in a lot of messages. Here, no news is good news:

```
>>> assert len(items) == 1
```

If you type `items<Enter>` in the REPL, Python will show you the contents:

```
>>> items  
['sammiches']
```

Cool, we added one element.

Adding many elements to a list

Let's try to add "chips" and "ice cream" to the `items`:

```
>>> items.append('chips', 'ice cream')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: append() takes exactly one argument (2 given)
```

Here is one of those pesky exceptions, and these will cause your programs to *crash*, something we want to avoid at all costs. We see that `append()` *takes exactly one argument*, and we gave it two. If you look at the `items`, you'll see that nothing was added:

```
>>> items
['sammiches']
```

OK, so maybe we were supposed to give it a `list` of items to add? Let's try that:

```
>>> items.append(['chips', 'ice cream'])
```

Well, that didn't cause an exception, so maybe it worked? We would expect there to be 3 `items`, so let's use an assertion to check that:

```
>>> assert len(items) == 3
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AssertionError
```

What is the length?

```
>>> len(items)
2
```

Only 2? Let's look at `items`:

```
>>> items
['sammiches', ['chips', 'ice cream']]
```

Check that out! Lists can hold any type of data like strings and numbers and even other lists. We asked `append` to add `['chips', 'ice cream']`, which is a `list`, and that's just what it did. Of course, it's not quite what we *wanted*.

Let's reset `items` so we can fix this. We can either use the `del` command to delete the element at index `1`:

```
>>> del items[1]
```

Or reassign it a new value:

```
>>> items = ['sammiches']
```

If you read further into the [help](#), you will find the [extend](#) method:

```
| extend(self, iterable, /)
|     Extend list by appending elements from the iterable.
```



Let's try that:

```
>>> items.extend('chips', 'ice cream')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: extend() takes exactly one argument (2 given)
```

Well that's frustrating! Now Python is telling us that [extend\(\)](#) [takes exactly one argument](#) which, if you refer to the [help](#), should be an [iterable](#). A [list](#) is something you can iterate (travel over from beginning to end), so that will work:

```
>>> items.extend(['chips', 'ice cream'])
```

Nothing happened. No exception, so maybe that worked? Let's check the length. It *should* be 3:

```
>>> assert len(items) == 3
```

Yes! Let's look at the [items](#) we've added:

```
>>> items
['sammiches', 'chips', 'ice cream']
```

Great! This is sounding like a pretty delicious outing.

If you know everything that will go into the [list](#), you can create it like so:

```
>>> items = ['sammiches', 'chips', 'ice cream']
```

The `append` and `extend` methods add new elements to the *end* of a given `list`. The `insert` method allows you to place new items at any position by specifying the index. I can use the index `0` to put a new element at the beginning of `items`:

```
>>> items.insert(0, 'soda')
>>> items
['soda', 'sammiches', 'chips', 'ice cream']
```

In addition to `help(list)`, you can also find lots of great documentation here:

<https://docs.python.org/3/tutorial/datastructures.html>

I recommend you read over all the `list` functions so you get an idea of just how powerful this data structure is!

Indexing lists

So now we have a list of `items`. We know how to use `len` to find how many `items` there are, and now we need to know how to get parts of the `list` to format. Indexing a `list` in Python looks exactly the same as indexing a `str`. (This actually makes me a bit uncomfortable, so I tend to imagine a `str` as a `list` of characters and then I feel somewhat better.)

0	1	2	3
['soda', 'sammiches', 'chips', 'ice cream']			
-4	-3	-2	-1

All indexing in Python is zero-offset, so the first element of `items` is at index `items[0]`:

```
>>> items[0]
'soda'
```

If the index is negative, Python starts counting backwards from the end of the list. The index `-1` is the last element of the list:

```
>>> items[-1]
'ice cream'
```

You should be very careful when using indexes to reference elements in a list. This is unsafe code:

```
>>> items[10]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
IndexError: list index out of range
```



Referencing an index that is not present will cause an exception.

We'll soon learn how to safely *iterate* or travel through a `list` so that we don't have to use indexes to get at elements.

Slicing lists

`[0:2]`



0 1 2 3

```
['soda', 'sammiches', 'chips', 'ice cream']
```

`[2:]`



You can extract "slices" (sub-lists)

of a `list` by using `list[start:stop]`. To get the first two items elements, you use `[0:2]`. Remember that the `2` is actually the index of the *third* element but *it's not inclusive*:

```
>>> items[0:2]
['soda', 'sammiches']
```

If you leave out `start`, it will be `0`, so this does the same thing:

```
>>> items[:2]
['soda', 'sammiches']
```

If you leave out `stop`, it will go to the end of the list:

```
>>> items[2:]
['chips', 'ice cream']
```

Oddly, it is completely *safe* for slices to use list indexes that don't exist. Here I can ask for all the elements from index `10` to the end even though there is nothing at index `10`. Instead of an exception, we get an empty list:

```
>>> items[10:]  
[]
```

For our exercise, you're going to need to get the word "and" into the list if there are three or more elements. Could you use a list index to do that?

Finding elements in a list

Did we remember to pack the chips?! Often we want to know if some items is in a `list`. The `index` method will return the location of an element in a `list`:

```
>>> items.index('chips')  
2
```



`list.index` is unsafe code because it will cause an exception if the argument is not present in the list!

See what happens if we check for the fog machine:

```
>>> items.index('fog machine')  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
ValueError: 'fog machine' is not in list
```

You should never use `index` unless you have first verified that an element is present. The `x in y` that we used in "Crow's Nest" to see if a letter was in the list of vowels can also be used for lists. We get back a `True` value if `x` is in the collection of `y`:

```
>>> 'chips' in items  
True
```



I hope they're salt and vinegar chips.

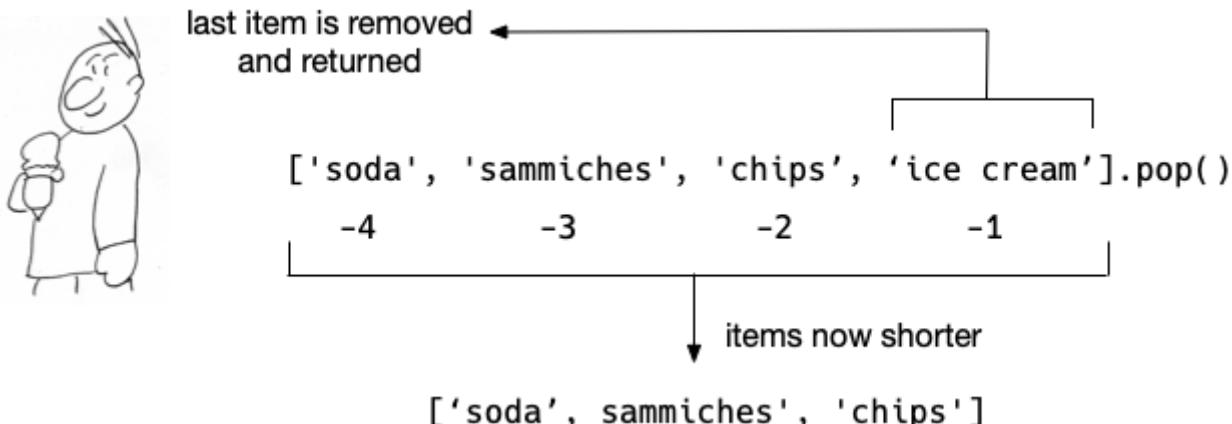
The same returns `False` if it is not present:

```
>>> 'fog machine' in items  
False
```

We're going to need to talk to the planning committee. What's a picnic without a fog machine?

Removing elements from a list

We've seen that we can use the `del` function to delete a `list` element by index. The `list.pop` method will remove *and return* the element at the index. By default it will remove the *last* item (`-1`):

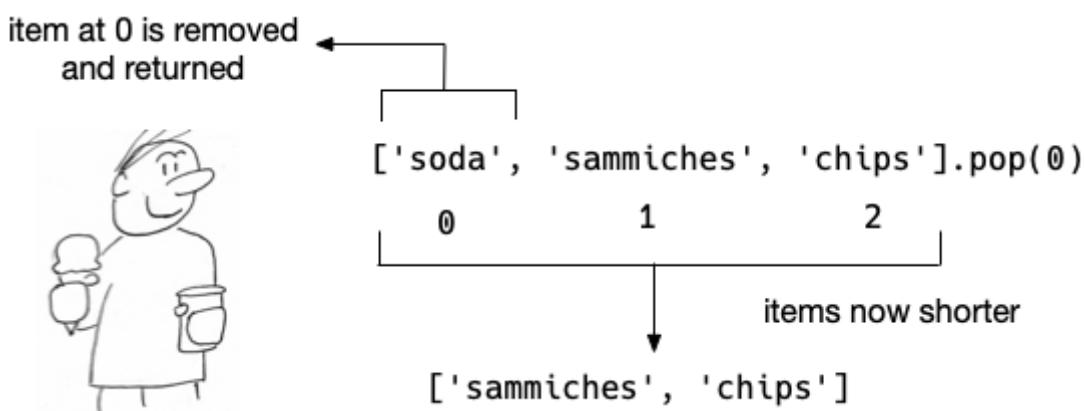


```
>>> items.pop()  
'ice cream'
```

If we look at the list, we will see it's shorter by one:

```
>>> items  
['soda', 'sammiches', 'chips']
```

You can use an item's index to remove an element at a particular location. For instance, we can use `0` to remove the first element:

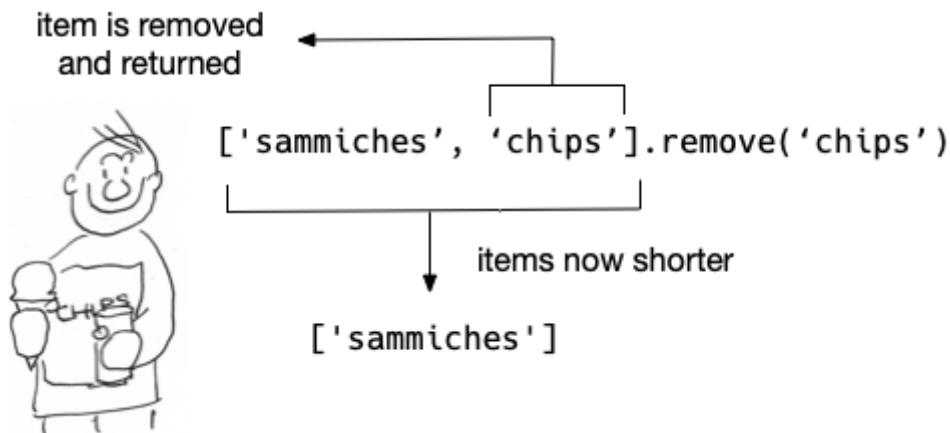


```
>>> items.pop(0)  
'soda'
```

And now our `list` is shorter still:

```
>>> items  
['sammiches', 'chips']
```

You can also use the `list.remove` method to remove the first occurrence of a given item:



```
>>> items.remove('chips')  
>>> items  
['sammiches']
```



The `list.remove` will cause an exception if the element is not present.



If we try to `remove` the chips again, we get an exception:

```
>>> items.remove('chips')  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
ValueError: list.remove(x): x not in list
```

So don't use this code unless you've verified that a given element is in the list:

```
item = 'chips'  
if item in items:  
    items.remove(item)
```

Sorting and reversing a list

If the `--sorted` flag is present, we're going to need to sort the `items`. You might notice in the help documentation that two methods, `list.reverse` and `list.sort` stress that they work *IN PLACE*. That means that the `list` itself will be either reversed or sorted and *nothing will be returned*. So, given this list:

```
>>> items = ['soda', 'sammiches', 'chips', 'ice cream']
```

`None ← items.sort()`



`items are sorted,
and nothing is returned` The `sort` method will return nothing:

```
>>> items.sort()
```

But if you inspect `items`, you will see they have been sorted alphabetically:

```
>>> items  
['chips', 'ice cream', 'sammiches', 'soda']
```

Note that Python will sort a `list` of numbers *numerically*, so we've got that going for us, which is nice:

```
>>> sorted([4, 2, 10, 3, 1])  
[1, 2, 3, 4, 10]
```

As with `list.sort`, we see nothing on the `list.reverse` call:

```
>>> items.reverse()
```

But the `items` are now in the opposite order:

```
>>> items  
['soda', 'sammiches', 'ice cream', 'chips']
```

The `list.sort` and `list.reverse` methods are easily confused with the the `sorted` and `reversed` functions. The `sorted` function accepts a `list` as an argument and *returns* a new `list`:

```
>>> items = ['soda', 'sammiches', 'chips', 'ice cream']
>>> sorted(items)
['chips', 'ice cream', 'sammiches', 'soda']
```

It's crucial to note that the `sorted` function *does not alter* the `list`:

```
>>> items
['soda', 'sammiches', 'chips', 'ice cream']
```

The `list.sort` method is a function that belongs to the `list`. It can take arguments that affect the way the sorting happens. Let's look at the `help(list.sort)`:

```
sort(self, /, *, key=None, reverse=False)
    Stable sort *IN PLACE*.
```

So we could also `sort` the `items` in `reverse` like so:

```
>>> items.sort(reverse=True)
```

And now they look like this:

```
>>> items
['soda', 'sammiches', 'ice cream', 'chips']
```



The `reversed` function works a bit differently:

```
>>> reversed(items)
<list_reverseiterator object at 0x10e012ef0>
```

I bet you expecting to see a new `list` with the `items` in reverse? This is an example of a *lazy* function in Python. The process of reversing a `list` might take a while, so Python is showing that it has

generated an "iterator object" that will provide the reversed list just as soon as we actually ask for the elements. We can do that in the REPL by using the `list` function to evaluate the iterator:

```
>>> list(reversed(items))
['ice cream', 'chips', 'sammiches', 'soda']
```

As with the `sorted` function, the original `items` itself remain unchanged:

```
>>> items
['soda', 'sammiches', 'chips', 'ice cream']
```

If you use the `list.sort` method instead of the `sorted` function, you might end up deleting your data. Imagine you wanted to set your `items` equal to the sorted list of `items` like so:

```
>>> items = items.sort()
```

What is in `items` now? If you print the `items` in the REPL, you won't see anything useful, so inspect the `type`:

```
>>> type(items)
<class 'NoneType'>
```

It's no longer a `list`. We just set it equal to the result of called `items.sort()` method that works on the `list` *in-place* and returns `None`. I would note that I tend to not use the `sort/reverse` methods because I don't generally like to mutate my data. I would tend to do something like this:

```
>>> items = ['soda', 'sammiches', 'chips', 'ice cream']
>>> sorted_items = sorted(items)
>>> sorted_items
['chips', 'ice cream', 'sammiches', 'soda']
```

Now I have explicitly named a `sorted_items` list, and the original `items` has not been altered.

If the `--sorted` flag is given to your program, you will need to sort your items in order to pass the test. Will you use `list.sort` or the `sorted` function?

Lists are mutable

As we've seen, we can change a `list` quite easily. The `list.sort` and `list.reverse` methods change the whole list, but you can also change any single element by referencing it by index. Maybe we make our picnic slightly healthier by changing out the chips for apples:

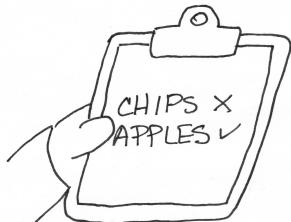
```
>>> items
['soda', 'sammiches', 'chips', 'ice cream']
>>> if 'chips' in items:           ①
...     idx = items.index('chips')  ②
...     items[idx] = 'apples'       ③
...

```

- ① See if the string '**chips**' is in the list of **items**.
- ② Assign the index of '**chips**' to the variable **idx**.
- ③ Use the index **idx** to change the element to '**apples**'.

Let's look at **items** to verify:

```
>>> items
['soda', 'sammiches', 'apples', 'ice cream']
```



We can also write a couple of tests:

```
>>> assert 'chips' not in items ①
>>> assert 'apples' in items   ②
```

- ① Make sure "chips" are no longer on the menu.
- ② Check that we now have some "apples."

Joining a list



In our exercise, you'll need to print a string based on the number of elements in the given list. The string will intersperse some other string like ', ' in between all the elements of the list. Oddly, this is the syntax to join a list on the string made of the comma and a space:

```
>>> ', '.join(items)
'soda, sammiches, chips, ice cream'
```

Here we use the **str.join** method and pass the **list** as an argument. It always feels backwards to me, but that's the way it goes. The result of **str.join** is a *new string*:

```
>>> type(', '.join(items))  
<class 'str'>
```

The original `list` remains unchanged:

```
>>> items  
['soda', 'sammiches', 'chips', 'apples']
```

There is quite a bit more that we can do with Python's `list`, but that should be enough for you to solve this problem.

Conditional branching with if/elif/else

You need to use the conditional branching based on the number of items to correctly format the output. In the Crow's Nest exercise, there were two conditions (a "binary" choice)—either a vowel or not—so we used `if/else` statements. Here we have three options to consider, so you will have to use `elif` (else-if). For instance, we want to classify someone by their age by three options:

1. If their age is greater than `0`, it is valid.
2. If their age is less than `18`, they are a minor.
3. Otherwise they are 18 year or older, which means they can vote:

Here is how I could write that code:

```
>>> age = 15
>>> if age < 0:
...     print('You are impossible.')
... elif age < 18:
...     print('You are a minor.')
... else:
...     print('You can vote.')
...
You are a minor.
```

See if you can use that to figure out how to write the three options for `picnic.py`. That is, first write the branch that handles one item. Then write the branch that handles two items. Then write the last branch for three or more items. Run the tests *after every change to your program*.

Now go write the program yourself before you continue to look at my solution.

Hints:

- Go into your `picnic` directory and run `new.py picnic.py` to start your program. Then run `make test` (or `pytest -xv test.py`) and you should pass the first two tests.
- Next work on getting your `--help` usage looking like the above. It's very important to define your arguments correctly. For the `items` argument, look at `nargs` in `argparse` as discussed in chapter 1's "One or more of the same positional arguments" section.
- If you use `new.py` to start your program, be sure to leave the "boolean flag" and modify it for your `sorted` flag.
- Solve the tests in order! First handle just one item, then handle two items, then handle three. Then handle the sorted items.

You'll get the best benefit from this book if you try writing the program and passing the tests before turning the page to read the solution!

Solution

```
1 #!/usr/bin/env python3
2 """Picnic game"""
3
4 import argparse
5
6
7 # -----
8 def get_args(): ①
9     """Get command-line arguments"""
10
11     parser = argparse.ArgumentParser(
12         description='Picnic game',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('item', ②
16                         metavar='str',
17                         nargs='+',
18                         help='Item(s) to bring')
19
20     parser.add_argument('-s', ③
21                         '--sorted',
22                         action='store_true',
23                         help='Sort the items')
24
25     return parser.parse_args() ④
26
27
28 # -----
29 def main(): ⑤
30     """Make a jazz noise here"""
31
32     args = get_args() ⑥
33     items = args.item ⑦
34     num = len(items) ⑧
35
36     if args.sorted: ⑨
37         items.sort() ⑩
38
39     bringing = '' ⑪
40     if num == 1: ⑫
41         bringing = items[0] ⑬
42     elif num == 2:
43         bringing = ' and '.join(items) ⑮
44     else: ⑯
45         items[-1] = ' and ' + items[-1] ⑰
46         bringing = ', '.join(items) ⑱
47
```

```
48     print('You are bringing {}'.format(bringing)) ⑯
49
50
51 # -----
52 if __name__ == '__main__': ⑰
53     main()
```

- ① The `get_args` function is placed first so that we can easily see what the program accepts when we read it. Note that the function order here is not important to Python, only to us, the reader.
- ② The `item` argument uses the `nargs='+'` so that it will accept *one or more* positional arguments which will be strings.
- ③ The dashes in the short (`-s`) and long (`--sort`) names make this an *option*. There is no value associated with this argument. It's either present (in which case it will be `True`) or absent (or `False`).
- ④ Process the command-line arguments and `return` them to the caller.
- ⑤ The `main` function is where the program will start.
- ⑥ Call the `get_args` function and put the returned value into the variable `args`. If there is a problem partsing the arguments, the program will fail before the values are returned.
- ⑦ Copy the `item` list from the `args` into the new variable `items`.
- ⑧ Use the length function `len` to get the number of `items` in the list. There can never be zero arguments because we defined the argument using `nargs='+'` which always requires at least one value.
- ⑨ The `args.sorted` value with either be `True` or `False` because it was defined as a "flag."
- ⑩ If we are supposed to sort the items, call the `items.sort()` method to sort them *in-place*.
- ⑪ Initialize the variable `bringing` with an empty string. We'll put the items we're bringing into this.
- ⑫ If the number of items is 1...
- ⑬ Then we will assign the one item to `bringing`.
- ⑭ If the number of items is 2...
- ⑮ Put the string ' `and` ' in between the `items`.
- ⑯ Otherwise...
- ⑰ Alter the last element in `items` to append the string ' `and` ' before whatever is already there.
- ⑱ Join the `items` on the string '`,` '.
- ⑲ Print the output string, using the `str.format` method to interpolate the `bringing` variable.
- ⑳ When Python runs the program, it will read all the lines to this point but will not run anything. Here we look to see if we are in the "main" namespace.

If we are, call the `main` function to make the program begin.

Discussion

How did it go? Did it take you long to write your version? How different was it from mine? Let's talk about my solution. It's perfectly fine if yours is really different from mine, just as long as you pass the tests!

Defining the arguments

This program can accept a variable number of arguments which are all the same thing (strings). In my `get_args`, I define an `item` like so:

```
parser.add_argument('item',  
                    metavar='str',  
                    nargs='+',  
                    help='Item(s) to bring') ④
```

- ① A single, required, positional (because no dashes in name) argument called `item`.
- ② An indicator to the user in the usage that this should be a string.
- ③ The number of arguments where '+' means *one or more*.
- ④ A longer help description that appears for the `-h` or `--help` options.

This program also accepts `-s` and `--sorted` arguments. Remember that the leading dashes makes them optional. They are "flags," which typically means that they are `True` if they are present and `False` if absent.

```
parser.add_argument('-s',  
                    '--sorted',  
                    action='store_true',  
                    help='Sort the items') ④
```

- ① The short flag name.
- ② The long flag name.
- ③ If the flag is present, store a `True` value. The default value will be `False`.
- ④ The longer help description.

Assigning and sorting the items

To get the arguments, in `main` I call `get_args` and assign them to the `args` variable. Then I create the `items` variable to hold the `args.item` value(s):

```
args = get_args()  
items = args.item
```

If `args.sorted` is `True`, then I need to sort my `items`. I chose the *in-place sort* method here:

```
if args.sorted:  
    items.sort()
```

Now I have the items, sorted if needed, and I need to format them for the output.

Formatting the items

I suggested you solve the tests in order. There are 4 conditions we need to solve:

1. Zero items
2. One item
3. Two items
4. Three or more items

The first test is actually handled by `argparse` — if the user fails to provide any arguments, they get a usage:

```
$ ./picnic.py  
usage: picnic.py [-h] [-s] str [str ...]  
picnic.py: error: the following arguments are required: str
```

Since `argparse` handles the case of no arguments, we have to handle the other three conditions. Here's one way to do that:

```
bringing = ''  
①  
if num == 1:  
    ②  
        bringing = items[0]  
    ③  
elif num == 2:  
    ④  
        bringing = ' and '.join(items)  
    ⑤  
else:  
    ⑥  
        items[-1] = 'and ' + items[-1] ⑦  
        bringing = ', '.join(items) ⑧
```

- ① Initialize a variable for what we are `bringing`.
- ② Check if the number of items is one.
- ③ If there is one item, then `bringing` is the one item.
- ④ Check if the number of items is two.
- ⑤ If two items, join the `items` on the string '`and`'.
- ⑥ Otherwise...
- ⑦ Insert the string '`and`' before the last item

⑧ Join all the `items` on the string `', '`.

Printing the items

Finally to `print` the output, I can use a format string where the `{}` indicates a placeholder for some value like so:

```
>>> print('You are bringing {}'.format(bringing))
You are bringing salad, soda, and cupcakes.
```

Or, if you prefer, you can use an `f'''`-string:

```
>>> print(f'You are bringing {bringing}.')
You are bringing salad, soda, and cupcakes.
```

They both get the job done, so whichever you prefer.

Testing

For this exercise, I've written the tests for you. Can you see how this is similar to the `hello.py` where we created a `greet` function inside the program and added a `test_greet` function to test it? If I were writing this code for myself, I would do the same by creating a function to format the items and placing the unit test for that inside the program itself.

I believe there is an art to testing. It's up to you to figure out how best to test your own code. Try copying some of the relevant `test_` functions from `test.py` into your `picnic.py` and then running `pytest picnic.py` to see how it works. What do you prefer?

Summary

- Python lists are ordered sequences of other Python data types such as strings and numbers.
- There are methods like `append` and `extend` to add elements to a `list` and `pop` and `remove` to remove them.
- You can use `x in y` to ask if element `x` is in the list `y`. You could also use `list.index` to find the index of an element, but this will cause an exception if the element is not present.
- Lists can be sorted and reversed, and elements within lists can be modified. Lists are useful when the order of the elements is important.
- Strings and lists share many features such as using `len` to find their lengths, using zero-based indexing where `0` is the first element and `-1` is the last, and using slices to extract smaller pieces from the whole.
- The `str.join` method can be used to make a new `str` from a `list`.
- `if/elif/else` can be used to branch code depending on conditions.

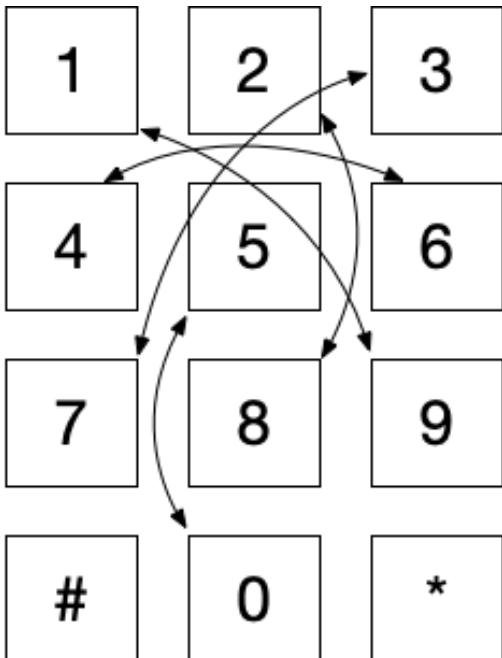
Going Further



- Add an option so the user can choose not to print with the Oxford comma.
- Add an option to separate items with some character passed in by the user (like a semicolon if the list of items needed to contain commas).

Chapter 4: Jump the Five: Working with dictionaries

"When I get up, nothing gets me down." - D. L. Roth



In an episode of the television show *The Wire*, drug dealers encode telephone numbers that they text in order to obscure them from the police who they assume are intercepting their messages. They use an algorithm we'll call "Jump The Five" where a number is changed to the one that is opposite on a US telephone pad if you jump over the 5. You feel me?

If we start with "1" and jump across the 5, we get to "9," then "6" jumps the 5 to become "4," and so forth. The numbers "5" and "0" will swap with each other. In this exercise, we're going to write a Python program called `jump.py` that will take in some text as a positional argument. Each number in the text will be encoded using this algorithm. All non-number will pass through unchanged, for example:

```
$ ./jump.py 867-5309  
243-0751  
$ ./jump.py 'Call 1-800-329-8044 today!'  
Call 9-255-781-2566 today!
```

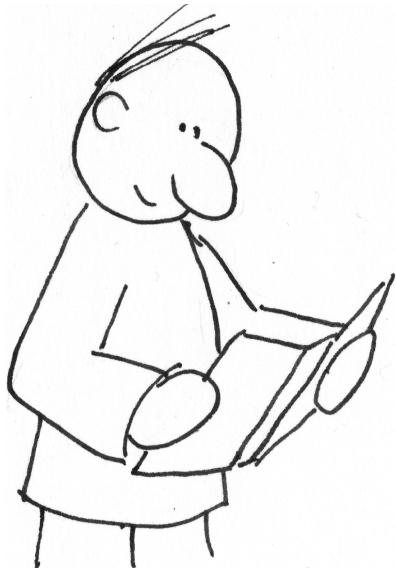
We will need some way to inspect each character in the input text and identify the numbers. We will learn how to use a `for` loop for this and how that relates to a "list comprehension." Then we will need some way to associate a number like "1" with the number "9," and so on for all the numbers. We'll learn about a data structure in Python called a "dictionary" type that allows us to do exactly that.

In this chapter, you will learn to:

- Create a dictionary.
- Use a `for` loop to process text character-by-character.
- Check if items exist in a dictionary.
- Retrieve values from a dictionary.
- Print a new string with the numbers substituted for their encoded values.

Before we get started with the coding, we need to learn about Python's dictionaries.

Dictionaries



A Python `dict` allows us to relate some *thing* (a "key") to some other *thing* (a "value"). An actual dictionary does this. If we look up a word like "quirky" in a dictionary (<https://www.merriam-webster.com/dictionary/quirky>), we can find a definition. We can think of the word itself as the "key" and the definition as the "value."

quirky \Leftrightarrow unusual, esp. in an interesting or appealing way

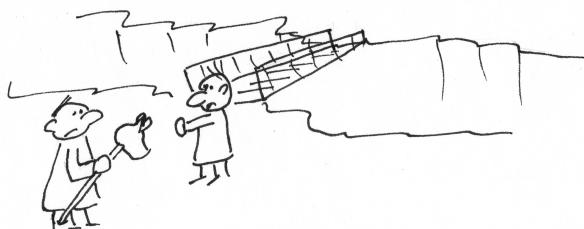
Dictionaries actually provide quite a bit more information such as pronunciation, part of speech, derived words, history, synonyms, alternate spellings, etymology, first known use, etc. (I really love dictionaries.) Each of those attributes has a value, so we could also think of the lookup itself as a dictionary:

quirky

definition	\Leftrightarrow unusual, esp. in an interesting or appealing way
pronunciation	\Leftrightarrow 'kwər-kē
part of speech	\Leftrightarrow adjective

Let's see how we can use Python's dictionaries to go beyond word definitions.

Creating a dictionary



In the film *Monty Python and the Holy Grail*, King Arthur and his knights must cross The Bridge of Death. Anyone who wishes to cross must correctly answer three questions from the Keeper. Those who fail are cast into the Gorge of Eternal Peril.

Let us ride to CAMEL...no, sorry, let us create and use a `dict` to keep track of the questions and answers as key/value pairs. Once again, I want you to fire up your `python3/ipython` REPL or Jupyter Notebook and type these out for yourself!

Lancelot goes first. We can use the `dict()` function to create an empty dictionary for his answers.

```
>>> answers = dict()
```

Or we can use empty curly brackets:

```
>>> answers = {}
```



The Keeper's first question: "What is your name?" Lancelot answers "My name is Sir Lancelot of Camelot." We can add the key "name" to our `answers` by using square brackets `[]`—not curlies! and the literal string '`'name'`:

```
>>> answers['name'] = 'Sir Lancelot'
```

`{'name': 'Sir Lancelot'}`

curly brackets
mean "dictionary"

If you type `answers<Enter>` in the REPL, Python will show you a structure in curlies to indicate this is a `dict`:

```
>>> answers
{'name': 'Sir Lancelot'}
```

You can verify with the `type` function:

```
>>> type(answers)
<class 'dict'>
```

Next the Keeper asks, "What is your quest?" to which Lancelot answers "To seek the Holy Grail." Let's add "quest" to the `answers`:

```
>>> answers['quest'] = 'To seek the Holy Grail'
```



There's no return value to let us know something happened, so type `answers` to inspect the variable again to ensure our new key/value was added:

```
>>> answers
{'name': 'Sir Lancelot', 'quest': 'To seek the Holy Grail'}
```

Finally the Keeper asks "What is your favorite color?," and Lancelot answers "blue."

```
>>> answers['favorite_color'] = 'blue'
>>> answers
{'name': 'Sir Lancelot', 'quest': 'To seek the Holy Grail', 'favorite_color': 'blue'}
```

If you knew all the answers beforehand, you could create `answers` using the `dict()` function with this syntax where you do *not* have to quote the keys and the keys are separate from the values with equal signs:

```
>>> answers = dict(name='Sir Lancelot', quest='To seek the Holy Grail',
favorite_color='blue')
```

Or this syntax using curly braces `{}` where the keys must be quoted and are followed by a colon (`:`):

```
>>> answers = {'name': 'Sir Lancelot', 'quest': 'To seek the Holy Grail',
'favorite_color': 'blue'}
```

answers

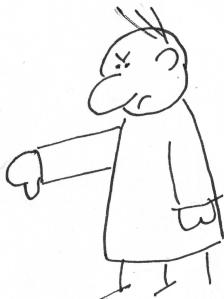
name	⇒ Sir Lancelot
quest	⇒ To seek the Holy Grail
favorite color	⇒ blue

It might be helpful to think of the dictionary `answers` as a box that inside holds the key/value pairs that describe Lancelot's answers just the way the "quirky" dictionary holds all the information about that word.

Accessing dictionary values

To retrieve the values, you use the key name inside square brackets (`[]`). For instance, I can get the `name` like so:

```
>>> answers['name']
'Sir Lancelot'
```



Let's request his "age":

```
>>> answers['age']
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
KeyError: 'age'
```



You will cause an exception if you ask for a dictionary key that doesn't exist!

Just as with lists, you can use the `x in y` to first see if a key exists in the `dict`:

```
>>> 'quest' in answers
True
>>> 'age' in answers
False
```



The `dict.get` method is a *safe* way to ask for a value:

```
>>> answers.get('quest')  
'To seek the Holy Grail'
```

When the requested key does not exist in the `dict`, it will return the special value `None`:

```
>>> answers.get('age')
```

That doesn't print anything because the REPL won't print a `None`, but we can check the `type`:

```
>>> type(answers.get('age'))  
<class 'NoneType'>
```

There is an optional second argument you can pass to `dict.get` which is the value to return *if the key does not exist*:



```
>>> answers.get('age', 'NA')  
'NA'
```

Other dictionary methods

If you want to know how "big" a dictionary is, the `len` (length) function on a `dict` will tell you how many key/value pairs are present:

```
>>> len(answers)
3
```

The `dict.keys` method will give you just the keys:

```
>>> answers.keys()
dict_keys(['name', 'quest', 'favorite_color'])
```

And `dict.values` will give you the values:

```
>>> answers.values()
dict_values(['Sir Lancelot', 'To seek the Holy Grail', 'blue'])
```

Often we want both together, so you might see code like this:

```
>>> for key in answers.keys():
...     print(key, answers[key])
...
name Sir Lancelot
quest To seek the Holy Grail
favorite_color blue
```

An easier way to write this would be to use the `dict.items` method which will return a `list` of the key/value pairs:

```
>>> answers.items()
dict_items([('name', 'Sir Lancelot'), ('quest', 'To seek the Holy Grail'),
('favorite_color', 'blue')])
```

The above `for` loop could also be written using the `dict.items` method:

```
>>> for key, value in answers.items(): ①
...     print(f'{key:15} {value}')      ②
...
name           Sir Lancelot
quest          To seek the Holy Grail
favorite_color blue
```

- ① For each key/value pair, unpack them into the variables `key` and `value`. Note that you don't have to call them `key` and `value`. You could use `k` and `v` or `question` and `answer`.
- ② Print the `key` in a left-justified field 15 characters wide. The `value` is printed normally.

In the REPL you can execute `help(dict)` to see all the methods available to you like `pop` to remove a

key/value or `update` to merge with another `dict`.

Each key in the `dict` is unique. That means if you set a value for a given key twice:

```
>>> answers = {}
>>> answers['favorite_color'] = 'blue'
>>> answers
{'favorite_color': 'blue'}
```

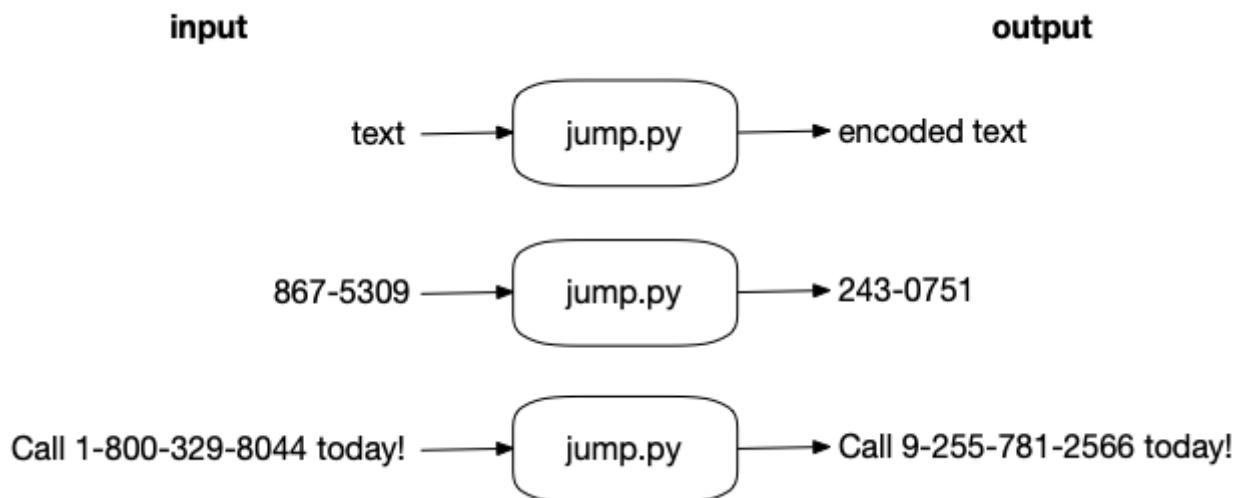
You will not have two entries but one entry with the *second* value:

```
>>> answers['favorite_color'] = 'red'
>>> answers
{'favorite_color': 'red'}
```

Keys don't have to be strings—you can also use numbers like `int` and `float`. Whatever value you use must be immutable. For instance, a `list` could not be a key because it is mutable.

Writing jump.py

Now let's get started with writing our program. Here is a diagram of the inputs and outputs. Note that your program will only affect the numbers in the text. Anything that is *not* a number is unchanged:



When run with no arguments, `-h`, or `--help`, your program should print a usage:

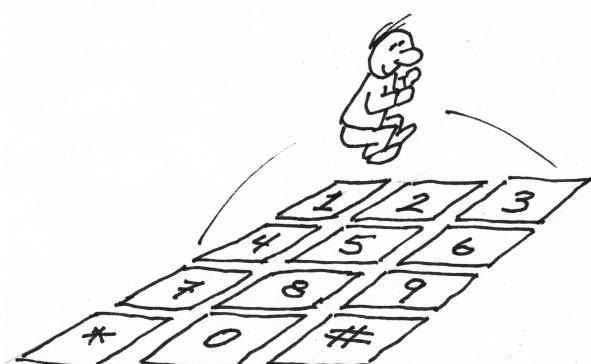
```
$ ./jump.py -h
usage: jump.py [-h] str

Jump the Five

positional arguments:
  str      Input text

optional arguments:
  -h, --help show this help message and exit
```

Here is the substitution table for the numbers:



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

How would you represent this using a `dict`? Try creating a `dict` called `jump` in the REPL and then using a test. Remember that `assert` will return nothing if the statement is `True`:

```
>>> assert jumper['1'] == '9'
>>> assert jumper['5'] == '0'
```

Next, you will need a way to visit each character in the given text. I suggest you use a `for` loop like so:

```
>>> for char in 'ABC123':
...     print(char)
...
A
B
C
1
2
3
```

Now, rather printing the `char`, print the value of `char` in the `jumper` table or print the `char` itself. Look at the `dict.get` method! Also, if you read `help(print)`, you'll see there is an `end` option to change the newline that gets stuck onto the end to something else.

Hints:

- The numbers can occur anywhere in the text, so I recommend you process the input character-by-character with a `for` loop.
- Given any one character, how can you look it up in your table?
- If the character is in your table, how can you get the value (the translation)?
- If how can you `print` the translation or the value without printing a newline? Look at `help(print)` in the REPL to read about the options to `print`.
- If you read `help(str)` on Python's `str` class, you'll see that there is a `replace` method. Could you use that?

Now spend the time to write the program on your own before you look at the solutions! Use the tests to guide you.

Solution

```
1 #!/usr/bin/env python3
2 """Jump the Five"""
3
4 import argparse
5
6
7 # -----
8 def get_args(): ①
9     """Get command-line arguments"""
10
11     parser = argparse.ArgumentParser(
12         description='Jump the Five',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('text', metavar='str', help='Input text') ②
16
17     return parser.parse_args()
18
19
20 # -----
21 def main(): ③
22     """Make a jazz noise here"""
23
24     args = get_args() ④
25     jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0', ⑤
26                 '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
27
28     for char in args.text: ⑥
29         print(jumper.get(char, char), end='') ⑦
30     print() ⑧
31
32
33 # -----
34 if __name__ == '__main__':
35     main() ⑨
```

- ① Define the `get_args` function first so it's easy to see when we read the program.
- ② We define one positional argument called `text`.
- ③ Define a `main` function where the program starts.
- ④ Get the command-line `args`.
- ⑤ Create a `dict` for the lookup table.
- ⑥ Process each character in the text.
- ⑦ Print either the value of the `char` in the `jumper` table or the `char` if it's not present, making sure *not* to print a newline by adding `end=''`.

- ⑧ Print a newline.
- ⑨ Call the `main()` function if the program is in the "main" namespace.

Discussion

Defining the arguments

If you look at the solution, you'll see that the `get_args` function is defined first. Our program needs to define one positional argument. Since it's some "text" I expect, I call the argument '`text`' and then assign that to a variable called `text`.

```
parser.add_argument('text', metavar='str', help='Input text')
```

While all that seems rather obvious, I think it's very important to name things *what they are*. That is, please don't leave the name of the argument as '`positional`' — that does not describe what it *is*. It may seem like overkill to use `argparse` for such a simple program, but it handles the validation of the correct *number* and *type* of arguments as well as the generation of help documentation, so it's well worth the effort.

Using a `dict` for encoding

I suggested you could represent the substitution table as a `dict` where each number `key` has its substitute as the `value` in the `dict`. For instance, I know that if I jump from `1` over the `5` I should land on `9`:

```
>>> jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
...           '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
>>> jumper['1']
'9'
```

Since there are only 10 numbers to encode, this is probably the easiest way to write this. Note that the numbers are written with quotes around them, so they are actually of the type `str` and not `int` (integers). I do this because I will be reading characters from a `str`. If we stored them as actual numbers, I would have to coerce the `str` types using the `int` function:

```
>>> type('4')
<class 'str'>
>>> type(4)
<class 'int'>
>>> type(int('4'))
<class 'int'>
```

Method 1: Using a `for` loop to print each character

As suggested in the introduction, I can process each character of the `text` using a `for` loop. To start, I might first just see if each character of the text is in the `jumper` table using the `x in y` construct.

```
>>> text = 'ABC123'
>>> for char in text:
...     print(char, jumper)
...
A False
B False
C False
1 True
2 True
3 True
```



When `print` is given more than one argument, it will put a space in between each of bit of text. You can change that with the `sep` argument. Read `help(print)` to learn more.

Now let's try to translate the numbers. I could use an `if` expression where I print the value from the `jumper` table if `char` is present, otherwise print the `char`:

```
>>> for char in text:
...     print(char, jumper[char] if char in jumper else char)
...
A A
B B
C C
1 9
2 8
3 7
```

It's a bit laborious to check for every character. The `dict.get` method allows us to safely ask for a value if it is present. For instance, the letter "A" is not in `jumper`. If we try to retrieve that value, we'll get an exception:

```
>>> jumper['A']
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
KeyError: 'A'
```

But if we use `jumper.get`, there is no exception:

```
>>> jumper.get('A')
```

When a key doesn't exist in the dictionary, the special `None` value is returned:

```
>>> for char in text:  
...     print(char, jumper.get(char))  
...  
A None  
B None  
C None  
1 9  
2 8  
3 7
```

We can provide a second, optional argument to `get` that is the default value to return when the key does not exist. In our case, if a character does not exist in the `jumper`, we want to print the character itself. If we had "A," then we'd want to print "A":

```
>>> jumper.get('A', 'A')  
'A'
```

But if we have, "5" then we want to print "0":

```
>>> jumper.get('5', '5')  
'0'
```

So we can use that to process all the characters:

```
>>> for char in text:  
...     print(jumper.get(char, char))  
...  
A  
B  
C  
9  
8  
7
```

I don't want that newline printing after every character, so I can use `end=''` to tell Python to put the empty string at the `end` instead of a newline. When I run this in the REPL, the output is going to look funny because I have to hit <Enter> after the `for` loop for it to run, then I'll be left with `ABC987` with no newline and then the `>>>` prompt:

```
>>> for char in text:  
...     print(jumper.get(char, char), end='')  
...  
ABC987>>>
```

And so in your code you have to add another `print()`. Mostly I wanted to point out a couple things you maybe didn't know about `print`. One is that you can have it put whatever you like for the `end`, and that you can `print()` with no arguments to print a newline. There are several other really cool things `print` can do, so I'd encourage you to read `help(print)` and try them out!

Method 2: Using a `for` loop to build a new string

There are several other ways we could solve this. I don't like all the `print` statements in the first solution, so here's another take:

```
def main():
    args = get_args()
    jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
              '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
    new_text = ''                      ①
    for char in args.text:             ②
        new_text += jumper.get(char, char) ③
    print(new_text)                   ④
```

- ① In this alternate solution, you create an empty `new_text` variable.
- ② Same `for` loop...
- ③ Append either the encoded number or the original `char` to the `new_text`
- ④ Print the `new_text`.

While it's fun to explore all the things we can do with `print`, that code is a bit ugly. I think it's cleaner to create a `new_text` variable and call `print` just once with that. To do this, we start off by setting a `new_text` equal to the empty string:

```
>>> new_text = ''
```

And we use our same `for` loop to process each character in the `text`. Each time through the loop, we use `+=` to append the right-hand side of the equation to the left-hand side. The `+=` adds the value on the right to the variable on the left:

```
>>> new_text += 'a'
>>> assert new_text == 'a'
>>> new_text += 'b'
>>> assert new_text == 'ab'
```

On the right, we're using the `jumper.get` method. Each character will be appended to the `new_text`:

```
>>> new_text = ''  
>>> for char in text:  
...     new_text += jumper.get(char, char)  
...
```

Now we can call `print` one time with our new value:

```
>>> print(new_text)  
ABC987
```

Method 3: Using a `for` loop to build a new `list`

This method is the same as above, but, rather than `new_text` being a `str`, it's a `list`:

```
def main():  
    args = get_args()  
    jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',  
              '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}  
    new_text = []  
    for char in args.text:  
        new_text.append(jumper.get(char, char))  
    print(''.join(new_text))
```

- ① Initialize `new_text` as an empty `list`.
- ② Iterate through each character of the `text`.
- ③ Append the results of the `jumper.get` call to the `new_text` variable.
- ④ Join the `new_text` on the empty string to create a new `str` to print.

As we go through the book, I'll keep reminding you how Python treats strings and lists similarly. It's easy to go back and forth between those two types. Here I'm using `new_text` exactly the same as above, starting off with it empty and then making it longer for each character. We could actually use the exact same `+=` syntax instead of the `list.append` method:

```
for char in args.text:  
    new_text += jumper.get(char, char)
```

After the `for` loop is done, we have all the new characters that need to be put back together into a new string to print.

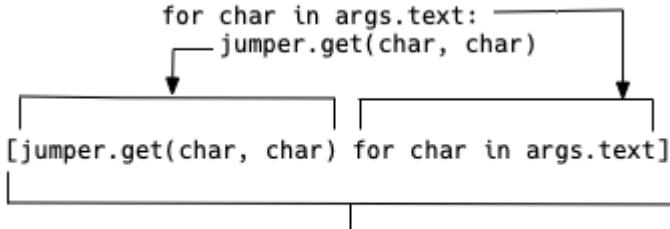
Method 4: List comprehension

A shorter solution uses a "list comprehension" which is basically a one-line `for` loop inside square brackets (`[]`) which results in a new `list`:

```

def main():
    args = get_args()
    jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
              '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
    print(''.join([jumper.get(char, char) for char in args.text]))

```



make a new list with the result A list comprehension is read backwards from a `for` loop, but it's all there. It's just one line of code instead four!

```

>>> text = '867-5309'
>>> [jumper.get(char, char) for char in text]
['2', '4', '3', '-', '0', '7', '5', '1']

```

You can `join` that new `list` on the empty string to create a new string you can `print`:

```

>>> print(''.join([jumper.get(char, char) for char in text]))
243-0751

```

Method 5: str.translate

This last method uses a really powerful method from the `str` class to change all the characters in one step:

```

def main():
    args = get_args()
    jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
              '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
    print(args.text.translate(str.maketrans(jumper)))

```

The argument to `str.translate` is a translation table that defines how each character should be translated. That's exactly what our `jumper` does!

```

>>> text = 'Jenny = 867-5309'
>>> text.translate(str.maketrans(jumper))
'Jenny = 243-0751'

```

We'll revisit this function in the "Apples and Bananas" exercise where I'll explain it in greater detail.

(Not) using `str.replace`

Note that you could *not* use `str.replace` to change each number 0-9. Watch how we start off with this string:

```
>>> text = '1234567890'
```

When you change "1" to "9," now you have two 9s:

```
>>> text = text.replace('1', '9')
>>> text
'9234567890'
```

Which means when you try to change all the 9s to 1s, you end up with two 1s:

```
>>> text = text.replace('9', '1')
>>> text
'1234567810'
```

You might try to write it like so:

```
>>> text = '1234567890'
>>> for n in jumper.keys():
...     text = text.replace(n, jumper[n])
...
>>> text
'1234543215'
```

But the correctly encoded string is "9876043215", which is exactly why `str.translate` exists!

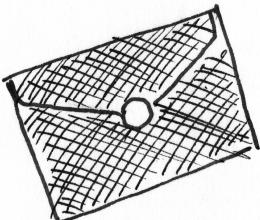
Summary

- You create a new dictionary using the `dict()` function or with empty curly brackets (`{}`).
- Dictionary values are retrieved using their keys inside square brackets or by using the `dict.get` method.
- For a `dict` called `x`, you can use `'key' in x` to determine if a key exists.
- You can use a `for` loop to iterate the characters of a `str` just like you can iterate through the elements of a `list`. You can think of strings as lists of characters.
- The `print` function takes optional keyword arguments like `end=''` which we can use to print a value to the screen without a newline.

Going Further

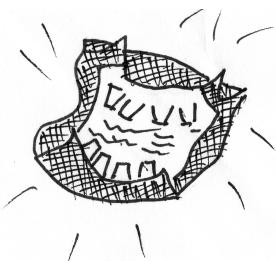
- Try creating a similar program that encodes the numbers with strings, e.g., "5" becomes "five", "7" becomes "seven."
- What happens if you feed the output of the program back into itself. For example, if you run `./jump.py 12345`, you should get `98760`. If you run `./jump.py 98760`, do you recover the original numbers? This is called "round-tripping," and it's a common operation with algorithms that encode and decode text.

Chapter 5: Howler: Working with files and STDOUT



In Harry Potter, a "Howler" is a nasty-gram that arrives by owl at Hogwarts. It will tear itself open, shout a blistering message at the recipient, and then combust. In this exercise, we're going to write a program that will transform text into a rather mild-mannered version of a Howler by MAKING ALL THE LETTERS UPPERCASE. The text that we'll process will be given as a single, positional argument.

For instance, if our program is given the input, "How dare you steal that car!", it should scream those back. Remember spaces on the command line delimit arguments, so multiple words need to be enclosed in quotes to be considered one argument:

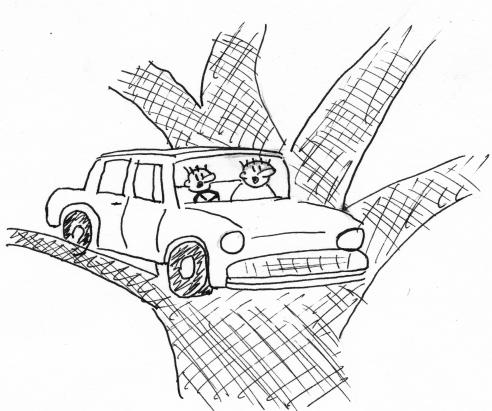


```
$ ./howler.py 'How dare you steal that car!'
HOW DARE YOU STEAL THAT CAR!
```

The argument to the program may also name a file, in which case we need to read the file for the input:

```
$ ./howler.py ../inputs/fox.txt
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```

Our program will also accept an `-o` or `--outfile` option that names an output file into which the output text should be written. In that case, *nothing* will be printed on the command line:



```
$ ./howler.py -o out.txt 'How dare you steal that car!'
```

And there should now be a file called **out.txt** that has the output:

```
$ cat out.txt  
HOW DARE YOU STEAL THAT CAR!
```

In this exercise, you will learn to:

- Accept text input from the command line or from a file.
- Change strings to uppercase.
- Print output either to the command line or to a file that needs to be created.

Reading files

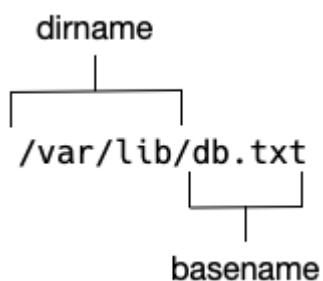


This will be our first exercise that will involve reading files. The argument to the program will be some text. That text might name an input file in which case you will open and read the file. If it's not the name of a file, then you'll use the text itself.

The built-in `os` (operating system) module has a method for detecting if a string is the name of a file. To use it, you must import the `os` module. For instance, there's probably not a file called "blargh" on your system:

```
>>> import os  
>>> os.path.isfile('blargh')  
False
```

The `os` module contains loads of useful submodules and functions. Consult the documentation at <https://docs.python.org/3/library/os.html> or use `help(os)` in the REPL. For instance, the `os.path` module has functions like `basename` and `dirname` for getting a file's name or directory from a path, for example:



```
>>> path = '/var/lib/db.txt'  
>>> os.path.dirname(path)  
'/var/lib'  
>>> os.path.basename(path)  
'db.txt'
```

In the `inputs` directory of the source code repo, there are several files. I'll use a file called `inputs/fox.txt`. Note you will need to be in the main directory of the repo for this to work:

```
>>> file = 'inputs/fox.txt'  
>>> os.path.isfile(file)  
True
```

Once you've determined that the argument is the name of a file, you must `open` it to `read` it. The return from `open` is a *file handle*, the thing we use to `read` the file. I usually call this variable `fh` to remind me that it's a file handle. If I have more than one open file handle like both input and output handles, I may call them `in_fh` and `out_fh`.

```
>>> fh = open(file)
```



If you try to `open` a file that does not exist, you'll get an exception.

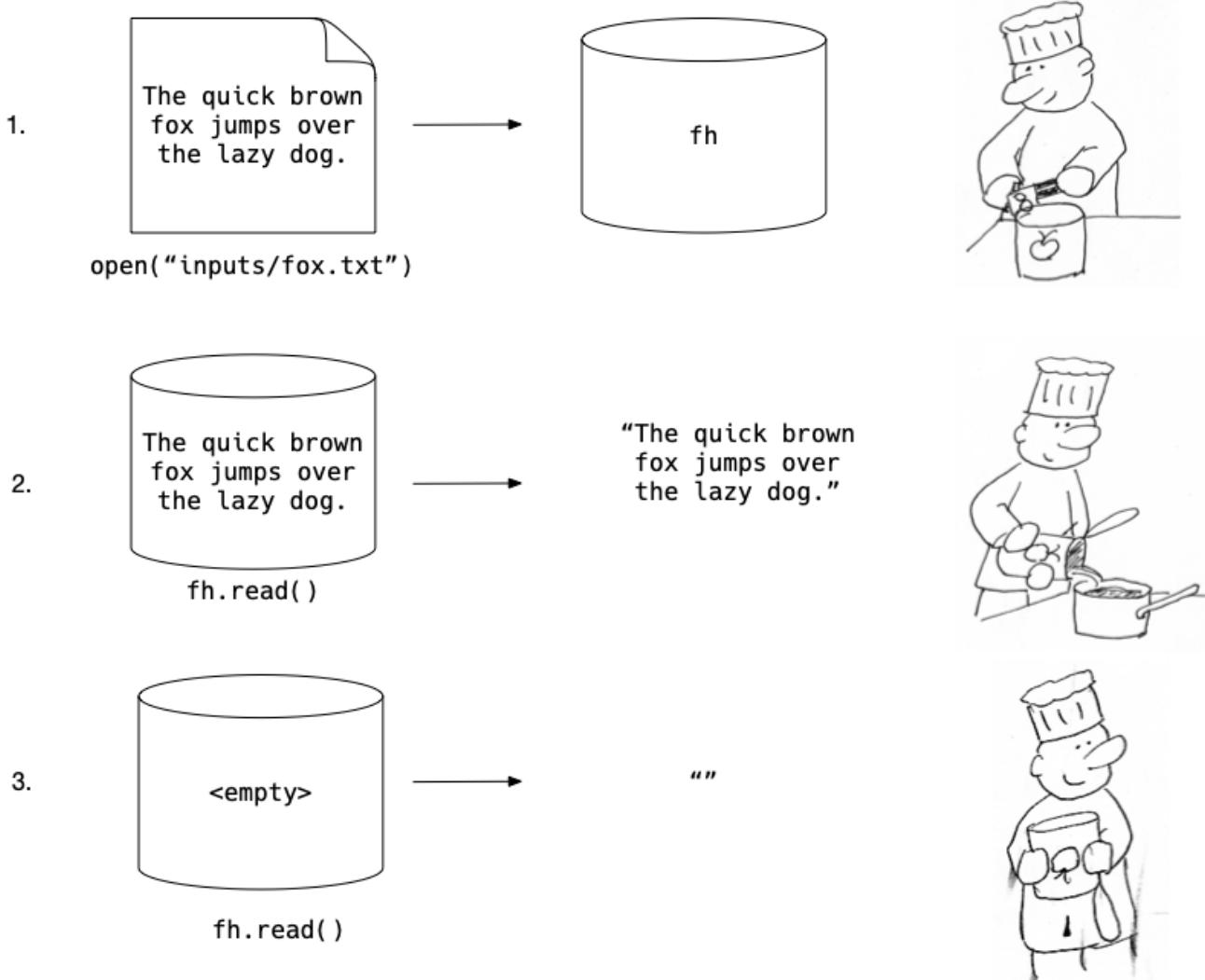
This is unsafe code:

```
>>> file = 'blargh'  
>>> open(file)  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
FileNotFoundError: [Errno 2] No such file or directory: 'blargh'
```

Always check that the file exists!

```
>>> file = 'inputs/fox.txt'  
>>> if os.path.isfile(file):  
...     fh = open(file)
```

I think of the `file` here ("inputs/fox.txt") as the *name of the file* on disk. It's a bit like a can of tomatoes.



1. The file handle (`fh`) is a mechanism I can use to get at the contents of the file. To get at the tomatoes, we need to `open` the can.
2. The `fh.read` method returns what is inside the `file`. With the can opened, we can get at the contents.
3. Once the file handle has been read, there's nothing left.

Let's see what `type` the `fh` is:

```
>>> type(fh)
<class '_io.TextIOWrapper'>
```

In computer lingo, "io" means "input/output." This is some sort of object that wraps I/O operations around text. You can use `help(fh)` (using the name of the variable itself) to read the docs on the `class TextIOWrapper`. The two methods you'll use quite often are `read` and `write`. Right now, we just care about `read`. Let's see what that gives us:

```
>>> fh.read()
'The quick brown fox jumps over the lazy dog.\n'
```



Do me a favor and execute that line one more time. What do you see?

```
>>> fh.read()  
''
```

A file handle is different from something like a `str`. Once you `read` a file handle, it's empty! It's like pouring the tomatoes out of the can. Now the can is empty, and you can't empty it again.

`open(file).read()`
 └─
 `fh.read()`

We can actually compress the `open` and `read` into one line of code by *chaining* those methods together. The `open` returns a file handle that can be used for the call to `read`. Run this:

```
>>> open(file).read()  
'The quick brown fox jumps over the lazy dog.\n'
```

And now run it again:

```
>>> open(file).read()  
'The quick brown fox jumps over the lazy dog.\n'
```

Each time you `open` the `file`, you get a fresh file handle to `read`.

If you want to preserve the contents, you need to copy them into a variable.

```
>>> text = open(file).read()  
>>> text  
'The quick brown fox jumps over the lazy dog.\n'
```

The `type` of the result is a `str`:

```
>>> type(text)  
<class 'str'>
```

```
open(file).read().rstrip()
```



```
fh.read().rstrip()
```



```
str.rstrip()
```

If I want, I can chain a `str` method onto the end of that. For instance, maybe I want to remove the trailing newline. The `str.rstrip` method will remove any whitespace (which includes newlines) from the *right* end of a string.

```
>>> text = open(file).read().rstrip()  
>>> text  
'The quick brown fox jumps over the lazy dog.'
```



Once you have your input text, whether it is from the command line or from a file, you need to UPPERCASE it. The `str.upper()` is probably what you want.

At this point, you have read your input, whether it is from the command-line or from a file.

Writing files

The output of the program should either appear on the command line or be written to a file. Command-line output is also called "standard out" or `STDOUT`. It's the *standard* or normal place for *output* to occur.

There's also an option to the program to write the output to a file, so let's look at how to do that. You still need to `open` a file handle, but we have to use an optional second argument, the string '`w`', that instructs Python to open it for *writing*. Other modes include '`r`' for *reading* (the default) and '`a`' for *Appending*.

Table 1. File writing modes

Mode	Meaning
w	write
r	read
a	append

You can additionally describe the kind of content, whether '`t`' for *text* (the default) or '`b`' for *binary*:

Table 2. File content modes

Mode	Meaning
t	text
b	bytes

You can combine these two tables like '`rb`' to read a binary file or '`at`' to append to a text file. Here we will use '`wt`' to write a text file. I'll call my variable `out_fh` to remind me that this is the "output file handle":



```
>>> out_fh = open('out.txt', 'wt')
```

If the file does not exist, it will be created. If the file does exist, then it will be *overwritten*. It's possible to `open` in a mode that will append text to an existing file. For this exercise, '`wt`' will suffice.

You can use the `write` method of the file handle to put text into the file. Whereas the `print` method will append a newline (`\n`) unless you instruct it not to, the `write` method will *not* add a newline, so you have to explicitly add one.

If you use the `fh.write` method in the REPL, you will see that it return the number of bytes written. Here each character is a byte, and remember that the newline (`\n`) is included:

```
>>> out_fh.write('this is some text\n')
18
```

You can check that this is correct:

```
>>> len('this is some text\n')
18
```

Most code tends to ignore this return value; that is, we don't bother to capture the results into a variable or check that we got a non-zero return. If `write` fails, there's usually some much bigger problem with your system. It's also possible to use the `print` function with the optional `file` argument. Notice that I don't include a newline with `print` because it will add one:

```
>>> print('this is some more text', file=out_fh)
```

When you are done writing to a file handle, you should `close` it so that Python can clean up the file and release the memory associate with it. This function returns no value:

```
>>> out_fh.close()
```

We can verify that our text made it:

```
>>> open('out.txt').read()
'this is some text\nthis is some more text\n'
```

When you `print` on an open file handle, the text will be appended to any previously written data. Look at this code, though:

```
>>> print("I am what I am an' I'm not ashamed", file=open('hagrid.txt', 'wt'))
```

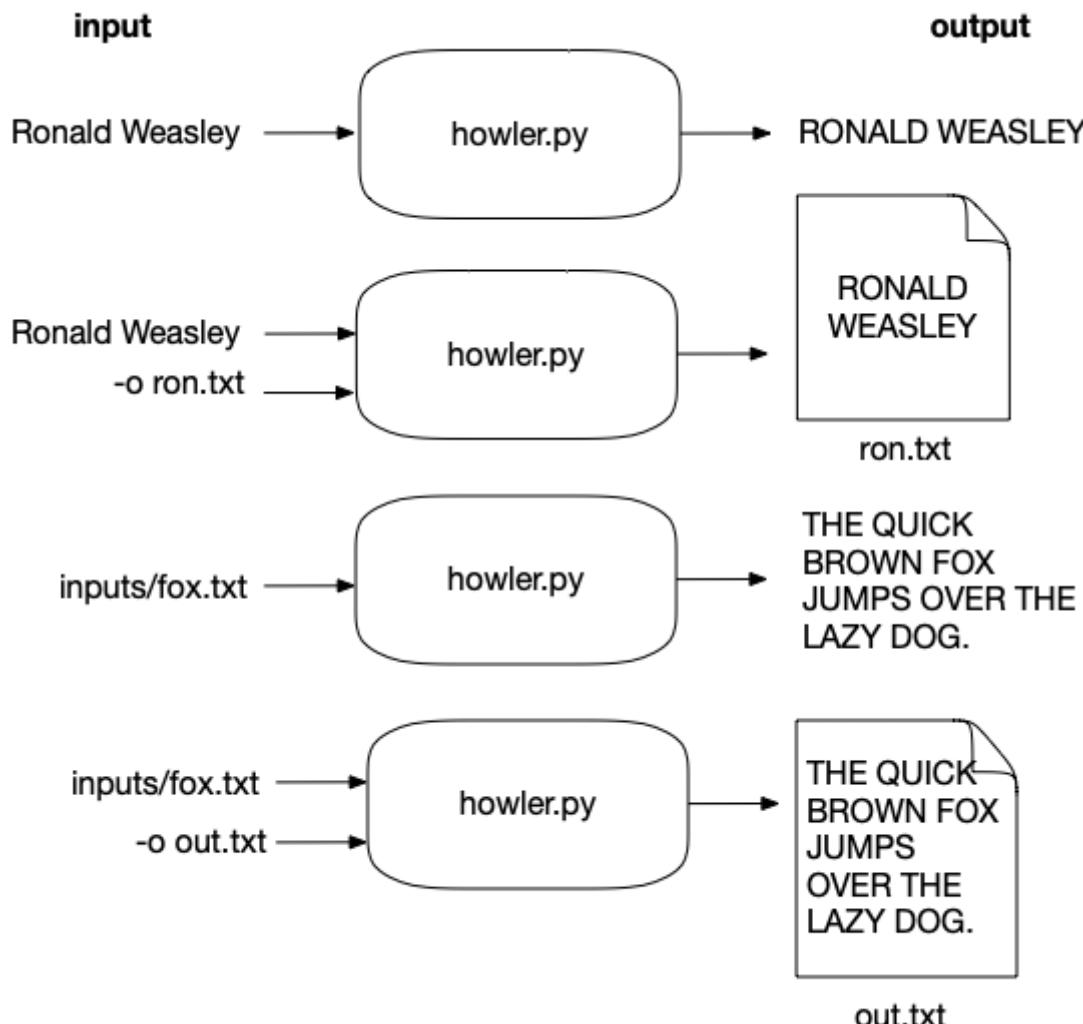
If you run that line twice, will the file called "hagrid.txt" have the line once or twice? Let's find out:

```
>>> open('hagrid.txt').read()
"I am what I am an' I'm not ashamed\n"
```

Just once! Why is that? Remember, each called `open` gives us a new file handle, so calling `open` twice results in new file handles. Both are opened in `write` mode, so that existing data is *overwritten*. It's important to understand how to `open` files properly or you may end up erasing important data files!

Writing `howler.py`

Here is a string diagram showing the overview of the program and some example inputs and outputs:



When run with no arguments, it should print a short usage:

```
$ ./howler.py
usage: howler.py [-h] [-o str] STR
howler.py: error: the following arguments are required: STR
```

When run with `-h` or `--help`, the program should print a longer usage statement:

```
$ ./howler.py -h
usage: howler.py [-h] [-o str] str

Howler (upper-cases input)

positional arguments:
  str                  Input string or file

optional arguments:
  -h, --help            show this help message and exit
  -o str, --outfile str
                        Output filename (default: )
```

If the argument is a regular string, it should uppercase that:

```
$ ./howler.py 'How dare you steal that car!'
HOW DARE YOU STEAL THAT CAR!
```

If the argument is the name of a file, it should uppercase the *contents of the file*:

```
$ ./howler.py ../inputs/fox.txt
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```

If given an `--outfile` filename, the uppercased text should be written to the indicated file and nothing should be printed to STDOUT:

```
$ ./howler.py -o out.txt ../inputs/fox.txt
$ cat out.txt
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```

Hints:

- Start with `new.py` and alter the `get_args` section until your usage statements match the ones above.
- Run the test suite and try to pass just the first test that handles text on the command line and output to `STDOUT`.
- The next test is to see if you can write the output to a given file. Figure out how to do that.
- The next test is for reading input from a file. Then work on that. Don't try to do all the things at once!
- There is a special file handle that always exists called "standard out" (often `STDOUT`). If you `print` without a `file` argument, then it defaults to `sys.stdout`. You will need to `import sys` in order to use it.

Be sure you really try to write the program and pass all the tests before moving on to read the

solution! If you get stuck, maybe whip up a batch of Polyjuice Potion and freak out your friends.

Solution

```
1 #!/usr/bin/env python3
2 """Howler"""
3
4 import argparse
5 import os
6 import sys
7
8
9 # -----
10 def get_args():
11     """get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='Howler (upper-case input)',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text', metavar='str', help='Input string or file') ①
18
19     parser.add_argument('-o', ②
20                         '--outfile',
21                         help='Output filename',
22                         metavar='str',
23                         type=str,
24                         default='')
25
26     args = parser.parse_args() ③
27
28     if os.path.isfile(args.text): ④
29         args.text = open(args.text).read().rstrip() ⑤
30
31     return args ⑥
32
33
34 # -----
35 def main():
36     """Make a jazz noise here"""
37
38     args = get_args() ⑦
39     out_fh = open(args.outfile, 'wt') if args.outfile else sys.stdout ⑧
40     out_fh.write(args.text.upper() + '\n') ⑨
41     out_fh.close() ⑩
42
43
44 # -----
45 if __name__ == '__main__':
46     main()
```

- ① The `text` argument is a string that may be the name of a file.
- ② The `--outfile` option is also a string that names a file.
- ③ Parse the command-line arguments into the variable `args` so that we can manually check the `text` argument.
- ④ Check if `args.text` names an existing file.
- ⑤ If so, overwrite the value of `args.text` with the results of reading the file.
- ⑥ Now that we've fixed up the `args`, we can `return` them to the caller.
- ⑦ Call `get_args` to get the arguments to the program.
- ⑧ Use an `if` expression to choose either `sys.stdout` or a newly opened file handle to write the output.
- ⑨ Use the opened file handle to write the output converted to `upper`.
- ⑩ Close the file handle.

Discussion

How did it go for you this time? I hope you avoided Snape's office. You really don't want more Saturday detentions.

Defining the arguments

The `get_args` function, as always, is the first. Here I define two arguments. The first is a positional `text` argument. Since it may or may not name a file, all I can know is that it will be a string.

```
parser.add_argument('text', metavar='str', help='Input string or file')
```

The other argument is an option, so I give it a short name of `-o` and a long name of `--outfile`. The default `type` for all arguments is `str`, but I go ahead and state that explicitly. The `default` value is the empty string. I could just as easily use the special `None` type which is also the default value, but I prefer to use a defined argument like the empty string.

```
parser.add_argument('-o',
                    '--outfile',
                    help='Output filename',
                    metavar='str',
                    type=str,
                    default='')
```

Reading input from a file or the command line

This is a deceptively simple program that demonstrates a couple of very important elements of file input and output. The `text` input might be a plain string or it might be the name of a file. This pattern will come up repeatedly in this book:

```
if os.path.isfile(args.text):
    args.text = open(args.text).read().rstrip()
```

```
open(args.text).read().rstrip()
  ↓
  fh.read().rstrip()
  ↓
str ← str.rstrip()
```

The function `os.path.isfile` will tell us if

there is a file with the name in `text`. If that returns `True`, then we can safely `open(file)` to get a *file handle* which has a method called `read` which will return *all* the contents of the file. This is usually safe, but be careful if you write a program that could potentially read gigantic files. For instance, in my day job we regularly deal with files with sizes in the 10s to 100s of gigabytes, so I would need to ensure my system has more memory than the size of the file!

The result of `open(file).read()` is a `str` which itself has a method called `rstrip` that will return a copy of the string *stripped* of the whitespace off the *right* side of the string (hence the name `rstrip`).

The longer way to write the above would be:

```
if os.path.isfile(text):
    fh = open(text)
    text = fh.read()
    text = text.rstrip()
```

In my version, I choose to handle this inside the `get_args` function. This is the first time I'm showing you that you can intercept and alter the arguments before passing them on to `main`. We'll use this idea quite a bit in more exercises. I like to do all the work to validate the user's arguments inside `get_args`. I could just as easily do this in `main` after the call to `get_args`, so this is entirely a style issue.

Choosing the output file handle

The line decides where to put the output of our program

```
out_fh = open(args.outfile, 'wt') if args.outfile else sys.stdout
```

The `if` expression will open `args.outfile` for writing text (`wt`) if the user provided that argument; otherwise, we will use `sys.stdout` which is a file handle to STDOUT. Note that we don't have to call `open` on `sys.stdout` because it is always there and always open for business.

```
open(args.outfile, 'wt') if args.outfile else sys.stdout
```

A diagram illustrating the structure of the if expression. It shows the expression `open(args.outfile, 'wt') if args.outfile else sys.stdout` with three horizontal brackets underneath. The first bracket spans from the start of the expression to the opening parenthesis of the if clause. The second bracket spans from the if keyword to the closing parenthesis of the if clause. The third bracket spans from the else keyword to the end of the expression. Below each bracket is a label: "this will be returned if the predicate is True" under the first, "predicate" under the second, and "this will be returned if the predicate is False" under the third.

Printing the output

To get uppercase, we can use the `text.upper` method, then we need to find a way to print it to the output file handle. I chose to do:

```
out_fh.write(text.upper())
```

But you could also do:

```
print(text.upper(), file=out_fh)
```

Finally I need to close the file handle with `out_fh.close()`.

Review



- To `read` or `write` to files, you must `open` them.
- The default mode for `open` is for reading a file.
- To write a text file, you must use '`wt`' as the second argument to `open`.
- By default, you `write` text to a file handle. You must use the '`b`' flag to indicate that you want to write binary data.
- The `os.path` module contains many useful functions such as `isfile` that will tell you if a file exists by the given name.
- `STDOUT` (standard output) is always available via the special `sys.stdout` file handle which is always open.
- The `print` function takes an optional `file` argument of where to put the output. That argument must be an open file handle such as `sys.stdout` (the default) or the result of `open`.

Going Further

- Add a flag that will lowercase the input instead.
- Alter the program handle multiple input files
- Change `--outfile` to `--outdir` and write each input file to the same file name in the output directory.

Chapter 6: Words Count: Reading files/STDIN, iterating lists, formatting strings

"I love to count!" — Count von Count



Counting things is a surprisingly important programming skill. Maybe you're trying to find how many pizzas were sold each quarter or how many times you see certain words in a set of documents. Usually the data we deal with in computing comes to us in files, so we're going to push a little further into reading files and manipulating strings by writing a Python version of the venerable Unix `wc` (word count) program. We're going to write a program called `wc.py` that will count the characters, words, and lines for each input argument which should be files.

Given one or more valid files, it should print the number of lines, words, and characters, each in columns 8 characters wide, followed by a space and then the name of the file. Here's what it looks like for one file:

```
$ ./wc.py ../inputs/scarlet.txt
 7035 68061 396320 ../inputs/scarlet.txt
```

When there are many files, print the counts for each file and then print a "total" line summing each column:

```
$ ./wc.py ../inputs/s*.txt
 7035 68061 396320 ../inputs/scarlet.txt
   17    118     661 ../inputs/sonnet-29.txt
     3      7      45 ../inputs/spiders.txt
 7055 68186 397026 total
```

There may also be *no* arguments, in which case we'll read from "standard in" (`STDIN`) which is the complement to `STDOUT` we used in "Howler." You can use `STDIN` to chain programs together where the output of one program becomes the input for the next. To pass text via `STDIN`, you can use the `<` redirect from a file:

```
$ ./wc.py < ../inputs/fox.txt
1      9      45 <stdin>
```

Or pipe (|) the output of one command into another:

```
$ cat ../inputs/fox.txt | ./wc.py
1      9      45 <stdin>
```

Given a non-existent file, it should print an error message and exit with a non-zero exit value:

```
$ ./wc.py foo
usage: wc.py [-h] FILE [FILE ...]
wc.py: error: argument FILE: can't open 'foo': \
[Errno 2] No such file or directory: 'foo'
```

So our program is going to have to:

- Take 0 or more positional arguments.
- Validate that any given arguments are actually files.
- Read each file and count the lines, words, and characters.
- If no files are present, read from **STDIN**.
- Print the total number of lines, words, and characters if there is more than one file.

Defining file inputs

The first step will be to define your arguments to `argparse`. The program takes *zero or more* positional arguments and nothing else. Remember that you never have to define the `-h` or `--help` arguments as `argparse` handles those automatically.

In "Picnic," we used `nargs='+'` to indicate one or more items for our picnic. Here we want to use `nargs='*' to indicate zero or more. For what it's worth, there's one other value that nargs can take and that is ? for zero or one. In all cases, the argument(s) will be returned as a list. Even if there are no arguments, you will still get an empty list ([]). For this program, if there are no arguments, we'll read STDIN.`

Table 3. Possible values for `nargs`

Symbol	Meaning
<code>?</code>	zero or one
<code>*</code>	zero or more
<code>+</code>	one or more

In "Howler," we used the "standard out" (`STDOUT`) file handle with `sys.stdout`. To read `STDIN`, we'll use Python's `sys.stdin` file handle which is similarly always open and available to you. Because you are using `nargs='*'`, the values will be a `list`, and so the `default` should be a `list` as well. Can you figure out how to make the `default` value for your file argument be a `list` with `sys.stdin`?

Lastly, we should discuss the `type` of the positional arguments. If they are provided, they should be *readable files*. We saw in "Howler" how to test if the input argument was a file by using `os.path.isfile`. In that program, the input might be either plain text or a file name, so we had to check this ourselves. In this program, however, we will require that any arguments should be files, and so we can define the `type=argparse.FileType('r')`. When you do this, `argparse` takes on all the work to validate the inputs from the user and produce useful error messages. If the user provides valid input, then `argparse` will provide you with a `list` of *open file handles*. All in all, this saves you quite a bit of time.

Iterating lists

Your program will end up with a **list** of file handles. In "Jump The Five," we used a **for** loop to iterate through the characters in the input text. Here we can use a **for** loop over the **file** inputs.

```
for fh in args.file:  
    # read each file
```

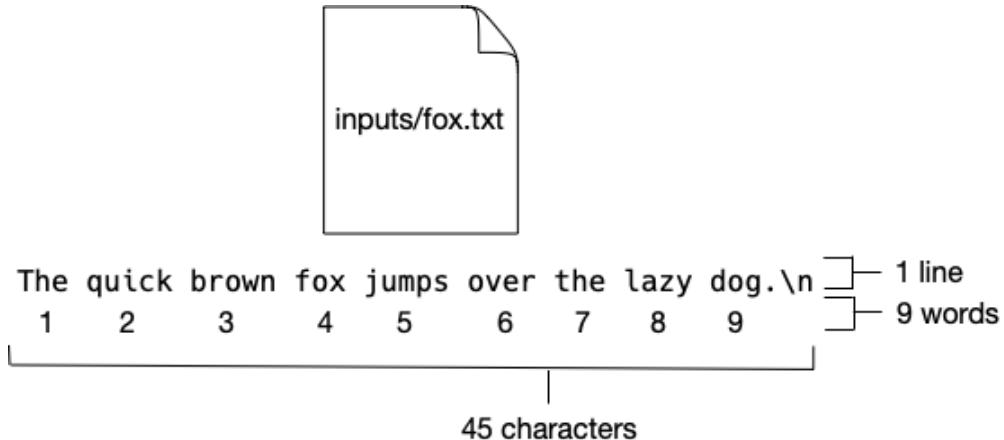
The **fh** is a "file handle." We saw in "Howler" how to manually **open** and **read** a file. Here the **fh** is already open, so we can just **read** it. There are many ways to read a file, however. The **read** method will give you the *entire contents* of the file in one go. If the file is large—say, if the size of the file exceeds your available memory on your machine—then your program will crash. I would recommend, instead, that you use another **for** loop on the **fh**. Python will understand this to mean that you wish to read each **line** of input, one-at-a-time.

```
for fh in args.file:  
    for line in fh:  
        # process the line
```

So that's two levels of **for** loops, one for each file handle and then another for each line in each file handle.

What you're counting

The output for each file will be the number of lines, words, and characters, each printed in a field 8 characters wide followed by the name of the file which will be available to you via `fh.name`. Let's take a look at the output from the standard `wc` program on my system. Notice that when run with just one argument, it produces counts only for that file:



```
$ wc fox.txt
 1      9      45 fox.txt
```

When run with multiple files, it also shows a "total" line:

```
$ wc fox.txt sonnet-29.txt
 1      9      45 fox.txt
 17    118    669 sonnet-29.txt
 18    127    714 total
```

For each file, you will need to create variables that hold the numbers for lines, words, and characters. For instance, if you use the `for line in fh` loop that I suggest, then you need to have a variable like `num_lines` to increment on each iteration. That is, somewhere in your code you will need to set a variable to `0` and then, inside the `for` loop, make it go up by one. The idiom in Python for this is:

```
num_lines = 0
for line in fh:
    num_lines += 1
```

You also need to count the number of words and characters, so you'll need similar `num_words` and `num_chars` variables. In "Picnic," we discussed how we can convert back and forth between strings and lists. For the purposes of this exercise, we'll use the `str.split` method to break each `line` on spaces. You can then use the length of the resulting `list` as the number of words. For the number of characters, you can use the same length function on the `line` and add that to a `num_chars` variable.

Formatting your results

This is the first exercise where the output needs to be formatted in a particular way. Don't try handle this part manually. That way lies madness. Instead, you need to learn the magic of the `str.format` method. The `help` doesn't have much in the way of documentation, so I'd recommend you read PEP3101 (<https://www.python.org/dev/peps/pep-3101/>).

We've seen that the curly braces (`{}`) inside the `str` part create placeholders that will be replaced by the values passed to the method:

```
>>> import math  
>>> 'Pi is {}'.format(math.pi)  
'Pi is 3.141592653589793'
```

You can put formatting information inside the curly braces to specify how you want the value displayed. If you are familiar with `printf` from C-type languages, this is the same idea. For instance, I can print just two numbers of `pi` after the decimal. The `:` introduces the formatting options, and the `0.02f` describes two decimal points of precision:

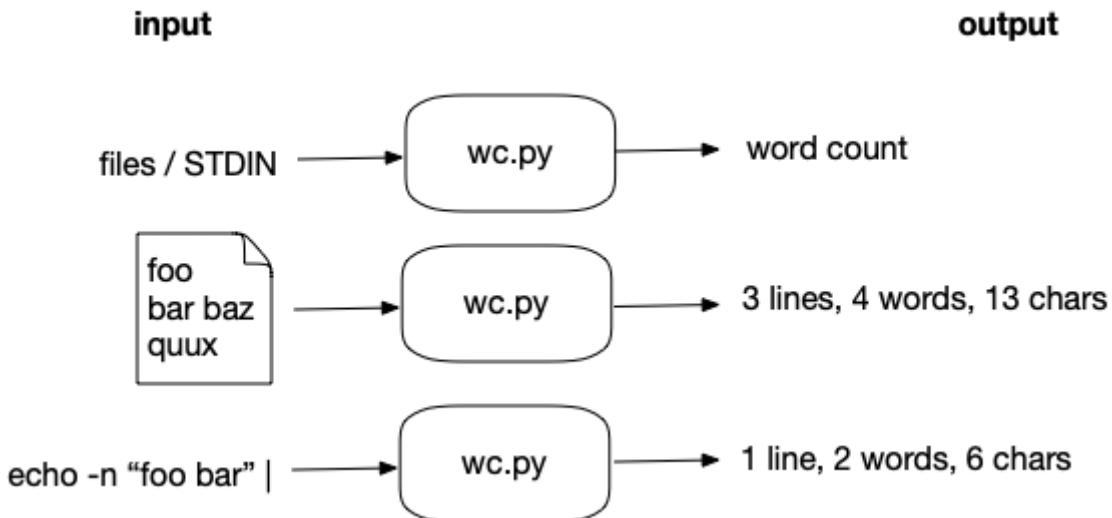
```
>>> 'Pi is {:.02f}'.format(math.pi)  
'Pi is 3.14'
```

The formatting information comes after the colon (`:`) inside the curly braces. You can also use the f-string method where the variable comes *before* the colon:

```
>>> f'Pi is {math.pi:0.02f}'  
'Pi is 3.14'
```

Here you need to use `{:8}` for each of lines, words, and characters so that they all line up in neat columns. The `8` describes the width of the field which is assumed to be a string. The text will be right-justified.

Writing wc



Now it's time to write out Python program called `wc.py` that will emulate the `wc` program in Unix that counts the lines, words, and characters in the given file arguments. If run with the `-h` or `--help` flags, the program should print usage:

```
$ ./wc.py -h
usage: wc.py [-h] [FILE [FILE ...]]
```

Argparse Python script

positional arguments:

```
FILE      Input file(s) (default: [<_io.TextIOWrapper name='<stdin>'  
                                mode='r' encoding='UTF-8']]
```

optional arguments:

```
-h, --help show this help message and exit
```

Hints:

- Start with `new.py` and delete all the non-positional arguments.
- Use `nargs='*'` to indicate zero or more positional arguments for your `file` argument.
- How could you use `sys.stdin` for the `default`? Remember that both `narg='*'` and `nargs='+'` mean that the arguments will be supplied as a `list`. How can you create a `list` that contains just `sys.stdin` for the `default` value?
- Remember that you are just trying to pass one test at a time. Create the program, get the help right, then worry about the first test.
- Compare the results of your version to the `wc` installed on your system. Note that not every Unix-like system has the same `wc`, so results may vary.

Time to write this yourself before you read the solution. Fear is the mind-killer. You can do this.

Solution

```
1 #!/usr/bin/env python3
2 """Emulate wc (word count)"""
3
4 import argparse
5 import sys
6
7
8 # -----
9 def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Emulate wc (word count)',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('file',
17                         metavar='FILE',
18                         nargs='*',
19                         default=[sys.stdin],          ①
20                         type=argparse.FileType('r'),  ②
21                         help='Input file(s)')
22
23     return parser.parse_args()
24
25
26 # -----
27 def main():
28     """Make a jazz noise here"""
29
30     args = get_args()
31
32     total_lines, total_chars, total_words = 0, 0, 0 ③
33     for fh in args.file:                          ④
34         lines, words, chars = 0, 0, 0 ⑤
35         for line in fh:                           ⑥
36             lines += 1                            ⑦
37             chars += len(line)                   ⑧
38             words += len(line.split())          ⑨
39
40         total_lines += lines                  ⑩
41         total_chars += chars
42         total_words += words
43
44     print(f'{lines:8}{words:8}{chars:8} {fh.name}') ⑪
45
46     if len(args.file) > 1:                  ⑫
47         print(f'{total_lines:8}{total_words:8}{total_chars:8} total') ⑬
```

```
48
49
50 # -----
51 if __name__ == '__main__':
52     main()
```

- ① If you set the `default` to a `list` with `sys.stdin`, then you have handled the `STDIN` option.
- ② If the user supplies any arguments, `argparse` will check if they are valid file inputs. If there is a problem, `argparse` will halt execution of the program and show the user an error message.
- ③ These are the variables for the "total" line, if we need them.
- ④ Iterate through the `list` of `arg.file` inputs. I use the variable `fh` to remind me that these are open file hanndles, even `STDIN`.
- ⑤ Initialize variables to count *just this file*.
- ⑥ Iterate through each `line` of `fh`.
- ⑦ For each line, we increment `lines` by 1.
- ⑧ The number of `chars` is incremented by the length of the `line`.
- ⑨ To get the number of words, we can `split` the `line` on spaces (the default). We length of that `list` is added to the `words`.
- ⑩ We add the numbers for this file to the `total_` variables.
- ⑪ Print the counts for this file using the `{:8}` option to print in a field 8 characters wide.
- ⑫ Check if we had more than 1 input.
- ⑬ Print the "total" line.

Discussion

Defining the arguments

This program is rather short and seems rather simple, but it's definitely not exactly easy. One part of the exercise is to really get familiar with `argparse` and the trouble it can save you. The key is in defining the `file` positional arguments. If you use `nargs='*'` to indicate zero or more arguments, then you know `argparse` is going to give you back a `list` with zero or more elements. If you use `type=argparse.FileType('r')`, then any arguments provided must be readable files. The `list` that `argparse` returns will be a `list` of *open file handles*. Lastly, if you use `default=[sys.stdin]`, then you understand that `sys.stdin` is essentially an open file handle to read from "standard in" (AKA `STDIN`), and you are letting `argparse` know that you want the default to be a `list` containing `sys.stdin`.

Reading a file using a `for` loop

I can create a list of open file handles in the REPL to mimic what I'd get from `args.file`:

```
>>> files = [open('../inputs/fox.txt')]
```

Before I use a `for` loop to iterate through them, I need to set up three variables to track the *total* number of lines, words, and characters:

```
>>> total_lines, total_chars, total_words = 0, 0, 0
```



Inside the `for` loop for each file handle, I initialize three more variables to hold the count of lines, characters, and words *for this particular file*. I then use another `for` loop to iterate over each line in the file handle (`fh`). For the `lines`, I can add `1` on each pass through the `for` loop. For the `chars`, I can add length of the line (`len(line)`) to track the number of characters. Lastly for the `words`, I can use `line.split()` to break the line on whitespace to create a `list` of "words." It's not actually a perfect way to count actual words, but it's close enough. I can use the `len` function on the `list` to add to the `words` variable. The `for` loop ends when the end of the file is reached, and that is when I can `print` out the counts and the file name using `{:8}` placeholders in the print template to indicate a text field 8 characters wide.

```
>>> for fh in files:  
...     lines, words, chars = 0, 0, 0  
...     for line in fh:  
...         lines += 1  
...         chars += len(line)  
...         words += len(line.split())  
...     print(f'{lines:8}{words:8}{chars:8} {fh.name}')  
...     total_lines += lines  
...     total_chars += chars  
...     total_words += words  
...  
1         9      45 ../inputs/fox.txt
```

Notice that the `print` statement lines up with the inner `for` loop so that it will run after we're done iterating over the lines in `fh`. I chose to use the f-string method to print each of `lines`, `words`, and `chars` in a space 8 characters wide. After printing, I can add the counts to my "total" variables to keep a running total.

Lastly, if the number of file arguments is greater than 1, I need to print my totals:

```
if len(args.files) > 1:  
    print(f'{total_lines:8}{total_words:8}{total_chars:8} total')
```

A low-memory version

There is a potentially serious problem waiting to bite us in this program. In the `get_args`, we're reading the entire file into memory with this line:

```
if os.path.isfile(args.text):  
    args.text = open(args.text).read().rstrip()
```

We could, instead, only `open` the file:

```
if os.path.isfile(args.text):  
    args.text = open(args.text)
```

And later read it line-by-line:

```
for line in args.text:  
    out_fh.write(line.upper())
```

The problem, though, is how to handle the times when the `text` argument is actually text and not the name of a file. The `io` (input-output) module in Python has a way to represent text as *stream*:

```
>>> import io  
      ①  
>>> text = io.StringIO('foo\nbar\nbaz\n') ②  
>>> for line in text:  
      ③  
...     print(line, end='')  
      ④  
...  
foo  
bar  
baz
```

- ① Import the `io` module.
- ② Use the `io.StringIO` function to turn the given `str` value into something we can treat like an open file handle.
- ③ Use a `for` loop to iterate the "lines" of text by separated by newlines.
- ④ Print the `line` using the `end=''` option to avoid having 2 newlines.

To make this work, we can change how we handle the `args.text` like so:

```
if os.path.isfile(args.text):  ①  
    args.text = open(args.text) ②  
else:  
    args.text = io.StringIO(args.text + '\n') ③
```

- ① If `args.text` is a file...
- ② Replace `args.text` with the file handle created by opening the file named by it.
- ③ Otherwise, replace `args.text` with an `io.StringIO` value that will act like an open file handle.
Note that we need to add a newline to the text so that it will look like the lines of input coming from an actual file.

Review

- The `nargs` (number of arguments) option to `argparse` allows you to validate the number of arguments from the user. The star ('*') means zero or more while '+' means one or more.
- If you define an argument using `type=argparse.FileType('r')`, then `argparse` will validate that the user has provided a readable file and will make the value available in your code as an open file handle.
- You can read and write from the Unix standard in/out file handles by using `sys.stdin` and `sys.stdout`.
- You can nest `for` loops to handle multiple levels of processing.
- The `str.split` method will split a string on spaces into words.
- The `len` function can be used on both strings and lists. For the latter, it will tell you the number of elements contained.
- The `str.format` and Python's f-strings both recognize the same printf-style formatting options to allow you to control how a value is displayed.

Going Further

- Add the same flags as your system `wc` and change your program to create the same output; i.e., print all the columns for no arguments but only print the "lines" when `-l` is present.
- Implement other system tools like `head`, `tail`, `cat`, and `tac` (the reverse of `cat`).

Chapter 7: Gashlycrumb: Looking items up in a dictionary

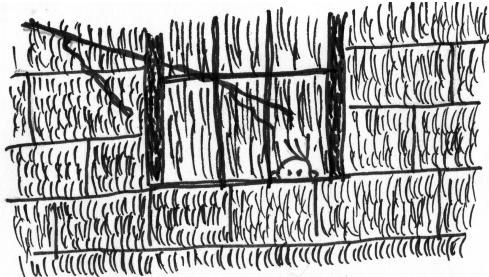


Figure 4. N is for Neville who died of ennui.

Every time you log into a website, the code behind it has to look up your username and password to compare to the values you put into the login form. Whenever you give your phone number at the hardware store or scan your library card to checkout a book, a computer program uses one piece of information to find other things like how often you buy compost or if you have any overdue books. Probably all these examples would be using a database to find that information. We're going to use a dictionary that we will fill with information from an input file.

In this exercise, we're going to look up a line of text from an input file that starts with the letter provided by the user. The text will come from an input file which will default to Edward Gorey's "The Gashlycrumb Tinies," an abecedarian book that describes various and ghastly ways in which children die. Instead of "A is for apple, B is for bear," we get:

A is for Amy who fell down the stairs. B is for Basil assaulted by bears.

Our `gashlycrumb.py` program will take a letter of the alphabet as a positional argument and will look up the line of text from an *optional* input file that starts with that letter. The input file will have each letter on a separate line:

```
$ head -2 gashlycrumb.txt
A is for Amy who fell down the stairs.
B is for Basil assaulted by bears.
```

When our unfortunate user runs our program, here is what they would see. Note that we will consider the letter in a *case-insensitive* fashion:

```
$ ./gashlycrumb.py a
A is for Amy who fell down the stairs.
```

Our program will need to:

- Accept a single positional argument we'll call `letter`
- Accept an optional `--file` argument which must be a file. The default value will be

'gashlycrumb.txt' (provided).

- Read the file, find the first letter of each line, and build a data structure that associates that letter to the line of text.
- Check if the given **letter** is a present in the given **file**.
- Print the line of text for the **letter** if present or an error if it isn't.



You can draw from several previous programs:

- From the "Word Count" program, you know how to take a file input and read it line-by-line.
- From the "Article" program, you know how to get the first letter of a bit of text.
- From the "Jump The Five" program, you know how to build a dictionary and lookup a value.

Now you'll put all those skills together to recite morbid poetry!

Writing `gashlycrumb.py`

Before you begin writing, I would encourage you to run the tests with `make test` or `pytest -xv test.py` in the `gashlycrumb` directory. The first test should fail:

```
test.py::test_exists FAILED
```

This is just a reminder that the first thing to do is to create the file called `gashlycrumb.py`. You can do this however you like, maybe by running `new.py gashlycrumb.py` in the `gashlycrumb` directory or by copying the `template/template.py` file, or by just starting a new file from scratch.

Run your tests again and you should pass the first and possibly the second tests if your program produces a usage statement. Next let's get the arguments straight. Modify your program's parameters in the `get_args` function so that it will produce the following usage when the program is run with *no arguments* or with the `-h` or `--help` flags:

```
$ ./gashlycrumb.py -h
usage: gashlycrumb.py [-h] [-f str] str

Gashlycrumb

positional arguments:
  str                  Letter ①

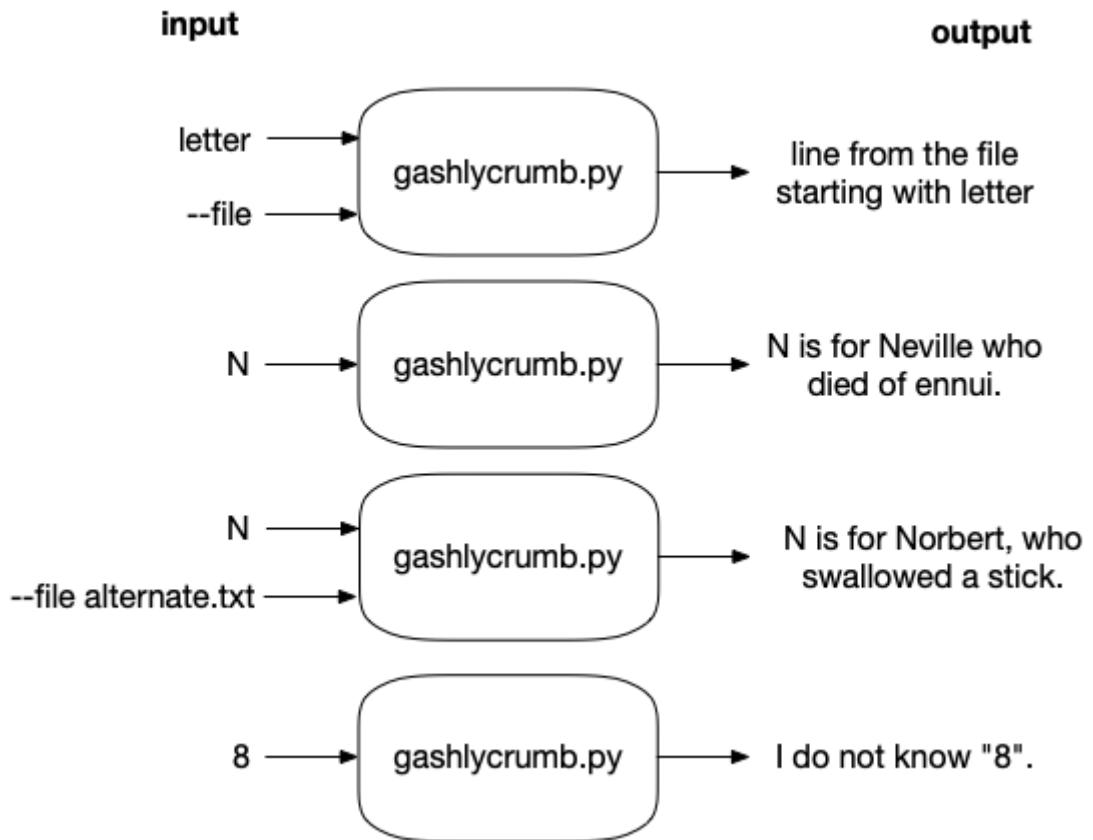
optional arguments:
  -h, --help            show this help message and exit. ②
  -f str, --file str   Input file (default: gashlycrumb.txt) ③
```

① The "letter" is a required positional argument.

② The `-h` and `--help` arguments are created automatically by `argparse`.

③ The `-f` or `--file` argument is an option with a default value of "gashlycrumb.txt"

Here is a string diagram showing how the program will work:



Once you have the correct usage, start off by echoing the `letter`:

```
def main():
    args = get_args()
    print(args.letter)
```

Try running it to make sure it works:

```
$ ./gashlycrumb.py a
a
```

Next read the `file` line-by-line using a `for` loop. Note that I'm using `end=''` with the `print` so that it won't print the newline that's already attached to each `line` of the `file`:

```
def main():
    args = get_args()
    print(args.letter)

    for line in args.file:
        print(line, end='')
```

Try running it to ensure you can read the input `file`:

```
$ ./gashlycrumb.py a | head -3
a
A is for Amy who fell down the stairs.
B is for Basil assaulted by bears.
```

Use the "alternate.txt" file, too:

```
$ ./gashlycrumb.py a -f alternate.txt | head -3
a
A is for Alfred, poisoned to death.
B is for Bertrand, consumed by meth.
```

If provided a `--file` argument that does not exist, your program should exit with an error message. Note that, if you use `type=argparse.FileType('r')`, this error should be produced automatically by `argparse`:

```
$ ./gashlycrumb.py -f blargh b
usage: gashlycrumb.py [-h] [-f str] str
gashlycrumb.py: error: argument -f/--file: can't open 'blargh': [Errno 2] No such file
or directory: 'blargh'
```

Now think about how you can use the first letter of each `line` to create an entry into a `dict`. Use the `print` command to look at your dictionary. Figure out how to check if the given `letter` is in (wink, nudge, nudge) your dictionary. If given a value that does not exist in the list of first characters on the lines from the input file (when searched without regard to case), you should print a message:



```
$ ./gashlycrumb.py 3
I do not know "3".
$ ./gashlycrumb.py CH
I do not know "CH".
```

A is for Amy who fell down the stairs

{ "A": "A is for Amy who fell down the stairs" } If the given `letter` is in the dictionary, print the value for it:

```
$ ./gashlycrumb.py a  
A is for Amy who fell down the stairs.  
$ ./gashlycrumb.py z  
Z is for Zillah who drank too much gin.
```

Run the test suite to ensure your program meets all the requirements. Read the errors closely and fix your program.

Hints:

- Start with `new.py` and remove everything but the positional and optional `argparse.FileType('r')` parameters.
- A dictionary is a natural data structure that you can use to associate some value like the letter "A" to some phrase like "A is for Amy who fell down the stairs." Create a new, empty `dict`.
- Once you have an open file handle, you can `read` the file line-by-line with a `for` loop.
- Each line of text is a string. How can you get the first character of a string?
- Using that first character, how can you set the value of a `dict` to be the key and the line itself to be the value?
- Once you have constructed the dictionary of letters to lines, how can you check that the user's `letter` argument is `in` the dictionary?
- Can you solve this without a `dict`?

No skipping ahead to the solution until you have written your own version! If you peek, you might be die a horrible death stamped by kittens.

Solution

```
1 #!/usr/bin/env python3
2 """Lookup tables"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """get command-line arguments"""
10
11     parser = argparse.ArgumentParser(
12         description='Gashlycrumb',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('letter', help='Letter', metavar='str', type=str) ①
16
17     parser.add_argument('-f', ②
18                         '--file',
19                         help='Input file',
20                         metavar='str',
21                         type=argparse.FileType('r'),
22                         default='gashlycrumb.txt')
23
24     return parser.parse_args()
25
26
27 # -----
28 def main():
29     """Make a jazz noise here"""
30
31     args = get_args()
32     letter = args.letter ③
33     lookup = {line[0].upper(): line.rstrip() for line in args.file} ④
34
35     if letter.upper() in lookup: ⑤
36         print(lookup[letter.upper()]) ⑥
37     else:
38         print(f'I do not know "{letter}".') ⑦
39
40
41 # -----
42 if __name__ == '__main__':
43     main()
```

① The required argument (a letter).

② An optional `--file` that must be a readable file, if provided. Defaults to a known value.

- ③ Copy the `args.letter` value into the variable `letter`.
- ④ Build the `lookup` using a dictionary comprehension that reads the given file. Use the `upper` function to disregard case.
- ⑤ See if the `letter` argument is in the `lookup` dictionary, checking the `upper` value to disregard case.
- ⑥ If so, `print` the line of text from the `lookup` for the `letter`.
- ⑦ Otherwise, print a message that the `letter` is unknown.

Discussion

Did the frightful paws of the kittens hurt much? Let's talk about how I solved this problem. Remember, mine is just one of many possible solutions.

Handling the arguments

I prefer to have all the logic for parsing and validating the command-line arguments in the `get_args` function. In particular, `argparse` can do a fine job verifying tedious things such as an argument being an existing, readable `--file` which is why I use `type=argparse.FileType('r')` for that argument. If the user doesn't supply a valid argument, then `argparse` will throw an error, printing a helpful message along with the short usage and exiting with an error code.

By the time I get to the line `args = get_args()`, I know that I have a valid, open file handle in the `args.file` slot. In the REPL, I can manually do what `argparse` has done by using `open` to get a file handle which I like to usually call `fh`. For copyright purposes, I'll use my alternate text:

```
>>> fh = open('alternate.txt')
```

Reading the input file

I know that I want to use a dictionary where the keys are the first letters of each line and the values are the lines themselves. That means I need to start by creating a new, empty dictionary either by using the `dict()` function or by setting a variable equal to an empty set of curly braces `{}`. I'll call my variable `lookup`:

```
>>> lookup = {}
```

I will use a `for` loop to read each `line` of text. From Crow's Nest, I know I can use `line[0].upper()` to get the first letter of `line` and uppercase it. I can use that as the key into `lookup`. Each `line` of text ends with a newline that I'd like to remove. I can use the `str.rstrip` method to strip whitespace from the right side of the `line` (`rstrip = right strip`). The result of that will be the value for my `lookup`:

```
for line in fh:  
    lookup[line[0].upper()] = line.rstrip()
```

I'd like to look at the resulting `lookup` dictionary. I can `print` it from the program or type `lookup` in the REPL, but it's going to be hard to read. I encourage you to try it. Luckily there is a lovely module called `pprint` to "pretty print" data structures. Here is how I can import the `pprint pprint` function as the shortcut `pp`:

```
>>> from pprint import pprint as pp
```

Now let's take a peek at `lookup`:

```
>>> pp(lookup)
{'A': 'A is for Alfred, poisoned to death.',
 'B': 'B is for Bertrand, consumed by meth.',
 'C': 'C is for Cornell, who ate some glass.',
 'D': 'D is for Donald, who died from gas.',
 'E': 'E is for Edward, hanged by the neck.',
 'F': 'F is for Freddy, crushed in a wreck.',
 'G': 'G is for Geoffrey, who slit his wrist.',
 'H': "H is for Henry, who's neck got a twist.",
 'I': 'I is for Ingrid, who tripped down a stair.',
 'J': 'J is for Jered, who fell off a chair.',
 'K': 'K is for Kevin, bit by a snake.',
 'L': 'L is for Lauryl, impaled on a stake.',
 'M': 'M is for Moira, hit by a brick.',
 'N': 'N is for Norbert, who swallowed a stick.',
 'O': 'O is for Orville, who fell in a canyon',
 'P': 'P is for Paul, strangled by his banyan',
 'Q': 'Q is for Quintanna, flayed in the night',
 'R': 'R is for Robert, who died of spite',
 'S': 'S is for Susan, stung by a jelly',
 'T': 'T is for Terrange, kicked in the belly',
 'U': "U is for Uma, who's life was vanquished",
 'V': 'V is for Victor, consumed by anguish',
 'W': "W is for Walter, who's socks were too long",
 'X': 'X is for Xavier, stuck through with a prong',
 'Y': 'Y is for Yoeman, too fat by a piece',
 'Z': 'Z is for Zora, smothered by a fleece.'}
```

Hey, that looks like a handy data structure. Hooray for us!

Looping with `for` versus a list comprehensions

```
lookup = {}
for line in fh:
    lookup[line[0].upper()] = line.rstrip()
```

```
lookup = { line[0].upper(): line.rstrip() for line in fh }
```

We used three lines of code to build our `lookup` dictionary. We can actually accomplish that in *one* line by using a "dictionary comprehension." We've created list comprehensions by sticking a `for` inside brackets `[]`. A dictionary comprehension is the same but the `for` loop is inside curly braces `{}`.

If you are following along by pasting code into the REPL, note that we have exhausted the file handle `fh` just above by reading it. (Refer back to the Howler to read about file handles.) I need to `open` it again for this next bit:

```
>>> fh = open('gashlycrumb.txt')
>>> lookup = { line[0].upper(): line.rstrip() for line in fh }
```

If you `pprint` it again, you should see the same output as above. It may seem like showing off to write one line of code instead of three, but it really does make a good deal of sense to write compact, idiomatic code. More code always means more chances for bugs, so I usually try to write code that is as simple as possible (but no simpler).

Dictionary lookups



Now that I have a `lookup`, I can ask if some value is `in` the keys. Note that I know the letters are in uppercase and I assume the user could give me lower, so I use `letter.upper()` to only compare that case:

```
>>> letter = 'a'
>>> letter.upper() in lookup
True
>>> lookup[letter.upper()]
'A is for Amy who fell down the stairs.'
```

If the letter is found, I can print the line of text for that letter; otherwise, I can print the message that I don't know that letter:

```
>>> letter = '4'
>>> if letter.upper() in lookup:
...     print(lookup[letter.upper()])
... else:
...     print('I do not know "{}".format(letter))
...
I do not know "4".
```

dict vs list of tuple

I don't have to use a `dict` to solve this problem. I could, for example, use a `list` of `tuple` values. We haven't talked about the `tuple` yet. It's an immutable `list`, and you can create a tuple by separating items with commas:

```
>>> x = 'foo', 'bar'
>>> type(x)
<class 'tuple'>
```

It's common to put parentheses around the values if only to make your intent clearer. For instance, this is valid syntax for a tuple with only one value:

```
>>> x = 'foo',
```

To me, that looks like an incomplete statement. I think this is easier to understand:

```
>>> x = ('foo',)
```

Note that the trailing comma is required or you end up with a single value:

```
>>> x = ('foo')
>>> type(x)
<class 'str'>
```

Let's compare and contrast a **tuple** and a **list**:

```
>>> x = ('foo', 'bar')
>>> y = ['foo', 'bar']
```

We can use the **len** function to find their lengths:

```
>>> len(x)
2
>>> len(y)
2
```

And we can use index values to get specific elements:

```
>>> x[0]
'foo'
>>> y[1]
'bar'
```

We can alter the **list** called **y** like maybe adding a new value:

```
>>> y.append('baz')
>>> y
['foo', 'bar', 'baz']
```

But we cannot do this to the **tuple** called **x**:

```
>>> x.append('baz')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'tuple' object has no attribute 'append'
```

We can also change any element inside a **list**:

```
>>> y[0] = 'quux'
>>> y
['quux', 'bar', 'baz']
```

But, again, a **tuple** is immutable:

```
>>> x[0] = 'quux'
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: 'tuple' object does not support item assignment
```

The **dict.items** method will return a **list** of **tuple** values. Our **lookup** is rather large, so here's a way to look at the first three values of that using **pp** (pretty print) function. Note the outermost square brackets **[]** show that this is a **list**. Each element of the list is enclosed in parentheses to indicate they are **tuple** values:

```
>>> pp(list(lookup.items())[:3])
[('A', 'A is for Amy who fell down the stairs.'),
 ('B', 'B is for Basil assaulted by bears.'),
 ('C', 'C is for Clara who wasted away.')]
```

So a **dict** can be thought of as a list of tuples which I could make directly using a list comprehension:

```
>>> fh = open('gashlycrumb.txt')
>>> lookup = [ (line[0].upper(), line.rstrip()) for line in fh ]
```

I can unpack **lookup** into **char** and **line** to extract, say, just the first 10 **char** elements:

```
>>> [char for char, line in lookup][:10]
['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J']
```

And then use **in** to see if my **letter** is present:

```
>>> letter = 'a'  
>>> letter.upper() in [char for char, line in lookup]  
True
```

And get the value like so:

```
>>> [line for char, line in lookup if char == letter.upper()]  
['A is for Amy who fell down the stairs.']
```

The problem is that the cost of the search is proportional to the number of values. That is, if we were searching a million keys in a list, then Python starts searching at the beginning of the list and goes until it finds the value. When you store items in a `dict`, the search time for a key can be much shorter, often nearly instantaneous. The method `dict.items()` will return a `list` of `tuple` values containing (`key, value`), so the line between these two is already blurry.

Review

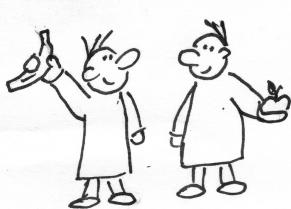
- A dictionary comprehension is a way to build a dictionary in a one-line `for` loop.
- Defining file input arguments using `argparse.FileType` saves you time and code.
- Looking up a key in a `dict` is faster than searching for an element in a `list`. The time to search a `list` is proportional to its length while dictionary keys are stored in a way optimized for finding them.
- Python's `pprint` module is used for "pretty printing" complex data structures.

Going Further

- Write an interactive version that takes input directly from the user. Use `while True` to set up an infinite loop and keep using `input` to get the user's next `letter`:

```
$ ./gashlycrumb_interactive.py
Please provide a letter [! to quit]: t
T is for Titus who flew into bits.
Please provide a letter [! to quit]: 7
I do not know "7".
Please provide a letter [! to quit]: !
Bye
```

Chapter 8: Apples and Bananas: Find and replace



Have you ever misspelled a word? I haven't, but I've heard that many other people often do. We can use computers to find and replace all instances of a misspelled word with the correction. Or maybe you'd like to replace all mentions of your ex's name in your poetry with your new love's name? Find and replace is your friend. To get us started, let us consider the children's song "Apples and Bananas" wherein we intone about our favorite fruits to consume:

I like to eat, eat, eat apples and bananas

Subsequent verses substitute the main vowel sound in the fruits for various other vowel sounds, such as the long "a" (as in "hay") sound:

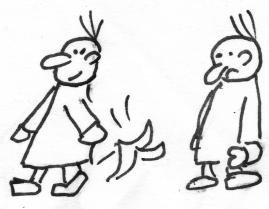
I like to ate, ate, ate ay-ples and ba-nay-nays



Or the ever-popular long "e" (as in "knee"):

I like to eat, eat, eat ee-ples and bee-nee-nees

And so forth. In this exercise, we'll write a Python program called `apples.py` takes some text given as a single positional argument and replaces all the vowels in the text with a given `-v` or `--vowel` options (default `a`).



The program should handle text on the command line:

```
$ ./apples.py foo  
faa
```

And accept the `-v` or `--vowel` option:

```
$ ./apples.py foo -v i  
fii
```



Your program should *preserve the case* of the input vowels:

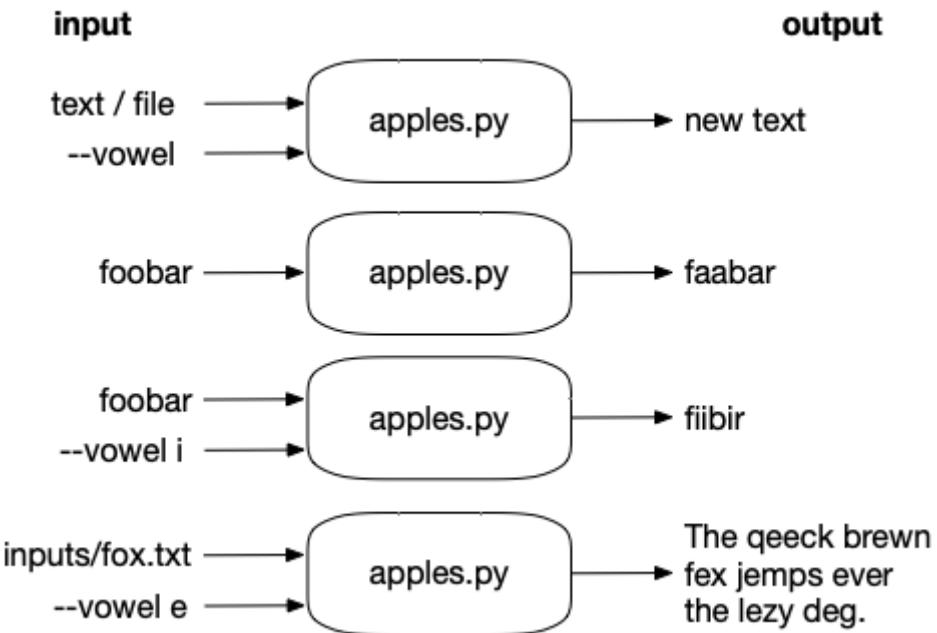
```
$ ./apples.py -v i "APPLES AND BANANAS"  
IPPLIS IND BININIS
```

As with the "Howler" program, the text argument may name a file in which case your program should read the contents of the file.



```
$ ./apples.py ../inputs/fox.txt  
Tha qaack brawn fax jamps avar tha lazy dag.  
$ ./apples.py --vowel e ../inputs/fox.txt  
The qeeck brewn fex jemps ever the lezy deg.
```

It might help to look at a diagram of the program's inputs and output:



Here is the usage that should print when there are *no arguments*:

```
$ ./apples.py
usage: apples.py [-h] [-v str] str
apples.py: error: the following arguments are required: str
```

And the program should always print usage for the `-h` and `--help` flags:

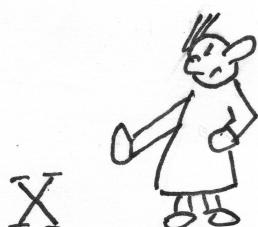
```
$ ./apples.py -h
usage: apples.py [-h] [-v str] str

Apples and bananas

positional arguments:
  str                  Input text or file

optional arguments:
  -h, --help            show this help message and exit
  -v str, --vowel str  The vowel to substitute (default: a)
```

Note that the program should complain if the `--vowel` argument is not a single, lowercase vowel:



```
$ ./apples.py -v x foo
usage: apples.py [-h] [-v str] str
apples.py: error: argument -v/--vowel: \
invalid choice: 'x' (choose from 'a', 'e', 'i', 'o', 'u')
```

So our program is going to need to:

- Take a positional argument that might be some plain text or may name a file.
- If the argument is a file, use the contents as the input text.
- Take an optional `-v` or `--vowel` argument that should default to the letter "a".
- Verify that the `--vowel` option is in the set of vowels "a," "e," "i," "o," and "u."
- Replace all instances of vowels in the input text with the specified (or default) `--vowel` argument.
- Print the new text to `STDOUT`.

Altering strings

So far in our discussions of Python strings, numbers, lists, and dictionaries, we've seen how easily we can change or *mutate* variables. There is a problem, however, in that *strings are immutable*. Suppose we have a `text` variable that holds our input text:



```
>>> text = 'The quick brown fox jumps over the lazy dog.'
```

If we wanted to turn the first "e" (at index 2) into an "i," we cannot do this:

```
>>> text[2] = 'i'  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
TypeError: 'str' object does not support item assignment
```

To change `text`, we need to set it equal to an entirely new value. In "Jump the Five" we saw that you can use a `for` loop to iterate over the characters in a string. For instance, I could laboriously uppercase the `text` like so:



```
new = ''          ①  
for char in text:  
    new += char.upper() ③
```

- ① Set a `new` variable equal to the empty string.
- ② Iterate through each character in the text.
- ③ Append the uppercase version of the character to the `new` variable.

We can inspect the `new` value to verify that it is all uppercase:

```
>>> new  
'THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.'
```

Using this idea, you could iterate through the characters of `text` and build up a new string. Whenever the character is a vowel, you can change it for the given `vowel`, otherwise you use the character itself. We had to identify vowels in "Crow's Nest," so you can refer back to how you did that.

Using `str.replace`

In "Jump The Five," we used the `str.replace` method that might work. Let's look at the `help`:



```
>>> help(str.replace)  
replace(self, old, new, count=-1, /)  
    Return a copy with all occurrences of substring old replaced by new.
```

`count`
Maximum number of occurrences to replace.
-1 (the default value) means replace all occurrences.

If the optional argument `count` is given, only the first `count` occurrences are replaced.

Let's play with that in the REPL. I could `replace` "T" for "X". Can you see a way to replace all the vowels using this idea?



```
>>> text.replace('T', 'X')  
'Xhe quick brown fox jumps over the lazy dog.'
```

Remember that this method never mutates the given string but instead returns a *new string*.

Using `str.translate`

We also looked at the `str.translate` method. The documentation is a bit more cryptic.



```
>>> help(str.translate)
translate(self, table, /)
    Replace each character in the string using the given translation table.
```

`table`
Translation table, which must be a mapping of Unicode ordinals to
Unicode ordinals, strings, or None.

The table must implement lookup/indexing via `__getitem__`, for instance a
dictionary or list. If this operation raises `LookupError`, the character is
left untouched. Characters mapped to `None` are deleted.

In "Jump The Five," we created a `dict` that associated the string '`1`' to the string '`9`' and so forth:

```
jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
           '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
```

And then we used that to change all the keys of the `jumper` to their corresponding values:

```
>>> '876-5309'.translate(str.maketrans(jumper))
'234-0751'
```

What should be the keys and values of the `dict` for this exercise?

Other ways to mutate strings



If you know about regular expressions, that's a strong solution. If you haven't heard of those, don't worry as I'll introduce them in the discussion. The point is for you to go *play*

with this and come up with a solution. I found 8 ways to change all the vowels to a new character, so there are many ways you could approach this.

How many *different* methods can you find on your own before you look at my solution?

Hints:

- Consider using the `choices` option in the `argparse` documentation for how to constrain the `--vowel` options.
- Be sure to change both lower- and uppercase versions of the vowel, preserving the case of the input characters.

Now is the time to dig in and see what you can do before you look at my solutions.

Solution

```
1 #!/usr/bin/env python3
2 """Apples and Bananas"""
3
4 import argparse
5 import os
6
7
8 # -----
9 def get_args():
10     """get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Apples and bananas',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('text', metavar='str', help='Input text or file') ①
17
18     parser.add_argument('-v',
19                         '--vowel',
20                         help='The vowel(s) allowed',
21                         metavar='str',
22                         type=str,
23                         default='a',
24                         choices=list('aeiou')) ②
25
26     args = parser.parse_args()
27
28     if os.path.isfile(args.text): ③
29         args.text = open(args.text).read().rstrip() ④
30
31     return args
32
33
34 # -----
35 def main():
36     """Make a jazz noise here"""
37
38     args = get_args()
39     text = args.text
40     vowel = args.vowel
41
42     # Method 1: Iterate every character
43     new_text = [] ⑤
44     for char in text:
45         if char in 'aeiou': ⑥
46             new_text.append(vowel) ⑦
47         elif char in 'AEIOU': ⑧
48             new_text.append(vowel) ⑨
```

```
48             new_text.append(vowel.upper()) ⑩
49     else:
50         new_text.append(char)           ⑪
51
52     print(''.join(new_text))        ⑫
53
54
55 # -----
56 if __name__ == '__main__':
57     main()
```

- ① The input might be text or a file name, so define as a string.
- ② Use the `choices` to restrict the user to one of the listed vowels.
- ③ Check if the `text` argument is a file.
- ④ If it is, read the file, using the `str.rstrip` to remove any trailing whitespace.
- ⑤ Create a new list to hold the characters we'll select.
- ⑥ Iterate through each character of the text.
- ⑦ See if the current character is in the list of lowercase vowels.
- ⑧ If it is, use the `vowel` instead of the character.
- ⑨ See if the current character is in the list of uppercase vowels.
- ⑩ If it is, use the `vowel.upper` instead of the character.
- ⑪ Otherwise, take the character itself.
- ⑫ Print a new string made by joining the `new_text` list on the empty string.

Discussion

I came up with eight ways to write my solution. All of them have the same `get_args`, so let's look at that first.

Defining the parameters

This is one of those problems that has many valid and interesting solutions. The first problem to solve is, of course, getting and validating the user's input. As always, I will use `argparse`. I usually define all my required parameters first. The `text` parameter is a positional string that *might* be a file name:

```
parser.add_argument('text', metavar='str', help='Input text or file')
```

The `--vowel` option is also a string, and I decided to use the `choices` option to have `argparse` validate that the user's input is in the `list('aeiou')`:

```
parser.add_argument('-v',
                    '--vowel',
                    help='The vowel to substitute',
                    metavar='str',
                    type=str,
                    default='a',
                    choices=list('aeiou'))
```

That is, `choices` wants a `list` of options. I could pass in `['a', 'e', 'i', 'o', 'u']`, but that's a lot of typing on my part. It's much easier to type `list('aeiou')` to have Python turn the `str` "aeiou" into a `list` of the characters. Both of these are the same:

```
>>> ['a', 'e', 'i', 'o', 'u']
['a', 'e', 'i', 'o', 'u']
>>> list('aeiou')
['a', 'e', 'i', 'o', 'u']
```

We can even write a test for this. No `AssertionError` error means that it's OK:

```
>>> assert ['a', 'e', 'i', 'o', 'u'] == list('aeiou')
```

```
open(args.text).read().rstrip()
  ↓
  fh.read().rstrip()
  ↓
str ← str.rstrip()
```

The next task is detecting if `text` is the name of a file that should be read for the text or is the text itself. This is the same code I used in "Howler," and again I choose to handle the `text` argument inside the `get_args` function so that, by the time I get `text` inside my `main`, it's all be handled:

```
if os.path.isfile(args.text):
    args.text = open(args.text).read().rstrip()
```

At this point, the user's arguments to the program have been fully vetted. We've got `text` either from the command line or from a file, and we've verified that the `--vowel` is actually one of the allowed characters. To me, this is a single "unit" where I've handled the arguments, and processing can now go forward by returning the arguments:

```
return args
```

Eight ways to replace the vowels

How many ways did you find to replace the vowels? You only needed one, of course, to pass the test, but I hope you probed the edges of the language to see how many different techniques there are. I know that the Zen of Python says:

There should be one — and preferably only one — obvious way to do it. -

<https://www.python.org/dev/peps/pep-0020/>

But I really come from the Perl mentality that "There Is More Than One Way To Do It" (TMTOWTDI or "Tim Toady").

Method 1: Iterate every character

The first method is similar to "Jump The Five" where we can use a `for` loop on a string to access each character. Here is code you can copy and paste into the `ipython` REPL

```
>>> text = 'Apples and Bananas!'          ①
>>> vowel = 'o'                          ②
>>> new_text = []                         ③
>>> for char in text:                   ④
...     if char in 'aeiou':              ⑤
...         new_text.append(vowel)        ⑥
...     elif char in 'AEIOU':            ⑦
...         new_text.append(vowel.upper()) ⑧
...     else:                           ⑨
...         new_text.append(char)
...
>>> text = ''.join(new_text)             ⑩
>>> text
'Opplos ond Bononos!'
```

- ① Set a `text` variable to the string "Apples and Bananas!"
- ② Set the `vowel` variable to the string "o."
- ③ Set the variable `new_text` to an empty `list`.
- ④ Use a `for` to iterate `text`, putting each character into the `char` variable.
- ⑤ If the character is in the set of lowercase vowels,
- ⑥ Substitute in the `vowel` to the `new_text`.
- ⑦ If the character is in the set of uppercase vowels,
- ⑧ Substitute the `vowel.upper()` version into `new_text`.
- ⑨ Otherwise, use the `char` as-is.
- ⑩ Turn the `list` called `new_text` into new `str` by joining it on the empty string ('').

Note that it would be just fine to start off making `new_text` an empty string and then concatenating

the new characters. Then you wouldn't have to `join` them at the end. Whatever you prefer:

```
new_text += vowel
```

Method 2: Using `str.replace`

For the alternate solutions, I'll just show the `main` function:

```
def main():
    args = get_args()
    text = args.text
    vowel = args.vowel

    for v in 'aeiou': ①
        text = text.replace(v, vowel).replace(v.upper(), vowel.upper()) ②

    print(text)
```

- ① Iterate through the list of vowels. We don't have to say `list('aeiou')` here because Python will automatically treat the string `'aeiou'` like a `list` because we are using it in a *list context* with the `for`.
- ② Use the `str.replace` method twice to replace both the lower- and upper-case versions of the vowel in the `text`.

I mentioned in the introduction the `str.replace` method that will return a new string with all instances of one string replaced by another.

```
>>> s = 'foo'
>>> s.replace('o', 'a')
'faa'
>>> s.replace('oo', 'x')
'fx'
```

Note that the original string remains unchanged:

```
>>> s
'foo'
```

You don't have to chain the two `replace` methods. It could be written as two separate statements:

```

text = text.replace(v, vowel).replace(v.upper(), vowel.upper())
      ↴           ↴           ↴
      |           |           |
text = text.replace(v, vowel)
text = text.replace(v.upper(), vowel.upper())

```

Method 3: Using `str.translate`

```

def main():
    args = get_args()
    vowel = args.vowel
    trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5) ①
    text = args.text.translate(trans) ②

    print(text)

```

① Create the translation table.

② Call the `str.translate` method on the `text` variable passing the `trans` table as the argument.

How can we use `str.translate` method to solve this? I showed in "Jump The Five" how the `jumper` dictionary could be used to create a translation table using the `str.maketrans` method to convert each number to another number. In this problem I need to change all the lower- and upper-case vowels (10 total) to some given `vowel`. For instance, to make all the vowels into the letter "o," I could create translation table `t` like so:

```

t = {'a': 'o',
     'e': 'o',
     'i': 'o',
     'o': 'o',
     'u': 'o',
     'A': 'O',
     'E': 'O',
     'I': 'O',
     'O': 'O',
     'U': 'O'}

```

I can use `t` with the `str.translate` method:

```

>>> 'Apples and Bananas'.translate(str.maketrans(t))
'Opplos ond Bononos'

```

If you read the documentation for `str.maketrans`, you will find that another way to make the translation is to supply two strings of equal lengths:

```
maketrans(x, y=None, z=None, /)
    Return a translation table usable for str.translate().
```

If there is only one argument, it must be a dictionary mapping Unicode ordinals (integers) or characters to Unicode ordinals, strings or None. Character keys will be then converted to ordinals.

If there are two arguments, they must be strings of equal length, and in the resulting dictionary, each character in x will be mapped to the character at the same position in y. If there is a third argument, it must be a string, whose characters will be mapped to None in the result.

The first string should contain the letters I want to replace which are the lower- and uppercase vowels '**aeiouAEIOU**'. The second string is composed of the letters to use for substitution. I would like to use '**ooooo**' for the '**aeiou**' and '**OOOOO**' for '**AEIOU**'. I can repeat my **vowel** five times using the ***** operator that you normally associate with numeric multiplication. This is (sort of) "multiplying" a string, so, OK, I guess:

```
>>> vowel * 5
'ooooo'
```

Now to handle the uppercase version, too:

```
>>> vowel * 5 + vowel.upper() * 5
'ooooooOOOOO'
```

Now to make the translation table:

```
>>> trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5)
```

Let's inspect the **trans** table. I want to "pretty print" the data structure so I can see it, so I will use the `pprint.pprint` (pretty print) function:

```
>>> from pprint import pprint as pp
>>> pp(trans)
{65: 79,
 69: 79,
 73: 79,
 79: 79,
 85: 79,
 97: 111,
 101: 111,
 105: 111,
 111: 111,
 117: 111}
```

The enclosing curly braces {} tell us that `trans` is a `dict`. Each character is represented by its *ordinal* value, which is the character's position in the ASCII table (<http://www.asciitable.com/>). (You will use this later in the "Gematria" program.) You can go back and forth from characters and their ordinal values by using `chr` and `ord`. Here are the `ord` values for the vowels:

```
>>> for char in 'aeiou':  
...     print(char, ord(char))  
...  
a 97  
e 101  
i 105  
o 111  
u 117
```

And here you can create the same output but starting with the `ord` values to get the `chr` values:

```
>>> for num in [97, 101, 105, 111, 117]:  
...     print(chr(num), num)  
...  
a 97  
e 101  
i 105  
o 111  
u 117  
>>>
```

If you'd like to inspect all the ordinal values for all the printable characters, you can run this:

```
>>> import string  
>>> for char in string.printable:  
...     print(char, ord(char))
```

I don't include the output because there are 100 printable characters:

```
>>> print(len(string.printable))  
100
```

So the `trans` table is a mapping. The lowercase vowels ("aeiou") all map to the ordinal value 111 which is "o." The uppercase vowels ("AEIOU") map to 79 which is "O." I can use the `dict.items` method to iterate over the key/value pairs of `trans` to verify this is the case:

```

>>> for x, y in trans.items():
...     print(f'{chr(x)} => {chr(y)}')
...
a => o
e => o
i => o
o => o
u => o
A => O
E => O
I => O
O => O
U => O

```

The original `text` will be unchanged by the `translate` method, so we overwrite `text` with the new version:

```

>>> text = 'Apples and Bananas!' ①
>>> trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5) ②
>>> text = text.translate(trans) ③
>>> text
'Opplos ond Bononos!'

```

① Initialize `text`.

② Make a translation table.

③ Use the translation table as the argument to `text.translate`. Overwrite the original value of `text` with the result.

That was a lot of explanation about `ord` and `chr` and dictionaries and such, but look how simple and elegant that solution is! This is much shorter than Method 1. Fewer lines of code (LOC) means fewer opportunities for bugs!

Method 4: List comprehension

```

def main():
    args = get_args()
    vowel = args.vowel
    text = [ ①
        vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c ②
        for c in args.text
    ]
    print(''.join(text)) ③

```

① Use a list comprehension to process all the characters in `args.text` to create a new `list` called `text`.

② Use a compound `if` expression to handle three cases (lowercase vowel, uppercase vowel, the

default).

③ Print a new string by joining `text` on the empty string.

Following up on Method 1, we can use a "list comprehension" to significantly shorten the `for` loop. In Gashlycrumb we looked at a "dictionary comprehension" as a one-line method to create a new dictionary using a `for` loop. Here we can do the same, creating a new `list`.

For example, let's generate a list of the squared values of the numbers 1 through 4. We can use the `range` function to get the numbers from a starting number to an ending number (not inclusive!). `range` is a *lazy* function, which means it won't actually produce values until your program actually needs them. In the REPL, I must use `list` to see the values, but most of the time your code doesn't need to do this:

```
>>> list(range(1, 5))
[1, 2, 3, 4]
```

I can write a `for` loop to `print` the squares:

```
>>> for num in range(1, 5):
...     print(num ** 2)
...
1
4
9
16
```

But what I really want is a `list` that contains those values. A simple way to do this is to create an empty `list` to which we will `append` those values:

```
>>> squares = []
>>> for num in range(1, 5):
...     squares.append(num ** 2)
```

And now I can verify that I have my squares:

```
>>> assert len(squares) == 4
>>> assert squares == [1, 4, 9, 16]
```

do this **for these things**

————— [num ** 2 for num in range(1, 5)]

new list ← [num ** 2 for num in range(1, 5)] A list comprehension is novel from a `for` loop in that it *returns a new list*:

```
>>> [num ** 2 for num in range(1, 5)]  
[1, 4, 9, 16]
```

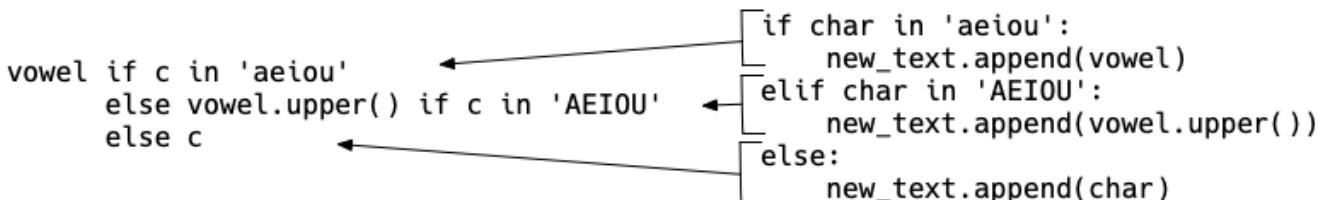
I can assign this to the `squares` variable and verify that I still have what I expected. Ask yourself which version of the code you'd rather maintain, the longer one with the `for` loop or the shorter one with the list comprehension?

```
>>> squares = [num ** 2 for num in range(1, 5)]  
>>> assert len(squares) == 4  
>>> assert squares == [1, 4, 9, 16]
```

For our program, we're going to condense the `if/elif/else` logic from Method 1 into a compound `if` expression. First let's see how we could shorten the `for` loop version:

```
>>> text = 'Apples and Bananas!'  
>>> new_text = []  
>>> for c in text:  
...     new_text.append(vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else  
c)  
...  
>>> ''.join(new_text)  
'Opplos ond Bononos!'
```

Here's a diagram that shows how the parts of the expression match up to the original `if/elif/else`:



And now to turn that into a list comprehension:

```
>>> text = 'Apples and Bananas!'  
>>> new_text = [  
...     vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c ①  
...     for c in text ] ②  
...  
>>> ''.join(new_text)  
'Opplos ond Bononos!'
```

① Do this

② For these things.

Method 5: List comprehension with function

The compound `if` expression inside the list comprehension borders is complicated enough that it probably should be a function. We can *define* a new function with the `def` statement and call it `new_char`. It accepts a character we'll call `c`. After that, it's the same compound `if` expression as before:

```
def main():
    args = get_args()
    vowel = args.vowel

    def new_char(c): ①
        return vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c ②

    text = ''.join([new_char(c) for c in args.text]) ③

    print(text)
```

① Define a function to choose a new character. Note that it uses the `vowel` variable because the function has been declared in the same scope. This is called a "closure" because `new_char` *closes over* the variable.

② Use the compound `if` expression to select the correct character.

③ Use a list comprehension to process all the characters in `text`.

You can play with the `new_char` function by putting this into your REPL:

```
vowel = 'o'
def new_char(c):
    return vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c
```

It should always return the letter "o" if the argument is a lowercase vowel:

```
>>> new_char('a')
'o'
```

And "O" if the argument is an uppercase vowel:

```
>>> new_char('A')
'0'
```

Otherwise it should return the given character:

```
>>> new_char('b')
'b'
```

We can use the `new_char` function to process all the characters in `text` using a list comprehension:

```
>>> text = 'Apples and Bananas!'
>>> text = ''.join([new_char(c) for c in text])
>>> text
'Opplos ond Bononos!'
```

A note about the fact that the `new_char` function is declared *inside* the `main` function. Yes, you can do that! The function is then only "visible" inside the `main` function. Here I define a `foo` function that has a `bar` function inside it. I can call `foo` and it will call `bar`, but from outside of `foo` the `bar` function does not exist ("is not visible" or "is not in scope"):

```
>>> def foo():
...     def bar():
...         print('This is bar')
...     bar()
...
>>> foo()
This is bar
>>> bar()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'bar' is not defined
```

I did this because I actually created a special type of function with `new_char` called a "closure" because it is "closing" around the `vowel`. If I had defined `new_char` outside of `main`, the `vowel` would not be visible to `new_char` because it only exists inside the `main` function. If we draw a box around the `main` function, that is where the `new_char` function can be seen. Outside of that, the function is considered undefined.

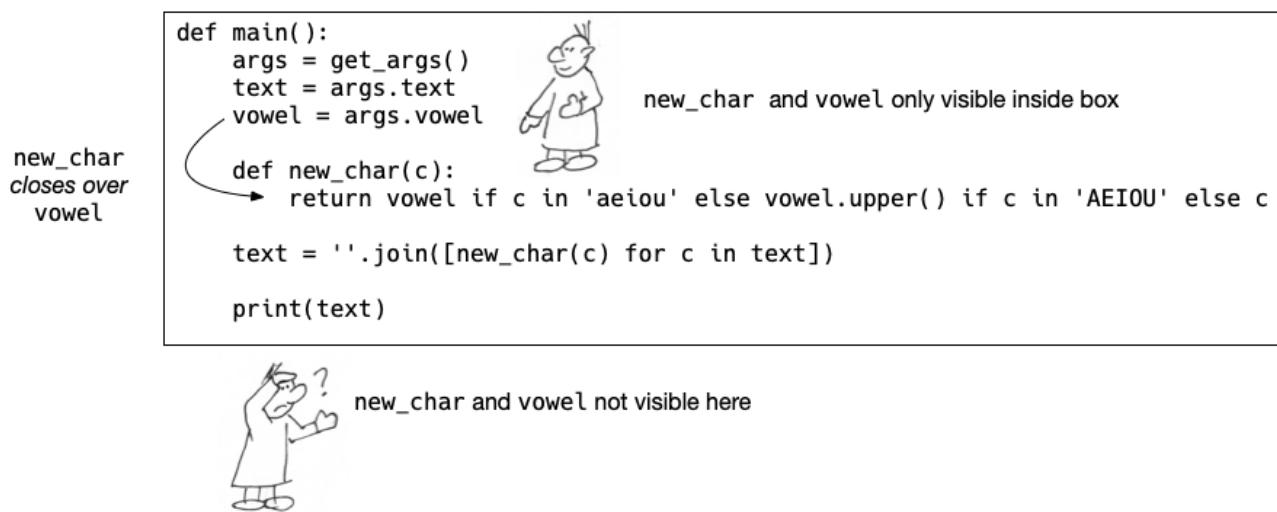


Figure 5. Visibility of the `new_char` function

There are many reasons why to write a function like this, even if it has just one line. I think of functions as *units* of code that describe some concept—ideally just *one* per function! We can formalize our understanding of our functions with assertions:

```
>>> assert all([new_char(v) == '0' for v in 'AEIOU'])
>>> assert all([new_char(v) == 'o' for v in 'aeiou'])
```

In two lines of code, I've just tested all 10 vowels by using list comprehensions plus the `all` function. Let's take a moment to understand `all` and `any`. If you read `help(all)`, you'll see "Return True if `bool(x)` is True for all values `x` in the iterable." So all the values need to be `True` for the entire expression to be `True`:

```
>>> all([True, True, True])
True
```

If any value is `False`, then the whole expression is `False`:

```
>>> all([True, False, True])
False
```

The `any` function returns `True` if *any* of the values are `True`:

```
>>> any([True, False, True])
True
```

And `False` if there are no `True` values:

```
>>> any([False, False, False])
False
```

We can check the values manually:

```
>>> [new_char(v) == '0' for v in 'AEIOU']
[True, True, True, True, True]
```

So then `all` should be `True` for the entire expression:

```
>>> all([new_char(v) == '0' for v in 'AEIOU'])
True
```

We can move these into formal `test_` functions that can be run with `pytest`. This is the idea of *unit testing* where I personally think of functions as units. Some people have other ideas about what is the best *unit* to test. I recommend you read further about testing to understand the range of opinions.

Method 6: The `map` function

For this next method, I want to introduce the `map` function as it's quite similar to a list comprehension. The `map` function accepts two arguments:

1. A function
2. An iterable like a list, lazy function, or generator.



I like to think of `map` like a paint booth: You load up the booth with, say, blue paint. Unpainted cars go in, blue paint is applied, and blue cars come out. I can create a function to "paint" my cars by adding the string "blue" to the beginning:

```
>>> list(map(lambda car: 'blue ' + car, ['BMW', 'Alfa Romeo', 'Chrysler']))  
['blue BMW', 'blue Alfa Romeo', 'blue Chrysler']
```



The first argument you see here starts with the keyword `Lambda` which is used to create an *anonymous* function. That is, with the regular `def` keyword, the function name follows. With `Lambda`, there is no name, only the list of parameters and the function body.

Think about regular named functions like `add1` that adds 1 to a value:

```
def add1(n):  
    return n + 1
```

It works as expected:

```
>>> assert add1(10) == 11  
>>> assert add1(add1(10)) == 12
```

Now compare to the `Lambda` version. We can assign it to the variable `add1` which then kinda sorta acts like the function's name:

```
>>> add1 = lambda n: n + 1
```

function name parameter(s) in parentheses

```
define → def add1(n):
    return n + 1 ← function body
      ↑
      return
```

like function body
"def" (no "return")

```
like
function → add1 = lambda n: n + 1
      ↑
      parameter(s)
      no parentheses
```

It works exactly the same as using `def` to define a function the normal way:

```
>>> assert add1(10) == 11
>>> assert add1(add1(10)) == 12
```

The function body for a `lambda` pretty much needs to fit on one line, and they don't have `return` at the end. In both versions, the argument to the function is `n`. In the usual `def add(n)`, the argument is defined in the parentheses just after the function name. In the `lambda n` version, there is no function name and no parentheses around the function's parameter, `n`.

There is no difference in how you can use them. They are both functions:

```
>>> type(lambda x: x)
<class 'function'>
```

If you are comfortable with using `add1` in a list comprehension:

```
>>> [add1(n) for n in [1, 2, 3]]
[2, 3, 4]
```

Then it's a short step to using the `map` function. Note that `map` is a lazy function that won't actually create the values until you actually need them. To see the results in the REPL, I need to use the `list` function to convince `map` to produce the `list` of results:

```
>>> list(map(add1, [1, 2, 3]))  
[2, 3, 4]
```

We can write the list comprehension with the `add1` code in-line:

```
>>> [n + 1 for n in [1, 2, 3]]  
[2, 3, 4]
```



```
map(lambda n: n + 1, [1, 2, 3])  
[2, 3, 4]
```

Which looks very similar to the `lambda` code:

```
>>> list(map(lambda n: n + 1, [1, 2, 3]))  
[2, 3, 4]
```

So here is how I could use a `map`:

```
def main():  
    args = get_args()  
    vowel = args.vowel  
    text = map(  
        lambda c: vowel if c in 'aeiou' else vowel.upper() ①  
        if c in 'AEIOU' else c, args.text) ②  
  
    print(''.join(text)) ④
```

- ① The `map` function wants a function for the first argument and a `list` for the second.
- ② Use `lambda` to create an *anonymous* function that accepts character, `c`.
- ③ `args.text` is the second argument to `map`. Because `map` expects this argument to be a `list`, it will automatically coerce it to a `list` (which is what we want).
- ④ The `map` returned a new `list` into the `text` variable, so we join it on the empty string to print it.

Higher-order functions

The `map` function is called a "higher-order function" (HOF) because it takes *another function* as an argument, which is wicked cool. Later we'll use another HOF called `filter`.

Method 7: Using `map` with a defined function

We are not required to use a `lambda` expression with `map`. Any function at all will work, so let's go back to using our `new_char` function:

```
def main():
    args = get_args()
    vowel = args.vowel

    def new_char(c): ①
        return vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c

    print(''.join(map(new_char, args.text))) ②
```

① Define a function that will return the proper character.

② Use `map` to apply this *named* function to all the characters in `text`. The result is a `list` which is joined into a new string to overwrite `text`.

Notice that `map` takes `new_char` *without parentheses* as the first argument. If you added the parens, you'd be *calling* the function and would see this error:

```
>>> text = ''.join(map(new_char(), text))
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: new_char() missing 1 required positional argument: 'c'
```

The diagram illustrates the mapping process. At the top, the code `map(new_char, 'Apples')` is shown. Below it, a curved arrow points from the word "Apples" down to a list of characters: `['A', 'p', 'p', 'l', 'e', 's']`. Another curved arrow points from the `map` function down to the list, indicating that `map` takes each character from `text` and passes it as an argument to the `new_char` function.

`map` takes each character from `text` and passes it as the argument to the `new_char` function which decides whether to return the `vowel` or the original character. The result of mapping these characters is a new list of characters that we `join` on the empty string to create a new version of `text`.

Method 8: Using regular expressions

A "regular expression" is a term from linguistics and computer science. It is a way of describing a "regular language" [2]. Essentially regular expressions give us a way to *describe patterns of text*, which is truly, outstandingly useful.

To use regular expressions, you must `import re` in your code. Here I will show how we can use the "substitute" function `re.sub` to find any of the lower- and uppercase vowels and replace them with the given `vowel`:

```
def main():
    args = get_args()
    text = args.text
    vowel = args.vowel
    text = re.sub('[aeiou]', vowel, text)           ①
    text = re.sub('[AEIOU]', vowel.upper(), text)  ②
    print(text)
```

- ① Substitute any of the lowercase vowels for the given vowel (which is lowercase because of the restrictions in `get_args`).
② Substitute any of the uppercase vowels for the uppercased vowel.

Regular expressions which are a separate domain-specific language (DSL). That is, they really have nothing whatsoever to do with Python. They have their own syntax and rules, and they are used in many places from command-line tools to databases. Regexes are incredibly powerful and well worth the effort to learn them.

The `re.sub` function will *substitute* all instances of text matching a given pattern for a new string. The square brackets around the vowels '`[aeiou]`' create a "character class," meaning anything matching one of the characters listed inside the brackets. The second argument is the string that will replace the found strings—here our `vowel` provided by the user. The third argument is the string we want to change, which is the `text` from the user.

```
>>> import re
>>> text = 'Apples and Bananas!'
>>> vowel = 'o'
>>> re.sub('[aeiou]', vowel, text)
'Applos ond Bononos!'
```

Note that `re.sub` returns a new string, and the original `text` remains unchanged by the operation:

```
>>> text
'Apples and Bananas!'
```

That almost worked, but it missed the uppercase vowel "A". I could overwrite the `text` in two steps to get both lower- and uppercase:

```
>>> text = re.sub('[aeiou]', vowel, text)
>>> text = re.sub('[AEIOU]', vowel.upper(), text)
>>> text
'Opplos ond Bononos!'
```

Or do it in one step just like the `str.replace` method above:

```
>>> text = 'Apples and Bananas!'
>>> text = re.sub('[AEIOU]', vowel.upper(), re.sub('[aeiou]', vowel, text))
>>> text
'Opplos ond Bononos!'
```

One of the biggest differences with this solution to all the others is how we use regular expressions to describe what we were looking for and didn't have to write the code to actually find the characters! This is more along the lines of *declarative* programming. We declare what we want, and the computer does the grunt work!

[2] https://en.wikipedia.org/wiki/Regular_language

Refactoring with tests

There are many ways to solve this problem. The most important step is to get your program to work properly. Tests let you know when you've reached that point. From there, you can explore other ways to solve the problem and keep using the tests to ensure you still have a correct program. Tests actually provide great freedom to be creative. Always be thinking about tests you can write for your own programs so that, when you change them later, they will always keep working!

Review

- You can use `argparse` to limit an argument's values to a `list` of `choices` that you define.
- Strings cannot be directly modified, but the `str.replace` and `str.translate` methods can create a *new, modified string* from an existing string.
- A `for` loop on a string will iterate the characters of the string.
- A list comprehension is a short-hand way to write a `for` loop inside `[]` to create a new `list`.
- Functions can be defined inside other functions. Their visibility is then limited to the enclosing function. * Function can reference variables declared within the same scope creating a closure.
- The `map` function is similar to a list comprehension. It will create a new, modified list by applying some function to every member of a given list. The original list will not be changed.
- Regular expressions provide a syntax for describing patterns of text with the `re` module. The `re.sub` method will substitute found patterns with new text. The original text will be unchanged.

Going Further

- Write a version that collapses multiple adjacent vowels into a single substituted value. For example, "quick" should become "qack" and not "qaack." = Chapter 9: Dial-A-Curse: Generating random insults from lists of words

"He or she is a slimy-sided, frog-mouthed, silt-eating slug with the brains of a turtle." — Dial-A-Curse



Random events are at the heart of interesting games and puzzles. Humans quickly grow bored of things that are always the same, so let's learn how to make our programs more interesting by having them behave differently each time they are run. This exercise will introduce how to randomly select one or more elements from lists of options. To explore randomness, we'll create a program called `abuse.py` that will insult the user by randomly selecting adjectives and nouns to create slanderous epithets.

In order to test randomness, though, we need to control it. It turns out that "random" events on computers are rarely actually random but only "pseudo-random," which means we can control them using a "seed."^[3] Each time you use the same seed, you get the same "random" choices!

Shakespeare had some of the best insults, so we'll draw from the vocabulary of his works. Here is the list of adjectives you should use:

bankrupt base caterwauling corrupt cullionly detestable dishonest false filthsome filthy foolish foul gross heedless indistinguishable infected insatiate irksome lascivious lecherous loathsome lubbery old peevish rascaly rotten ruinous scurilous scurvy slanderous sodden-witted thin-faced toad-spotted unmannered vile wall-eyed

And these are the nouns:

Judas Satan ape ass barbermanger beggar block boy braggart butt carbuncle coward coxcomb cur dandy degenerate fiend fishmonger fool gull harpy jack jolthead knave liar lunatic maw milksop minion ratcatcher recreant rogue scold slave swine traitor varlet villain worm

For instance, it might produce the following:

```
$ ./abuse.py  
You slanderous, rotten block!  
You lubbery, scurilous ratcatcher!  
You rotten, foul liar!
```

In this exercise, you will learn to:

- Use `parser.error` from `argparse` to throw errors
- Learn about random seeds to control randomness
- Take random choices and samples from Python lists
- Iterate an algorithm a specified number of times with a `for` loop
- Format output strings

[3] "The generation of random numbers is too important to be left to chance." — Robert R. Coveyou

Writing abuse.py

The arguments to this program are options that have default values, meaning it can run with no arguments at all. The `-n` or `--number` option will default to `3` and will determine how many insults are created:

```
$ ./abuse.py --number 2  
You filthsome, cullionly fiend!  
You false, thin-faced minion!
```

And the `-a` or `--adjectives` option should default to `2` and will determine how many adjectives are used in each insult:

```
$ ./abuse.py --adjectives 3  
You caterwauling, heedless, gross coxcomb!  
You sodden-witted, rascaly, lascivious varlet!  
You dishonest, lecherous, foolish varlet!
```

Lastly, your program should accept a `-s` or `--seed` argument (default `None`) that will control the randomness of the program. The following should be exactly reproducible, no matter who runs the program on any machine at any time:

```
$ ./abuse.py --seed 1  
You filthsome, cullionly fiend!  
You false, thin-faced minion!  
You sodden-witted, rascaly cur!
```



When run with no arguments, the program should generate insults using the defaults:

```
$ ./abuse.py  
You foul, false varlet!  
You filthy, insatiate fool!  
You lascivious, corrupt recreant!
```

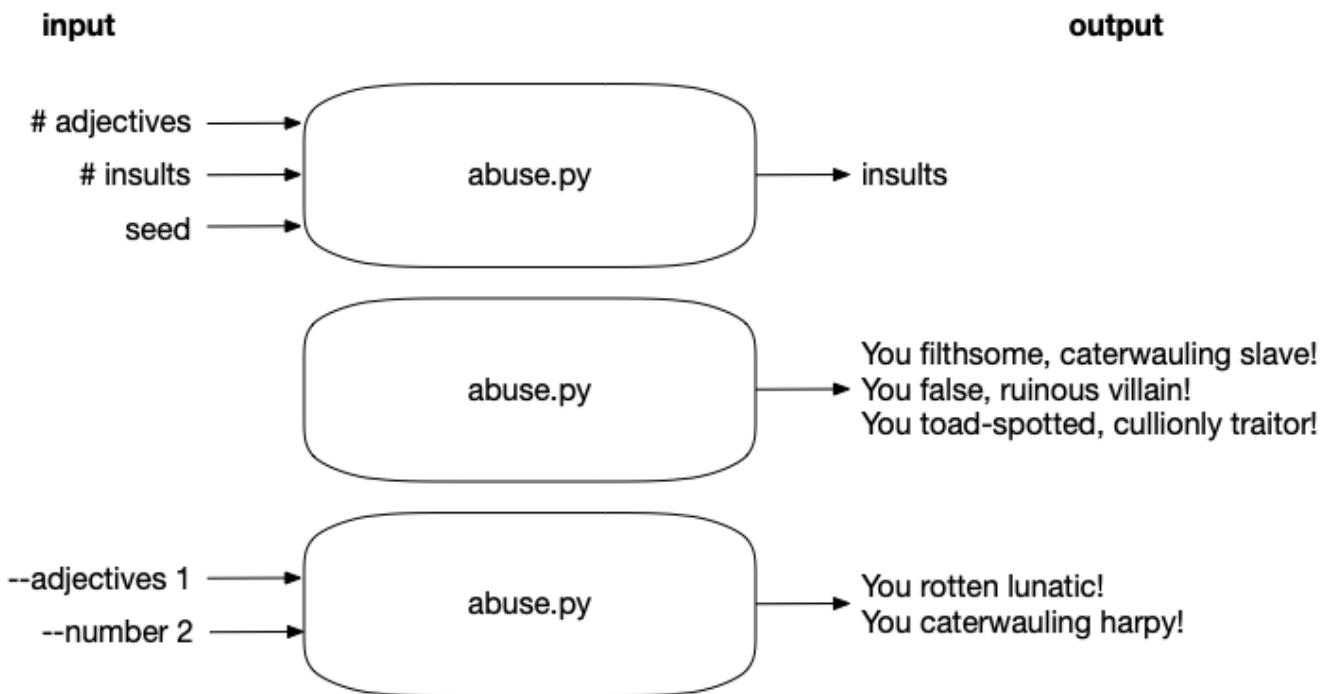
I recommend you start by copying the `template/template.py` to `abuse/abuse.py` or by using `new.py` to create the `abuse.py` program in the `abuse` directory of your repository. The next step is to make your program produce the following usage for `-h` or `--help`:

```
$ ./abuse.py -h
usage: abuse.py [-h] [-a int] [-n int] [-s int]

Heap abuse

optional arguments:
  -h, --help            show this help message and exit
  -a int, --adjectives int
                        Number of adjectives (default: 2)
  -n int, --number int  Number of insults (default: 3)
  -s int, --seed int    Random seed (default: None)
```

Here is a string diagram to help you see the program:



Validating arguments

All of the options for number of insults and adjectives as well as the random seed should all be `int` values. If you define each using `type=int` (remember there are no quotes around the `int`), then `argparse` will handle the validation and conversion of the arguments to an `int` value for you. That is, just by defining `type=int`, the following error will be generated for you:

```
$ ./abuse.py -n foo
usage: abuse.py [-h] [-a int] [-n int] [-s int]
abuse.py: error: argument -n/--number: invalid int value: 'foo'
```

Additionally, if either `--number` or `--adjectives` less than 1, your program should exit with an error code and message:

```
$ ./abuse.py -a -4
usage: abuse.py [-h] [-a int] [-n int] [-s int]
abuse.py: error: --adjectives "-4" must be > 1
$ ./abuse.py -n -4
usage: abuse.py [-h] [-a int] [-n int] [-s int]
abuse.py: error: --number "-4" must be > 1
```

Let's take a look at one of the tests in `test.py` to understand what is expected:

```
def test_bad_adjective_num():  
    """bad_adjectives"""\n\n    n = random.choice(range(-10, 0))  
    rv, out = getstatusoutput(f'{prg} -a {n}')  
    assert rv != 0  
    assert re.search(f"--adjectives \"{n}\" must be > 0", out)
```

- ① The name of the function must start with `test_` in order for `pytest` to find and run it.
- ② Use the `random.choice` function to randomly select a value from the `range` of numbers from -10 to 0. We will use this same function in our program, so note here how it is called!
- ③ Run the program with the bad `-a` argument, getting back a return value (`rv`) and the standard out (`out`).
- ④ Assert that the return value is not 0 where "0" would indicate success (or "zero errors").
- ⑤ Assert that the output somewhere contains the statement that the `--adjectives` argument must be greater than 0.

I would recommend you look at the section on "Manually checking arguments" in chapter 1. There I introduce the `parser.error` function that you can call inside the `get_args` function to do the following:

1. Print the short usage statement.
2. Print an error message to the user.
3. Stop execution of the program.
4. Exit with a non-zero exit value to indicate an error.

That is, your `get_args` normally finishes with:

```
return args.parse_args()
```

Instead, put the `args` into a variable and check the `args.adjectives` value to see if it's less than 1. If it is, call `parser.error` with an error message to report to the user:

```
args = parser.parse_args()

if args.adjectives < 1:
    parser.error(f"--adjectives '{args.adjectives}' must be > 0')
```

Also do this for the `args.number`. If they are both fine, then you can `return` the arguments to the calling function:

```
return args
```

Importing and seeding the `random` module

Once you have defined and validated all the program's arguments, you are ready to heap scorn upon the user. We need to add `import random` to our program so we can use functions from that module to select adjectives and nouns. I recommend you add it just after the `import argparse` statement near the top of your program.

As usual first thing we need to do in the `main` is to call `get_args` to get our arguments, and the very next step is to pass the `args.seed` value to the `random.seed` function:

```
def main():
    args = get_args()
    random.seed(args.seed)
```

We can read about the `random.seed` function in the REPL:

```
>>> import random
>>> help(random.seed)
```

There we learn that the function will "[i]nitialize internal state from hashable object." That is, we set an initial value from some *hashable* Python type. Both `int` and `str` types are hashable, and I suggest you define the `seed` argument as an `int`. The default value for `args.seed` is `None`, so, if the user has not indicated any seed, this is the same as not setting it at all.

If you look at the `test.py` program, you will notice that all the tests that expect a particular output will pass a `-s` or `--seed` argument. Here is the first test for output:

```
def test_01():
    out = getoutput(f'{prg} -s 1 -n 1') ①
    assert out.strip() == 'You filthsome, cullionly fiend!' ②
```

- ① Run the program with the seed value 1, requesting 1 insult.
- ② Verify that the entire output is the one expected insult.

Defining the adjectives and nouns

Above I've given you a long list of adjectives and nouns that you should use in your program. In order to pass the tests, they must be in the same order as I have provided. (You may notice that they are alphabetically sorted.) You could create a `list` by individually quoting each word:

```
>>> adjectives = ['bankrupt', 'base', 'caterwauling']
```

Or you could save yourself a good bit of typing if you use the `str.split` function to create a new `list` from a `str` by splitting on spaces:

```
>>> adjectives = 'bankrupt base caterwauling'.split()  
>>> adjectives  
['bankrupt', 'base', 'caterwauling']
```

If you try to make one giant string of all the adjectives, it will wrap around and look ugly. I'd recommend you use triple quotes (either single or double quotes) that allow you to include newlines:

```
>>> """  
... bankrupt base  
... caterwauling  
... """.split()  
['bankrupt', 'base', 'caterwauling']
```

Once you have variables for `adjectives` and `nouns`, you might check that you have the right number:

```
>>> assert len(adjectives) == 36  
>>> assert len(nouns) == 39
```

Taking random samples and choices



harpy
jack
jolthead
knave
liar

In addition to the `random` module's `seed` function, we will also use the `choice` and `sample` functions. In the `test_bad_adjective_num` function above, we saw one example of using `random.choice`. We can use it similarly to select a noun from the `list` of `nouns`. Notice that this function returns a single item, so, given a `list` of `str` values, it will return a single `str`:

```
>>> random.choice(nouns)
'braggart'
>>> random.choice(nouns)
'milksop'
```

For the `adjectives`, you should use `random.sample`. If you read the `help(random.sample)`, you will see this function takes the `list` of `adjectives` and a `k` parameter for how many items to sample:

```
sample(population, k) method of random.Random instance
    Chooses k unique random elements from a population sequence or set.
```

Note that this function returns a new `list`:

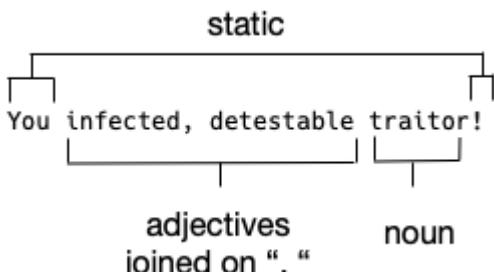
```
>>> random.sample(adjectives, 2)
['detestable', 'peevish']
>>> random.sample(adjectives, 3)
['slanderous', 'detestable', 'base']
```

There is also a `random.choices` that works similarly to `sample` but which might select the same items twice because it samples "with replacement," so we will not use that.

Formatting the output

The output of the program is some `--number` of insults, which you could generate using a `for` loop and the `range` function. It doesn't matter here that `range` starts at zero. What's important is that it generates three values:

```
>>> for n in range(3):
...     print(n)
...
0
1
2
```



You can loop the `--number` of times needed, select your sample of adjectives and your noun, and then format the output. Each insult should start with the string "You ", then have the adjectives joined on a comma and a space, then the noun, and finish with an exclamation point. You could use either an f-string or the `str.format` function to `print` the output to

STDOUT.

Hints:

- Perform the check for positive values for `--adjectives` `--number` *inside* the `get_args` function and use `parser.error` to throw the error while printing a message and the usage.
- If you set the default for `args.seed` to `None` while using a `type=int`, you should be able to directly pass the argument's value to `random.seed` to control testing.
- Use a `for` loop with the `range` function to create a loop that will execute `--number` of times to generate each insult.
- Look at the `sample` and `choice` functions in the `random` module for help in selecting some adjectives and a noun.
- You can use three single ('') or double quotes ("") to create a multi-line string and then `split()` that to get a list of strings. This is easier than individually quoting a long list of shorter strings (e.g., the list of adjectives and nouns).
- To construct an insult string to print, you can use the `+` operator to concatenate strings, use the `str.join` method, or use format strings.

Now give this your best shot before reading ahead to the solution!

Solution

```
1 #!/usr/bin/env python3
2 """Heap abuse"""
3
4 import argparse
5 import random
6
7
8 # -----
9 def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Heap abuse',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('-a',                                ②
17                         '--adjectives',
18                         help='Number of adjectives',
19                         metavar='int',
20                         type=int,
21                         default=2)
22
23     parser.add_argument('-n',                                ③
24                         '--number',
25                         help='Number of insults',
26                         metavar='int',
27                         type=int,
28                         default=3)
29
30     parser.add_argument('-s',                                ④
31                         '--seed',
32                         help='Random seed',
33                         metavar='int',
34                         type=int,
35                         default=None)
36
37     args = parser.parse_args()                            ⑤
38
39     if args.adjectives < 1:                            ⑥
40         parser.error('--adjectives "{}" must be > 0'.format(args.adjectives))
41
42     if args.number < 1:                                ⑦
43         parser.error('--number "{}" must be > 0'.format(args.number))
44
45     return args                                         ⑧
46
47
```

```

48 # -----
49 def main():
50     """Make a jazz noise here"""
51
52     args = get_args()                                     ⑨
53     random.seed(args.seed)                               ⑩
54
55     adjectives = """"                                    ⑪
56     bankrupt base caterwauling corrupt cullionly detestable dishonest false
57     filthsome filthy foolish foul gross heedless indistinguishable infected
58     insatiate irksome lascivious lecherous loathsome lubbery old peevish
59     rascaly rotten ruinous scurilous scurvy slanderous sodden-witted
60     thin-faced toad-spotted unmannered vile wall-eyed
61     """".strip().split()
62
63     nouns = """"                                       ⑫
64     Judas Satan ape ass barbemonger beggar block boy braggart butt
65     carbuncle coward coxcomb cur dandy degenerate fiend fishmonger fool
66     gull harpy jack jolthead knave liar lunatic maw milksop minion
67     ratcatcher recreant rogue scold slave swine traitor varlet villain worm
68     """".strip().split()
69
70     for _ in range(args.number):                      ⑬
71         adjs = ', '.join(random.sample(adjectives, k=args.adjectives)) ⑭
72         print(f'You {adjs} {random.choice(nouns)}!') ⑮
73
74
75 # -----
76 if __name__ == '__main__':
77     main()

```

- ① Bring in the `random` module so we can use functions.
- ② Define the parameter for the number of adjectives, setting the `type=int` and the default value.
- ③ Similarly define the parameter for the number insults.
- ④ The random seed default should be `None`.
- ⑤ Get the `args` from parsing the argument. Error such as non-integer values will be handled at this point by `argparse`.
- ⑥ If we make it to this point, we can perform additional checking such as verifying that the `args.adjectives` is greater than 0. If there is a problem, call `parser.error` with the error message to print.
- ⑦ Similarly check for the `args.number`.
- ⑧ If we make it to this point, then all the user's arguments have been validated.
- ⑨ This is where the program actually begins as it is the first action inside our `main`. We always start off by getting the arguments.
- ⑩ Set the `random.seed` using whatever value was passed by the user. Any `int` value is valid, and we know that `argparse` has handled the validation and conversion of the argument to an `int`.

- ⑪ Create a `list` of `adjectives` by splitting the very long string contained in the triple quotes.
- ⑫ Do the same for the `list` of `nouns`.
- ⑬ Use a `for` loop over the `range` of the `args.number`. Since we don't actually need the value from the `range`, we can use the `_` to disregard it.
- ⑭ Use the `random.sample` function to select the correct number of adjectives and join them on the comma-space `str`.
- ⑮ Use an f-string to format the output to `print`.

Discussion

Defining the arguments

More than half of my solution is just in defining the program's arguments to `argparse`. The effort is well worth the result, because `argparse` will ensure that each argument is a valid integer value because I set `type=int`. Notice there are no quotes around the `int`—it's not the string '`int`' but a reference to the class in Python.

```
parser.add_argument('-a',  
                    '--adjectives',  
                    help='Number of adjectives',  
                    metavar='int',  
                    type=int,  
                    default=2)
```

- ① The short flag.
- ② The long flag.
- ③ The help message.
- ④ A type hint for the user, note that this is a `str` value.
- ⑤ The actual Python `type` for converting the input, note that this is the bareword `int` for the integer class.
- ⑥ The default value for the number of adjectives per insult.

For `--adjectives` and `--number`, I can set reasonable defaults so that no input is required from the user. This makes your program dynamic, interesting, and testable. How do you know if your values are being used correctly unless you change them and test that the proper change was made in your program? Maybe you started off hardcoding the number of insults and forgot to change the `range` to use a variable. Without changing the input value and testing that the number of insults changed accordingly, it might be a user who discovers your bug, and that's somewhat embarrassing.

Using `parser.error`

I really love the `argparse` module for all the work it saves me. For instance, the `type=int` saves me all the trouble of verifying that the input is an integer, creating an error message, and then converting the user's input to an actual `int` value.

Something else I enjoy about `argparse` is that, if I find there is a problem with an argument, I can use `parser.error` to do four things:

1. Print the short usage of the program to the user
2. Print a specific message about the problem
3. Halt execution of the program
4. Return an error code to the operating system

I can't very easily tell `argparse` that the `--number` should be a positive integer, only that it must be of type `int`. I can, however, inspect the value myself and call `parser.error('message')` if there is a problem. I do all this inside `get_args` so that, by the time I call `args = get_args()` in my `main` function, I know that all the arguments have been validated.

I highly recommend you tuck this tip into your back pocket. It can prove quite handy, saving you loads of time validating user input and generating useful error messages. (And it's really quite likely that the future user of your program will be *you*, and you will really appreciate your efforts!)

Program exit values and `STDERR`

I would like to highlight the exit value of your program. Under normal circumstances, your program should exit with a value of `0`. In computer science, we often think of `0` as a `False` value, but here it's quite positive. In this instance we should think of it like "zero errors." If you use `sys.exit()` in your code to exit a program prematurely, the default exit value is `0`. If you want to indicate to the operating system or some calling program that your program exited with an error, you should return *any value other than 0*. The `parser.error` function does this for you automatically.

Additionally, it's common for all error messages to be printed not to `STDOUT` (standard out) but to `STDERR` (standard error). On Unix command line, we can segregate these two output channels using `1` for `STDOUT` and `2` for `STDERR`. Notice how I can use `2>err` to redirect `STDERR` to the file called `err` so that nothing appears on `STDOUT`:

```
$ ./abuse.py -a -1 2>err
```

And now we can verify that the expected error messages are in the `err` file:

```
$ cat err
usage: abuse.py [-h] [-a int] [-n int] [-s int]
abuse.py: error: --adjectives "-1" must be > 0
```

If you were to handle all of this yourself, you would need to write something like:

```
if args.adjectives < 1:
    parser.print_usage() ①
    print('--adjectives "{}" must be > 0'.format(args.adjectives), file=sys.stderr) ②
    sys.exit(1) ③
```

- ① Print the short "usage". You can also `parser.print_help()` to print the more verbose output for `-h`.
- ② Print the error message to the `sys.stderr` file handle. This is similar to the `sys.stdout` file handle we used in the "Howler" exercise.
- ③ Exit the program with a value that is not `0` to indicate an error.

Writing Pipelines



As you write more programs in your career, you may eventually start chaining them together. We often call these "pipelines" as the output of one program is "piped" to become the input for the next program. If there is an error in any part of the pipeline, we want the entire operation to stop so that the problems can be fixed. A non-zero return value from any program is a warning flag to halt operations.

Controlling randomness with `random.seed`

Once I'm in `main` and have my arguments, I can control the randomness of the program by calling `random.seed(args.seed)` because:

1. The default value of the `seed` is `None`, and setting `random.seed` to `None` is the same as not setting it at all.
2. The `type` of `args.seed` is `int` which is the proper type for `random.seed`. I do not have to validate the argument further. Negative integers are valid values.

Iterating `for` loops with `range`

To generate some `--number` of insults, I use the `range` function. Because I don't need the number of the insult, I can use the underscore (`_`) as a throwaway value:

```
>>> num_insults = 2
>>> for _ in range(num_insults):
...     print('An insult!')
...
An insult!
An insult!
```

The underscore is a way to unpack a value and indicate that you do not intend to use it. That is, it's not possible to write this:

```
>>> for in range(num_insults):
File "<stdin>", line 1
    for in range(num_insults):
```

You have to put *something* after the `for` that looks like a variable. If you put a named variable like `n`

and then don't use it in the loop, some tools like `pylint` will detect this as a possible error (and well it could be). The `_` shows that you won't use it, which is good information for your future self, some other user, or external tools to know.

For instance, you can use multiple `_`, e.g., here I can unpack a 3-tuple so as to get the middle value:

```
>>> x = 'Jesus', 'Mary', 'Joseph'  
>>> _, name, _ = x  
>>> name  
'Mary'
```

Constructing the insults

To create my list of adjectives, I used the `str.split` method on a long, multi-line string I created using three quotes:

```
>>> adjectives = """  
... bankrupt base caterwauling corrupt cullionly detestable dishonest  
... false filthsome filthy foolish foul gross heedless indistinguishable  
... infected insatiate irksome lascivious lecherous loathsome lubbery old  
... peevish rascaly rotten ruinous scurilous scurvy slanderous  
... sodden-witted thin-faced toad-spotted unmannered vile wall-eyed  
... """.strip().split()  
>>> nouns = """  
... Judas Satan ape ass barbemonger beggar block boy braggart butt  
... carbuncle coward coxcomb cur dandy degenerate fiend fishmonger fool  
... gull harpy jack jolthead knave liar lunatic maw milksop minion  
... ratcatcher recreant rogue scold slave swine traitor varlet villain worm  
... """.strip().split()  
>>> len(adjectives)  
36  
>>> len(nouns)  
39
```

To select some number of adjectives, I chose to use `random.sample` function since I needed more than one:

```
>>> import random  
>>> random.sample(adjectives, k=3)  
['filthsome', 'cullionly', 'insatiate']
```

For just one randomly selected value, I use `random.choice`:

```
>>> random.choice(nouns)  
'boy'
```

To concatenate them together, I need to put ', ' (a comma and a space) between each of the adjectives, and I can use `str.join` for that:

```
>>> adjs = random.sample(adjectives, k=3)
>>> adjs
['thin-faced', 'scurvy', 'sodden-witted']
>>> ', '.join(adjs)
'thin-faced, scurvy, sodden-witted'
```



And feed all this to a format string:

```
>>> adjs = ', '.join(random.sample(adjectives, k=3))
>>> print(f'You {adjs} {random.choice(nouns)}!')
You heedless, thin-faced, gross recreant!
```

And now you have a handy way to make enemies and influence people.

Review

- To indicate a problem to `argparse`, use the `parser.error` function to print a short usage, report the problem, and exit the program with an error value to indicate a problem.
- The `str.split` method is a useful way to create a `list` of string values from a long string.
- The `random.seed` function can be used to make reproducible "random" selections each time a program is run.
- The `random` module's `choice` and `sample` functions are useful for randomly selecting one or several items from a list of choices, respectively.

Going Further

- Read your adjective and nouns from files that are passed as arguments.
- Add tests to verify that the files are processed correctly and new insults are still stinging.

Chapter 10: Telephone: Randomly mutating strings

"What we have here is a failure to communicate." — Captain



Now that we've played with randomness, let's apply the idea to randomly mutating a string. This is interesting because strings are actually *immutable* in Python, so we'll have to figure out a way around that. To explore these ideas, we'll write a version of the game of "Telephone." In that game, a secret message is whispered through a line or circle of people. Each time the message is transmitted, it's usually changed in some way. When the last person receives the message, they say it out loud to compare it to the original message.

We will write a program called `telephone.py` that will mimic this game. It will mutate some `text` by some varying amount and then print "You said: " plus the original text followed by "I heard: " with the modified message. The text may come from the command line:

```
$ ./telephone.py 'The quick brown fox jumps over the lazy dog.'  
You said: "The quick brown fox jumps over the lazy dog."  
I heard : "TheMquick brown fox jumps ovMr t:e lamy dog."
```

Or from a file:

```
$ ./telephone.py ../inputs/fox.txt  
You said: "The quick brown fox jumps over the lazy dog."  
I heard : "The quick]b'own fox jumps ovek the la[y dog."
```

In this exercise, you will learn to:

- Round numbers
- Use the `string` module
- Modify strings and lists to introduce random point mutations

More briefing

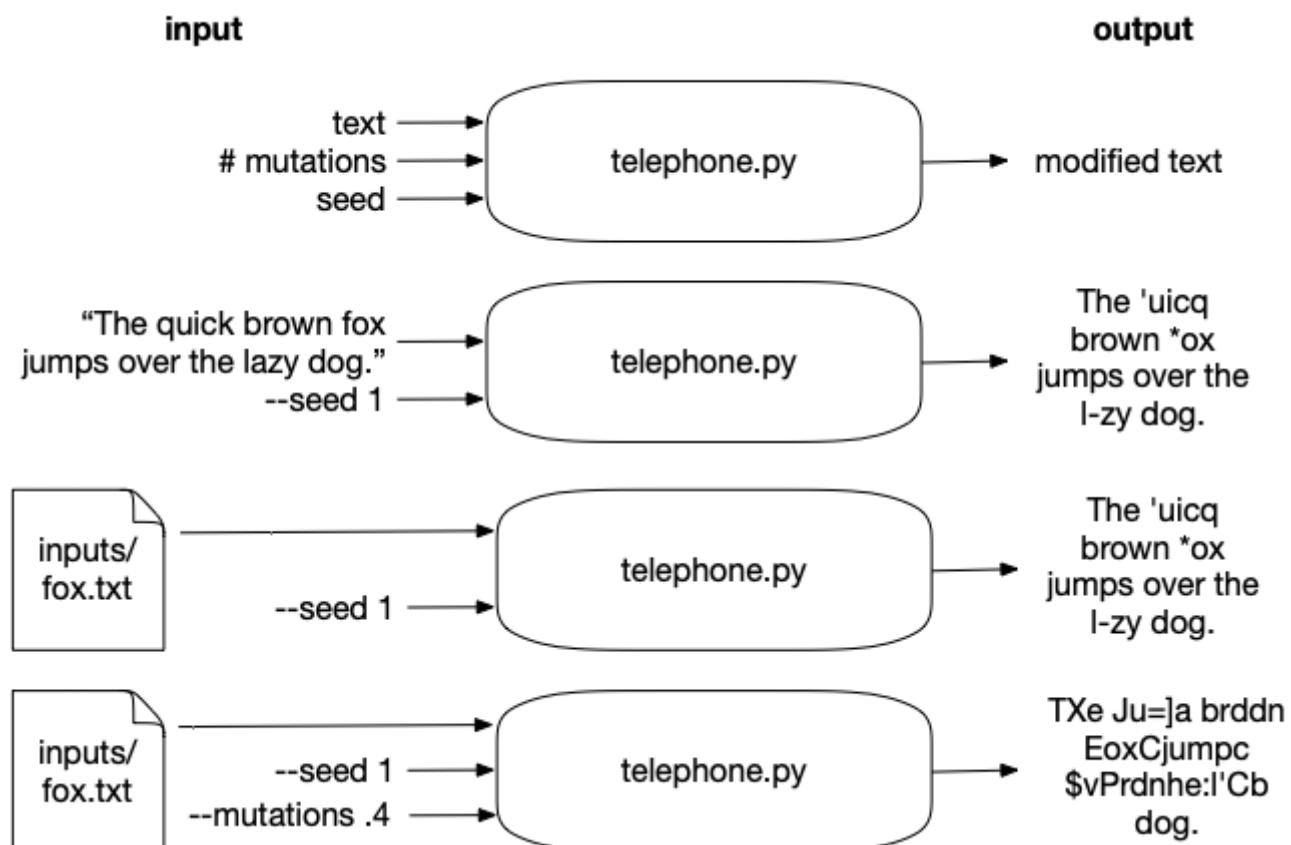
The program should accept a `-m` or `--mutations` option which should be a number between 0 and 1 with a default value of `0.1` (10%). This will be a percentage of the number of letters that should be altered. For instance, `.5` means that 50% of the letters should be changed:

```
$ ./telephone.py ..../inputs/fox.txt -m .5
You said: "The quick brown fox jumps over the lazy dog."
I heard : "F#eYquJsY ZrHnna"o. Muz/$ Nver t/Relazy dA!."
```

Because we are using the `random` module, we'll accept an `int` value for the `-s` or `--seed` option so that we can reproduce our "random" selections:

```
$ ./telephone.py ../inputs/fox.txt -s 1
You said: "The quick brown fox jumps over the lazy dog."
I heard : "The 'uicq brown *ox jumps over the l-zy dog."
```

Here is a string diagram of the program:



Writing telephone.py

I recommend you start by with `new.py telephone.py` to create the program in the `telephone` directory. You could also copy the `template/template.py` to `telephone/telephone.py`. Modify the `get_args` function until your `-h` output matches the following:

```
$ ./telephone.py -h
usage: telephone.py [-h] [-s int] [-m float] str

Telephone

positional arguments:
  str                  Input text or file

optional arguments:
  -h, --help            show this help message and exit
  -s int, --seed int    Random seed (default: None)
  -m float, --mutations float
                        Percent mutations (default: 0.1)
```

Now run the test suite. You should pass at least the first two tests (the `telephone.py` program exists and prints something like a "usage" when run with `-h` or `--help`).

The next two tests check that your `--seed` and `--mutations` options both reject non-numeric values. This should happen automatically if you define these parameter using `type` of `int` and `float`, respectively. Your program should behave like this:

```
$ ./telephone.py -s blargh foo
usage: telephone.py [-h] [-s int] [-m float] str
telephone.py: error: argument -s/--seed: invalid int value: 'blargh'
$ ./telephone.py -m blargh foobar
usage: telephone.py [-h] [-s str] [-m float] str
telephone.py: error: argument -m/--mutations: invalid float value: 'blargh'
```

The next test checks if the program rejects `--mutations` outside of the range 0-1 (where both bounds are inclusive). This is not a check that you can easily describe to `argparse`, so I suggest that you look at how we handled the validation of the arguments in the `abuse.py` program. In the `get_args` function of that program, we manually checked the value of the arguments and used the `parser.error` function to throw an error. Your program should do this:

```
$ ./telephone.py -m -1 foobar
usage: telephone.py [-h] [-s str] [-m float] str
telephone.py: error: --mutations "-1.0" must be between 0 and 1
```

Note that a `--mutations` value of `0` is acceptable, in which case we will print the input text back without modifications. This is another program that accepts the input text either from the

command line directly or from a file. I suggest you look at the "Howler" solution where, inside the `get_args` function, I suggest that you use `os.path.isfile` to detect if the text argument is a file. If it is a file, read the contents of the file for the `text` value.

Once you have taken care of all the program parameters, start off your `main` function with setting the `random.seed` and echoing back the given text:

```
def main():
    args = get_args()
    random.seed(args.seed)
    print(f'You said: "{args.text}"')
    print(f'I heard : "{args.text}"')
```

Your program should handle command-line text:

```
$ ./telephone.py 'The quick brown fox jumps over the lazy dog.'
You said: "The quick brown fox jumps over the lazy dog."
I heard : "The quick brown fox jumps over the lazy dog."
```

Or an input file:

```
$ ./telephone.py ../inputs/fox.txt
You said: "The quick brown fox jumps over the lazy dog."
I heard : "The quick brown fox jumps over the lazy dog."
```

At this point, your code should pass up to `test_for_echo`. The next tests start asking you to mutate the input, so let's discuss how to do that.

Calculating the number of mutations

The number of letters that need to be changed can be calculated by multiplying the length of the input text by the `args.mutations` value. If we want to change 20% of the characters in "[t]he quick brown fox..." string, we'll find that is not a whole number:

```
>>> text = 'The quick brown fox jumps over the lazy dog.'  
>>> mutations = .20  
>>> len(text) * mutations  
8.8
```

We can use the `round` function to give us the nearest integer value. Read `help(round)` to understand how to round floating point numbers to a specific number of digits:

```
>>> round(len(text) * mutations)  
9
```

Note that you could also convert a `float` to an `int` by using the `int` function, but this truncates the fractional part of the number rather than rounding it:

```
>>> int(len(text) * mutations)  
8
```

You will need this value for later:

```
>>> num_mutations = round(len(text) * mutations)  
>>> assert num_mutations == 9
```

The mutation space



When we change a character, what will we change it to? For this, I'd like to introduce you to the `string` module. I'd encourage you to take a look at the documentation by importing the module and reading the `help`:

```
>>> import string  
>>> help(string)
```

We can, for instance, get all the lowercase ASCII letters. Note that this is not a method call as there are no parentheses `()` at the end:

```
>>> string.ascii_lowercase  
'abcdefghijklmnopqrstuvwxyz'
```

Note that this returned a `str`:

```
>>> type(string.ascii_lowercase)  
<class 'str'>
```

For our program, I would like to use `string.ascii_letters` and `string.punctuation`. How do we concatenate two strings together? We can use the `+` operator. These are the characters we will draw from when replacing:

```
>>> alpha = string.ascii_letters + string.punctuation  
>>> alpha  
'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ!#$%&\'()*+, -.<./:;<=>?@[\\]^_`{|}~'
```

Selecting the characters to mutate

There are at least two approaches we could take to choosing which characters to change.

Non-deterministic selection

One way would be to mimic solution 1 in "Apples". We could iterate `for char in text` and flip a coin to decide whether to take the original character or randomly select a new character:

```
new_text = '' ①
for char in args.text: ②
    new_text += random.choice(alpha) if random.random() <= args.mutations else char ③
print(new_text) ④
```

① Initialize `new_text` as an empty string.

② Iterate through each character in the text.

③ Use `random.random` to generate a value between 0 and 1. If that value is less than or equal to the `args.mutation` value, then randomly choose from the `alpha`; otherwise, use the character.

④ Print the resulting `new_text`.

We used the `random.choice` function in `abuse.py` to randomly select *one* value from a list of choices. Here we select a character from the `alpha` if the `random.random` value falls within the range of the `args.mutation` value (which we know is also a `float`).

The problem with this approach is that it is *non-deterministic*. That is, by the end of the `for` loop, we are not guaranteed of having made the correct number of changes, which we calculated above as `num_mutations`. If the mutation value was 40%, you could expect that you'd end up changing about 40% of the string because a random value between 0 and 1 should be less than or equal to 0.4 about 40% of the time. However, since the decision whether to change each character is random, you might end up with more or less mutation than this, especially with a short text. Since we can't be certain — since we can't determine beforehand — that we will end up with exactly 40% mutations, this is *non-deterministic*.

Still, this is a really useful technique that you should note. Imagine you have an input file with millions to potentially billions of lines of text and you want to randomly sample approximately 10% of the lines. The above approach would be a reasonably fast and accurate way to select which lines to take and which to discard. A larger sample size helps to get closer to the desired number of mutations.

Randomly sampling characters

A deterministic approach to the million-line file would require first reading the entire input to count the number of lines. Then you would have to choose which lines to take, and then you would go back through the file a second time to take those lines. This approach would take *much* longer than the method described above. Depending on how large the input file is, how the program is written, and how much memory your computer has, the program could possibly even crash your

computer!



Because our input is rather small, we will use this algorithm because it has the advantage of being exact and testable. Rather than lines of text, though, we'll consider indexes of characters. We've seen the `str.replace` method (in "Apples") that allows us to change, for instance, all instances of one string to another:

```
>>> 'foo'.replace('o', 'a')  
'faa'
```

We can't use that method here, however, because we only want to change some characters. Instead we need to use the indexes of the characters in the text. We can use the `random.sample` function to select the indexes for us.

The first argument to `random.sample` needs to be something like a `list`. We should give it a `range` of numbers up to the length of our `text`. So, if our `text` is 44 characters long:

```
>>> text  
'The quick brown fox jumps over the lazy dog.'  
>>> len(text)  
44
```

We can use the `range` function to make a `list` of numbers up to 44. Note that, in the REPL, we see that `range` is a lazy function:

```
>>> range(len(text))  
range(0, 44)
```

It won't actually produce the 44 values until we force it, which we can do in the REPL using the `list` function. You rarely have to do this in your actual programs, however:

```
>>> list(range(len(text)))
```

We calculated above that the `num_mutations` for altering 20% of `text` is 9. Here are the indexes that need to be changed:

```
>>> indexes = random.sample(range(len(text)), num_mutations)
>>> indexes
[13, 6, 31, 1, 24, 27, 0, 28, 17]
```

I suggest you use a `for` loop to iterate through each of these index values:

```
>>> for i in indexes:
...     print(f'{i}: {text[i]}')
...
13 w
6 i
31 t
1 h
24 s
27 v
0 T
28 e
17 o
```

You should replace the character at that position with a randomly selected character from `alpha`.

```
>>> for i in indexes:
...     print(f'{i}: {text[i]} -> {random.choice(alpha)}')
...
13 w -> b
6 i -> W
31 t -> B
1 h -> #
24 s -> d
27 v -> :
0 T -> C
28 e -> %
17 o -> ,
```

I will introduce one other twist—we don't want the replacement value to ever be the same as the character it is replacing. Can you figure out how to get a subset of `alpha` that *does not* include the character at the position?

Mutating a string

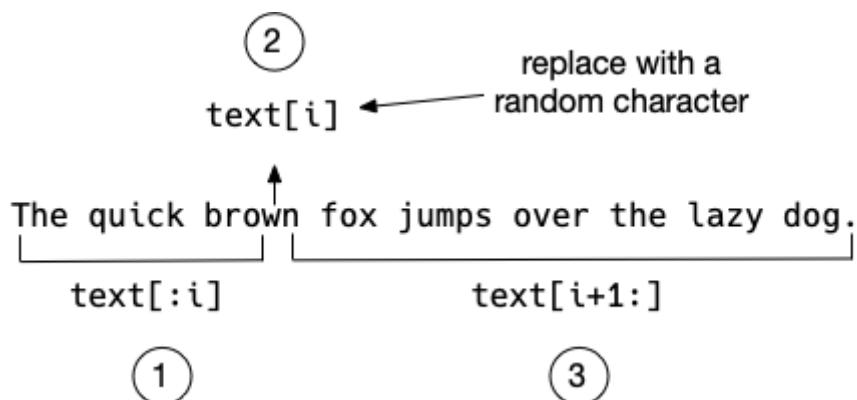
Python `str` variables are *immutable*, meaning we cannot directly modify them. For instance, we want to change the character '`w`' at position `13` to a '`b`'. It would be handy to directly modify `text[13]`, but that will create an exception:

```
>>> text[13] = 'b'  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
TypeError: 'str' object does not support item assignment
```

The only way to modify the `str` value `text` is to overwrite with a new `str`. So, you need to create a new `str` with

1. the part of `text` before a given index
2. the randomly selected value from `alpha`
3. the part of `text` after a given index

For 1 and 3, you can use string *slices*.



For example, if the index `i` is 13, the slice before it is:

```
>>> text[:13]  
'The quick bro'
```

And the part after:

```
>>> text[14:]  
'n fox jumps over the lazy dog.'
```

Using the 1, 2, and 3 above, your `for` loop should be:

```
for i in index:  
    text = 1 + 2 + 3
```

Can you figure that out?

Time to write

OK, the lesson is over. You have to go write this now. Use the tests. Solve them one at a time. You can do this.

Solution

```
1 #!/usr/bin/env python3
2 """Telephone"""
3
4 import argparse
5 import os
6 import random
7 import string ①
8
9
10 # -----
11 def get_args():
12     """Get command-line arguments"""
13
14     parser = argparse.ArgumentParser(
15         description='Telephone',
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18     parser.add_argument('text', metavar='str', help='Input text or file') ②
19
20     parser.add_argument('-s',
21                         '--seed',
22                         help='Random seed',
23                         metavar='int',
24                         type=int,
25                         default=None)
26
27     parser.add_argument('-m', ④
28                         '--mutations',
29                         help='Percent mutations',
30                         metavar='float',
31                         type=float,
32                         default=0.1)
33
34     args = parser.parse_args() ⑤
35
36     if not 0 <= args.mutations <= 1: ⑥
37         parser.error(f"--mutations '{args.mutations}' must be between 0 and
1')
38
39     if os.path.isfile(args.text): ⑦
40         args.text = open(args.text).read().rstrip()
41
42     return args ⑧
43
44
45 # -----
46 def main():
```

```

47     """Make a jazz noise here"""
48
49     args = get_args()
50     text = args.text
51     random.seed(args.seed)                                     ⑨
52     alpha = string.ascii_letters + string.punctuation ⑩
53     len_text = len(text)                                    ⑪
54     num_mutations = round(args.mutations * len_text) ⑫
55     new_text = text                                       ⑬
56
57     for i in random.sample(range(len_text), num_mutations): ⑭
58         new_char = random.choice(alpha.replace(new_text[i], '')) ⑮
59         new_text = new_text[:i] + new_char + new_text[i + 1:] ⑯
60
61     print(f'You said: "{text}"\nI heard : "{new_text}"')    ⑰
62
63
64 # -----
65 if __name__ == '__main__':
66     main()

```

- ① We need to import the `string` module here.
- ② The program takes one positional argument for the `text`. This could be either a string of text or a file which needs to be read.
- ③ The `--seed` parameter is an `int` value with a default of `None`.
- ④ The `--mutations` parameter must be a `float` with a default of `0.1`.
- ⑤ Here we gather the `args` from the command line. If `argparse` detects failures such as non-numeric values for `seed` or `mutations`, then the program dies here and the user sees an error message. If this call succeeds, then `argparse` has validated the arguments and converted the values, e.g., `args.mutations` is actually a `float` value.
- ⑥ If `args.mutations` is *not* in the acceptable range of 0-1, then use `parser.error` to halt the program and print the given message. Note the use of feedback to echo the bad `args.mutation` value to the user.
- ⑦ If `args.text` names an existing file, then read that file for the contents, overwriting the original value of `args.text`.
- ⑧ Return `args` to the caller (line 49).
- ⑨ Be sure to set the `random.seed` to the value provided by the user. Remember that the default value for `args.seed` is `None`, which is the same as not setting the seed.
- ⑩ Set `alpha` to be the characters we'll use for replacements.
- ⑪ Since I use the `len(text)` more than once, I put it into a variable.
- ⑫ Figure the `num_mutations`.
- ⑬ Make copy of `text`.
- ⑭ Use `random.sample` to get `num_mutations` indexes to change. That function returns a `list` that we can iterate using the `for` loop.

- ⑯ Select a `new_char` using `random.choice` to select from a string we create by replacing in `alpha` the current character (`text[i]`) with nothing. This ensures that our new character cannot be the same as the one we are replacing.
- ⑰ Overwrite the `text` by concatenating the slice before the current index, the `new_char`, and the slice after the current index.
- ⑱ Print the `text`.

Discussion

Defining the arguments

There's nothing in `get_args` that we haven't seen before. The `--seed` argument is an `int` that we will pass to the `random.seed` function so as to control the randomness for testing. The `default` should be `None` so that we can call `random.seed(args.seed)` where `None` is the same as not setting it. (Note that we could also define the seed as a `str` value as both are acceptable seeds.) The `--mutations` is a `float` with a reasonable default, and I test that `args.mutations` is in the proper range and use `parser.error` if it is not. As in other programs, I test if the `text` argument is a file and read the contents if it is.

Mutating a string

Now we saw earlier that we can't just change the `text`:

```
>>> text[13] = 'b'  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
TypeError: 'str' object does not support item assignment
```

So we're going to have to create a *new* string using the text before and after `i` which we can get with string slices using `text[start:stop]`. If you leave out `start`, Python starts at `0` (the beginning of the string), and if you leave out `stop` then it goes to the end, so `text[:]` is a copy of the entire string.

If `i` is `13`, then bit before `i` is:

```
>>> i = 13  
>>> text[:i]  
'The quick bro'
```

And the bit after `i + 1`:

```
>>> text[i+1:]  
'n fox jumps over the lazy dog.'
```

Now for what to put in the middle. I noted that we should use `random.choice` to select a character from `alpha`, which is the combination of all the ASCII letters and punctuation *without* the current character. I can use the `str.replace` method to get rid of the current letter:

```
>>> alpha = string.ascii_letters + string.punctuation  
>>> alpha.replace(text[i], '')  
'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ!"#$%&\()'*)+,--.  
.:/;=>?@[\\]^_`{|}~'
```

And use that to get a new letter that won't include what it's replacing:

```
>>> new_char = random.choice(alpha.replace(text[i], ''))  
>>> new_char  
'E'
```

There are many ways to join strings together into new strings, and the `+` operator is perhaps the simplest:

```
>>> text = text[:i] + new_char + text[i+1:]  
>>> text  
'The quick broEn fox jumps over the lazy dog.'
```

We do this for each index in the `random.sample` of indexes, each time overwriting `text`. After the `for` loop is done, we have mutated all the positions of the input string and can `print` it.

Using a `list` instead of a `str`

Strings are immutable, but lists are not. We see above that a move like `text[13] = 'b'` creates an exception, but we can change `text` into a list and directly modify it with the same syntax:

```
>>> text = list(text)  
>>> text[13] = 'b'
```

And then we can turn that `list` back into a `str` by joining it on the empty string:

```
>>> ''.join(text)  
'The quick brobn fox jumps over the lazy dog.'
```

Here is a version of `main` that uses this approach:

```
def main():  
    args = get_args()  
    text = args.text  
    random.seed(args.seed)  
    alpha = string.ascii_letters + string.punctuation  
    len_text = len(text)  
    num_mutations = round(args.mutations * len_text)  
    new_text = list(text) ①  
  
    for i in random.sample(range(len_text), num_mutations):  
        new_text[i] = random.choice(alpha.replace(new_text[i], '')) ②  
  
    print('You said: "{}"\nI heard : "{}"'.format(text, ''.join(new_text))) ③
```

- ① Set the `new_text` to be a `list` of characters from the `text` value.
- ② Now you can directly modify a value in `new_text`.
- ③ Join `new_list` on the empty string to make a new `str`.

There's no particular advantage of one way over the other, but I would personally choose the second method because I don't like messing around with slicing strings. To me, modifying a `list` in-place makes much more sense than repeatedly chopping up and piecing together a `str`.

Mutations in DNA

For what it's worth, this is (kind of, sort of) how DNA changes over time. The machinery to copy DNA makes mistakes, and mutations randomly occur. Many times the change has no deleterious affect on the organism. Our example only changes characters to other characters, what are called "point mutations" or "single nucleotide variations" (SNV) or "single nucleotide polymorphisms" (SNP) in biology, but we could write a version that would also randomly delete or insert new characters which are called "in-dels" (insertion-deletions) in biology. Mutations (that don't result in the demise of the organism) occur at a fairly standard rate, so counting the number of mutations between a conserved region of any two organisms can allow an estimate of how long ago they diverged from a common ancestor!

Review



- A string cannot be directly modified, but the variable containing the string can be repeatedly overwritten with new values.
- Lists can be directly modified, so it can sometimes help to use `list` on a string to turn it into a `list`, modify that, then use `''.join` to change it back to a `str`.
- The `string` module has handy functions for dealing with strings.

Going Further

- Apply the mutations to randomly selected words instead of the whole string.
- Add insertions and deletions in addition to mutations; maybe create arguments for the percentage of each and choose to add or delete characters at the indicated frequency.
- Add an option for `-o` or `--output` that names a file to write the output. The default should be to print to `STDOUT`.
- Add an flag to limit the replacements to character values only (no punctuation).
- Add tests to `test.py` for every new feature and ensure your program works properly.

Chapter 11: Bottles of Beer Song: Writing and testing functions



Few songs are as annoying as "100 Bottles of Beer on the Wall." Hopefully you've never had to ride for hours in a van with middle school boys who like to sing this. The author has. It's a fairly simple song that we can write an algorithm to generate. This gives us an opportunity to play with counting up and down, formatting strings, and—new to this exercise—writing functions and tests for those functions!

Our program will be called `bottles.py` and will take one option `-n` or `--num` which must be a *positive int* (default `10`). The program should print all the verses from `--num` down to `1`. There should be two newlines between each verse to visually separate them, but there must be only one newline after the last verse (for one bottle) which should print "No more bottles of beer on the wall" rather than "0 bottles":

```
$ ./bottles.py --num 2
2 bottles of beer on the wall,
2 bottles of beer,
Take one down, pass it around,
1 bottle of beer on the wall!

1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
No more bottles of beer on the wall!
```

In this exercise, you will:

- Learn how to produce a list of numbers decreasing in value.
- Write a function to create a verse of the song using a test to verify when the verse is correct.
- Explore how `for` loops can be written as list comprehensions which in turn can be written with the `map` function.

Writing `bottles.py`

As always, the program should respond to `-h` or `--help` with a usage statement:

```
$ ./bottles.py -h
usage: bottles.py [-h] [-n int]

Bottles of beer song

optional arguments:
  -h, --help            show this help message and exit
  -n int, --num int    How many bottles (default: 10)
```

Start off by copying the `template.py` or using `new.py` to create your `bottles.py` program. Then modify the `get_args` function until your usage matches the above. Since the `-h` and `--help` flags are automatically handled by `argparse`, you need define only the `--num` option with `type=int` and `default=10`.

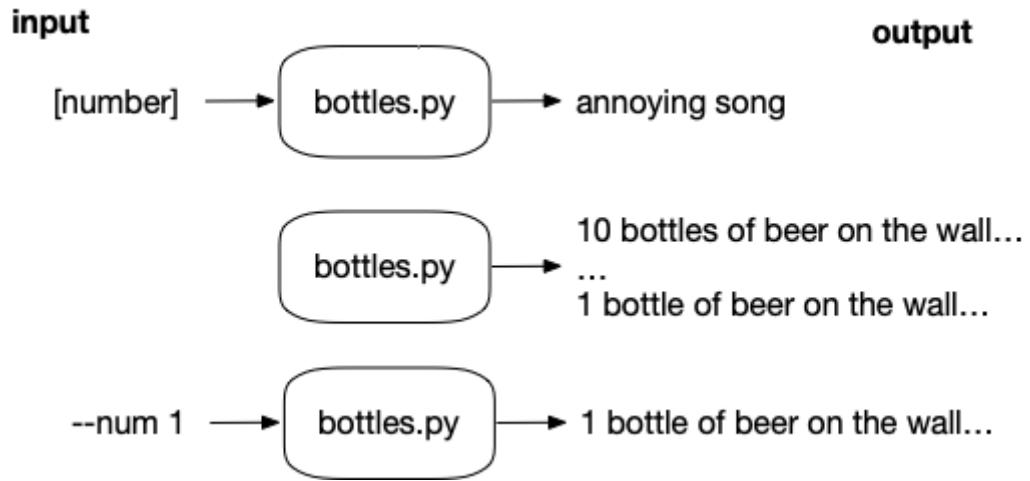
If the `--num` argument is not an `int` value, your program should print an error message and exit with an error value. This should happen automatically if you define your parameter to `argparse` properly:

```
$ ./bottles.py -n foo
usage: bottles.py [-h] [-n int]
bottles.py: error: argument -n/--num: invalid int value: 'foo'
$ ./bottles.py -n 2.4
usage: bottles.py [-h] [-n int]
bottles.py: error: argument -n/--num: invalid int value: '2.4'
```

It should do likewise when the `--num` is a negative number. To handle this, I suggest you manually check this and call `parser.error` inside the `get_args` functions as in previous exercises:

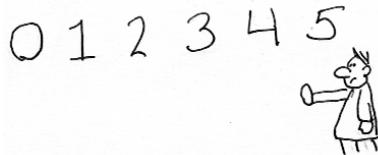
```
$ ./bottles.py --num -4
usage: bottles.py [-h] [-n int]
bottles.py: error: --num (-4) must > 0
```

Here is a string diagram of the inputs and outputs:



Counting down

The song starts at the given `--num` value like 10 and needs to count down to 9, 8, 7, and so forth. So how do we do that in Python? We've seen how to use `range(start, stop)` to get a list of integers that go *up* in value. If you give it just one number, it will be consider the `stop` and will assume `0` as the `start`. Because this is a lazy function, I must use `list` in the REPL to force it to produce the numbers:



```
>>> list(range(5))
[0, 1, 2, 3, 4]
```

Remember that the `stop` value is never included in the output, so the above stopped at 4 and not 5. If you give `range` two numbers, they are considered to be `start` and `stop`:

```
>>> list(range(1, 5))
[1, 2, 3, 4]
```

To reverse this sequence, you might be tempted to swap the `start` and `stop` values. Unfortunately, if the `start` is greater than the `stop`, you get an empty list:

```
>>> list(range(5, 1))
[]
```

We saw in "Picnic" that we can use the `reversed` function to reverse a `list`. This is another lazy function, so, again, I use the `list` function to force the values in the REPL:

```
>>> list(reversed(range(1, 5)))
[4, 3, 2, 1]
```



The `range` function can also take an optional third argument for a `step` value. For instance, you could use this to count by 5s:

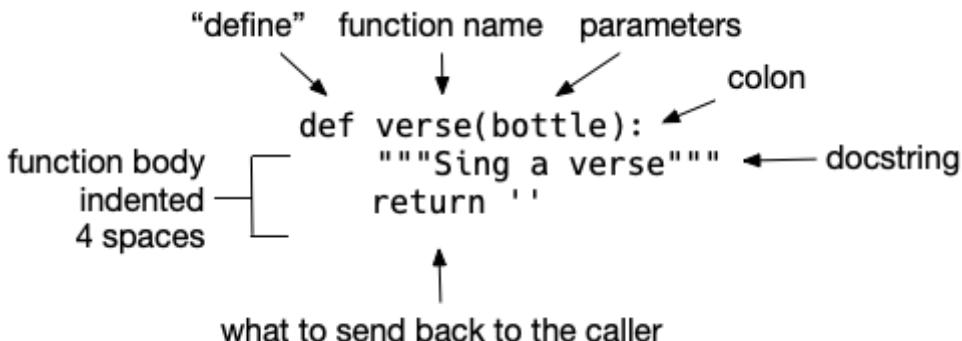
```
>>> list(range(0, 50, 5))
[0, 5, 10, 15, 20, 25, 30, 35, 40, 45]
```

So another way to count down is to swap the `start` and `stop` and use `-1` for the step:

```
>>> list(range(5, 0, -1))
[5, 4, 3, 2, 1]
```

Writing a function

Up to this point, I've suggested that all your code go into the `main` function. This is the first exercise where I suggest you write a function. I would like you to consider how to write the code to sing one verse. The function could take the number of the verse and return the text for that verse.



You can start off with something like the diagram. The `def` keyword "defines" a function, and the name of the function follows. Function names should have only letters, numbers and the underscore and cannot start with a number. After the name comes parentheses that hold any parameters that the function accepts. Here our function should be called `verse` and has the parameter `bottle` (or `number` or whatever you want to call it). After the parameters comes a colon to show the end of the `def` line. The function body comes next, with all lines being indented at least 4 spaces.

The "docstring" is a comment just after the function definition and will show up in the `help` for your function. If you enter this into the REPL:

```
>>> def verse(bottle):
...     """Sing a verse"""
...     return ''
...
>>> help(verse)
```

Then you will see:

```
Help on function verse in module __main__:

verse(bottle)
    Sing a verse
```

The `return` statement tells Python what to send back from the function. It's not very interesting because it will only ever send back the empty string right now:

```
>>> verse(10)
''
```

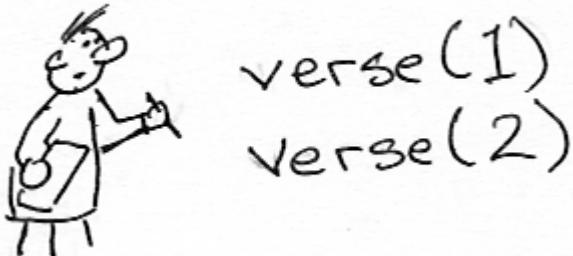
Writing a test for `verse`

In the spirit of *test-driven development*, let us write a test for `verse` before we go any further. Here is a test you can use. Copy and paste this code as a new function into your `bottles.py` program:

```
# -----
def test_verse():
    """Test verse"""

    one = verse(1)
    assert one == '\n'.join([
        '1 bottle of beer on the wall,', '1 bottle of beer,', 
        'Take one down, pass it around,', 
        'No more bottles of beer on the wall!'
    ])

    two = verse(2)
    assert two == '\n'.join([
        '2 bottles of beer on the wall,', '2 bottles of beer,', 
        'Take one down, pass it around,', '1 bottle of beer on the wall!'
    ])
```



There are many, many ways you could write this program. I have in mind that my `verse` function will produce a single verse of the song, returning a `str` value which is the four lines of the verse joined on newlines. You don't have to write it this way, but I would like you to consider what it means to write a function and a *unit test*. If you read about software testing, you'll find that there are different definitions of what a "unit" of code is. In this book, I consider a *function* to be a *unit*, and so my unit tests are tests of individual functions.

By the way, even though the song has potentially hundreds of verses, these two tests should cover everything you need to check. The first test shows that we are looking for "1 bottle" (singular) and not "1 bottles" (plural). We also check that the last line says "No more bottles" instead of "0 bottles." The test for "2 bottles of beer" is making sure that the numbers are "2 bottles" and then "1 bottle." Presumably if you've managed to pass those two tests, your program ought to be able to handle any value.

Here I've written `test_verse` to test just the `verse` function. The name of the function matters because I am using the `pytest` module to find all the functions in my code that start with `test_` and run them. If your `bottles.py` program has the above functions for `verse` and `test_verse`, you can run `pytest bottles.py`. Try it, and you should see something like this:

```

$ pytest bottles.py
=====
test session starts =====
platform darwin -- Python 3.7.3, pytest-4.6.5, py-1.8.0, pluggy-0.12.0
rootdir: /Users/kyclark/work/manning/playful_python/bottles_of_beer
plugins: openfiles-0.3.2, arraydiff-0.3, doctestplus-0.3.0, remotedata-0.3.1, cov-2.7.1
collected 1 item

bottles.py F [100%]

=====
FAILURES =====
----- test_verse -----
----- test_verse() -----
def test_verse():
    """Test verse"""

    one = verse(1) ①
>     assert one == '\n'.join([
        '1 bottle of beer on the wall,', '1 bottle of beer,', 'Take one down, pass it around,', 'No more bottles of beer on the wall!'
    ])
E     AssertionError: assert '' == '1 bottle of beer on the wal...ottles of beer on the wall!' ③
E         + 1 bottle of beer on the wall,
E         + 1 bottle of beer,
E         + Take one down, pass it around,
E         + No more bottles of beer on the wall!

bottles.py:49: AssertionError
=====
1 failed in 0.10 seconds =====

```

- ① Call the `verse` function with the argument `1` to get the `one` verse of the song.
- ② The `>` at the beginning of this line indicates this is the source of the error. The test checks if the value of `one` is equal to an expected `str` value. Since it's not, this line throws an exception causing the assertion to fail.
- ③ The `E` lines show the difference between what was received and what was expected. The value of `one` is the empty string (`''`) which does not match the expected string "1 bottle of beer..." and so on.

To pass the first test, you could copy the code for the expected value of `one` directly from the test. Change your `verse` function to be this:

```

def verse(bottle):
    """Sing a verse"""

    return '\n'.join([
        '1 bottle of beer on the wall,', '1 bottle of beer,', 
        'Take one down, pass it around,', 
        'No more bottles of beer on the wall!'
    ])

```

And run your test again. Now the first test should pass and the second one should fail. I'll only show the relevant error lines:

```

=====
 FAILURES =====
----- test_verse -----
----- test_verse -----



def test_verse() -> None:
    """Test verse"""

    one = verse(1)
    assert one == '\n'.join([ ①
        '1 bottle of beer on the wall,', '1 bottle of beer,', 
        'Take one down, pass it around,', 
        'No more bottles of beer on the wall!'
    ])

    two = verse(2) ②
>     assert two == '\n'.join([ ③
        '2 bottles of beer on the wall,', '2 bottles of beer,', 
        'Take one down, pass it around,', '1 bottle of beer on the wall!'
    ])

E     AssertionError: assert '1 bottle of ... on the wall!' == '2 bottles of ... on
the wall!' ④
E         - 1 bottle of beer on the wall,
E         ? ^
E         + 2 bottles of beer on the wall,
E         ? ^      +
E         - 1 bottle of beer,
E         ? ^
E         + 2 bottles of beer, ...
E
E     ...Full output truncated (7 lines hidden), use '-vv' to show

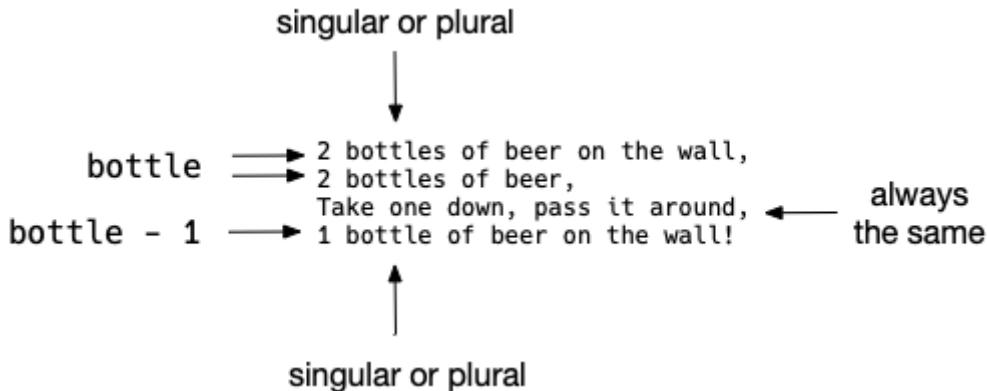
```

① This assertion passes this time.

② Call the `verse` with the value of 2.

③ Assert that `two` is equal to an expected string.

④ These E lines are showing you the problem. It got '`1 bottle`' but expected '`2 bottles`', etc.



Go back and look at your `verse` definition. Think about which parts need to change—the first, second, and fourth lines. The third line is always the same. You’re given a value for `bottle` which needs to be used in the first two lines along with either “bottle” or “bottles,” depending on the value of `bottle` (Hint: it’s only singular for the value 1; otherwise, it’s plural). The fourth line needs the value of `bottle - 1` and, again, the proper singular or plural depending on that value. Can you figure how to write this?

Focus on passing those two tests before you move to the next stage of printing the whole song. That is, do not attempt anything until you see this:

```
$ pytest bottles.py
=====
test session starts =====
platform darwin -- Python 3.7.3, pytest-4.6.5, py-1.8.0, pluggy-0.12.0
rootdir: /Users/kyclark/work/manning/playful_python/bottles_of_beer
plugins: openfiles-0.3.2, arraydiff-0.3, doctestplus-0.3.0, remotedata-0.3.1, cov-2.7.1
collected 1 item

bottles.py .
[100%]

=====
1 passed in 0.05 seconds =====
```

Using the `verse` function

At this point, you know:

1. That the `--num` value is a valid integer value greater than 0.
2. How to count from that `--num` value backwards down to 0.
3. That the `verse` function will print any one verse properly.

Now you need to put them together. I suggest you start by using a `for` loop with the `range` function to count down. Use each value from that to produce a `verse`. There should be 2 newlines after every verse except for the last.

You will use the regular `pytest -xv test.py` (or `make test`) to test the program at this point. In the parlance of testing, the `test.py` is an *integration test* because it checks that the program *as a whole* is working. From this point on, we'll focus more on how to write *unit* tests to check individual functions as well as *integration* tests to ensure that all the functions work together.

Once you can pass the test suite using a `for` loop, try to rewrite it using either a list comprehension or a `map`. I would suggest commenting out your working code by adding `#` to the beginnings of the lines and then try other ways to write the algorithm. Use the tests to verify that your code still passes. If is at all motivating, my solution is one line long. Can you write a single line of code that combines the `range` and `verse` functions to produce the expected output?

Hints:

- Define the `--num` argument as an `int` with a default value of `10`.
- Use `parser.error` to get `argparse` to print an error message for a negative `--num` value.
- Write the `verse` function. Use the `test_verse` function and `pytest` to make that work properly.
- Combine the `verse` function with the `range` to create all the verses.

Do try your best to write the program before reading the solution. Also feel free to solve the problem in a completely different way, even writing your own unit tests!

Solution

```
1 #!/usr/bin/env python3
2 """Bottles of beer song"""
3
4 import argparse
5
6
7 # -----
8 def get_args():
9     """Get command-line arguments"""
10
11     parser = argparse.ArgumentParser(
12         description='Bottles of beer song',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('-n', ①
16                         '--num',
17                         metavar='int',
18                         type=int,
19                         default=10,
20                         help='How many bottles')
21
22     args = parser.parse_args() ②
23
24     if args.num < 1: ③
25         parser.error('--num {} must > 0'.format(args.num))
26
27     return args
28
29
30 # -----
31 def verse(bottle): ④
32     """Sing a verse"""
33
34     next_bottle = bottle - 1 ⑤
35     s1 = '' if bottle == 1 else 's' ⑥
36     s2 = '' if next_bottle == 1 else 's' ⑦
37     num_next = 'No more' if next_bottle == 0 else next_bottle ⑧
38     return '\n'.join([ ⑨
39         f'{bottle} bottle{s1} of beer on the wall,',
40         f'{bottle} bottle{s1} of beer,',
41         f'Take one down, pass it around,',
42         f'{num_next} bottle{s2} of beer on the wall!',
43     ])
44
45
46 # -----
47 def test_verse(): ⑩
```

```

48     """Test verse"""
49
50     one = verse(1) ⑪
51     assert one == '\n'.join([
52         '1 bottle of beer on the wall,', '1 bottle of beer,', 
53         'Take one down, pass it around,', 
54         'No more bottles of beer on the wall!'
55     ])
56
57     two = verse(2) ⑫
58     assert two == '\n'.join([
59         '2 bottles of beer on the wall,', '2 bottles of beer,', 
60         'Take one down, pass it around,', '1 bottle of beer on the wall!'
61     ])
62
63
64 # -----
65 def main():
66     """Make a jazz noise here"""
67
68     args = get_args()
69     print('\n\n'.join(map(verse, range(args.num, 0, -1)))) ⑬
70
71
72 # -----
73 if __name__ == '__main__':
74     main()

```

- ① Define the `--num` argument as an `int` with a default value of `10`.
- ② Parse the command-line argument into the variable `args`.
- ③ If the `args.num` is less than 1, use `parser.error` to display an error message and exit the program with an error value.
- ④ Define a function that can create a single `verse`.
- ⑤ Define a `next_bottle` that is one less than the current bottle.
- ⑥ Define a `s1` (the first "s") that is either the character '`'s'` or the empty string, depending on the value of `bottle`.
- ⑦ Do the same for `s2` (the second "s"), depending on the value of `next_bottle`.
- ⑧ Define a value for `next_num` depending on whether the next value is `0` or not.
- ⑨ Create a return string by joining the four lines of text on the newline. Substitute in the variables to create the correct verse.
- ⑩ Define a unit test called `test_verse` for the `verse` function. The prefix `test_` means that the `pytest` module will find this function and execute it.
- ⑪ Test the last `verse` with the value `1`.
- ⑫ Test a `verse` with the value `2`.
- ⑬ Use a `map` function to send all the values from counting down into the `verse` function. The `map`

function returns a new `list` of verses that can be joined on the string `\n\n` (two newlines).

Discussion

Defining the arguments

There isn't anything new with the `get_args` function in this program. By this point, you have had several opportunities to define an optional integer parameter with a default argument as well as using `parser.error` to halt your program if the user provides a bad argument. By relying on `argparse` to handle so much busy work, you are saving yourself loads of time as well as ensuring that you have good data to work with.

Counting down

We know how to count down from the given `--num`, and we know we can use a `for` loop to iterate:

```
>>> for n in range(3, 0, -1):
...     print(f'{n} bottles of beer')
...
3 bottles of beer
2 bottles of beer
1 bottles of beer
```

Instead of directly making up each verse in the `for` loop, I suggested in the introduction that you create a function called `verse` to create any given verse and use that with the `range` of numbers. Up to this point, we've been doing all our work right in the `main` function. As you grow as a programmer, your programs will become longer, hundreds to even thousands of lines of code (LOC). Long programs and functions can get very difficult to test and maintain, so you should try to break ideas into small, functional units that you can understand and test. Ideally, functions should do *one* thing. If you understand and trust your smaller functions, then you know you can put them together to achieve larger tasks.

Test-Driven Development

I wanted you to add the `test_verse` function to your program to use with `pytest` to create a working `verse` function. This idea follows the principles of *Test-Driven Development* described in that book by Kent Beck (2002):

1. Add a new test for an unimplemented unit of functionality.
2. Run all previously written tests and see the newly added test fails.
3. Write code that implements the new functionality.
4. Run all tests and see them succeed.
5. Refactor (rewrite to improve readability or structure).
6. Start at the beginning (repeat).

For instance, assume we want a function that adds 1 to any given number. We'll call it `add1` and

define the function body as `pass` to tell Python "nothing to see here":

```
def add1(n):
    pass
```

Now write a `test_add1` function where you pass some arguments to the function and use `assert` to verify that you get back the value that you expect.

```
def test_add1():
    assert add1(0) = 1
    assert add1(1) = 2
    assert add1(-1) = 0
```

Run `pytest` (or whatever testing framework you like) and verify that the function *does not work* (of course it won't because it just executes `pass`). Then go fill in the function code that *does work* (`return n + 1` instead of `pass`). Pass all manner of arguments you can imagine, including nothing, one thing, and many things. ^[4]

The `verse` function

I provided you with a `test_verse` function that shows you exactly what is expected for the arguments of 1 and 2. What I like about writing my tests first is that it gives me an opportunity to think about how I'd like to use the code, what I'd like to give as arguments, and what I expect to get back in return. For instance, what *should* the function `add1` return if given:



- no arguments?
- more than one argument?
- the value `None`?
- anything other than a numeric type (`int`, `float`, or `complex`) like a `str` value or a `dict`?

You can write tests to pass both good and bad values and decide how you want your code to behave under both favorable and adverse conditions.

Here's the `verse` function I wrote which passes the `test_verse` function:

```

def verse(bottle):
    """Sing a verse"""

    next_bottle = bottle - 1
    s1 = '' if bottle == 1 else 's'
    s2 = '' if next_bottle == 1 else 's'
    num_next = 'No more' if next_bottle == 0 else next_bottle
    return '\n'.join([
        f'{bottle} bottle{s1} of beer on the wall,',
        f'{bottle} bottle{s1} of beer,',
        f'Take one down, pass it around,',
        f'{num_next} bottle{s2} of beer on the wall!',
    ])

```

It's annotated above, but essentially I isolated all the parts of the return string that vary and created variables to substitute into those places. I need both the `bottle` and the `next_bottle` which I can then use to decide if there should be an "s" or not after the the "bottle" strings. I also need to figure out whether to print the next bottle as a number or the string "No more" when the `next_bottle` is `0`. Choosing the values for `s1`, `s2`, and `num_next` all involve *binary* decisions meaning they are a choice between *two* values, so I find it best to use an `if` expression.

This function passes `test_verse`, and so we can move on to using it to generate the song.

Iterating through the verses

I can use a `for` loop to count down and `print` each `verse`:

```

>>> for n in range(3, 0, -1):
...     print(verse(n))
...
3 bottles of beer on the wall,
3 bottles of beer,
Take one down, pass it around,
2 bottles of beer on the wall!
2 bottles of beer on the wall,
2 bottles of beer,
Take one down, pass it around,
1 bottle of beer on the wall!
1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
No more bottles of beer on the wall!

```

That's *almost* correct. I want an two newlines in between all the verses. I could use the `end` option to `print` to use 2 newlines for all values greater than 1:

```

>>> for n in range(3, 0, -1):
...     print(verse(n), end='\n' * (2 if n > 1 else 1))
...
3 bottles of beer on the wall,
3 bottles of beer,
Take one down, pass it around,
2 bottles of beer on the wall!

2 bottles of beer on the wall,
2 bottles of beer,
Take one down, pass it around,
1 bottle of beer on the wall!

1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
No more bottles of beer on the wall!

```

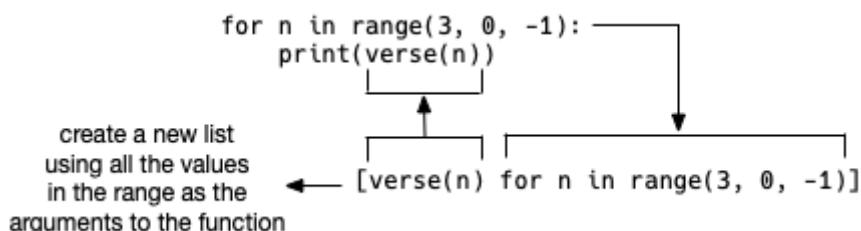


Figure 6. A `for` loop compared to a list comprehension.

I know that I could also use the `str.join` method to put 2 newlines in between items in a `list`. My items are the verses, and I know I can turn a `for` loop into a list comprehension:

```

>>> print('\n\n'.join([verse(n) for n in range(3, 0, -1)]))
3 bottles of beer on the wall,
3 bottles of beer,
Take one down, pass it around,
2 bottles of beer on the wall!

2 bottles of beer on the wall,
2 bottles of beer,
Take one down, pass it around,
1 bottle of beer on the wall!

1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
0 bottles of beer on the wall!

```

That is a fine solution; however, I would like you to start noticing a pattern we will see repeatedly. That is applying some function to every element of a sequence.

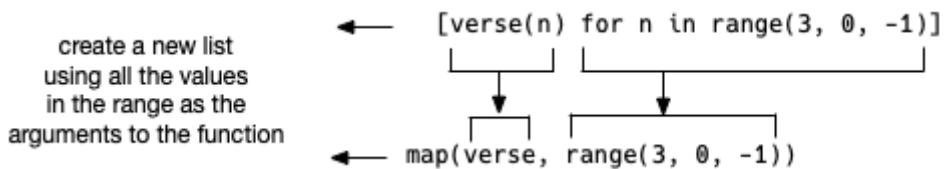


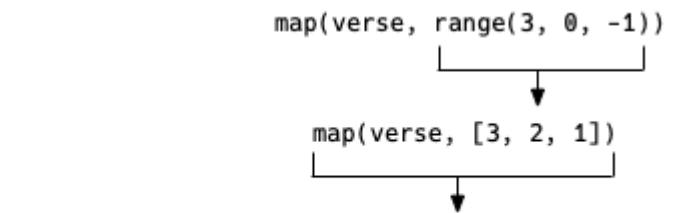
Figure 7. A list comprehension can be replaced with `map`. They both return a new list.

Here we have a descending `range` of numbers, and we want to send each number through the `verse` function to collect the results verses on the other end. It's like the paint booth idea in "Apples and Bananas" where the function "painted" the cars "blue" by adding the word "blue" to the front. When we want to apply a function to every element in a sequence, we might consider refactoring the code using `map`:

```
>>> print('\n\n'.join(map(verse, range(3, 0, -1))))
3 bottles of beer on the wall,
3 bottles of beer,
Take one down, pass it around,
2 bottles of beer on the wall!

2 bottles of beer on the wall,
2 bottles of beer,
Take one down, pass it around,
1 bottle of beer on the wall!

1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
No more bottles of beer on the wall!
```



a new list with
three verses

[verse(3), verse(2), verse(1)]

Whenever I need to transform some sequence of items by some function, I like to start off by thinking about I'll handle just *one* of the items. I find it's much easier to write and test one function with one input rather than some possibly huge list of operations. I tend to favor `map` or list comprehensions because they so naturally favor a *functional* approach—that is, one using functions. I really like writing and testing small functions, and so this approach works well for me. You may prefer to program differently, and that's fine, too.

If you want to use `map`, remember that it wants a *function* as the first argument and then a sequence of elements that will become arguments to the function. Our `verse` function (which we've tested!) is the first argument, and the `range` provides the `list`. Map will make each element of the `range` an argument to the `verse` function and will return a new `list` with the results of all those function calls. Many are the `for` multi-line loops that can be better written as mapping a function over a list of arguments!

1500 other solutions



There are literally hundreds of ways to solve this problem. The website <http://www.99-bottles-of-beer.net/> claims to have 1500 variations in various languages. Compare your solution to others there. Trivial as the actual program may be, it has allowed us to explore some really interesting ideas in Python, testing, and algorithms!

[4] A CS professor once told me in office hours to handle the cases of 0, 1, and n (infinity), and that has always stuck with me.

Review

- Test-Driven Development (TDD) is central to developing dependable, reproducible code. Tests also give you the freedom to refactor (reorganize and improve for speed or clarity) your code knowing that you can always verify your new version still works the same. As you write your code, always write tests!
- The `range` function will count backwards if you swap `start` and `stop` and supply the optional third `step` value of `-1`.
- A `for` loop can often be replaced with a list comprehension or a `map` for shorter, more concise code.

Going Further

- Replace the Arabic numbers (1, 2, 3) with text (one, two, three).
- Add a `--step` option (positive `int`, default `1`) that allows the user to skip, like by 2s or 5s.
- Add a `--reverse` flag to reverse the order of the verses, counting up instead of down.

Chapter 12: Ransom: Randomly captilizing text

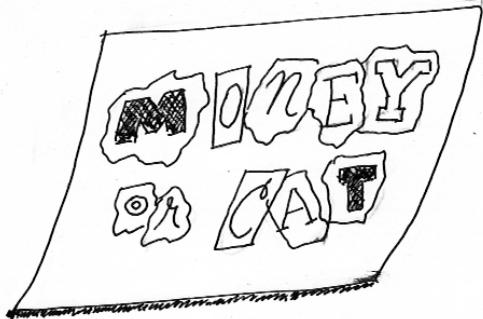


All this hard work writing code is getting on my nerves. I'm ready to turn to a life of crime! I've kidnapped (cat-napped?) the neighbor's cat. I want to send a ransom note to tell them my demands. In the good old days, I'd cut letters from magazines and paste them onto a piece of paper to spell out my demands. That sounds like too much work. Instead, I'm going to write a Python program called `ransom.py` that will encode text into randomly capitalized letters:

```
$ ./ransom.py 'give us 2 million dollars or the cat gets it!'
gIVE uS 2 MILLioN dollaRS OR tHe cAt GETS It!
```

As you can see, my diabolical program accepts the heinous input text as a positional argument. Since this program uses the `random` module, I want to accept a `-s` or `--seed` option so I can replicate the vile output.

```
$ ./ransom.py --seed 3 'give us 2 million dollars or the cat gets it!'
gIVE uS 2 MILLioN dollaRS OR tHe cAt GETS It!
```



The dastardly positional argument might name a vicious file, in which case that should be read for the demoniac input text:

```
$ ./ransom.py --seed 2 ./inputs/fox.txt  
the qUIck BROWN fOX JUmPs ovEr ThE LAZY D0g.
```

If the unlawful program is run with no arguments, it should print a short, infernal usage:

```
$ ./ransom.py
usage: ransom.py [-h] [-s int] str
ransom.py: error: the following arguments are required: str
```

If the nefarious program is run with `-h` or `--help` flags, it should print a longer, fiendish usage:

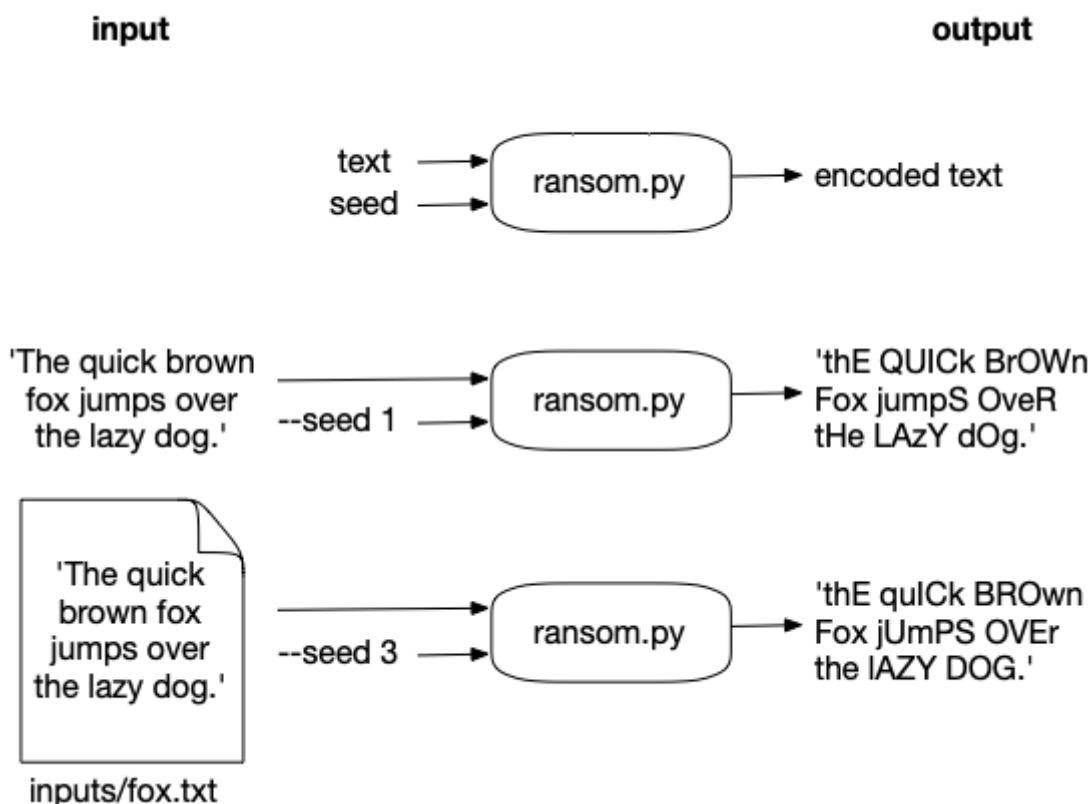
```
$ ./ransom.py -h
usage: ransom.py [-h] [-s int] str

Ransom Note

positional arguments:
  str                  Input text or file

optional arguments:
  -h, --help            show this help message and exit
  -s int, --seed int    Random seed (default: None)
```

Here is a noxious string diagram to visualize the inputs and outputs:



In this chapter, you will:

- Learn how to use the `random` module to figuratively "flip a coin" to decide between two choices.
- Explore ways to generate new strings from existing one, incorporating random decisions.
- Study the similarities of `for` loops, list comprehensions, and the `map` function

Writing `ransom.py`

I would suggest starting with `new.py` or copying the `template/template.py` file to create `ransom.py` in the `ransom` directory. This program, like several before it, accepts a required, positional string for the `text` and an optional integer (default `None`) for the `--seed`. Also as in previous exercises, the `text` argument may name a file that should be read for the `text` value.

To start out, use this for your `main` code:

```
def main():
    args = get_args()      ①
    random.seed(args.seed) ②
    print(args.text)       ③
```

- ① Get the processed command-line arguments.
- ② Set the `random.seed` with the value from the user. The default is `None` which is the same as not setting it.
- ③ Start off by echoing back the input.

If you run this program, it should echo the input from the command line:

```
$ ./ransom.py 'your money or your life!'
your money or your life!
```

Or from an input file:

```
$ ./ransom.py ../inputs/fox.txt
The quick brown fox jumps over the lazy dog.
```

The important thing when writing a program is to take baby steps. You should run your program *after every change*, checking manually and with the tests to see if you are progressing. Once you have this working, we can think about how to randomly capitalize this awful message.

Mutating the text

We've seen before that we can't directly modify a `str` value:

```
>>> text = 'your money or your life!'
>>> text[0] = 'Y'
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: 'str' object does not support item assignment
```

So how can we randomly change the case of some of the letters? I'd like to suggest that, instead of

thinking about how to change many letters, you should think about how to change *one* letter. That is, given a single letter, randomly return that the upper- or lowercase version of the letter. Maybe let's call this function `choose`. Here's a test for it:

```
def test_choose():
    random.seed(1)          ①
    assert choose('a') == 'a' ②
    assert choose('b') == 'b'
    assert choose('c') == 'c'
    assert choose('d') == 'd'
    random.seed(None).       ③
```

- ① The test sets the `random.seed` to `1` to start in order to ensure that the same "random" choices (which aren't really random) are made each time we run the test.
- ② The `choose` function is given a series of letters and uses the `assert` function to test if the value returned by the function is the expected letter.
- ③ Set the `random.seed` to `None` to unset it so that this won't affect the rest of the program.

Note that any time you set `random.seed`, it is a *GLOBAL* change. Every call to a `random` function will be affected by the seed, even if it's in a different function or a different module!

Random seeds

Have you wondered how I knew what would be the result of `choose` for a given random seed? Well, I confess that I wrote the function, then set the seed and ran it with given inputs. I recorded the results as the assertions you see. In the future, these results should still be the same. If they are not, I've changed something and probably broken my program.

Flipping a coin



You need to `choose` between return the upper- or lowercase version of the character you are given. It's a *binary* choice, meaning we have two options, so we can use the analogy of flipping a coin. Heads or tails? Or, for our purposes, 0 or 1.

```
>>> import random  
>>> random.choice([0, 1])  
1
```

Or `True` or `False` if you prefer:

```
>>> random.choice([False, True])  
True
```

Think about using an `if` expression where you return the uppercase answer when the `0` or `False` option is selected and the lowercase version otherwise. My entire `choose` function is this one line.

Creating a new string

I'd encourage you to start by mimicing the first approach from "Apples and Bananas." You can iterate through the characters of `text` using a `for` loop, using them as the argument to your `choose` function, and building a new `list` or `str` from the results. Once you can pass the test with a `for` loop, try to rewrite it as a list comprehension and then a `map`.

Now off you go! Write the program, pass the tests.

Solution

```
1 #!/usr/bin/env python3
2 """Ransom note"""
3
4 import argparse
5 import os
6 import random
7
8
9 # -----
10 def get_args():
11     """get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='Ransom Note',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text', metavar='str', help='Input text or file') ①
18
19     parser.add_argument('-s',      ②
20                         '--seed',
21                         help='Random seed',
22                         metavar='int',
23                         type=int,
24                         default=None)
25
26     args = parser.parse_args()    ③
27
28     if os.path.isfile(args.text): ④
29         args.text = open(args.text).read().rstrip()
30
31     return args                  ⑤
32
33
34 # -----
35 def choose(char): ⑥
36     """Randomly choose an upper or lowercase letter to return"""
37
38     return char.upper() if random.choice([0, 1]) else char.lower() ⑦
39
40
41 # -----
42 def test_choose(): ⑧
43     """Test choose"""
44
45     random.seed(1)            ⑨
46     assert choose('a') == 'a'. ⑩
47     assert choose('b') == 'b'
```

```

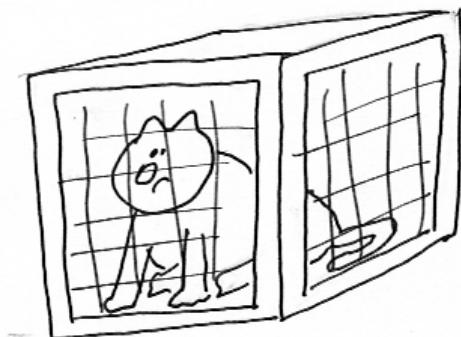
48     assert choose('c') == 'C'
49     assert choose('d') == 'd'
50     random.seed(None)          ⑪
51
52
53 # -----
54 def main():
55     """Make a jazz noise here"""
56     args = get_args()
57     text = args.text
58     random.seed(args.seed)      ⑫
59
60     # Method 1: Iterate each character, add to a list
61     ransom = []                ⑬
62     for char in args.text:      ⑭
63         ransom.append(choose(char)) ⑮
64
65     print''.join(ransom)        ⑯
66
67
68 # -----
69 if __name__ == '__main__':
70     main()

```

- ① The `text` argument is a positional string value.
- ② The `--seed` option is an integer that defaults to `None`
- ③ Process the command-line arguments into the `args` variable.
- ④ If the `args.text` is a file, use the contents of that as the new `args.text value.
- ⑤ Return the `args` to the caller.
- ⑥ Define a function to randomly return the upper- or lowercase version of a character.
- ⑦ Use the `random.choice` to select either `0` or `1` which, in the Boolean context of the `if` expression, evaluate to `False` and `True`, respectively.
- ⑧ Define a `test_choose` function that will be run by `pytest`. It takes no arguments.
- ⑨ Set the `random.seed` to a known value for the purposes of the test.
- ⑩ Use the `assert` function to verify that we get the expected result from the `choose` for a known argument.
- ⑪ Set the `random.seed` back to `None` so that our changes won't affect any other part of the program.
- ⑫ Set the `random.seed` to the given `args.seed` value. The default is `None`, which is the same as not setting it. That means the program will appear random when no seed is given but will be testable when we do provide one.
- ⑬ Create an empty list to hold the new `ransom` message.
- ⑭ Use a `for` loop to iterate through each character of `args.text`.
- ⑮ Append the chosen letter to the `ransom` list.

- ⑯ Join the `ransom` list on the empty string to create a new `str` to print.

Discussion



I like this problem because there are so many interesting ways to solve it. I know, I know, Python likes there to be "one obvious way" to solve it, but let's explore, shall we? There's nothing in the `get_args` that we haven't seen several times by now, so let's skip that.

Iterating through elements in a sequence

Assume that we have the following cruel message:

```
>>> text = '2 million dollars or the cat sleeps with the fishes!'
```

We want to randomly upper- and lowercase the letters. As suggested in the description of the problem, we can use a `for` loop to iterate over each character. One way to print an uppercase version of the `text` is to print an uppercase version *of each letter*:

```
for char in text:  
    print(char.upper(), end='')
```



That would give us "2 MILLION DOLLARS OR THE CAT SLEEPS WITH THE FISHES!" Now, instead of always printing `char.upper`, we could randomly choose between `upper` and `lower`. For that, let's use `random.choice` to choose between two values like `True` and `False` or `0` and `1`:

```
>>> import random  
>>> random.choice([True, False])  
False  
>>> random.choice([0, 1])  
0  
>>> random.choice(['blue', 'green'])  
'blue'
```

Following the first solution from "Apples and Bananas," we could create a new `list` to hold our ransom message and add these random choices:

```
ransom = []  
for char in text:  
    if random.choice([False, True]):  
        ransom.append(char.upper())  
    else:  
        ransom.append(char.lower())
```

Then we can join the new characters on the empty string to print a new string:

```
print''.join(ransom))
```

It's much less code to write this with an `if` expression to select whether to take the upper- or lowercase character.

```
ransom = []  
for char in text:  
    ransom.append(char.upper() if random.choice([False, True]) else char.lower())
```

The diagram illustrates the flow of the ternary operator in the assignment statement. It shows the expression `char.upper() if random.choice([False, True]) else char.lower()` being broken down into two parallel branches. The left branch, `char.upper()`, is connected by a line to the `char.upper()` part of the original expression. The right branch, `char.lower()`, is connected by a line to the `char.lower()` part of the original expression. Both branches converge at a point where they are connected to the `random.choice([False, True])` part of the expression, which is enclosed in brackets and has arrows pointing to it from both sides.

We don't have to use actual Boolean values (`False` and `True`). We could use `0` and `1` instead:

```
ransom = []  
for char in text:  
    ransom.append(char.upper() if random.choice([0, 1]) else char.lower())
```

When numbers are evaluated *in a Boolean context* (that is, in a place where Python expects to see a Boolean value), `0` is considered `False` and every other number is `True`.

Writing a function to choose the letter

The `if` expression is a bit of code that we could put into a function. I find it hard to read shoved inside the `ransom.append`. By putting it into a function, we can give it a descriptive name and write a test for it:

```
def choose(char):
    """Randomly choose an upper or lowercase letter to return"""

    return char.upper() if random.choice([0, 1]) else char.lower()
```

The name is short and descriptive, and now we can run the `test_choose` function to test that our function does what we think. This code is much easier to read:

```
ransom = []
for char in text:
    ransom.append(choose(char))
```

Another way to write `list.append`

`ransom.append(choose(char))`



`ransom += choose(char)`



add the result
of the function
to `ransom`

The above annotated solution creates an empty `list` to

which we `append` the return from `choose`. Another way to write `list.append` is using the `+=` operator to add the right-hand value (the element to add) to the left-hand side (the list):

```
def main():
    args = get_args()
    random.seed(args.seed)

    ransom = []
    for char in args.text:
        ransom += choose(char)

    print(''.join(ransom))
```

Using a `str` instead of a `list`

The `list` solution requires that `ransom` must be joined on the empty string to make a new string to print. We could, instead, start off with an empty string and build that up, one character at a time using the `+=` operator:

```
def main():
    args = get_args()
    random.seed(args.seed)

    ransom = ''
    for char in args.text:
        ransom += choose(char)

    print(ransom)
```

Using a `list comprehension`

We can shorten this to one line of code if we use a list comprehension:

```
def main():
    args = get_args()
    random.seed(args.seed)
    ransom = [choose(char) for char in args.text]
    print(''.join(ransom))
```

Or skip creating the `ransom` variable altogether:

```
def main():
    args = get_args()
    random.seed(args.seed)
    print(''.join([choose(char) for char in args.text]))
```

As a general rule, I only assign a value to a variable if I use it more than once or if I feel it makes my code more readable.

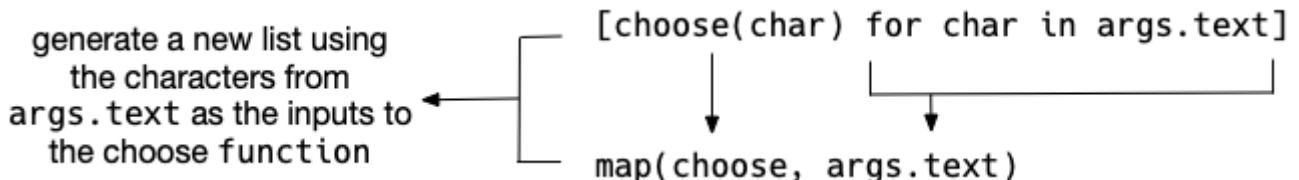
Using a `map` function

The `map` solution is fairly elegant and is a good bit less typing than the list comprehension. Remember that `map` returns a new `list` built by supplying each element of `args.text` to the `choose` function:

```
def main():
    args = get_args()
    random.seed(args.seed)
    ransom = map(choose, args.text)
    print(''.join(ransom))
```

Or, again, leave out the `ransom` assignment and use the `list` that comes back from `map` directly:

```
def main():
    args = get_args()
    random.seed(args.seed)
    print(''.join(map(choose, args.text)))
```



Comparing methods

It may seem silly to spend so much time working through so many ways to solve what is an essentially trivial problem, but one of the goals in this book is to explore the various ideas available in Python. The first method is a very imperative solution that a C or Java programmer would probably write. The version using a list comprehension is very idiomatic to Python—it is "Pythonic," as Pythonistas would say. The `map` solution would look very familiar to someone from a purely functional language like Haskell.

They all accomplish the same goal but embody different aesthetics and programming paradigms. My preferred solution would be the last one using `map`, but you should choose an approach that makes the most sense to you.

MapReduce

In 2004, Google released a paper on their "MapReduce" algorithm. The "map" phase applies some transformation to all the elements in a collection such as all the pages of the Internet that need to be indexed for searching. These operations can happen in *parallel*, meaning you can use many machines to process the pages separately from each other and in any order. The "reduce" phase then brings all the processed elements back together, maybe to put the results into a unified database.

In our `ransom.py` program, the "map" part selected a randomized case for the given letter, and the "reduce" part was putting all those bits back together into a new string. Conceivably, `map` could make use of multiple processors to run the functions *in parallel* as opposed to *sequentially* (like with a `for` loop), possibly cutting the time to produce the results.

Learning about map/reduce was, to me, a bit like learning the name of a new bird. I never even noticed that bird before, but, once I was told its name, I saw it everywhere. Once you understand this pattern, you'll begin see recognize it in many places!

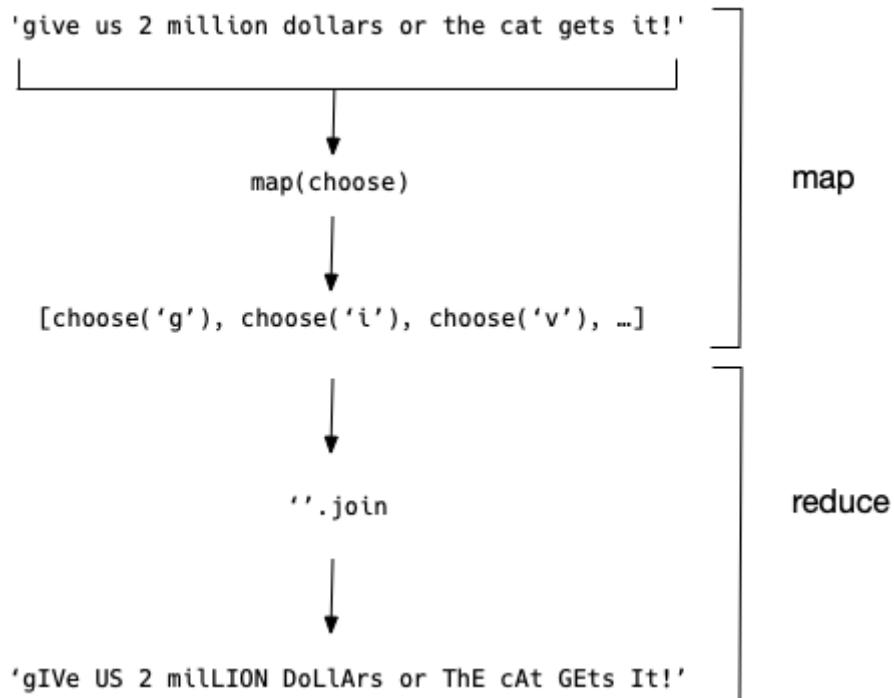


Figure 8. MapReduce

Review



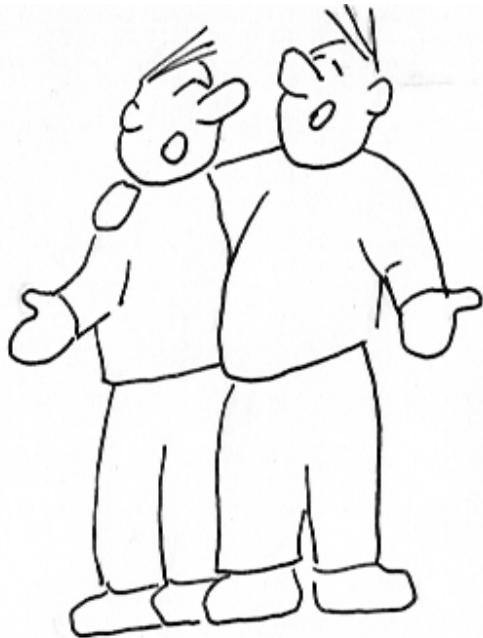
- Whenever you have lots of things to process, try to think about how you'd process just one of them.
- Write a test that helps you imagine how you'd like to use the a function to process one item. What will you pass in, and what do you expect back?
- Write your function to pass your test. Be sure to think about what you'll do with both good and bad input.
- To apply your function to each element in your input, use a `for` loop, a list comprehension, or a `map`.

Going Further

- Write a version that uses other ways of representing letters by combining ASCII characters such as the following. Feel free to make up your own substitutions. Be sure to update your tests.

A	4	K	<
B	3	L	_
C	(M	V
D)	N	\\
E	3	P	`
F	=	S	5
G	(-	T	+
H	-	V	\ /
I	1	W	\ /\ /
J	_		

Chapter 13: Twelve Days of Christmas: Algorithm design



Perhaps one of the worst songs of all time and the one that is sure to ruin my Christmas spirit is "The Twelve Days of Christmas." WILL IT EVER STOP!? AND WHAT IS WITH ALL THE BIRDS?! Still, it's pretty interesting to write an algorithm to generate the song starting from any given day because you have to count *up* as you add each verse (day) and then count *down* inside the verses (recapitulating the previous days' gifts). We'll be able to build off what you learned writing the "99 Bottles of Beer" song.

Our program will be called `twelve_days.py` and will generate the "Twelve Days of Christmas" song up to a given `-n` or `--num` argument (default `12`). Note that there should be two newlines in between each verse but only one at the end:

```
$ ./twelve_days.py -n 3
On the first day of Christmas,
My true love gave to me,
A partridge in a pear tree.
```

```
On the second day of Christmas,
My true love gave to me,
Two turtle doves,
And a partridge in a pear tree.
```

```
On the third day of Christmas,
My true love gave to me,
Three French hens,
Two turtle doves,
And a partridge in a pear tree.
```

The text will be printed to `STDOUT` unless there is a `-o` or `--outfile` argument, in which case the text

should be placed inside a file by that name. Note there should be 113 lines of text for the entire song:

```
$ ./twelve_days.py -o song.txt  
$ wc -l song.txt  
113 song.txt
```

In this exercise, you will:

- Create an algorithm to generate "The Twelve Days of Christmas" from any given day in the range 1-12.
- Reverse a list
- Use the `range` function
- Write text to a file or to `STDOUT`

Writing twelve_days.py

As always, I suggest you start by running `new.py twelve_days.py` in the `twelve` directory or by copying the `template/template.py` file. Your program should will take two options:

1. `-n` or `--num`: an `int` with a default of 12
2. `-o` or `--outfile`: a `str` with a default of nothing/`STDOUT`

When run with the `-h` or `--help` flag, the program should print a usage:

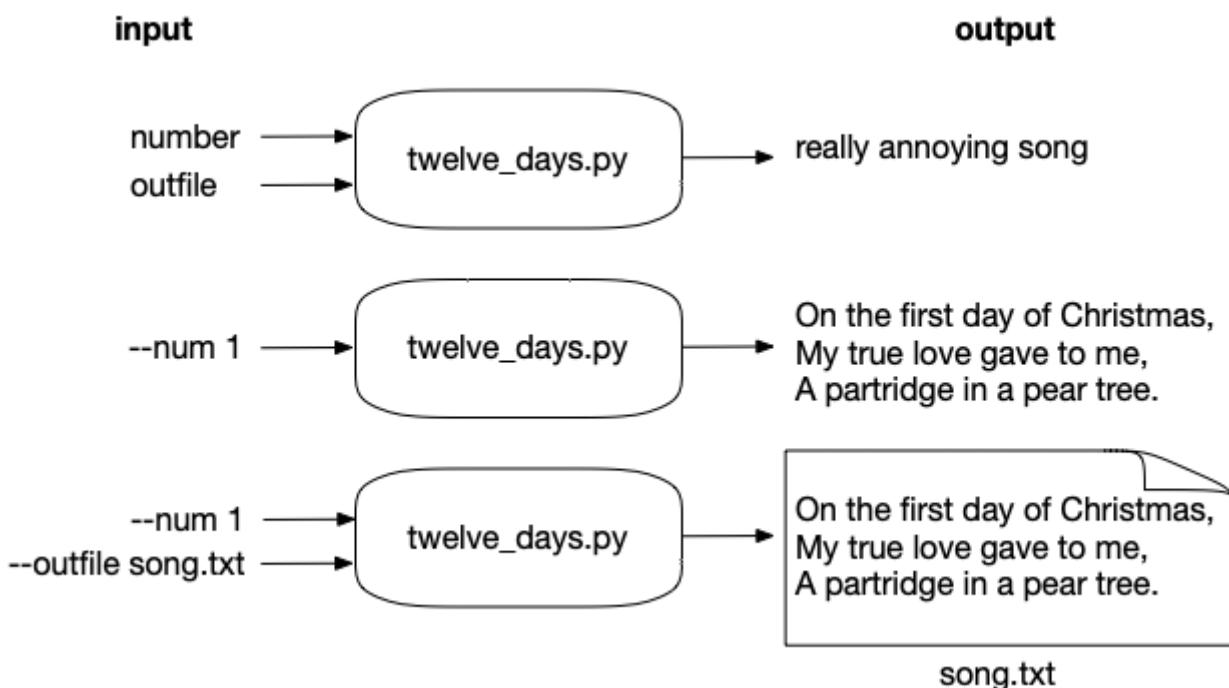
```
$ ./twelve_days.py -h
usage: twelve_days.py [-h] [-n int] [-o str]

Twelve Days of Christmas

optional arguments:
  -h, --help            show this help message and exit
  -n int, --num int    Number of days to sing (default: 12)
  -o str, --outfile str
                        Outfile (STDOUT) (default: )
```

Before trying to write the song, make your usage match the above. At this point, you should pass the first two tests when you run the test suite.

Here is a holly, jolly string diagram to get you in the mood for writing:



The program should complain if the `--num` value is not in the range 1-12. I suggest you check this inside the `get_args` function and use `parser.error` to halt with an error and usage:

```
$ ./twelve_days.py -n 21
usage: twelve_days.py [-h] [-n int] [-o str]
twelve_days.py: error: --num "21" must be between 1 and 12
```

Once you've handled the bad `--num`, you should pass the first three tests.

Counting

In the "99 Bottles of Beer" song, we need to count down from a given number. Here we need to count up to the `--num`. The `range` function will give us what we need, but we must remember to start at `1` (the default is `0`) and that *the upper bound is not included*:

```
>>> num = 3
>>> list(range(1, num))
[1, 2]
```

You need to add `1` to whatever you're given for `--num`:

```
>>> list(range(1, num + 1))
[1, 2, 3]
```

Let's start by printing something like the first line of each verse:

```
>>> for day in range(1, num + 1):
...     print(f'On the {day} day of Christmas,')
...
On the 1 day of Christmas,
On the 2 day of Christmas,
On the 3 day of Christmas,
```

At this point, I'm starting to think about how we wrote "99 Bottles of Beer." There we ended up creating a `verse` function that would generate any *one* verse. Then we used `str.join` to put them all together with two newlines. I suggest we try the same approach here.

```
def verse(day):
    """Create a verse"""
    return f'On the {day} day of Christmas,'
```

Let's see how we can use it:

```
>>> for day in range(1, num + 1):
...     print(verse(day))
...
```

```
On the 1 day of Christmas,
On the 2 day of Christmas,
On the 3 day of Christmas,
```

Notice that the function will not `print` the string but will `return` the verse so that we can test it:

```
>>> assert verse(1) == 'On the 1 day of Christmas,'
```

Here's a simple (but incorrect) `test_verse` function we could start off with:

```
def test_verse():
    """ Test verse """
    assert verse(1) == 'On the 1 day of Christmas,'
    assert verse(2) == 'On the 2 day of Christmas,'
```

Add the the `verse` and `test_verse` functions to your `twelve_days.py` program and then run `pytest twelve_days.py` to verify your code.

Creating the ordinal value

Maybe the first thing to do is to change the numeric value to its ordinal position, that is "1" to "first," "2" to "second." You could use a dictionary like we used in "Jump The Five" to associate each `int` value 1-12 with its `str` value. That is, you might create a new `dict` called `ordinal`:

```
>>> ordinal = {} # what goes here?
```

So that you can do:

```
>>> ordinal[1]
'first'
>>> ordinal[2]
'second'
```

You could also use a `list` if you think about how you could use the each `day` in the `range` to index into a `list` of the ordinal strings.

```
>>> ordinal = [] # what goes here?
```

How would you use the `int` value of a given number to access one of the values in the `list` called `ordinal`?

Now your `verse` might look something like:

```
def verse(day):
    """Create a verse"""
    ordinal = [] # something here!
    return f'On the {ordinal[day]} of Christmas,'
```

You can update your test with your expectations:

```
def test_verse():
    """ Test verse """
    assert verse(1) == 'On the first day of Christmas,'
    assert verse(2) == 'On the second day of Christmas,'
```

Once you have this working, you should be able to replicate something like this:

```
>>> for day in range(1, num + 1):
...     print(verse(day))
...
On the first day of Christmas
On the second day of Christmas
On the third day of Christmas
```

If you put the `test_verse` function inside your `twelve_days.py` program, you can verify that your `verse` function works by running `pytest twelve_days.py`. The `pytest` module will run any function that has a name starting with `test_`.

Shadowing

You might be tempted to use the variable name `ord`, and you would be allowed by Python to do this. The problem is that Python has function called `ord` that returns "the Unicode code point for a one-character string":

```
>>> ord('a')  
97
```

Python will not complain if you define a variable or another function with the name `ord`:

```
>>> ord = {}
```

Such that you could do this:

```
>>> ord[1]  
'first'
```

But then it overwrites the actual `ord` function, and so breaks a function call:

```
>>> ord('a')  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
TypeError: 'dict' object is not callable
```

This is called "shadowing," and it's quite dangerous. Any code in the scope of the shadowing would be affected by the change.

Tools like `pylint` can help you find problems like this in your programs. Assume the following code:

```
$ cat shadow.py  
#!/usr/bin/env python3  
  
ord = {}  
print(ord('a'))
```

Here is what `pylint` has to say:

```
$ pylint shadow.py
*****
Module shadow
shadow.py:3:0: W0622: Redefining built-in 'ord' (redefined-builtin)
shadow.py:1:0: C0111: Missing module docstring (missing-docstring)
shadow.py:4:6: E1102: ord is not callable (not-callable)
```

```
-----
Your code has been rated at -25.00/10
```

It's good to double-check your code with tools like [pylint](#), [flake8](#), and [mypy](#)!

Making the verses

Now that we have the basic structure of the program, let's focus on creating the *correct* output. We'll update our `test_verse` with the actual values for the first two verses. You can, of course, add more tests, but presumably if we can manage the first two, then we can handle all the other days:

```
def test_verse():
    """Test verse"""

    assert verse(1) == '\n'.join([
        'On the first day of Christmas,', 'My true love gave to me,', 
        'A partridge in a pear tree.'
    ])

    assert verse(2) == '\n'.join([
        'On the second day of Christmas,', 'My true love gave to me,', 
        'Two turtle doves,', 'And a partridge in a pear tree.'
    ])
```

If you add this to your `twelve_days.py` program, you can run `pytest twelve_days.py` to see how your `verse` function is failing:

```

===== FAILURES =====
----- test_verse -----

def test_verse():
    """Test verse"""

>     assert verse(1) == '\n'.join([
        'On the first day of Christmas,', 'My true love gave to me,',
        'A partridge in a pear tree.')
E     AssertionError: assert 'On the first...of Christmas,' == 'On the first ... a
pear tree.'
E         - On the first day of Christmas, ②
E         + On the first day of Christmas, ③
E         ?
E             +
E         + My true love gave to me,
E         + A partridge in a pear tree.

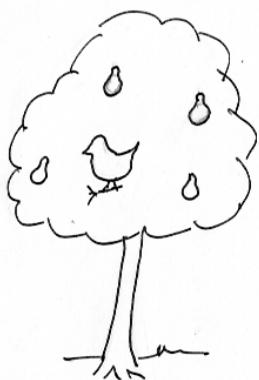
twelve_days.py:88: AssertionError
===== 1 failed in 0.11 seconds =====

```

- ① The leading > shows this is the code that is creating an exception. We are running `verse(1)` and asking if it's equal to the expected verse.
- ② This is the text that `verse(1)` actually produced, which is only the first line of the verse.
- ③ The lines following are what was expected.

Now we need to supply the rest of the lines for each verse. They all start off the same:

On the {ordinal[day]} day of Christmas,
My true love gave to me,



And then we need to add these gifts for each day:

1. A partridge in a pear tree
2. Two turtle doves
3. Three French hens
4. Four calling birds

5. Five gold rings
6. Six geese a laying
7. Seven swans a swimming
8. Eight maids a milking
9. Nine ladies dancing
10. Ten lords a leaping
11. Eleven pipers piping
12. Twelve drummers drumming

Note that for every day greater than 1, the last line changes "A partridge..." to "And a partridge in a pear tree."

Each verse needs to count backwards from the given `day`. For example, if the `day` is 3, then:

3. Three French hens
2. Two turtle doves
1. And a partridge in a pear tree

We've talked before in the "Picnic" chapter about how to reverse a `list` either with the `list.reverse` method or the `reversed` function, and we also used these ideas in "99 Bottles," so this code should not be unfamiliar:

```
>>> day = 3
>>> for n in reversed(range(1, day + 1)):
...     print(n)
...
3
2
1
```

Try to make the function return the first two lines and then the count down of the days:

```
>>> print(verse(3))
On the third day of Christmas,
My true love gave to me,
3
2
1
```

And then, instead of `3 2 1`, add the actual gifts:

```
>>> print(verse(3))
On the third day of Christmas,
My true love gave to me,
Three French hens,
Two turtle doves,
And a partridge in a pear tree.
```

If you can get that to work, then you ought to be able to pass the `test_verse` test.

Using the `verse` function

Once you have that working, think about a final structure that calls your `verse`. It could be a `for` loop:

```
verses = []
for day in range(1, args.num + 1):
    verses.append(verse(day))
```

A list comprehension:

```
verses = [verse(day) for day in range(1, args.num + 1)]
```

Or a `map`:

```
verses = map(verse, range(1, args.num + 1))
```



Printing

Once you have all the verses, you can use the `str.join` method to print the output. The default is to print this to "standard out" (`STDOUT`), but the program will also take an optional `--outfile` that names a file to write the output. I suggest you go back to the "Howler" solution for how to handle that. You ought to be able to copy the relevant code exactly to determine the proper file handle. If you do need to write to an `--outfile`, be sure to `open` it with the proper permissions. Again, see the "Howler" exercise for a review of that.

Time to write

It's not at all mandatory that you solve the problem the way that I describe. The "correct" solution is the one that you write and understand which passes the test suite. It's fine if you like the idea of creating a functions for `verse` and using the provided test. It's fine if you want to go another way, but do try to think of writing small functions *and tests* to solve small parts of your problem and combining them to solve the larger problem.

If you need more than one sitting or even several days, take your time. Sometimes a good walk or a nap can do wonders for solving problems. Don't neglect your hammock or a nice cup of tea.

Solution

```
1 #!/usr/bin/env python3
2 """Twelve Days of Christmas"""
3
4 import argparse
5 import sys
6
7
8 # -----
9 def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Twelve Days of Christmas',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('-n', ①
17                         '--num',
18                         help='Number of days to sing',
19                         metavar='int',
20                         type=int,
21                         default=12)
22
23     parser.add_argument('-o', ②
24                         '--outfile',
25                         help='Outfile (STDOUT)',
26                         metavar='str',
27                         type=str,
28                         default='')
29
30     args = parser.parse_args() ③
31
32     if args.num not in range(1, 13): ④
33         parser.error(f"--num '{args.num}' must be between 1 and 12") ⑤
34
35     return args
36
37
38 # -----
39 def main():
40     """Make a jazz noise here"""
41
42     args = get_args() ⑥
43     out_fh = open(args.outfile, 'wt') if args.outfile else sys.stdout ⑦
44     out_fh.write('\n\n'.join(map(verse, range(1, args.num + 1))) + '\n') ⑧
45     out_fh.close() ⑨
46
47
```

```

48 # -----
49 def verse(day): ⑩
50     """Create a verse"""
51
52     ordinal = [ ⑪
53         'first', 'second', 'third', 'fourth', 'fifth', 'sixth', 'seventh',
54         'eighth', 'ninth', 'tenth', 'eleventh', 'twelfth'
55     ]
56
57     gifts = [ ⑫
58         'A partridge in a pear tree.',
59         'Two turtle doves,',
60         'Three French hens,',
61         'Four calling birds,',
62         'Five gold rings,',
63         'Six geese a laying,',
64         'Seven swans a swimming,',
65         'Eight maids a milking,',
66         'Nine ladies dancing,',
67         'Ten lords a leaping,',
68         'Eleven pipers piping,',
69         'Twelve drummers drumming,',
70     ]
71
72     lines = [ ⑬
73         f'On the {ordinal[day - 1]} day of Christmas,',
74         'My true love gave to me,'
75     ]
76
77     lines.extend(reversed(gifts[:day])) ⑭
78
79     if day > 1: ⑮
80         lines[-1] = 'And ' + lines[-1].lower() ⑯
81
82     return '\n'.join(lines) ⑰
83
84
85 # -----
86 def test_verse(): ⑱
87     """Test verse"""
88
89     assert verse(1) == '\n'.join([
90         'On the first day of Christmas,', 'My true love gave to me,', 
91         'A partridge in a pear tree.'
92     ])
93
94     assert verse(2) == '\n'.join([
95         'On the second day of Christmas,', 'My true love gave to me,', 
96         'Two turtle doves,', 'And a partridge in a pear tree.'
97     ])
98

```

```
99
100 # -----
101 if __name__ == '__main__':
102     main()
```

- ① The `--num` option is an `int` with a default of 12.
- ② The `--outfile` option is a `str` with a default of either `None` or the empty string.
- ③ Capture the results of parsing the command-line arguments (`parser.parse_args`) into the `args` variable.
- ④ Check that the given `args.num` is in the allowed range (1-12, inclusive).
- ⑤ If `args.num` is invalid, use `parser.error` to print a short usage and the error message to `STDERR` and exit the program with an error value. Note that the error message includes the bad value for the user and explicitly states that a good value should be in the range 1-12.
- ⑥ Get the command-line arguments. Remember that all argument validation happens inside `get_args`. If this call succeeds, then we have good arguments from the user.
- ⑦ The output file handle (`out_fh`) is either the result of opening the given `args.outfile` for writing or is `sys.stdout`.
- ⑧ Use `map` to generate all the verses using the `range` function to generate a list of the days up to the given number. The verses are joined on two newlines. We need to include a final (single) newline because the `fh.write` method does not automatically add a newline (unlike `print`, which does).
- ⑨ Close the output file handle. This is also the last statement when the program runs.
- ⑩ Define a function to create any one verse from a given number 1-12.
- ⑪ The `ordinal` values for the numbers are a `list` of `str` values.
- ⑫ The `gifts` for the days is a `list` of `str` values.
- ⑬ The `lines` of each verse start off the same, substituting in the ordinal value of the given `day`.
- ⑭ I use the `list.extend` method to add the `gifts`, which are a slice from the given `day` and then `reversed`.
- ⑮ Check if this is for a `day` greater than 1.
- ⑯ Change the last of the `lines` to add "And " to the beginning appended to the lowercased version of the line.
- ⑰ Return the `lines` joined on the newline.
- ⑱ The unit test for the `verse` function.

Discussion

Defining the arguments

Nothing in the `get_args` is new, so we'll throw a sidelong, cursory glace. The `--num` option is an `int` value with a default value of 12. The `--outfile` option is a `str` with a default of either `None` or the empty string because both will evaluate to `False` when we check later. If the `--num` value is not in `range(1, 13)` (remembering that 13 is not included, so just the numbers 1-12), I use `parser.error` to show an error message and the short usage and exit with a non-zero exit value to indicate a failure. Notice that the error message provides feedback in showing the bad value and explaining exactly what is allowed.

Defining the output handle

If the user provides a value for `--outfile`, we'll use that as the name of a file to open. If they do not, we want to use `sys.stdout`. Because this is a binary decision (there are two options), I prefer to use an `if` expression. I always name variables that are file handles with `fh` somewhere in the name, so here the output file handle is `out_fh`.

```
out_fh = open(args.outfile, 'wt') if args.outfile else sys.stdout
```

Note that, if we `open` the output file, we must use `'wt'` for "write" and "text" modes. This is exactly the same code as from "Howler," so I would recommend you review that discussion.

Making one verse

I chose to make a function called `verse` to create any one verse given an `int` value of the `day`:

```
def verse(day):
    """Create a verse"""
```

	index	day
ordinal = [
'first',	0	1
'second',	1	2
'third',	2	3
'fourth',	3	4
'fifth',	4	5
'sixth',	5	6
'seventh',	6	7
'eighth',	7	8
'ninth',	8	9
'tenth',	9	10
'eleventh',	10	11
'twelfth'	11	12
]		

I decided to use a **list** to represent the **ordinal** value of the **day**:

```
ordinal = [
    'first', 'second', 'third', 'fourth', 'fifth', 'sixth', 'seventh',
    'eighth', 'ninth', 'tenth', 'eleventh', 'twelfth'
]
```

Since the **day** is based on counting from 1 but Python lists start from 0, I have to subtract 1:

```
>>> day = 3
>>> ordinal[day - 1]
'third'
```

I could have just as easily used a **dict**:

```
ordinal = {
    1: 'first', 2: 'second', 3: 'third', 4: 'fourth',
    5: 'fifth', 6: 'sixth', 7: 'seventh', 8: 'eighth',
    9: 'ninth', 10: 'tenth', 11: 'eleventh', 12: 'twelfth',
}
```

Then you don't have to subtract 1:

```
>>> ordinal[3]
'third'
```

Whatever works for you. I also used a **list** for the **gifts**:

```

gifts = [
    'A partridge in a pear tree.',
    'Two turtle doves,',
    'Three French hens,',
    'Four calling birds,',
    'Five gold rings,',
    'Six geese a laying,', 
    'Seven swans a swimming,', 
    'Eight maids a milking,', 
    'Nine ladies dancing,', 
    'Ten lords a leaping,', 
    'Eleven pipers piping,', 
    'Twelve drummers drumming,', 
]

```

gifts[:3]

'A partridge in a pear tree.',	0
'Two turtle doves,',	1
'Three French hens,',	2
'Four calling birds,',	3
'Five gold rings,',	4
'Six geese a laying,',	5
'Seven swans a swimming,',	6
'Eight maids a milking,',	7
'Nine ladies dancing,',	8
'Ten lords a leaping,',	9
'Eleven pipers piping,',	10
'Twelve drummers drumming,'	11



This makes a bit more sense as I

can use a list slice to get the `gifts` for a given `day`:

```

>>> gifts[:3]
['A partridge in a pear tree.', 
 'Two turtle doves,', 
 'Three French hens,']

```

But I want them in reverse order. The `reversed` function is lazy, so I need to use the `list` function to coerce the values in the REPL:

```

>>> list(reversed(gifts[:3]))
['Three French hens,', 
 'Two turtle doves,', 
 'A partridge in a pear tree.']

```

The first two lines of any verse are the same, substituting in the `ordinal` value for the `day`.

```
lines = [
    f'On the {ordinal[day - 1]} day of Christmas,',
    'My true love gave to me,'
]
```

I need to put these two `lines` together with the `gifts`. Since each verse is made of some number of lines, I think it will make sense to use a `list` to represent the entire verse. So I need to add the `gifts` to the `lines`, and I can use the `list.extend` method to do that:

```
>>> lines.extend(reversed(gifts[:day]))
```

And now there are 5 `lines`:

```
>>> lines
['On the third day of Christmas,',
 'My true love gave to me,',
 'Three French hens,',
 'Two turtle doves,',
 'A partridge in a pear tree.']
>>> assert len(lines) == 5
```

Note that I cannot use the `list.append` method. It's easy to confuse these two methods. The `list.extend` method takes another `list` as the argument, expands it, and adds all of the individual elements to the original `list`. The `list.append` method is meant to add one element to the `list`, so, if you give it a `list`, it will tack that entire `list` on to the end of the original list.

Here the `reversed` iterator will be added the end of `lines` such that it would have three elements rather than the desired five:

```
>>> lines.append(reversed(gifts[:day]))
>>> lines
['On the third day of Christmas,',
 'My true love gave to me,',
 <list_reverseiterator object at 0x105bc8588>]
```

Maybe you're thinking you could coerce the `reversed` with a `list`? Thinking you are, young Jedi, but, alas, that will still add a new `list` to the end:

```
>>> lines.append(list(reversed(gifts[:day])))
>>> lines
['On the third day of Christmas,',
 'My true love gave to me,',
 ['Three French hens,', 'Two turtle doves,', 'A partridge in a pear tree.']]
```

And we still have 3 `lines` rather than 5:

```
>>> len(lines)
3
```

If the `day` is greater than 1, I need to change the last line to say "And a" instead of "A":

```
if day > 1:
    lines[-1] = 'And ' + lines[-1].lower()
```

Note that this is another good reason to represent the `lines` as a `list` because the elements of a `list` are *mutable*. I could have represented the `lines` as a `str`, but strings are *immutable*, so it would be much harder to change the last line.

I want to return a single `str` value from the function, so I join the `lines` on a newline:

```
>>> print('\n'.join(lines))
On the third day of Christmas,
My true love gave to me,
Three French hens,
Two turtle doves,
A partridge in a pear tree.
```

My function returns the joined `lines` and will pass the `test_verse` function I provided.

Generating the verses

Given the `verse` function, I can create all the needed verses by iterating from 1 to the given `--num`. I could collect them in `list` of `verses`:

```
day = 3
verses = []
for n in range(1, day + 1):
    verses.append(verse(n))
```

I can check that I have the right number of verses:

```
>>> assert len(verses) == day
```

Whenever you see this pattern of creating an empty `str` or `list` and then using a `for` loop to add to it, consider instead using a list comprehension:

```
>>> verses = [verse(n) for n in range(1, day + 1)]  
>>> assert len(verses) == day
```

I personally prefer using `map` over list comprehensions. I will use the `list` function to coerce the lazy `map` function here, but it's not necessary in the program code:

```
>>> verses = list(map(verse, range(1, day + 1)))  
>>> assert len(verses) == day
```

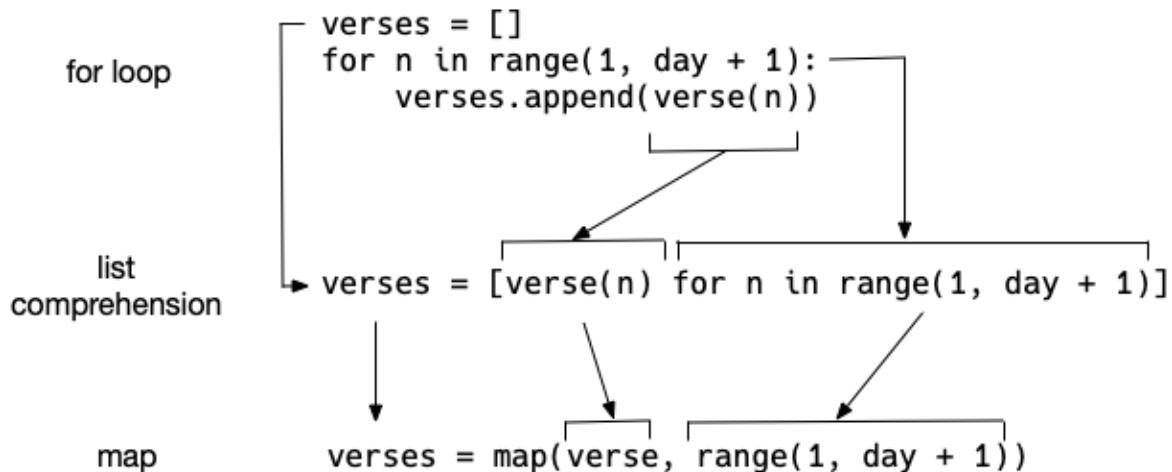


Figure 9. Building a `list` using an `for loop`, a `list comprehension`, and `map`.

All of these methods will produce the correct number of verses. Choose whichever one makes the most sense to you.

Printing the verses

Just like with the "99 Bottles of Beer," I want to `print` them with two newlines in between. The `str.join` method is a good choice:

```
>>> print('\n\n'.join(verses))  
On the first day of Christmas,  
My true love gave to me,  
A partridge in a pear tree.
```

```
On the second day of Christmas,  
My true love gave to me,  
Two turtle doves,  
And a partridge in a pear tree.
```

```
On the third day of Christmas,  
My true love gave to me,  
Three French hens,  
Two turtle doves,  
And a partridge in a pear tree.
```

You can use the `print` method with the optional `file` argument to put the text into an open filehandle. The `out_fh` will be either the file indicated by the user or `sys.stdout`:

```
print('\n\n'.join(map(verse, range(1, args.num + 1))), file=out_fh)
```



Or you can use the `out_fh.write` method, but you need to remember to add the trailing newline that `print` adds for you:

```
out_fh.write('\n\n'.join(map(verse, range(1, args.num + 1))) + '\n')
```

Finally, you need to close the file handle:

```
out_fh.close()
```

There are dozens to hundreds of ways to write this algorithm just as there are for "99 Bottles of Beer." If you came up with an entirely different approach which passed the test, that's terrific! Please share it with me. I wanted to stress the idea of how to write, test, and use a single `verse` function, but I'd love to see other approaches!

Review

- There are many ways to encode algorithms to perform repetitive tasks. In my version, I wrote and tested a function to handle one task and then mapped a range of input values over that.
- The `range` function will return `int` values between a given start and stop value, the latter of which is not included.
- You can use the `reversed` function to reverse the values returned by `range`.
- To `open` a file for writing, you must include the '`w`' flag. We were writing text (not binary data), so we also included '`t`' to make the open mode '`wt`'.
- The `sys.stdout` file handle is always open and available for writing.
- Modeling the `gifts` as a `list` allowed me to use a list slice to get all the gifts for a given day. I used the `reversed` function to put them into the right order for the song.
- I modeled the `lines` as a `list` because a `list` is mutable which I needed in order to change the last line when the day is greater than 1.
- Shadowing a variable or function is reusing an existing variable or function name. If, for instance, you create a variable with the name of an existing function, then the function is effectively hidden because of the shadow. Avoid shadowing by using tools like `pylint` to find these and many other common coding problems.

Going Further

- Install the `emoji` module (<https://pypi.org/project/emoji/>) and print various emojis for the gifts rather than text. For instance, you could use '`:bird:`' to print `\u260d` for every "bird" like a hen or dove. I also used '`:man:`', '`:woman:`', and '`:drum:`', but you can use whatever you like:

```
On the twelfth day of Christmas,  
My true love gave to me,  
Twelve \u260d drumming,  
Eleven \u260d piping,  
Ten \u260d a leaping,  
Nine \u260d dancing,  
Eight \u260d a milking,  
Seven \u260d a swimming,  
Six \u260d a laying,  
Five gold \u260d,  
Four calling \u260d,  
Three French \u260d,  
Two turtle \u260d,  
And a \u260d in a pear tree.
```

- Write the "Wheels on the Bus" song. Play "golf" with your code. In the actual game of golf, you want to score as low as possible. Each time you swing the club is a stroke you count, so when you "golf" your code, you try to use the fewest number of characters. For fun, search the Internet for "Python golf Tetris" to see how other programmers can implement a working game of Tetris in astoundingly small programs. = Chapter 14: Rhymer: Using regular expressions to create rhyming words



Did you write your own input file for the "Gashlycrumb" exercise? When I was writing mine, I would come up with a word like maybe "cyanide" and wonder what I could rhyme with that. Mentally I start with the first consonant sound of the alphabet and start substituting "b" for "byanide," skip "c" because that's already the first character, then "d" for "dynanide," and so forth. This is effective but tedious, so I decided to write a program to do this for me, as one does.

This is basically another find-and-replace type of program that we've seen before like swapping all the numbers in a string in "Jump The Five" or all the vowels in a string in "Apples and Bananas." We wrote those programs using very manual, *imperative* methods like iterating through all the characters of a string, comparing them to some wanted value, and possibly returning a new value. In the solution for "Apples and Bananas," we briefly touched on "regular expressions" (AKA "regexes" [\[5\]](#)) which provide us a *declarative* way to describe a pattern of text. Now we're going to really dig into regexes to see what we can do!

In this exercise, we're going to take a given word and create "words" that rhyme. For instance, the word "bake" rhymes with words like "cake," "make," and "thrake," the last of which isn't actually a dictionary word but just a new string we create by replacing the "b" in "bake" with "thr." The algorithm we'll use is to split a word into any initial consonants and the rest of the word, so "bake" is split into "b" and "ake." We replace the "b" with all the other consonants from the alphabet plus these consonant clusters:

```
bl br ch cl cr dr fl fr gl gr pl pr sc sh sk sl sm sn sp st  
sw th tr tw thw wh wr sch scr shr sph spl spr squ str thr
```

Be sure the output is sorted. For instance, these are the first three words our program will produce for "cake":

```
$ ./rhymer.py cake | head -3  
bake  
blake  
brake
```

And the last three:

```
$ ./rhymer.py cake | tail -3  
xake  
yake  
zake
```

We'll replace any leading consonants with a list of other consonant sounds to create a total of 56 words:

```
$ ./rhymer.py cake | wc -l  
56
```

Note that we'll replace *all* the leading consonants, not just the first one. For instance, with the word "chair" we need to replace "ch":

```
$ ./rhymer.py chair | tail -3  
xair  
yair  
zair
```

If a word like "apple" does not start with a consonant, then we'll append all the consonant sounds to the beginning to create words like "bapple" and "shrapple."

```
$ ./rhymer.py apple | head -3  
bapple  
blapple  
brapple
```

Because there is no consonant to *replace*, words that start with a vowel will produce 57 rhyming words:

```
$ ./rhymer.py apple | wc -l  
57
```



To make this a bit easier, the output should always be all lowercase even if the input has uppercase letters:

```
$ ./rhymer.py GUITAR | tail -3  
xuitar  
yuitar  
zuitar
```

If a word contains *nothing but consonants*, then print a message that the word cannot be rhymed:

```
$ ./rhymer.py RDNZL  
Cannot rhyme "RDNZL"
```

The task of finding these initial consonants is made significantly easier with regexes.

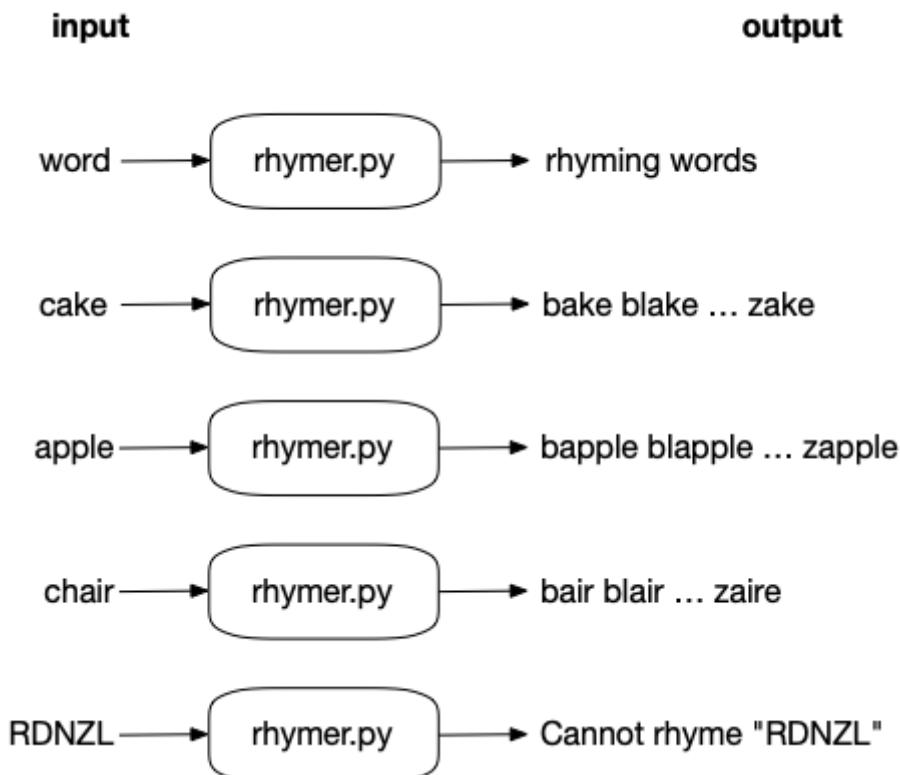
In this program, you will:

- Learn to write and use regular expressions
- Use a guard with a list comprehension
- Explore the similarities of list comprehension and guard to the `filter` function
- Entertain ideas of "truthiness" when evaluating Python types in a boolean context

[5] Pronounced with a hard "g" like in "George"

Writing `rhymer.py`

The program takes a single, positional argument which is the string to rhyme. Here is a snazzy, jazzy, frazzy, thwazzy string diagram:



If given no arguments or the `-h` or `--help` flags, it should print a usage statement:

```
$ ./rhymer.py -h
usage: rhymer.py [-h] str

Make rhyming "words"

positional arguments:
  str      A word to rhyme

optional arguments:
  -h, --help show this help message and exit
```

Breaking a word

To me, the main problem of the program is breaking the given word into the leading consonant sounds and the rest (something like the "stem") of the word. I wrote a function called `stemmer`, and this is my test for it:

```

def test_stemmer():
    """ Test stemmer """
    assert stemmer('') == ('', '')           ①
    assert stemmer('cake') == ('c', 'ake')    ②
    assert stemmer('chair') == ('ch', 'air')  ③
    assert stemmer('APPLE') == ('', 'apple') ④
    assert stemmer('RDNZL') == ('rdnzl', '') ⑤

```

The tests cover the following good and bad inputs:

- ① The empty string.
- ② A word with a single leading consonant.
- ③ A word with a leading consonant cluster.
- ④ A word with no initial consonants. Also an uppercase word, so checking that lowercase is returned.
- ⑤ A word with no vowels.

My `stemmer` function always returns a 2-tuple of the `(start, rest)` of the word, so the argument "cake" produces a tuple with two values of `('c', 'ake')`. The argument "chair" is split into the leading "ch" and the "air" strings. The argument "APPLE" has no `start` and only the `rest` of the word, which is lowercase. A string with no vowels like "RDNZL" returns the (lowercased) string as the `start` and nothing for the `rest`.

The first and last tests use "bad" values that can't be rhymed. It's up to me as the programmer to decide what to do with those. Should my `stemmer` function throw an exception or return `None` to indicate a failure? In this case, I decided to always return the same tuple and let the calling code deal with the problem. If there is nothing in the second position of the tuple, then there is nothing to rhyme. You don't have to make the same decisions I made. If you prefer to throw an exception for unrhymable words, then change your test to handle an exception.

Using regular expressions

It's certainly *possible* to write this program without regular expressions, but I hope you'll see how radically different using regexes can be from manually writing your own search-and-replace code. To start off, we need to bring in the `re` module:

```
>>> import re
```

I would encourage you to then read `help(re)` to get a feel for all that you can do with regexes. They are a deep subject with books ^[6] and whole branches of academia devoted to them. We will only scratch the surface of what they can do.

We need to write a regex that will find consonants at the beginning of a string. We can define consonants as the characters of the English which are not the vowels, "a," "e," "i," "o," and "u." Our `stemmer` will only return lowercase letters, and so there are only 21 consonants we need to define.

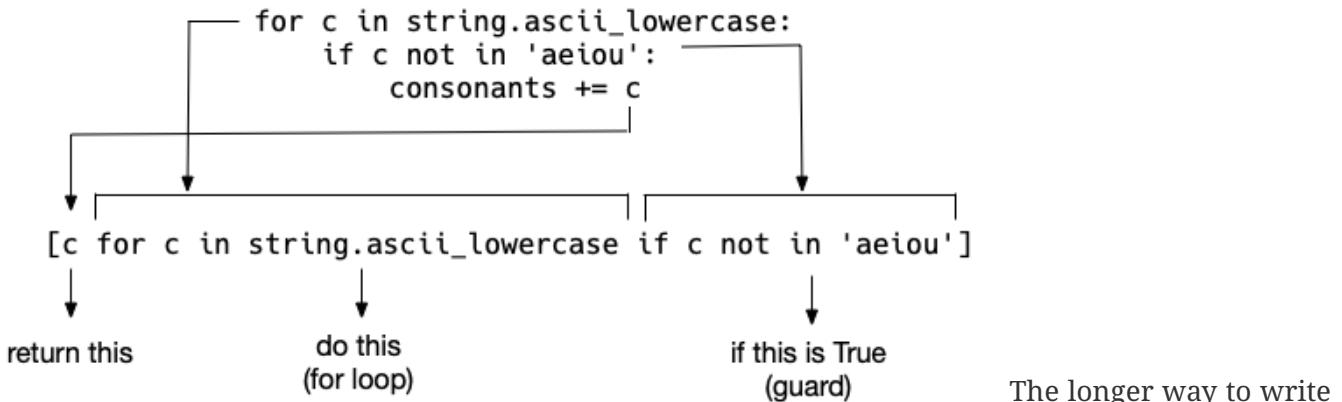
You could write them out, but I'd rather write a bit of code!

I can start with `string.ascii_lowercase`:

```
>>> import string  
>>> string.ascii_lowercase  
'abcdefghijklmnopqrstuvwxyz'
```

And use a list comprehension with a "guard" clause to filter out the vowels. As I want a `str` of consonants and not a `list`, I use `str.join` to make a new `str` value:

```
>>> consonants = ''.join([c for c in string.ascii_lowercase if c not in 'aeiou'])  
>>> consonants  
'bcdfghjklmnpqrstvwxyz'
```



this with a `for` loop and an `if` statement is this:

```
consonants = ''  
for c in string.ascii_lowercase:  
    if c not in 'aeiou':  
        consonants += c
```

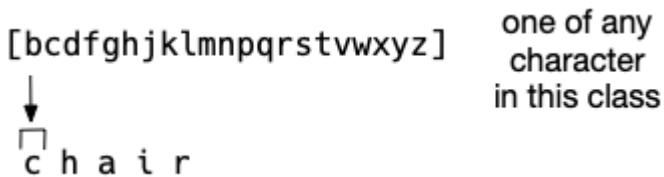
To use the `consonants` in a regex, I need to create a "character class" that includes all these values, which I can do by putting the characters inside square brackets:

```
>>> pattern = '[' + consonants + ']'  
>>> pattern  
'[bcdfghjklmnpqrstvwxyz]'
```

The `re` module has two search-like functions called `match` and `search`, and I always get them confused. They both look for a `pattern` (the first argument) in some `text`, but the `re.match` function starts *from the beginning* of the `text` while the `re.search` function will match starting *anywhere* in the `text`.

As it happens, `re.match` is just fine because we are looking for consonants at the beginning of a

string.



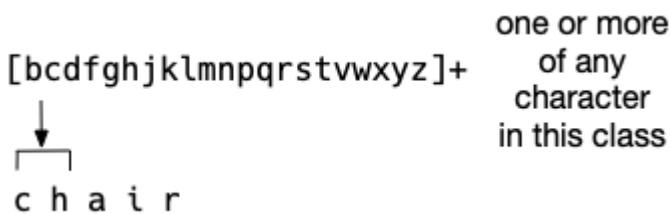
```
>>> text = 'chair'  
>>> re.match(pattern, text)           ①  
<re.Match object; span=(0, 1), match='c'> ②
```

- ① Try to match the given `pattern` in the given `text`. If this succeeds, we get an `re.Match` object; otherwise, the value `None` is returned.
- ② The match was successful, so we see a "stringified" version of the `re.Match` object.

The `match='c'` shows us that the regular expression found the string '`c`' at the beginning. Both the `re.match` and `re.search` functions will return an `re.Match` object on success. You can read `help(re.Match)` to learn more about all the cool things you can do with them:

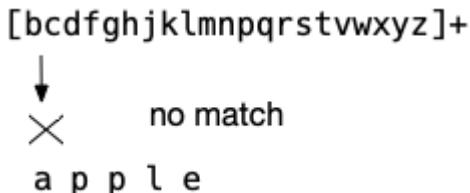
```
>>> match = re.match(pattern, text)  
>>> type(match)  
<class 're.Match'>
```

How do we get our regex to match the letters '`ch`'? We can put a '`+`' sign after the character class to say we want *one or more*. (Does this sound a bit like `nargs='+'` to say one or more arguments?) I will use an f-string here to create the pattern:



```
>>> re.match(f'[{consonants}]+', 'chair')  
<re.Match object; span=(0, 2), match='ch'>
```

What does it give us for a string with no leading consonants like "apple"?



```
>>> re.match(f'[{consonants}]+', 'apple')
```

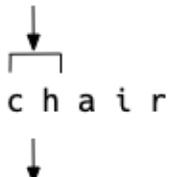
It appears like we got nothing back from that. What is the **type** of that return value?

```
>>> type(re.match(f'[{consonants}]+', 'apple'))
<class 'NoneType'>
```

Both the `re.match` and `re.search` functions return `None` to indicate a failure to match any text. We know that only some words will have a leading consonant sound, so this is not surprising. We'll see in a moment how to make this an optional match.

Using capture groups

([bcdfghjklmnpqrstvwxyz]+)



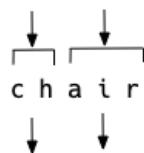
('ch',)

It's all well and good to have found (or not) the leading consonants, but the goal here is to split the `text` into two parts: the consonants (if any) and the rest of the word. We can wrap parts of the regex in parentheses to create "capture groups." If the regex matches successfully, we can recover the parts using the `re.Match` object's `groups` method:

```
>>> match = re.match(f'([{consonants}]+)', 'chair')
>>> match.groups()
('ch', )
```

1 or more
consonants 1 or more of
anything

([bcdfghjklmnpqrstvwxyz]+)(.+)



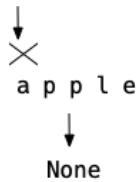
('ch', 'air') To capture everything that comes after the `consonants`, we can use the `.` to match anything and add `+` to mean one or more. We put that into parens to capture it:

```
>>> match = re.match(f'([{consonants}]+)(.+)', 'chair')
>>> match.groups()
('ch', 'air')
```

What happens when we try to use this on "apple"? It fails to make the first match on the consonants, and so *the whole match fails* and returns `None`:

1 or more consonants 1 or more of anything

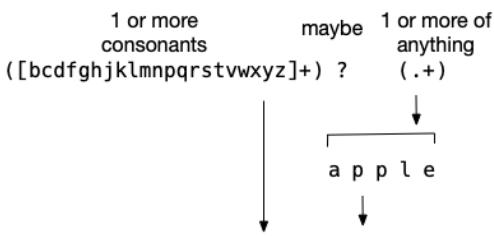
([bcdfghjklmnpqrstvwxyz]+)(.+)



```
>>> match = re.match(f'({{consonants}})+(.+)', 'apple')
>>> match.groups()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'groups'
```

Remember that `re.match` returns `None` when it fails to find the pattern. We can add `?` at the end of the `consonants` pattern to make it optional:

```
>>> match = re.match(f'({{consonants}})+?(.+)', 'apple')
>>> match.groups()
(None, 'apple')
```



The `match.groups` function returns a `tuple` containing the matches for each grouping created by the parentheses. You can also use the `match.group` (singular) with a group number to get a specific group. Note that these start numbering from 1:

```
>>> match.group(1) ①
>>> match.group(2) ②
'apple'
```

- ① There was no match for the first group on "apple," so this is a `None`.
- ② The second group captured the entire word.

If you match on the "chair," there are values for both groups:

```
>>> match = re.match(f'([{consonants}]+)?(.+)', 'chair')
>>> match.group(1)
'ch'
>>> match.group(2)
'air'
```

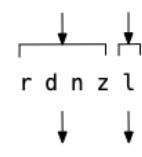
So far we've only dealt with lowercase text because our program will always emit lowercase values. Still, let's explore what happens when we try to match uppercase text:

```
>>> match = re.match(f'([{consonants}]+)?(.+)', 'CHAIR')
>>> match.groups()
(None, 'CHAIR')
```

Not surprisingly, that fails. Our pattern only defines lowercase characters. We could add all the uppercase consonants, but it's a bit easier to use a third optional argument to `re.match` to tell indicate case-insensitive searching:

```
>>> match = re.match(f'([{consonants}]+)?(.+)', 'CHAIR', re.IGNORECASE)
>>> match.groups()
('CH', 'AIR')
```

maybe 1 or 1 or more of
 more consonants anything
 ([bcd...uvwxyz]+)? (.+)



('rdn', 'z l')

What do you get when you search on text that has nothing but consonants?

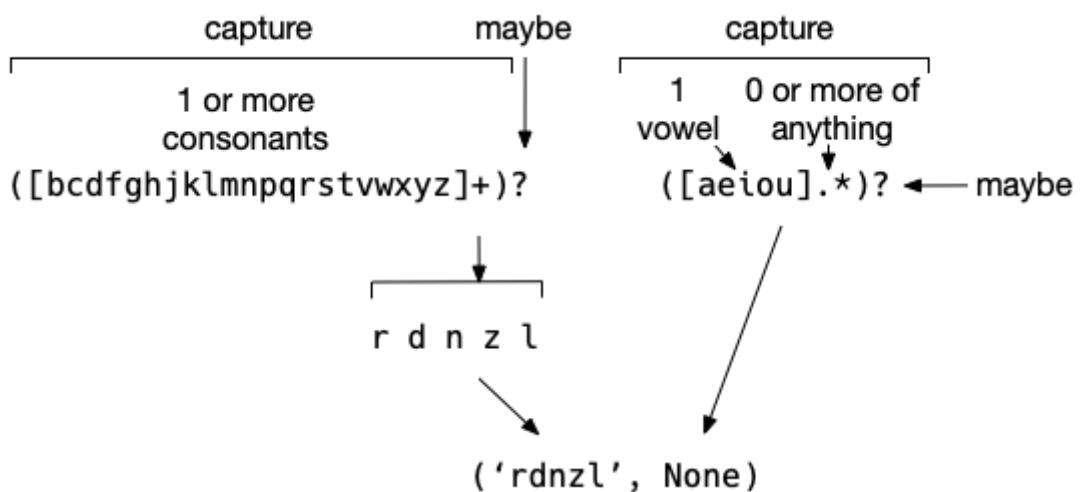
```
>>> match = re.match(f'([{consonants}]+)?(.+)', 'rdnzl')
>>> match.groups()
('rdn', 'z l')
```

Were you expecting the first group to include *all* the consonants and the second group to have nothing? It might seem a bit odd that it decided to split off the "l" into the last group, but we have to think *extremely literally* about how the regex engine is working. We described an optional group of one or more consonants that *must be followed* by one or more of anything else. The "l" counts as one or more of anything else, so the regex matched exactly what we requested. If we change the `(.+)` to `(.*?)` to make it *zero or more*, then it works as expected:

```
>>> match = re.match(f'(\{consonants\}]+)?(.*)', 'rdnzl')
>>> match.groups()
('rdnzl', '')
```

Our regex is not quite complete. We want to indicate that there should be a vowel after the consonants, and then anything else. We can use another character class to describe the vowels, followed by *zero or more* (*) of anything (.). The whole second capture group is optional, so we add ?:

```
>>> match = re.match(f'(\{consonants\}]+)?([aeiou].*)?', 'rdnzl')
>>> match.groups()
('rdnzl', None)
```



Truthiness

The last thing is to return the correct values from the `stemmer` function. I decided that I would always return a 2-tuple of (`start`, `rest`), and that I would always use the empty string to denote a missing value rather than a `None`:

```
return (match.group(1) or '', match.group(2) or '') if match else ('', '')
```

I'm using the `or` operator to decide between something on the left *or* something on the right. The `or` will return the first "truthy" value, the one that sort of-kind of evaluates to `True` in a boolean context:

```
>>> True or False      ①
True
>>> False or True     ②
True
>>> 1 or 0             ③
1
>>> 0 or 1              ④
1
>>> 0.0 or 1.0          ⑤
1.0
>>> '0' or ''            ⑥
'0'
>>> 0 or False           ⑦
False
>>> [] or ['foo']         ⑧
['foo']
>>> {} or dict(foo=1)    ⑨
{'foo': 1}
```

- ① It's easiest to see with literal `True` and `False` values.
- ② No matter the order, the `True` value will be taken.
- ③ In a boolean context, the `int` value `0` is "falsey" and any other value is "truthy."
- ④ The number values behave exactly like actual boolean values.
- ⑤ `float` values also behave like `int` values.
- ⑥ With `str` values, the empty string is the "falsey" and so anything else is "truthy." It may look odd because it returns `'0'` but that's not the *numeric* value zero but the *string* we use to represent the value of zero. Wow, such philosophical.
- ⑦ If no value is "truthy," then the last value is returned.
- ⑧ The empty `list` is "falsey," and so any non-empty `list` is "truthy."
- ⑨ The same is true of dictionaries. The empty `dict` is "falsey," and any non-empty `dict` is "truthy."

If you use the above regex and `return` in your `stemmer`, it should pass the `test_stemmer` test.

Creating the output

Let's review what the program should do:

- Take a positional string argument.
- Try to split it into two parts: any leading consonants and the rest of the word.
- If the split is successful, combine the "rest" of the word (which might actually be the entire word if there are no leading consonants) with all the other consonant sounds listed above. Be sure to *not* include the original consonant sound and to sort the rhyming strings.
- If you are unable to split the word, then print the message `Cannot rhyme "<word>"`.

OK, time to step off onto the dance floor. Do your best to write this yourself before looking at the solution.

[6] *Mastering Regular Expressions* by Jeffrey Friedl is one I would recommend

Solution

```
1 #!/usr/bin/env python3
2 """Make rhyming words"""
3
4 import argparse
5 import re      ①
6 import string
7
8
9 # -----
10 def get_args():
11     """get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='Make rhyming "words"',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('word', metavar='str', help='A word to rhyme')
18
19     return parser.parse_args()
20
21
22 # -----
23 def main():
24     """Make a jazz noise here"""
25
26     args = get_args()          ②
27     prefixes = list('bcdfghjklmnpqrstvwxyz') + (      ③
28         'bl br ch cl cr dr fl fr gl gr pl pr sc '
29         'sh sk sl sm sn sp st sw th tr tw thw wh wr '
30         'sch scr shr sph spl spr squ str thr').split()
31
32     start, rest = stemmer(args.word)          ④
33     if rest:                                ⑤
34         print('\n'.join(sorted([p + rest for p in prefixes if p != start])))
⑥
35     else:
36         print(f'Cannot rhyme "{args.word}"')    ⑦
37
38
39 # -----
40 def stemmer(word):
41     """Return leading consonants (if any), and 'stem' of word"""
42
43     vowels = 'aeiou'                      ⑧
44     consonants = ''.join(                ⑨
45         [c for c in string.ascii_lowercase if c not in vowels])
46     pattern = (                            ⑩
```

```

47      '([' + consonants + ']+)?' # capture one or more consonants, optional
48      '('                      # start capture
49      '[' + vowels + ']'       # at least one vowel
50      '.*'                   # zero or more of anything else
51      ')')?'                 # end capture, optional group
52      match = re.match(pattern, word.lower())           ⑪
53      return (match.group(1) or '', match.group(2) or '') if match else ('', '')
⑫
54
55
56 # -----
57 def test_stemmer(): ⑬
58     """test the stemmer"""
59
60     assert stemmer('') == ('', '')
61     assert stemmer('cake') == ('c', 'ake')
62     assert stemmer('chair') == ('ch', 'air')
63     assert stemmer('APPLE') == ('', 'apple')
64     assert stemmer('RDNZL') == ('rdnzl', '')
65
66
67 # -----
68 if __name__ == '__main__':
69     main()

```

- ① The `re` module is for regular expressions.
- ② Get the command-line arguments.
- ③ Define all the prefixes that will be added to create rhyming words.
- ④ Split the `word` argument into two possible parts. Because the `stemmer` function always returns a 2-tuple, we can unpack the values into two separate values.
- ⑤ Check if there is a part of the word that we can use to create rhyming strings.
- ⑥ If there is, use a list comprehension to iterate through all the prefixes and add them to the stem of the word. Use a guard to ensure that any given prefix is not the same as the beginning of the word. Sort all the values and print them, joined on newlines.
- ⑦ If there is nothing we can use to create rhymes, let the user know.
- ⑧ Since I will use the `vowels` more than once, I assign them to a variable.
- ⑨ The `consonants` are the letters that are not `vowels`. I will only match to lowercase letters.
- ⑩ The `pattern` is defined using consecutive literal strings that Python will join together into one string. By breaking up the pieces onto separate lines, I can comment on each part of the regular expression.
- ⑪ Use the `re.match` function to start matching *at the beginning* of the the given `word` which I convert to lowercase.
- ⑫ Even though I would expect this pattern to never fail, I conservatively use an `if` expression to ensure that I have a value for the `match`. If I do, I return a 2-tuple consisting of the two capture groups, but I ensure that neither group is a `None` but rather an empty string to denote a missing

value. If there is no `match`, I return a tuple of two empty strings.

- ⑬ The afore-described test function for the `stemmer` function.

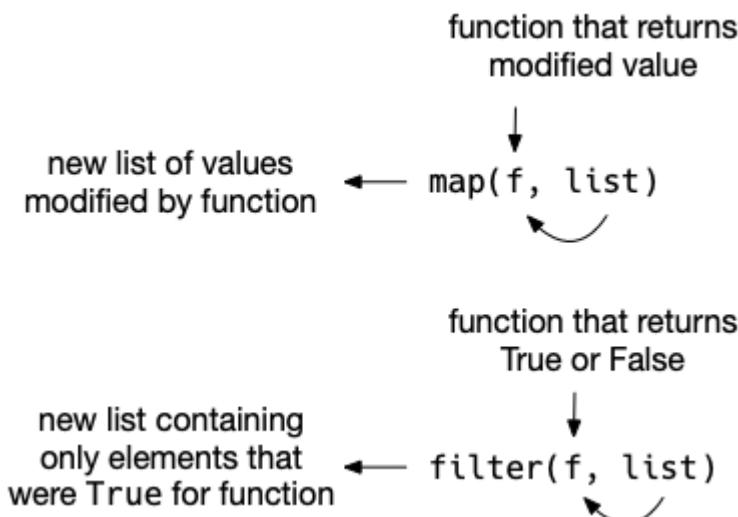
Discussion

There are many ways you could have written this, but, as always, I wanted to break the problem down into some units I could write and test. For me, this came down to splitting the word into a possible leading consonant sound and the rest of the word. If I can manage that, I can create rhyming strings; if I cannot, then I need to alert the user.

Stemming a word

For the purposes of this program, the "stem" of a word is the part after any initial consonants which I define using a list comprehension with a guard to take only the letters that are not vowels:

```
>>> vowels = 'aeiou'  
>>> consonants = ''.join([c for c in string.ascii_lowercase if c not in vowels])
```



I showed how this is a more concise way to write a `for` loop with an `if` statement. We've looked at `map` several times now and talked about how it is a *higher-order function* (HOF) because it takes *another function* as the first argument, applying it to all the members of a sequence to produce a new, transformed sequence (like painting cars blue). Here I'd like to introduce another HOF called `filter` which takes a function and some *iterable* (something that can be *iterated* like a `list`). As with `map`, the function is applied to all the elements of the sequence. The return value of function will be evaluated by its "truthiness." Only those elements that evaluate as "truthy" will be returned by `filter`.

Here is another way to write the idea of the list comprehension using a `filter`:

```
>>> consonants = ''.join(filter(lambda c: c not in vowels, string.ascii_lowercase))
```

Just as with `map`, I use the `lambda` keyword to create an *anonymous function*. The `c` is the name of the variable that will hold the argument which, in this case, will be each character from `string.ascii_lowercase`. The entire body of the function is the evaluation `c not in vowels`. Each of the vowels will return `False` for this:

```
>>> 'a' not in vowels  
False
```

And each of the consonants will return **True**:

```
>>> 'b' not in vowels  
True
```

Therefore only the consonants will be allowed to pass through the **filter**. To think back to our "blue" cars, let's write a **filter** that only accepts cars that start with the string "blue ":

```
>>> list(filter(lambda car: car.startswith('blue '), ['blue Honda', 'red Chevy', 'blue Ford']))  
['blue Honda', 'blue Ford']
```

When the **car** variable has the value "red Chevy," the **lambda** returns **False**, and so that value is rejected:

```
>>> car = 'red Chevy'  
>>> car.startswith('blue ')  
False
```

Note that if none of the elements from the original iterable are accepted, then **filter** will produce an empty **list** `[]`. For example, I could **filter** for numbers greater than 10. Note that **filter** is another *lazy* function that I must coerce using the **list** function in the REPL:

```
>>> list(filter(lambda n: n > 10, range(0, 5)))  
[]
```

A list comprehension would also return an empty list:

```
>>> [n for n in range(0, 5) if n > 10]  
[]
```

Here is a diagram showing the relationship of creating a new **list** called **consonants** using an imperative, **for**-loop approach, an idiomatic list comprehension with a guard, and a purely functional approach using **filter**. All of these are perfectly acceptable, though the most Pythonic way is probably the list comprehension. The **for** loop would be very familiar to a C or Java programmer, while the **filter** would be immediately recognizable to the Haskeller or even someone from a Lisp-like language. The **filter** might be slower than the list comprehension, especially if the iterable were large. Choose whichever way makes more sense for your style and application.

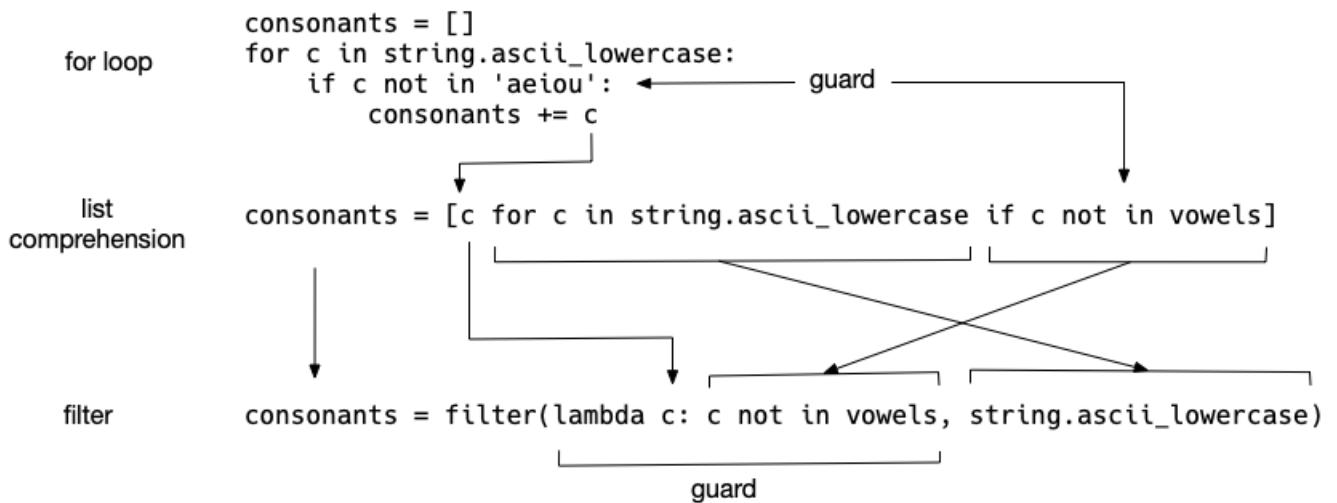


Figure 10. The `filter` function is similar to a list comprehension with a guard.

Writing a regular expression

We talked in the introduction about the individual parts of the regular expression I ended up using. I'd like to take a moment to mention the way I formatted the regex in the code using an interesting trick of the Python interpreter that will stitch together string literals into one long string if they are enclosed in a grouping like parentheses:

```
>>> this_is_just_to_say = ('I have eaten '
... 'the plums '
... 'that were in '
... 'the icebox')
>>> this_is_just_to_say
'I have eaten the plums that were in the icebox'
```

Note that there are no commas after each string as that would create a `tuple` with 4 individual strings:

```
>>> this_is_just_to_say = ('I have eaten ',
... 'the plums ',
... 'that were in ',
... 'the icebox')
>>> this_is_just_to_say
('I have eaten ', 'the plums ', 'that were in ', 'the icebox')
```

The advantage of writing out the regular expression like this is to add comments to each important part:

```
pattern = (
    '([' + consonants + ']+)?' # capture one or more consonants, optional
    '('                      # start capture
    '[' + vowels + ']'       # at least one vowel
    '.*'                     # zero or more of anything else
    ')?')'                   # end capture, optional group
```

I could have written the entire regex on one line. Ask yourself which version would you rather read and maintain:^[7]

```
pattern = '([' + consonants + ']+)?([' + vowels + '].*)?'
```

I really worked hard on creating the regular expression, and I have a good bit of confidence that this regex will never fail to match something, even the empty string, because both elements (the leading consonant cluster and the rest of the word) are optional. Still, I will very conservatively call the `re.match` so I can check the return value which I called `match`. Note that I lowercase the `word`:

```
match = re.match(pattern, word.lower())
```

It's possible my function might receive some value which might cause the `re.match` to return a `None`. I use the `match` value in an `if` expression to be absolutely sure that I can safely call `match.group`. As stated in the introduction, I would prefer to use the empty string rather than `None` to indicate a missing group, so I use the `or` operator to choose the first "truthy" value for both groups. If there is no `match`, I return a 2-tuple of empty strings:

```
return (match.group(1) or '', match.group(2) or '') if match else ('', '')
```

Testing and using the `stemmer` function

I first introduced the `test_stemmer` so we could think about what values we might pass in, both good and bad, and what we might expect to receive:

```
def test_stemmer():
    """test the stemmer"""
    assert stemmer('') == ('', '')
    assert stemmer('cake') == ('c', 'ake')
    assert stemmer('chair') == ('ch', 'air')
    assert stemmer('APPLE') == ('', 'apple')
    assert stemmer('RDNZL') == ('rdnzl', '')
```

I really recommend that you always write a `test_x` function *before* you write the `x` function as it can help you think of the function as this black box. Something goes in, something comes out, and what happens inside is a separate concern.

One of the very interesting things about Python code is that your `rhymers.py` program is also—kind of, sort of—a *module* of code. That is, you haven’t explicitly written it to be this container of reusable functions, but it is. You can even run the functions from inside the REPL. For this to work, be sure you run `python3` inside the same directory as the `rhymers.py` code:

```
>>> from rhymers import stemmer
```

And now you can run and test your `stemmer` function manually:

```
>>> stemmer('apple')
('', 'apple')
>>> stemmer('banana')
('b', 'anana')
```

If you don’t explicitly `import` a function, then you can use the fully qualified function name by adding the module name to the front:

```
>>> import rhymers
>>> rhymers.stemmer('apple')
('', 'apple')
>>> rhymers.stemmer('banana')
('b', 'anana')
```



There are many advantages to writing many small functions rather than long, sprawling programs. One is that small functions are much easier to write, understand, and test. Another is that you can put your tidy, tested functions into modules and share them across different programs you write. As you write more and more programs, you will find yourself solving some of the same problems repeatedly. It’s far better to create modules with reusable code rather than copying pieces from one program to another. If you ever find a bug in a shared function, you can fix it once and all the programs sharing the function get the fix. The other way is to find the duplicated code in every program and change it (and hoping that this doesn’t introduce even more problems because the code is entangled with other code!).

Creating rhyming strings

The `stemmer` function should always return a 2-tuple of the (`start`, `rest`) of a given word. As such, I can unpack the two values into separate variables:

```
>>> start, rest = stemmer('cat')
>>> start
'c'
>>> rest
'at'
```

If there is a value for `rest`, I can add all my `prefixes` to the beginning:

```
>>> prefixes = list('bcdfghjklmnpqrstvwxyz') + (
...     'bl br ch cl cr dr fl fr gl gr pl pr sc '
...     'sh sk sl sm sn sp st sw th tr tw wh wr'
...     'sch scr shr sph spl spr squ str thr').split()
```

I decided to use another list comprehension with a guard to skip any prefix that is the same as the `start` of the word. The result will be a new `list` which I pass to the `sorted` function to get the correctly ordered strings:

```
>>> sorted([p + rest for p in prefixes if p != start])
['bat', 'blat', 'brat', 'chat', 'clat', 'crat', 'dat', 'drat', 'fat', 'flat', 'frat',
 'gat', 'glat', 'grat', 'hat', 'jat', 'kat', 'lat', 'mat', 'nat', 'pat', 'plat',
 'prat',
 'qat', 'rat', 'sat', 'scat', 'schat', 'scrat', 'shat', 'shrat', 'skat', 'slat',
 'smat',
 'snat', 'spat', 'sphat', 'splat', 'sprat', 'squat', 'stat', 'strat', 'swat', 'tat',
 'that',
 'thrat', 'thwat', 'trat', 'twat', 'vat', 'wat', 'what', 'wrat', 'xat', 'yat', 'zat']
```

I then `print` that `list`, joined on newlines. If there is no `rest` of the given word, I `print` a message that the word cannot be rhymed:

```
if rest:
    print('\n'.join(sorted([p + rest for p in prefixes if p != start])))
else:
    print(f'Cannot rhyme "{args.word}"')
```

Writing `stemmer` without regular expressions

It is certainly possible to write a solution that does not use regular expressions. My idea was to find the first position of a vowel in the given string. If one is present, use a list slice to return the portion of the string up to that position and the portion starting at that position:

```

def stemmer(word):
    """Return leading consonants (if any), and 'stem' of word"""
    word = word.lower()          ①
    pos = list()                 ②
        filter(lambda v: v >= 0, ③
            map(lambda c: word.index(c) if c in word else -1, 'aeiou'))) ④
    if pos:                      ⑤
        first = min(pos)         ⑥
        return (word[:first], word[first:]) ⑦
    else:
        return (word, '')        ⑧

```

- ① Lowercase the given word to avoid dealing with uppercase letters.
- ② Coerce the lazy `filter` to a `list`.
- ③ Only allow values greater than 0. The value `-1` indicates "not found" as 0 is a valid index for a string.
- ④ `map` over the vowels `'aeiou'` to get their index if they are present in the given word. Use `-1` to indicate the vowel is not present.
- ⑤ The `pos` variable is a `list`. In a boolean context, the empty list evaluates to "falsey" and anything else is "truthy."
- ⑥ The `first` position of a vowel will be the *minimum* from the positions (`min(pos)`).
- ⑦ Use the `first` to slice the portion of `word` up to (but not including) `first` and then the portion from `first` to the end.
- ⑧ If `pos` is the empty `list` (meaning no vowels were present), then return a 2-tuple of the `word` and the empty string to indicate there is no `rest` of the word to use for rhyming.

This function will also pass the `test_stemmer` function. By writing a test just for the idea of this one function and exercising it with all the different values I would expect, I'm free to *refactor* my code. As stated before, the `stemmer` is a black box. What goes on inside the function is of no concern to the code that calls it. As long as the function passes the tests, then it is "correct" (for certain values of "correct").

Small functions and their *tests* will set you free to improve your programs! First make something work, and make it beautiful. Then try to make it better, using your tests to ensure it keeps working as expected.

[7] "Looking at code you wrote more than two weeks ago is like looking at code you are seeing for the first time." - Dan Hurvitz

Review

- Regular expressions allow us to declare a pattern that we wish to find. The regex *engine* will sort out whether the pattern is found or not. This is a *declarative* approach to program rather than the *imperative* method of manually seeking out patterns by writing code ourselves.
- We can wrap parts of the pattern in parentheses to "capture" them into groups that we can fetch from the result of `re.match` or `re.search`.
- You can add a guard to a list comprehension to avoid taking some elements from an iterable.
- The `filter` function is another way to write a list comprehension with a guard. Like `map`, it is a lazy, higher-order function that takes a function that will be applied to every element of an iterable. Only those elements which are "truthy" are returned.
- Python can evaluate many types including strings, numbers, lists, and dictionaries in a *boolean context* to arrive at a sense of "truthiness." That is, you are not restricted to just `True` and `False` in `if` expressions. The empty string `''`, the `int 0`, the `float 0.0`, the empty `list []`, and the empty `dict {}` are all considered "falsey," and so any value from those types like the non-empty `str`, `list` or `dict`, or any numeric value not zero-ish is considered "truthy."
- You can break long string literals into shorter strings in your code and then group them with parentheses to have Python join them into one long string. It's advisable to break long regexes into shorter strings and add comments on each line to document the function of each pattern.
- Write small functions *and tests* and share them in modules. Every `.py` file can be a module from which you can `import` functions. Sharing small, tested functions is better than writing long programs and copying/pasting code as needed.

Going Further



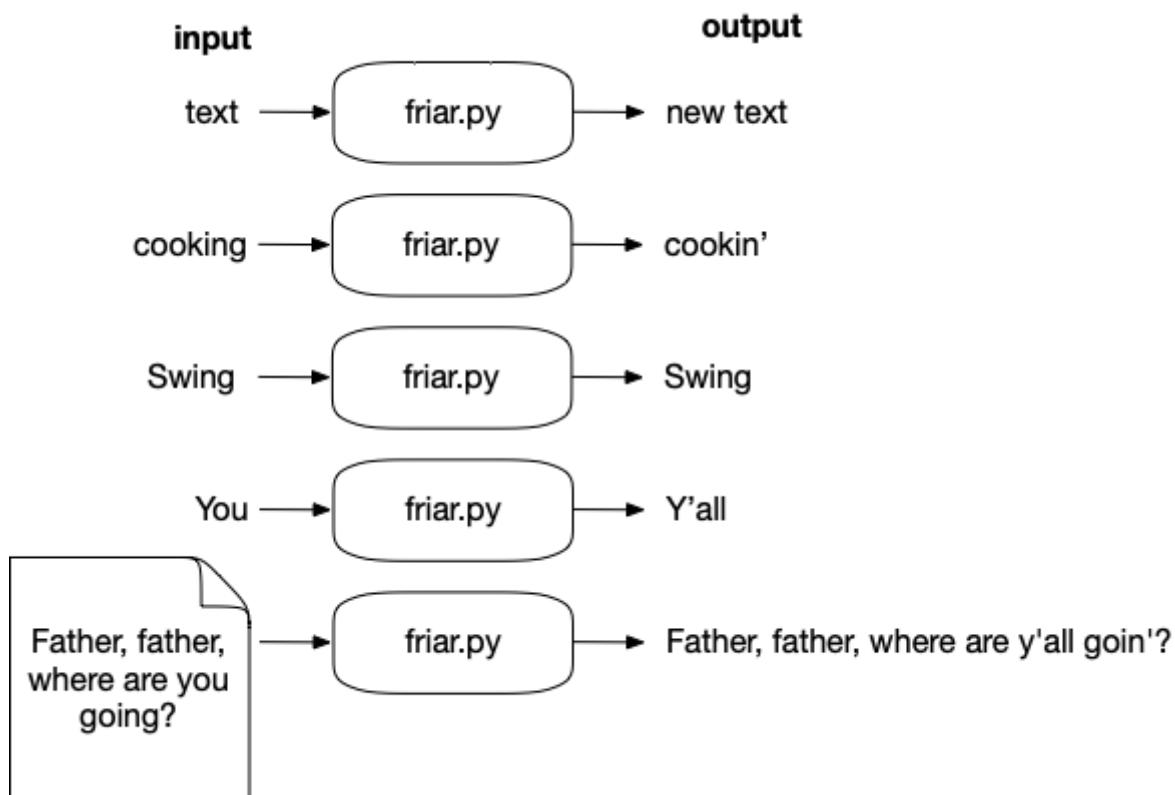
- Add an `--output` option to write the words to a given file. The default should be to write to `STDOUT`.
- Read an input file and create rhyming words for all the words in the file. You can borrow from the "Words Count" program to read a file and break it into words, then iterate each word and create an output file for each word with the rhyming words.
- Write a new program that finds all unique consonant sounds in a dictionary of English words like the file `/usr/share/dict/words` that usually exists on most Unix systems. Print them out in alphabetical order and use those to expand this program's consonants.
- Write a program to create Pig Latin where you move the initial consonant sound from the beginning of the word to the end and add "-ay" so that "cat" becomes "at-cay." If a word starts with a vowel, then add "-yay" to the end so that "apple" becomes "apple-yay."
- Write a program to create spoonerisms where the initial consonant sounds of adjacent words are switched so you get "blushing crow" instead of "crushing blow."

Chapter 15: The Kentucky Friar: More regular expressions



Your author grew up in the American Deep South where we tend to drop the final "g" of words ending in "ing," like "cookin'" instead of "cooking." We also tend to say "y'all" for the second-person, plural pronoun, which makes sense because Standard English is missing a distinctive word for this. In this exercise, we'll write a program called `friar.py` that will accept some input as a single positional argument and transforms the text by replacing the "g" with an apostrophe ('') words ending in "-ing" and also changes "you" to "y'all". Granted, we have no way to know if we're changing the first- or second-person "you," but it makes for a fun challenge, nonetheless.

Here is a string diagram to help you see the inputs and outputs:



When run with no arguments or the `-h` or `--help` flags, your program should present usage:

```
$ ./friar.py -h
usage: friar.py [-h] str

Southern fry text

positional arguments:
  str      Input text or file

optional arguments:
  -h, --help show this help message and exit
```

We will only change "-ing" words with *two syllables*, so "cooking" becomes "cookin'" but "swing" will stay the same. Our heuristic for identifying two-syllable "-ing" words is to inspect the part of the word before the "-ing" ending to see if it contains a vowel, which in this case will include "y." We can split "cooking" into "cook" and "ing"; since there is an "o" in "cook," we should drop the final "g":

```
$ ./friar.py Cooking
Cookin'
```

When we remove "ing" from "swing," though, we're left with "sw" which contains no vowel, so it will remain the same:

```
$ ./friar.py swing
swing
```

Be mindful to keep the case the same on the first letter, e.g, "You" should become "Y'all":

```
$ ./friar.py you
y'all
$ ./friar.py You
Y'all
```

As in several previous exercises, the input may name a file, in which case you should read the file for the input text. To pass the tests, you will need to preserve the line structure of the input, so I recommend you read the file line-by-line. So given this input:

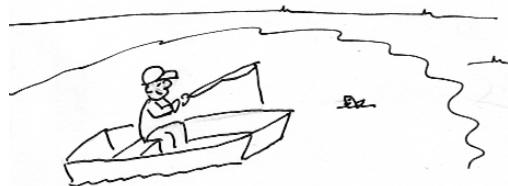
```
$ head -2 tests/banner.txt
O! Say, can you see, by the dawn's early light,
What so proudly we hailed at the twilight's last gleaming -
```

The output should have the same line breaks:

```
$ ./friar.py tests/banner.txt | head -2  
O! Say, can y'all see, by the dawn's early light,  
What so proudly we hailed at the twilight's last gleamin' -
```

To me, it's quite amusing to transform texts this way, but maybe I'm just weird:

```
$ ./friar.py tests/raven.txt  
Presently my soul grew stronger; hesitatin' then no longer,  
"Sir," said I, "or Madam, truly your forgiveness I implore;  
But the fact is I was nappin', and so gently y'all came rappin',  
And so faintly y'all came tappin', tappin' at my chamber door,  
That I scarce was sure I heard y'all" - here I opened wide the door: -  
Darkness there and nothin' more.
```



In this exercise, you will:

- Learn more about using regular expressions
- Use both `re.match` and `re.search` to find patterns anchored to the beginning of a string or anywhere in the string, respectively.
- Learn how the `$` symbol in a regex anchors a pattern to the *end* of a string.
- Learn how to use `re.split` to split a string.
- Explore how to write a manual solution for finding two-syllable "ing" words or the word "you."

Writing friar.py

As usual, I would recommend you start with `new.py friar.py` or copy the `template/template.py` to `friar/friar.py`. I recommend you start with a simple version that echos back the input from the command line:

```
$ ./friar.py cooking  
cooking
```

Or a file:

```
$ ./friar.py tests/blake.txt  
Father, father, where are you going?  
Oh do not walk so fast!  
Speak, father, speak to your little boy,  
Or else I shall be lost.
```

We need to process the input line-by-line, then word-by-word. You can use the `str.splitlines()` method to get each line of the input, and then the `str.split()` method to break the line on spaces into word-like units. This code:

```
for line in args.text.splitlines():  
    print(line.split())
```

Will create this output:

```
$ ./friar.py tests/blake.txt  
['Father,', 'father,', 'where', 'are', 'you', 'going?']  
['Oh', 'do', 'not', 'walk', 'so', 'fast!']  
['Speak,', 'father,', 'speak', 'to', 'your', 'little', 'boy,']  
['Or', 'else', 'I', 'shall', 'be', 'lost.']
```

If you look closely, it's going to be difficult to handle some of these word-like units as the adjacent punctuation is still attached to the words as in '`Father,`' and '`going?`'. We could write a regular expression to find just the parts of each "word" composed of ASCII letters, but I'd like to show you another way to split the text *using a regular expression!*

Splitting text using regular expressions

As in "Rhymer," we need to `import re` to use regexes:

```
>>> import re
```

For demonstration purposes, I'm going to set `text` to the first line:

```
>>> text = 'Father, father, where are you going?'
```

By default, `str.split` breaks text on spaces. Note that whatever text is used for splitting is missing from the result, so here there are no spaces:

```
>>> text.split()  
['Father,', 'father,', 'where', 'are', 'you', 'going?']
```

You can pass an optional value to `str.split` to indicate the string you want to use for splitting. If we choose the comma, then we end up with 3 strings instead of 6. Note that the commas which are used to split are missing from the result:

```
>>> text.split(',')  
['Father', ' father', ' where are you going?']
```

The `re` module has a function called `split` that works similarly. I recommend you read `help(re.split)` as this is a very powerful and flexible function. Like `re.match` that we used in "Rhymer," this function wants at least a `pattern` and a `string`. We can use `re.split` with a comma to get the same output as `str.split`, and, as before, the commas are missing from the result:

```
>>> re.split(',', text)  
['Father', ' father', ' where are you going?']
```

Shorthand classes

We are after the things that look like "words" in that they are composed of the characters that normally occur in words. The characters that *don't* normally occur in words (things like punctuation) are what we want to use for splitting. We've seen before that we can create a *character class* by putting literal values inside square brackets, like `'[aeiou]'` for the vowels. What if we create a character class where we enumerate all the non-letter characters? We could do something like this:

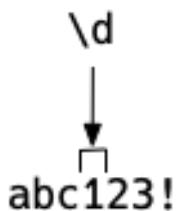
```
>>> import string  
>>> ''.join([c for c in string.printable if c not in string.ascii_letters])  
'0123456789!"#$%&\()'*)+,-./:;<=>?@[\\\]^_`{|}~ \t\n\r\x0b\x0c'
```

That won't be necessary because almost every implementation of regular expression engines define shorthand character classes.^[8] Here is a table of some of the most common shorthand classes and how they can be written longhand:

Table 4. Regex shorthand classes

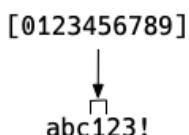
Character class	Shorthand	Other ways to write
Digits	\d	[0123456789], [0-9]
Whitespace	\s	[\t\n\r\x0b\x0c], same as string.whitespace
Word characters	\w	[a-zA-Z0-9_-]

The shorthand `\d` means any *digit* and is equivalent to '`[0123456789]`'. I can use the `re.search` method to look anywhere in a string for any digit. Here it will find the character '`1`' in the string '`abc123!`' because this is the *first* digit it finds:



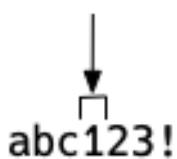
```
>>> re.search('\d', 'abc123!')
<re.Match object; span=(3, 4), match='1'>
```

Which is the same as using the longhand version:



```
>>> re.search('[0123456789]', 'abc123!')
<re.Match object; span=(3, 4), match='1'>
```

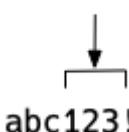
`[0-9]`



Or the version that uses the range of characters '`[0-9]`':

```
>>> re.search('[0-9]', 'abc123!')
<re.Match object; span=(3, 4), match='1'>
```

`\d+`



`abc123!` To have it find *one or more digits in a row*, add the `+`:

```
>>> re.search('\d+', 'abc123!')
<re.Match object; span=(3, 6), match='123'>
```

\w
↓
abc123!

Similarly, \w means "any word-like character." It includes all the Arabic numbers, the letters of the English alphabet, the dash ('-'), and underscore ('_'). The first match in the string is 'a':

```
>>> re.search('\w', 'abc123!')
<re.Match object; span=(0, 1), match='a'>
```

\w+
↓
abc123!

If you add the +, then it matches one or more word characters in a row which includes abc123 but not the !:

```
>>> re.search('\w+', 'abc123!')
<re.Match object; span=(0, 6), match='abc123'>
```

Negated shorthand classes

You can complement or "negate" a character class by putting the caret ^ *immediately inside* the character class. One or more of any character *not* a digit is '[^0-9]+', and so 'abc' is found:

[^0-9]+
↓
abc123!

```
>>> re.search('[^0-9]+', 'abc123!')
<re.Match object; span=(0, 3), match='abc'>
```

The class [^0-9]+ can also be written as \D+:

\D+
↓
abc123!

```
>>> re.search('\D+', 'abc123!')
<re.Match object; span=(0, 3), match='abc'>
```

\W+
↓
abc123!

Therefore, anything *not* a word character is \W:

```
>>> re.search('\W', 'abc123!')
<re.Match object; span=(6, 7), match='!'>
```

Table 5. Negated regex shorthand classes

Character class	Shorthand	Other ways to write
Not a digit	\D	[^0123456789], [^0-9]
Not whitespace	\S	[^ \t\n\r\x0b\x0c]
Word characters	\W	[^a-zA-Z0-9_-]

Using `re.split` with a captured regex

We can use \W as the argument to `re.split`:

```
>>> re.split('\W', 'abc123!')
['abc123', '']
```

There is a problem, though, because the result of `re.split` will omit those strings matching the pattern. Here we've lost the non-word character '!'. If we read `help(re.split)` closely, we can find the solution:

If capturing parentheses are used in pattern, then the text of all groups in the pattern are also returned as part of the resulting list.

We used capturing parentheses in the "Rhymer" to tell the regex engine to "remember" certain patterns that we then use `re.Match.groups` to retrieve after a match. Here it means that we need to put the split pattern in parens:

```
>>> re.split('(\W)', 'abc123!')
['abc123', '!', '']
```

Let's try that on our `text`:

```
>>> re.split('(\W)', text)
['Father', ',', '', ' ', 'father', ',', '', ' ', 'where', ' ', 'are', ' ', 'you', ' ',
'going', '?', '']
```

I'd like to group all the non-word characters together by adding `+` to the regex:

```
>>> re.split('(\W+)', text)
['Father', ' ', ' ', 'father', ' ', ' ', 'where', ' ', 'are', ' ', 'you', ' ', 'going', '?',
'']
```

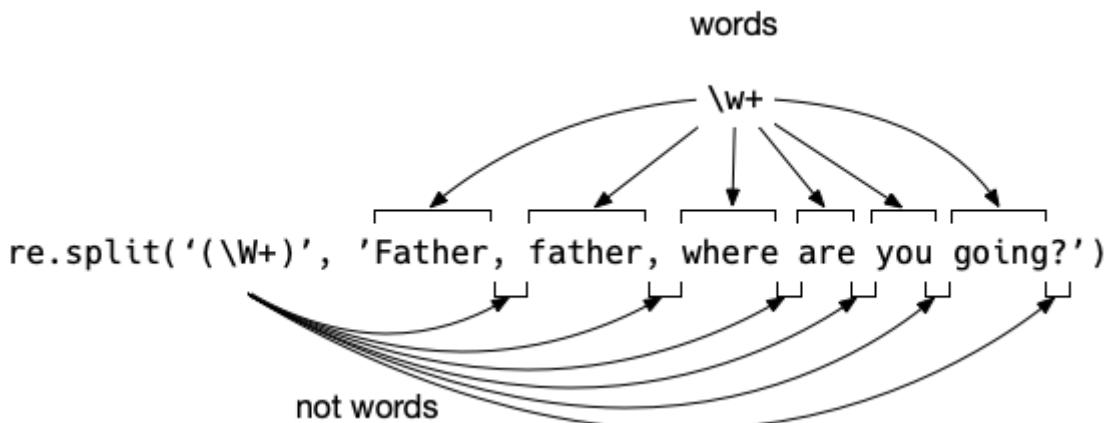


Figure 11. The `re.split` function can use a captured regex to return both the parts that match the regex and the things that did not.

That is so cool! Now we have a way to process each *actual* word and the bits in between them.

Writing the `fry` function

For me, the next step is to write function called `fry` to decide whether to modify *just one word*. That is, rather than thinking about how to handle all the text at once, I just want to think about how I will handle one word at a time. To help me think about how this function should work, I'll start off by writing the `test_fry` function and a stub for the actual `fry` function that has just the single command `pass` which tells Python to do nothing. If you like this idea, feel free to paste this into your program:

```
def fry(word):
    pass

def test_fry():
    assert fry('you') == "y'all"
    assert fry('You') == "Y'all"
    assert fry('fishing') == "fishin'"
    assert fry('Aching') == "Achin''"
    assert fry('swing') == "swing"
```

Then run `pytest friar.py` to see that, as expected, the test will fail:

```

=====
 FAILURES =====
----- test_fry -----

def test_fry():
>     assert fry('you') == "y'all" ①
E     assert None == "y'all"       ②
E     + where None = fry('you')

friar.py:47: AssertionError
===== 1 failed in 0.08 seconds =====

```

① The first test is failing.

② The result of `fry('you')` was `None` which does not equal "y'all."

Let's change our `fry` program to handle that string:

```

def fry(word):
    if word == 'you':
        return "y'all"

```

And run our tests:

```

=====
 FAILURES =====
----- test_fry -----

def test_fry():
    assert fry('you') == "y'all" ①
>     assert fry('You') == "Y'all" ②
E     assert None == "Y'all"      ③
E     + where None = fry('You')

friar.py:49: AssertionError
===== 1 failed in 0.16 seconds =====

```

① Now the first test passes.

② The second test fails because the "You" is capitalized.

③ The function returned `None` but should have returned "Y'all."

Let's handle that:

```

def fry(word):
    if word == 'you':
        return "y'all"
    elif word == 'You':
        return "Y'all"

```

If you run the test, you'll see the first two tests pass now; however, I'm definitely not happy with that solution. There is already a good bit of duplicated code. Can we write a more elegant way to match both "you" and "You" and which returns the correctly capitalized answer?

```
def fry(word):
    if word.lower() == 'you':
        return word[0] + "'all"
```

either
"y" or "Y"

[yY]

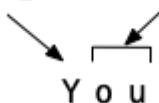


Better still, we can write a regular expression! There is one difference between "you" and "You"—the "y" or "Y"—that we can represent using the character class '**[yY]**'. This will match the lowercase version:

```
>>> re.match('[yY]ou', 'you')
<re.Match object; span=(0, 3), match='you'>
```

either literal
"y" or "Y" characters

[yY] ou



And the capitalized:

```
>>> re.match('[yY]ou', 'You')
<re.Match object; span=(0, 3), match='You'>
```

Now I'd like to reuse the initial character (either "y" or "Y") in the return value. I could *capture* it by placing it into parentheses. Try to rewrite your **fry** function using this idea and getting it to pass the first two tests again before moving on:

```
>>> match = re.match('([yY])ou', 'You')
>>> match.group(1) + "'all"
"Y'all"
```

The next step is to handle a word like "fishing":

```

=====
 FAILURES =====
-----
----- test_fry -----

def test_fry():
    assert fry('you') == "y'all"
    assert fry('You') == "Y'all"
>   assert fry('fishing') == "fishin'" ①
E       assert None == "fishin'"          ②
E           + where None = fry('fishing')

friar.py:52: AssertionError
=====
 1 failed in 0.10 seconds =====

```

① The third test fails.

② The return from `fry('fishing')` was `None` but the value "fishin'" was expected.

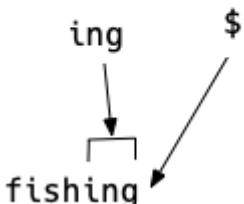
How can we identify a word that ends with "ing." There is actually a `str.endswith` function:

```

>>> 'fishing'.endswith('ing')
True

```

literal end of
characters the string



A regular expression to find "ing" at the end of a string would use \$ (pronounced "dollar") to *anchor* an expression:

```

>>> re.search('ing$', 'fishing')
<re.Match object; span=(4, 7), match='ing'>

```

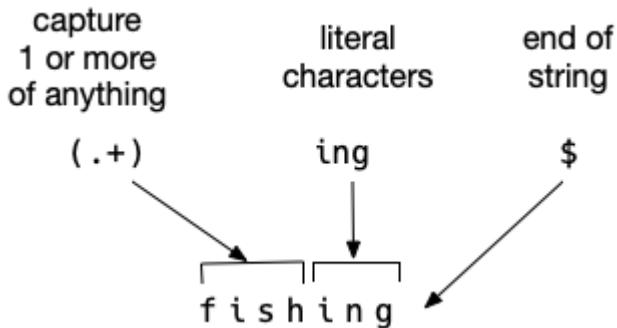
f i s h i n g
→ -1

`word[:-1] + "''" → fishin'` We can use a string slice to get all the characters up to the last at index `-1` and then append an apostrophe. Add this to your `fry` function and see how many tests you pass:

```

if word.endswith('ing'):
    return word[:-1] + "'"

```



Or you could use a groups with a regex to capture the first part of the word:

```
>>> match = re.search('(.+)ing$', 'fishing')
>>> match.group(1) + "in'"
"fishin'"
```

You should be able to get results like this:

```
===== FAILURES =====
----- test_fry -----
def test_fry():
    assert fry('you') == "y'all"
    assert fry('You') == "Y'all"
    assert fry('fishing') == "fishin'"
    assert fry('Aching') == "Achin'"
>     assert fry('swing') == "swing" ①
E     assert "swin'" == 'swing'      ②
E         - swin' ③
E         ^
E         + swing
E         ?

friar.py:59: AssertionError
===== 1 failed in 0.10 seconds =====
```

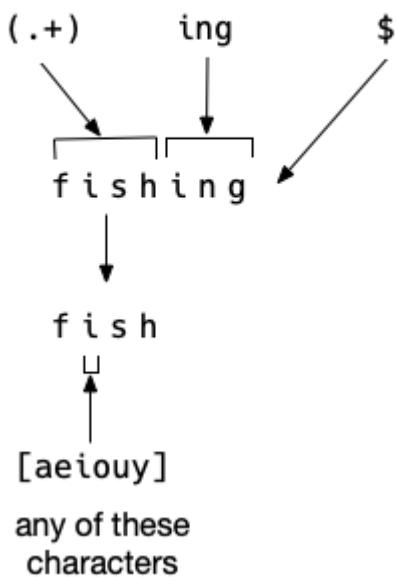
① This test failed.

② The result of `fry('swing')` was "swin'" but should have been "swing."

③ Sometimes the test results will be able to highlight the exact point of failure. Here you are being shown that there is a '`'` where there should be a `g`.

We need a way to identify words that have two syllables. I mentioned before that we'll use a heuristic that looks for a vowel `[aeiouy]` in the part of the word *before* the "ing" ending. Another regex could do the trick:

capture
1 or more literal characters end of string



```
>>> match = re.search('(.+)ing$', 'fishing') ①
>>> first = match.group(1) ②
>>> re.search('[aeiouy]', first) ③
<re.Match object; span=(1, 2), match='i'> ④
```

① The `(.+)` will match and capture one or more of anything followed by the characters `ing`. The return from `re.search` will either be an `re.Match` object if the pattern was found or `None` to indicate it was not.

② Here we know there will be a `match` value, so we can use `match.group(1)` to get the first capture group which will be anything immediately before `ing`. In actual code, we should check that `match` is not `None` or we'd trigger an exception by trying to execute the `group` method on a `None`.

③ We can use `re.search` on the `first` part of the string to look for a vowel.

④ As the return from `re.search` is an `re.Match` object, we know there is a vowel in the first part, so the word looks to have two syllables.

If the word matches this test, then return the word with the final "g" replaced with an apostrophe; otherwise, return the word unchanged. I would suggest you not proceed until you are passing all of `test_fry`.

Using the `fry` function

Now your program should be able to:

1. Read input from the command line or a file
2. Read the input line-by-line
3. Split each line into words and non-words
4. `fry` any individual word

The next step is to apply the `fry` function to all the word-like units. I hope you can see a familiar pattern emerging—applying a function to all elements of a list! You can use a `for` loop:

```
for line in args.text.splitlines():          ①
    words = []                                ②
    for word in re.split(r'(\W+)', line.rstrip()): ③
        words.append(fry(word))                ④
    print(''.join(words))                     ⑤
```

- ① Preserve the structure of the newlines in `args.text` by using `str.splitlines()`.
- ② Create a `words` variable to hold the transformed words.
- ③ Split each `line` into words and non-words.
- ④ Add the "fried" word to the `words` list.
- ⑤ Print a new string of the joined `words`.

That (or something like it) should work well enough to pass the tests. Once you have a version that works, see if you can rewrite the `for` loop as a list comprehension and a `map`.

Alrighty! Time to bear down and write this.

[8] There is a basic flavor of regular expression syntax that is recognized by everything from Unix command-line tools like `awk` to regex support inside of languages like Perl, Python, and Java. Some tools add extensions to their regexes that may not be understood by other tools. For example, there was a time when Perl's regex engine added many new ideas that eventually became a dialect known as "PCRE" (Perl-Compatible Regular Expressions). Not every tool that understands regexes will understand every flavor of regex, but, in all my years of writing and using regexes, I've rarely found this to be a problem.

Solution

```
1 #!/usr/bin/env python3
2 """Kentucky Friar"""
3
4 import argparse
5 import os
6 import re
7
8
9 # -----
10 def get_args():
11     """get command-line arguments"""
12     parser = argparse.ArgumentParser(
13         description='Southern fry text',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('text', metavar='str', help='Input text or file')
17
18     args = parser.parse_args()
19
20     if os.path.isfile(args.text): ①
21         args.text = open(args.text).read()
22
23     return args
24
25
26 # -----
27 def main():
28     """Make a jazz noise here"""
29
30     args = get_args() ②
31
32     for line in args.text.splitlines(): ③
33         print('.join(map(fry, re.split(r'(\W+)', line.rstrip())))) ④
34
35
36 # -----
37 def fry(word): ⑤
38     """Drop the 'g' from '-ing' words, change 'you' to 'y'all'"""
39
40     ing_word = re.search('(.+)ing$', word) ⑥
41     you = re.match('([Yy])ou$', word) ⑦
42
43     if ing_word: ⑧
44         prefix = ing_word.group(1) ⑨
45         if re.search('[aeiouy]', prefix, re.IGNORECASE): ⑩
46             return prefix + "in" ⑪
47     elif you: ⑫
```

```

48         return you.group(1) + "'all"      ⑬
49
50     return word                         ⑭
51
52
53 # -----
54 def test_fry():                      ⑮
55     """Test fry"""
56
57     assert fry('you') == "y'all"
58     assert fry('You') == "Y'all"
59     assert fry('fishing') == "fishin'"
60     assert fry('Aching') == "Achin'"
61     assert fry('swing') == "swing"
62
63
64 # -----
65 if __name__ == '__main__':
66     main()

```

- ① If the argument is a file, replace the `text` value with the contents from the file.
- ② Get the command-line arguments. The `text` value will either be the command-line text or the contents of a file by this point.
- ③ Use the `str.splitlines()` method to preserve the line breaks in the input text.
- ④ `map` the pieces of text split by the regular expression through the `fry` function which will return the words modified as needed. Use the `str.join` method to turn that resulting `list` back into a `str` to print.
- ⑤ Define a `fry` function that will handle one `word`.
- ⑥ Search for "ing" anchored to the end of `word`. Use a capture to remember the part of the string *before* the "ing."
- ⑦ Search for "you" or "You" starting from the beginning of `word`. Capture the [yY] alternation in a group.
- ⑧ If the search for "ing" returned a match...
- ⑨ Get the `prefix` (the bit before the "ing") which is `group` number 1.
- ⑩ Perform a case-insensitive search for a vowel (plus "y") in the `prefix`. If nothing is found, then `None` will be returned which evaluates to `False` in this boolean context. If a match is returned, then the not-`None` value will evaluate to `True`.
- ⑪ Append "in'" to the `prefix` and return it to the caller.
- ⑫ If the match for "you" succeeded...
- ⑬ Then return the captured first character plus "'all."
- ⑭ Otherwise, return the `word` unaltered.
- ⑮ The tests for `fry`.

Discussion

Breaking text into lines

In several previous exercises, I used a technique of reading an input file into the `args.text` value. If the input is coming from a file, then there will be newlines separating each line of text. I had suggested a `for` loop to handle each line of input text returned by `str.splitlines()` to preserve the newlines in the output. I suggested you also start with a second `for` loop to handle each word-like unit returned by the `re.split()`:

```
for line in args.text.splitlines():
    words = []
    for word in re.split('(\W+)', line.rstrip()):
        words.append(fry(word))
    print(''.join(words))
```

That's a total of 5 LOC that could be written in 2 if we replace the second `for` with a list comprehension:

```
for line in args.text.splitlines():
    print(''.join([fry(w) for w in re.split(r'(\W+)', line.rstrip())]))
```

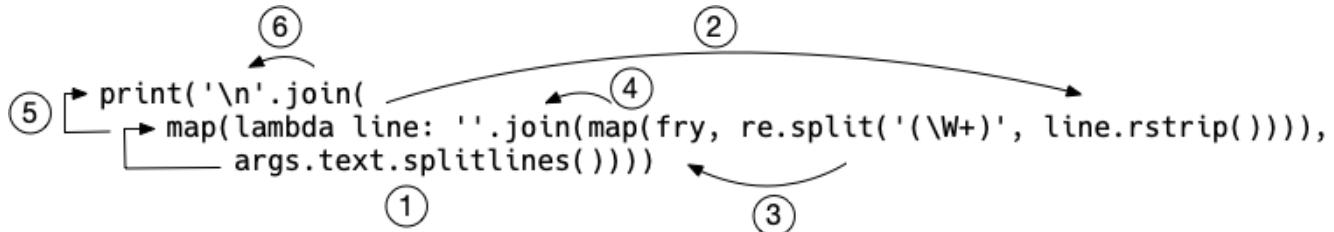
Or slightly shorter using a `map`:

```
for line in args.text.splitlines():
    print(''.join(map(fry, re.split(r'(\W+)', line.rstrip()))))
```

Or, if you're just a little bit weird, you can replace both `for` loops with `map`:

```
print('\n'.join(
    map(lambda line: ''.join(map(fry, re.split(r'(\W+)', line.rstrip()))),
        args.text.splitlines())))
```

Here's annotation to help you follow that:



1. `args.text` is split on newlines, each of which is passed to as `line` to the `lambda`.
2. The `line` is stripped of whitespace on the right and fed as the argument to `re.split`.

3. Each word-like unit from `re.split` is passed by `map` through the `fry` function.
4. The new `list` returned by `map` is joined on the empty string to create a new `str`.
5. The `map` returns a `list` of new `str` values created from the processed lines of text.
6. These new strings are joined back together on newlines to reconstitute the original line endings.

I find that last one a little hard to write and even harder to read, so I'd tend to use the version before it. Remember that code is *read* far more often than it is *written*. The last version is probably too clever by half.

Writing the `fry` function manually

You were not required, of course, to write a `fry` function. However you wrote your solution, I hope you wrote tests for it. Here is a version that is fairly close to some of the suggestions I made in the introduction. This version uses no regular expressions:

```
def fry(word):
    """Drop the 'g' from '-ing' words, change 'you' to 'y'all'"""

    if word.lower() == 'you':          ①
        return word[0] + "all"         ②

    if word.endswith('ing'):          ③
        if any(map(lambda c: c.lower() in 'aeiouy', word[:-3])): ④
            return word[:-1] + "'" ⑤
        else:
            return word             ⑥

    return word                      ⑦
```

- ① Force the `word` to lowercase and see if it matches "you."
- ② If so, return the first character (to preserve the case) plus "all".
- ③ If the `word` ends with "ing"...
- ④ Check if it's `True` that any of the characters in our vowel set are in the `word` up to the "ing" suffix.
- ⑤ If so, return the `word` up to the last index plus the apostrophe.
- ⑥ Else return the `word` unchanged.
- ⑦ If the `word` is neither an "ing" or "you" word, then return it unchanged.

Let's take a moment to appreciate the `any` function as it's one of my favorites. Here I'm using a `map` to check if each of the vowels exists in the portion of the `word` before the "ing" ending:

```
>>> word = "cooking"
>>> list(map(lambda c: (c, c.lower() in 'aeiouy'), word[:-3]))
[('c', False), ('o', True), ('o', True), ('k', False)]
```

The first character of "cooking" is "c," and it does not appear in the string of vowels. The next two characters ("o") do appear in the vowels, but "k" does not.

Let's reduce this to just the **True/False** values:

```
>>> list(map(lambda c: c.lower() in 'aeiou', word[:-3]))  
[False, True, True, False]
```

And now we can use **any** to tell us if *any* of the values are **True**:

```
>>> any([False, True, True, False])  
True
```

It's exactly the same as joining the values with **or**:

```
>>> False or True or True or False  
True
```

The **all** function returns **True** only if *all* the values are true:

```
>>> all([False, True, True, False])  
False
```

Which is the same as joining those values on **and**:

```
>>> False and True and True and False  
False
```

If it's **True** that one of the vowels appears in the first part of the **word**, we have determined this is (probably) a two-syllable word and can return the **word** with the final "g" replaced with an apostrophe. Otherwise, we return the unaltered **word**:

```
if any(map(lambda c: c.lower() in 'aeiou', word[:-3])):  
    return word[:-1] + "'"  
else:  
    return word
```

This approach works fine, but it's all quite manual as we have to write quite a bit of code to find our patterns.

Writing the `fry` function with regular expressions

Let's revisit the version of the `fry` function that uses regular expressions:

```
def fry(word):
    """Drop the 'g' from '-ing' words, change 'you' to 'y'all"""

    ing_word = re.search('(.+)ing$', word) ①
    you = re.match('([Yy])ou$', word)        ②

    if ing_word:                            ③
        prefix = ing_word.group(1)          ④
        if re.search('[aeiou]', prefix, re.IGNORECASE): ⑤
            return prefix + "in'"           ⑥
    elif you:                                ⑦
        return you.group(1) + "'all"         ⑧

    return word                             ⑨
```

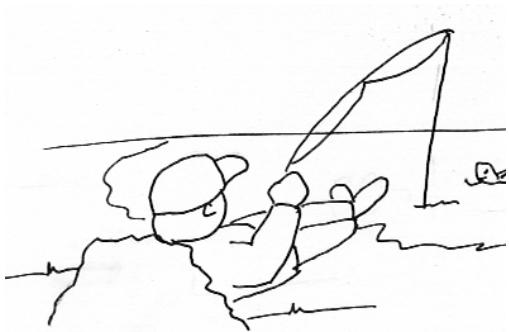
- ① The pattern '`(.+)ing$`' matches one or more of anything followed by "ing." The `$` anchors the pattern to the end of the string, so this is looking for a string that ends in "ing," but the string cannot just be "ing" as it has to have at least one of something before it. The parentheses capture the part before the "ing."
- ② The `re.match` starts matching at the beginning of the given `word` and is looking for either an upper- or lowercase "y" followed by "ou" and then the end of the string (`$`).
- ③ If `ing_word` is `None`, that means it failed to match. If it is not `None` (so it is "truthy"), that means it is an `re.Match` object we can use.
- ④ The `prefix` is the bit before the "ing" that we wrapped in parentheses. Because it is the first set of parens, we can fetch it with `ing_word.group(1)`.
- ⑤ Use `re.search` to look anywhere in the `prefix` for any of the vowels (plus "y") in a case-insensitive fashion. Remember that `re.match` would look at the beginning of `word`, which is not what we want.
- ⑥ Return the prefix plus the string "in'" so as to drop the final "g."
- ⑦ If the `re.match` for the "you" pattern fails, then `you` will be `None`. If it is not `None`, then it matched and `you` is an `re.Match` object.
- ⑧ We used parens to capture the first character so as to maintain the case. That is, if the word was "You," then we want to return "Y'all." Here we return that first group plus the string "'all."
- ⑨ If the `word` matched neither a two-syllable "ing" patter nor the word "you," then return the `word` unchanged.

I've been using regexes for maybe 20 years, so this version seems much simpler to me than the manual version. You may feel differently. If you are completely new to regexes, trust me that they are so very worth the effort to learn. I absolutely would not be able to do much of my work without them!

Review

- Regular expressions can be used to find patterns in text. The patterns can be quite complicated, like a grouping of non-word characters in between groupings of word characters.
- The `re` module has the functions `re.match` to find pattern at the beginning of some text, `re.search` to find a pattern anywhere inside some text, and `re.split` to break text on a pattern.
- If you use capturing parens on the pattern for `re.split`, then the captured split pattern will be included in the returned values. This allows you to reconstruct the original string with the separators.

Going Further



- You could also replace "your" with "y'all's." For instance, "Where are your britches?" could become "Whare are y'all's britches?"
- Change "getting ready" or "preparing" to "fixin'," as in "I'm getting ready to eat" to "I'm fixin' to eat."
- Change the string "think" to "reckon," as in "I think this is funny" to "I reckon this is funny." You should also change "thinking" to "reckoning," which then should become "reckonin'." That means you either need to make two passes for the changes or find both "think" and "thinking" in the one pass.
- Make a version of the program for another regional dialect. I lived in Boston for a while and really enjoyed saying "wicked" all the time instead of "very" as in "IT'S WICKED COLD OUT!" = Chapter 16: The Scrambler: Randomly reordering the middles of words



Yuor brian is an azinamg cmiobiaontn of hdarware and sftraowe. You'e rdineag tihs rhgit now eevn thgouh the wrdos are a mses, but yuor biran can mkae snese of it bceause the frsit and lsat ltrtees of ecah wrod hvae saeytd the smae. Yuor biran de'onst atlaulcy raed ecah lteetr of ecah wrod but rades wlohe wdors. The scamrbeld wrdos difteienly solw you dwon, but y'roue not rlleay eevn tyinrg to ulsrmbance the lrttees, are you? It jsut hnaepps!

In this exercise, we will write a program called `scrambler.py` that will take a single position positional argument that is text or a file and then scramble each "word" of the text. The scrambling should only work on words with 4 characters or more and should scramble the letters in the middle

of the word, leaving the first and last characters unchanged. The program should take a `-s` or `--seed` option (an `int` with default `None`) to pass to `random.seed`.

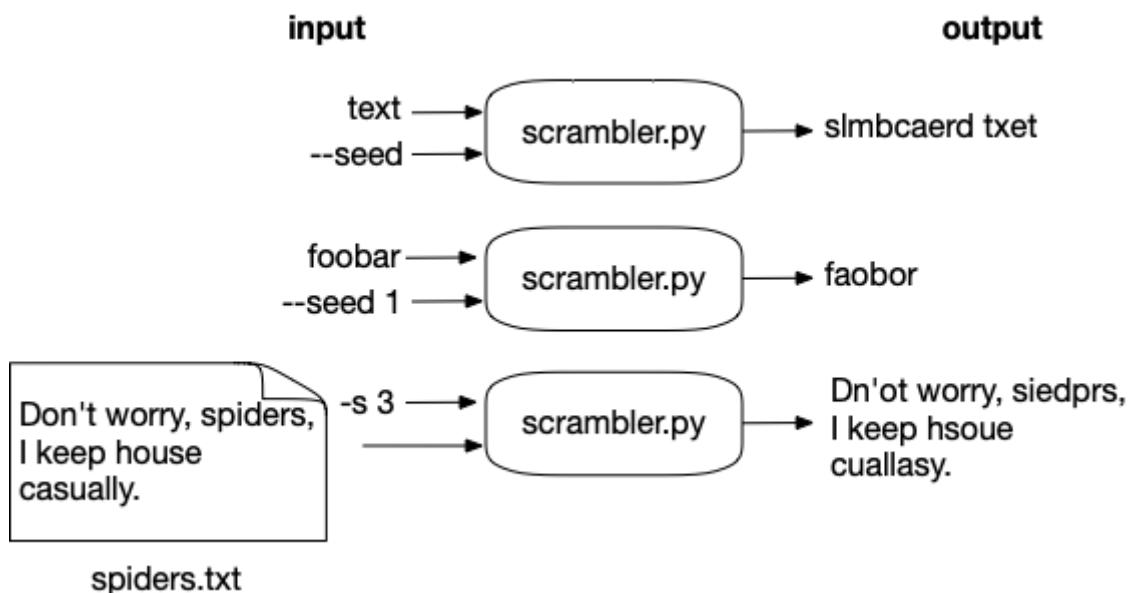
It should handle text on the command line:

```
$ ./scrambler.py --seed 1 "foobar bazquux"  
faobor buuzaqx
```

Or from a file:

```
$ cat ../inputs/spiders.txt  
Don't worry, spiders,  
I keep house  
casually.  
$ ./scrambler.py ../inputs/spiders.txt  
D'not wrory, sdireps,  
I keep hsuoe  
csalluay.
```

Here's a string diagram to help you think about it:



In this exercise, you will:

- Use a regular expression to split text into words.
- Use the `random.shuffle` function to shuffle a `list`.
- Create scrambled versions of words by shuffling the middle letters while leaving the first and last unchanged.

Writing `scrambler.py`

I recommend you start by using `new.py scrambler.py` to create the program in the `scrambler` directory. You can also copy `template/template.py` to `scrambler/scrambler.py`. You can refer to previous exercises like "Howler" to remember how to handle our positional argument that might be text or might a file of text to read. When run with no arguments or the flags `-h` or `--help`, it should present usage:

```
$ ./scrambler.py -h
usage: scrambler.py [-h] [-s int] str

Scramble the letters of words

positional arguments:
  str           Input text or file

optional arguments:
  -h, --help      show this help message and exit
  -s int, --seed int Random seed (default: None)
```

Once your program's usage statement matches this, then change your `main` definition to this:

```
def main():
    args = get_args()
    print(args.text.rstrip())
```

And verify that your program can echo text from the command line:

```
$ ./scrambler.py hello
hello
```

Or from an input file:

```
$ ./scrambler.py ../inputs/spiders.txt
Don't worry, spiders,
I keep house
casually.
```

Breaking the text into lines and words

As in the "Kentucky Friar," we want to preserve the line breaks of the input text by using `str.splitlines()`:

```
for line in args.text.rstrip().splitlines():
    print(line)
```

If we are reading the `spiders.txt` haiku, the first line is:

```
>>> line = "Don't worry, spiders,"
```

We need to break the `line` into "words." In the "Words Count" exercise, it was enough to use `str.split`. This leaves punctuation stuck to our words—both `worry`, and `spiders`, have commas:

```
>>> line.split()
['Don\'t', 'worry,', 'spiders,']
```

In the "Friar," we used the `re.split` function with the regular expression `(\W+)` to split text on one or more non-word characters. Let's try that:

```
>>> re.split('(\W+)', line)
['Don', "'", 't', ' ', 'worry', ' ', ' ', 'spiders', ',', '']
```

That won't work because it splits `Don't` into three parts, `Don`, `'`, and `t`. When I was working through this, my next idea was to use `\b` to break on *word boundaries*. Note I have to put an `r` in front of the first quote `r'\b'` to denote that it is a "raw" string. This still won't work as it splits the contracted word:

```
>>> re.split(r'\b', line)
['', 'Don', "'", 't', ' ', 'worry', ' ', ' ', 'spiders', ',', '']
```

While searching the Internet for a regex to split the text properly, I found this pattern on a Java discussion board which works just perfectly to separate *words* from *non-words*:

```
>>> re.split("[a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?", "Don't worry, spiders,")
['', "Don't", ' ', 'worry', ' ', ' ', 'spiders', ',']
```

The beautiful thing about regular expressions is that they are their own language, one that is used inside many other languages from Perl to Haskell. Let's try to understand this pattern.

```
(          ①
[a-zA-Z]    ②
(?:        ③
[a-zA-Z']*  ④
[a-zA-Z]    ⑤
)?        ⑥
)
```

- ① Start a capture group.
- ② A character class composed of the English alphabet, both lower- and uppercase.
- ③ Start a *non-capturing* group. That is, we want to group everything inside the parens so that we can make the entire group *optional* with a ?, but we don't want to capture the grouping.
- ④ Zero or more of any of the class composed of letters plus the apostrophe.
- ⑤ A character class of just the letters.
- ⑥ Make the second (non-capturing) group optional.
- ⑦ End the first capture group.

You will often see the `re.compile` is used to store a regular expression into a variable that has some meaningful name. I often do this, especially if I use the same regex at different points in my code:

```
>>> splitter = re.compile("([a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?)")
>>> splitter.split("Don't worry, spiders,")
['', "Don't", ' ', 'worry', ' ', 'spiders', ',']
```

Scrambling a word



Now that we have a way to process the *lines* and then *words* of our text, let's think about how we'll scramble our words by starting with just *one word*. You and I will need to use the same algorithm for scrambling the words, so here are the rules:

- If the word is 3 characters or shorter, return the word unchanged.

- Use a string slice to copy the characters not including the first and last.
- Use the `random.shuffle` method to mix up the letters in the middle.
- Return the new "word" by combining the first, middle, and last parts.

I recommend you create a function called `scramble` that will do all this and a test for it. Feel free to add this to your program:

```
def scramble(word):
    """Scramble a word"""

    pass

def test_scramble():
    """Test scramble"""

    random.seed(1)
    assert scramble("a") == "a"
    assert scramble("ab") == "ab"
    assert scramble("abc") == "abc"
    assert scramble("abcd") == "acbd"
    assert scramble("abcde") == "acbde"
    assert scramble("abcdef") == "aecbdf"
    assert scramble("abcde'f") == "abcd'ef"
    random.seed(None)
```

Inside your `scramble` function, you will have a word like "worry." We can use list slices to extract part of a string. Since Python starts numbering at `0`, we use `1` to indicate the *second* character:

```
>>> word = 'worry'
>>> word[1]
'o'
```

The last index of any string is `-1`:

```
>>> word[-1]
'y'
```

If we want a slice, we use the `list[start:stop]` syntax. Since the `stop` position is not included, we can get the `middle` like so:

```
>>> middle = word[1:-1]
>>> middle
'orr'
```

We can `import random` to get access to the `shuffle` method. Like the `list.sort` and `list.reverse` functions, the argument will be shuffled **in-place**, and the function will return `None`. That is, you might be tempted to write code like this:

```
>>> import random  
>>> x = [1, 2, 3]  
>>> shuffled = random.shuffle(x)
```

What is the value of `shuffled`? Is it something like `[3, 1, 2]` or is it `None`?

```
>>> type(shuffled)  
<class 'NoneType'>
```

`x = [1, 2, 3]
shuffled = random.shuffle(x)`

returns `None`

shuffles
in-place

`[2, 3, 1]`

The `x` list has been shuffled:

```
>>> x  
[2, 3, 1]
```

And the return from `random.shuffle` was `None`, so `shuffled` was assigned `None`.

If you've been following along, it turns out that we cannot shuffle the `middle` like this:

```
>>> random.shuffle(middle)  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
  File "/Users/kyclark/anaconda3/lib/python3.7/random.py", line 278, in shuffle  
    x[i], x[j] = x[j], x[i]  
TypeError: 'str' object does not support item assignment
```

The `middle` variable is a `str`:

```
>>> type(middle)  
<class 'str'>
```

The `random.shuffle` function is trying to directly modify a `str` value in-place, but `str` values in Python are *immutable*! One workaround is to make `middle` into a new `list` of the characters from `word`:

```
>>> middle = list(word[1:-1])
>>> middle
['o', 'r', 'r']
```

And that is something we can `shuffle`:

```
>>> random.shuffle(middle)
>>> middle
['r', 'o', 'r']
```

Then it's a matter of creating a new string with the original first letter, the shuffled middle, and the last letter. I'll leave that for you to work out.

Use `pytest scrambler.py` to have `pytest` execute the `test_scramble` function to see if it works correctly. Run this command *after every change to your program*. Ensure that your program always compiles and runs properly. Only make *one change at a time*, then save your program and run the tests!

Scrambling all the words

As in several previous exercises, we're now down to applying the `scramble` function to all the words. Can you see a familiar pattern?

```
for line in args.text.splitlines():
    for word in re.split("[a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?", line):
        # what goes here?
```

We've talked about how to apply a function to each element in a sequence. You might try a `for` loop, a list comprehension, or maybe a `map`. Think about how you can split the text into words and feed them to the `scramble` function, then join them back together to reconstruct the structure of the original text.

That should be enough to go on. Write your solution and use the included tests to check your program.

Solution

```
1 #!/usr/bin/env python3
2 """Scramble the letters of words"""
3
4 import argparse
5 import os
6 import re
7 import random
8
9
10 # -----
11 def get_args():
12     """Get command-line arguments"""
13
14     parser = argparse.ArgumentParser(
15         description='Scramble the letters of words',
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18     parser.add_argument('text', metavar='str', help='Input text or file') ①
19
20     parser.add_argument('-s',                         ②
21                         '--seed',
22                         help='Random seed',
23                         metavar='int',
24                         type=int,
25                         default=None)
26
27     args = parser.parse_args()          ③
28
29     if os.path.isfile(args.text):      ④
30         args.text = open(args.text).read()
31
32     return args                      ⑤
33
34
35 # -----
36 def main():
37     """Make a jazz noise here"""
38
39     args = get_args()                ⑥
40     random.seed(args.seed)          ⑦
41     splitter = re.compile("([a-zA-Z](?:[a-zA-Z]*[a-zA-Z])?)") ⑧
42
43     for line in args.text.splitlines(): ⑨
44         print('.join(map(scramble, splitter.split(line))))') ⑩
45
46
47 # -----
```

```

48 def scramble(word):          ⑪
49     """For words over 3 characters, shuffle the letters in the middle"""
50
51     if len(word) > 3 and re.match(r'\w+', word):    ⑫
52         middle = list(word[1:-1])                  ⑬
53         random.shuffle(middle)                   ⑭
54         word = word[0] + ''.join(middle) + word[-1] ⑮
55
56     return word      ⑯
57
58
59 # -----
60 def test_scramble(): ⑰
61     """Test scramble"""
62
63     random.seed(1)
64     assert scramble("a") == "a"
65     assert scramble("ab") == "ab"
66     assert scramble("abc") == "abc"
67     assert scramble("abcd") == "acbd"
68     assert scramble("abcde") == "acbde"
69     assert scramble("abcdef") == "aecbdf"
70     assert scramble("abcde'f") == "abcd'ef"
71     random.seed(None)
72
73
74 # -----
75 if __name__ == '__main__':
76     main()

```

- ① The `text` argument may be plain text on the command line or the name of a file to read.
- ② The `seed` option is an `int` that defaults to `None`.
- ③ Get the `args` so we can check the `text` value.
- ④ If `args.text` names an existing file, replace the value of `args.text` with the result of opening and reading the file's contents.
- ⑤ Return the arguments to the caller.
- ⑥ Get the command-line arguments.
- ⑦ Use the `args.seed` to set the `random.seed` value. If `args.seed` is the default `None`, then this is the same as not setting the seed.
- ⑧ Save the compiled regex into a variable.
- ⑨ Use `str.splitlines` to preserve the line breaks in `args.text`.
- ⑩ Use the `splitter` to break the `line` into a new `list` that is `map'd` into the `scramble` function. Join the resulting `list` on the empty string to create a new `str` to print.
- ⑪ Define a function to `scramble` a single `word`.
- ⑫ Only scramble words with 4 or more characters if they contain word characters.

- ⑯ Copy the second to second-to-last characters of the `word` into a new `list` called `middle`.
- ⑰ Shuffle the `middle`.
- ⑱ Set the `word` equal to the first character plus the `middle` plus the last character.
- ⑲ Return the `word` which may have been altered if it met the criteria.
- ⑳ The test for the `scramble` function.

Discussion

There is nothing new in the `get_args`, so I trust you understand that code. Refer to "Howler" if you want to revisit how to handle the `args.text` coming from the command line or from a file.

Processing the text

As mentioned in the introduction, I often will assign a *compiled* regex to a variable. Here I did it with the `splitter`:

```
splitter = re.compile("([a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?)")
```

Whenever you use something like `re.search` or `re.split`, the regex engine must parse the `str` value you provide for the regex into something it understands and can use. This parsing step must happen *each time* you call the function, so we can *compile* the regex just once and reuse it over and over.

The other reason I like to use `re.compile` is because I feel it can make my code more readable. Without this, I would have to write:

```
for line in args.text.splitlines():
    print(''.join(map(scramble, re.split("([a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?)", line))))
```

I find this version much easier to read:

```
splitter = re.compile("([a-zA-Z](?:[a-zA-Z']*[a-zA-Z])?)")
for line in args.text.splitlines():
    print(''.join(map(scramble, splitter.split(line))))
```

Maybe you still find that code somewhat confusing. Let's diagram the flow of the data:

```

print(''.join(map(scramble, splitter.split("Don't worry, spiders,"))))
    ↓
    ①
    ↓
print(''.join(map(scramble, ['', "Don't", ' ', 'worry', ' ', 'spiders', ','])))

    ↓
    ②
    ↙
    ③
    ↓
    ④
    ↓
print(''.join(['', "Dn'ot", ' ', 'worry', ' ', 'seirpds', ',']))
    ↓
    ⑤
    ↓
print("D'ont wrroy, srpdeis,")

```

1. The first action is that Python will call `splitter.split` with the string `"Don't worry, spiders,"`.
2. The result of splitting the text is a new `list` of words and non-words which mapped as the inputs to the `scramble` function.
3. The `map` essentially creates a `list` where the `scramble` function is called once for each element in the given `list`.
4. The result of `map` is a new `list` with the results of each application of the `scramble` function. This `list` is the argument to the `''.join` function.
5. The result of the `join` is a new `str` value which is the argument to `print`.

A longer way to write this with a `for` loop might look like this:

```

for line in args.text.splitlines(): ①
    words = [] ②
    for word in splitter.split(line): ③
        words.append(scramble(word)) ④
    print(''.join(words)) ⑤

```

- ① Use `str.splitlines()` to preserve the original line breaks.
- ② For each `line` of input, create a `words` variable to hold the scrambled words.
- ③ Use the `splitter` to split the `line`.
- ④ Add the result of `scramble(word)` to the `words` list.
- ⑤ Join the `words` on the empty string and pass to `print`.

You could write this with a list comprehension, too:

```
for line in args.text.splitlines():
    words = [scramble(word) for word in splitter.split(line)]
    print(''.join(words))
```

Or you could go quite the opposite direction and replace all `for` loops with `map`:

```
print('\n'.join(
    map(lambda line: ''.join(map(scramble, splitter.split(line))),
        args.text.splitlines())))
```

The last one reminds me of a programmer I used to work who would jokingly say "If it was hard to write, it should be hard to read!" It becomes somewhat clearer if you rearrange the code:

```
scrambler = lambda line: ''.join(map(scramble, splitter.split(line)))
print('\n'.join(map(scrambler, args.text.splitlines())))
```

Writing code that is correct, tested, and understandable is as much an art as it is a craft.

Scrambling a word

Let's take a closer look at my `scramble` function because I wrote it in a way that would make it easy to incorporate into a `map`:

```
def scramble(word):
    """For words over 3 characters, shuffle the letters in the middle"""

    if len(word) > 3 and re.match(r'\w+', word):      ①
        middle = list(word[1:-1])                      ②
        random.shuffle(middle)                         ③
        word = word[0] + ''.join(middle) + word[-1]  ④

    return word                                     ⑤
```

① Check if the given `word` is one we ought to scramble. First, it must be longer than 3 characters. Second, it must contain one or more word characters because the function will be passed both "word" and "non-word" strings. If either check returns `False`, we will return the `word` unchanged. The `r'\w+'` is used to create a "raw" string. The regex still works properly without this, but `pylint` complains about an "invalid escape character." Raw strings prevent Python from trying to interpret the string for escape sequences like `\s` for "whitespace" or `\b` for "word boundary."

② Copy the middle of the word to a new `list` called `middle`.

③ Shuffle the `middle` *in-place*. Remember that this function returns `None`.

④ Reconstruct the `word` by joining together the first character, the shuffled `middle`, and the last character.

⑤ Return the `word` which may or may not have been shuffled.

One key mistake you could make would be writing this code:

```
middle = random.shuffle(middle)
```

Honestly, I would prefer for `random.shuffle` to return a new, shuffled `list` rather than mutating the given argument, but c'est la vie. The function was written to mutate the data, and that's that.

Review



- The regex we used to split the text into words was quite complex but also gave us exactly what we needed. Writing the program without this piece would have made it significantly more difficult. Regexes, while complex and deep, are wildly powerful black magic that can make your programs incredibly flexible and useful.
- The `random.shuffle` function accepts a `list` which is mutated in-place.
- List comprehensions and `map` can often lead to more compact code, but going too far can reduce readability. Choose wisely.

Going Further

- Write a version where the `scramble` function sorts the middle letters into alphabetical order rather than shuffling them.
- Write a version that reverses each word rather than scrambles.
- Write a program to *unscramble* the text. For this, you'd probably need to have a dictionary of English words—most Unix systems have a file at `/usr/share/dict/words` for this. You would have to split the scrambled text into words/non-words, then compare each "word" to the words in your dictionary. As English has many anagrams (like "evil" and "vile"), you might not be able to make a definitive decision. = Mad Libs (Using regular expressions)

[mad libs string] | *mad-libs-string.png*

Write a Python program called `mad_lib.py` that will read a file given as a positional argument and find all the placeholders noted in `<>`, e.g., `<verb>`, prompt the user for the part of speech being requested, e.g., a "verb", and then substitute that into the text of the file, finally printing out all the placeholders replaced by the user's inputs. By default, this is an interactive program that will use the `input` prompt to ask the user for their answers, but, for testing purposes, you will have an option for `-i|--inputs` so the test suite can pass in all the answers and bypass the `input` calls.

Given no arguments or the `-h|--help` flag, the program should print a usage:

```
$ ./mad_lib.py
usage: mad_lib.py [-h] [-i str [str ...]] FILE
mad_lib.py: error: the following arguments are required: FILE
$ ./mad_lib.py -h
usage: mad_lib.py [-h] [-i str [str ...]] FILE
```

Mad Libs

```
positional arguments:
  FILE           Input file

optional arguments:
  -h, --help      show this help message and exit
  -i str [str ...], --inputs str [str ...]
                  Inputs (for testing) (default: None)
```

The structure of the input file has a part of speech enclosed in angle brackets, e.g., `<verb>`:

```
$ cat help.txt
<exclamation>! I need <noun>!
<exclamation>! Not just <noun>!
<exclamation>! You know I need <noun>!
<exclamation>!
```

When this is the input for the program, you should ask for each part of speech in order using the `input` command to ask the user for some text. When you've gotten all the text you need, print out the result of putting the user's answers into the placeholders:

```
$ ./mad_lib.py help.txt
exclamation: Hey
noun: tacos
exclamation: Oi
noun: fish
exclamation: Ouch
noun: pie
exclamation: Dang
Hey! I need tacos!
Oi! Not just fish!
Ouch! You know I need pie!
Dang!
```

The default mode is to be interactive, which is difficult to test, so take all the `--inputs` from the command line, skip the `input` prompts, and just show the resulting text:

```
$ ./mad_lib.py romeo_juliet.txt -i cars Detroit oil pistons \
> "stick shift" furious accelerate 42 foot hammer
Two cars, both alike in dignity,
In fair Detroit, where we lay our scene,
From ancient oil break to new mutiny,
Where civil blood makes civil hands unclean.
From forth the fatal loins of these two foes
A pair of star-cross'd pistons take their life;
Whose misadventur'd piteous overthrows
Doth with their stick shift bury their parents' strife.
The fearful passage of their furious love,
And the continuance of their parents' rage,
Which, but their children's end, nought could accelerate,
Is now the 42 hours' traffic of our stage;
The which if you with patient foot attend,
What here shall hammer, our toil shall strive to mend.
```

Hints:

- Definitely look into using regular expressions!

Solution

```
1 #!/usr/bin/env python3
2 """Mad Libs"""
3
4 import argparse
5 import re
6
7
8 # -----
9 def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Mad Libs',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('file',
17                         metavar='FILE',
18                         type=argparse.FileType('r'),
19                         help='Input file')
20
21     parser.add_argument('-i',
22                         '--inputs',
23                         help='Inputs (for testing)',
24                         metavar='str',
25                         type=str,
26                         nargs='+',
27                         required=False)
28
29     return parser.parse_args()
30
31
32 # -----
33 def main():
34     """Make a jazz noise here"""
35
36     args = get_args()
37     inputs = args.inputs
38     regex = re.compile('(<[^>]+>)')
39     text = args.file.read().rstrip()
40     blanks = list(regex.finditer(text))
41
42     if not blanks:
43         print('File "{}" has no placeholders'.format(args.file.name))
44         sys.exit(1)
45
46     for match in blanks:
47         placeholder, name = match.groups()
```

```
48     answer = inputs.pop(0) if inputs else input('{}: '.format(name))
49     text = re.sub	placeholder, answer, text, count=1)
50
51     print(text)
52
53
54 # -----
55 if __name__ == '__main__':
56     main()
```

Discussion

Defining the arguments

If you define the `file` with `type=argparse.FileType('r')`, then `argparse` will verify that the value is a file, creating an error and usage if it is not, and then will `open` it for you. Quite the time saver. I also define `--inputs` with `nargs='+'` so that I can get any number of strings as a `list`. If none are provided, the default value will be `None`, so be sure you don't assume it's a list and try doing list operations on a `None`.

Reading the input file

The first thing we need to do is `read` the input file:

```
>>> from pprint import pprint as pp
>>> text = open('help.txt').read()
>>> pp(text)
('<exclamation>! I need <noun>!\n'
 '<exclamation>! Not just <noun>!\n'
 '<exclamation>! You know I need <noun>!\n'
 '<exclamation>!\n')
```

Using regular expressions to find the pointy bits

We need to find all the `<...>` bits, so let's use a regular expression. We can find a literal `<` character like so:

```
>>> import re
>>> re.search('<', text)
<re.Match object; span=(0, 1), match='<'>
```

Now let's find its mate pair. The `.` means "anything," and we can add a `+` after it to mean "one or more":

```
>>> re.search('<.+>', text)
<re.Match object; span=(0, 28), match='<exclamation>! I need <noun>'>
```

Hmm, that matched all the way to the end of the string instead of stopping at the first available `>`. Often we find that regexes are "greedy" in that they will keep matching beyond where we want them to. If we add a `?` after the `+`, that will make it "non-greedy":

```
>>> re.search('<.+?>', text)
<re.Match object; span=(0, 13), match='<exclamation>'>
```

Another way is to say that we want to match one or more of anything that is **not** a `>`. We can create a character class `[>]` and then put a caret (`^`) **inside** it to negate the class. We'll add parens `()` to capture the whole mess as well as parens to capture the bit inside the `<>`

```
>>> re.search('<(([^\>]+)>) ', text)
<re.Match object; span=(0, 13), match='<exclamation>'>
```

Now we can `re.findall` if we just want the matching text:

```
>>> pp(re.findall('<(([^\>]+)>) ', text))
['<exclamation>',
 '<noun>',
 '<exclamation>',
 '<noun>',
 '<exclamation>',
 '<noun>',
 '<exclamation>']
```

But I wanted to use each match object, so I used `re.finditer` to return an iterator over the matches:

```
>>> it = re.finditer('<(([^\>]+)>) ', text)
>>> type(it)
<class 'callable_iterator'>
```

Iterables are "lazy", so if I `print` it, I'll just get a message about it being an object:

```
>>> pp(it)
<callable_iterator object at 0x1015a30f0>
```

In the REPL, I can call `list` to force Python to evaluate the iterable to the end:

```
>>> pp(list(it))
[<re.Match object; span=(0, 13), match='<exclamation>',>
 <re.Match object; span=(22, 28), match='<noun>',>
 <re.Match object; span=(30, 43), match='<exclamation>',>
 <re.Match object; span=(54, 60), match='<noun>',>
 <re.Match object; span=(62, 75), match='<exclamation>',>
 <re.Match object; span=(93, 99), match='<noun>',>
 <re.Match object; span=(101, 114), match='<exclamation>'>]
```

But note that it has not been exhausted and cannot be iterated again:

```
>>> pp(list(it))
[]
```

Which is why I convert it to a `list` right away so that I can evaluate if I found any (by inspecting the length) and then iterate over them:

```
>>> blanks = list(re.finditer('(<([>]+)>) ', text))
>>> len(blanks)
7
>>> for match in blanks:
...     print(match.group(1))
...
<exclamation>
<noun>
<exclamation>
<noun>
<exclamation>
<noun>
<exclamation>
```

The `groups` are defined in the order that you create them by counting the opening parenthesis. In our case, the second group is inside the first group, which is fine. If you start getting lots of groups, it might be best to name them:

```
>>> blanks = list(re.finditer(r'(?P<placeholder><(?P<name>[^>]+)>) ', text))
>>> for match in blanks:
...     print(match.group('name'))
...
exclamation
noun
exclamation
noun
exclamation
noun
exclamation
```

Getting `input` from the user

Now to get the values from the user. We can use the `name` as the prompt for `input`:

```
>>> blanks = list(re.finditer(r'(<([^\>]+)>) ', text))
>>> for match in blanks:
...     name = match.group(2)
...     answer = input('{}: '.format(name))
...
exclamation: Dude!
```

But I will make it a bit more flexible by using an `if` expression to take `pop` a value from the `inputs` list if that is available, otherwise I will use the `input`:

```
>>> inputs = ['Wow']
>>> answer = inputs.pop(0) if inputs else input('{}: '.format(name))
>>> answer
'Wow'
```

Using `re.sub` to replace the values

Now we need to put the user's `answer` into the original `text` which we can do with `re.sub` (substitute):

```
>>> pp(text)
('<exclamation>! I need <noun>!\n'
 '<exclamation>! Not just <noun>!\n'
 '<exclamation>! You know I need <noun>!\n'
 '<exclamation>!\n')
>>> placeholder = '<exclamation>'
>>> text = re.sub(placeholder, answer, text, count=1)
>>> pp(text)
('Wow! I need <noun>!\n'
 '<exclamation>! Not just <noun>!\n'
 '<exclamation>! You know I need <noun>!\n'
 '<exclamation>!\n')
```

The `count=1` is necessary to prevent `re.sub` from replacing **every** instance of the pattern. After doing that for every placeholder in the `text`, we can `print(text)` and we are done!

Going Further

- Learn more about regular expressions!

Gematria (Numeric encoding of text using ASCII values)

"This is where it calls for wisdom: Let the one who has insight calculate the number of the wild beast, for it is a man's number, and its number is 666." — Revelation 13:18

[gematria string] | *gematria-string.png*

Write a Python program called `gematria.py` that will numerically encode each word in a given text. The program should print a usage if given no arguments or the `-h|--help` flag:

```
$ ./gematria.py
usage: gematria.py [-h] str
gematria.py: error: the following arguments are required: str
$ ./gematria.py -h
usage: gematria.py [-h] str

Gematria

positional arguments:
  str      Input text or file

optional arguments:
  -h, --help show this help message and exit
```

The text may be given on the command line:

```
$ ./gematria.py 'foo bar baz'
324 309 317
```

Or in a file:

```
$ ./gematria.py ../inputs/fox.txt
289 541 552 333 559 444 321 448 314
```

The name of this program comes from gematria, a system for assigning a number to a word by summing the numeric values of each of the letters as defined by the Mispar godol (<https://en.wikipedia.org/wiki/Gematria>). For English characters, we can use the ASCII table (<https://en.wikipedia.org/wiki/ASCII>). Python provides these value through the `ord` function to convert a character to its "ordinal" (order in the ASCII table) value as well as the `chr` function to convert a number to its "character."

```
>>> ord('A')
65
>>> ord('a')
97
>>> chr(88)
'X'
>>> chr(112)
'p'
```

To implement an ASCII version of gematria in Python, for each word in a text we need to turn each letter into a number and add them all together. So, to start, note that Python can use a `for` loop to cycle through all the characters in a string:

```
>>> for char in 'abc':
...     print(ord(char))
...
97
98
99
```

Note that upper- and lowercase letters have different `ord` values:

```
>>> for char in 'ABC':
...     print(ord(char))
...
65
66
67
```

This is why you keep seeing `[A-Za-z0-9]` in regular expressions. If you look at the ASCII table above, you'll see that there are other characters in between the ranges for "0-9," "A-Z," and "a-z":

```
>>> ord('0'), ord('9')
(48, 57)
>>> ord('A'), ord('Z')
(65, 90)
>>> ord('a'), ord('z')
(97, 122)
>>> [chr(n) for n in range(58, 65)]
[':', ';', '<', '=', '>', '?', '@']
>>> [chr(n) for n in range(91, 97)]
['[', '\\\\', ']', '^', '_', '''']
```

For the purposes of this exercise, you can use `str.split` to split a line of text into "words"; however, this will leave punctuation attached to actual words:

```
>>> "Don't worry, spiders,".split()  
["Don't", 'worry,', 'spiders,']
```

Therefore, you should use a regular expression to remove anything that is *not* in the range [A-Za-z0-9]:

```
>>> import re  
>>> text = "Don't worry, spiders,"  
>>> [re.sub('[^A-Za-z0-9]', '', w) for w in text.split()]  
['Dont', 'worry', 'spiders']
```

For any given word, you can `sum` the `ord` values. I created a `word2num` function with the following test which I suggest you use:

```
def test_word2num():  
    """Test word2num"""  
  
    assert word2num("") == "0"  
    assert word2num("a") == "97"  
    assert word2num("abc") == "294"  
    assert word2num("ab'c") == "294"  
    assert word2num("4a-b'c,)") == "346"
```

If you put this inside your `gematria.py` file, then you will run `pytest -v gematria.py` to verify that your function works properly.

Hints:

- You'll want to read the input line-by-line because the tests are expecting lines of output where each word has been encoded.
- If you write a function like `word2num`, you're left with how to apply a function to a sequence of values. We've done this a few times now. How will you choose to do this? A `for` loop? A list comprehension? A `map`?

Solution

```
1 #!/usr/bin/env python3
2 """Gematria"""
3
4 import argparse
5 import os
6 import re
7
8
9 # -----
10 def get_args():
11     """Get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='Gematria',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text', metavar='str', help='Input text or file')
18
19     args = parser.parse_args()
20
21     if os.path.isfile(args.text):
22         args.text = open(args.text).read()
23
24     return args
25
26
27 # -----
28 def word2num(word):
29     """Sum the ordinal values of all the characters"""
30
31     return str(sum(map(ord, re.sub('[^a-zA-Z0-9]', ' ', word))))
32
33
34 # -----
35 def test_word2num():
36     """Test word2num"""
37
38     assert word2num("") == "0"
39     assert word2num("a") == "97"
40     assert word2num("abc") == "294"
41     assert word2num("ab'c") == "294"
42     assert word2num("4a-b'c,)") == "346"
43
44
45 # -----
46 def main():
47     """Make a jazz noise here"""


```

```
48
49     args = get_args()
50
51     for line in args.text.splitlines():
52         print(' '.join(map(word2num, line.split())))
53
54
55 # -----
56 if __name__ == '__main__':
57     main()
```

Discussion

The `text` argument for the program might be taken directly from the command line or from a named file. I chose to handle inside the `get_args` the reading of a file argument so that by the time I call `args = get_args()` I have the actual `text` I need to process.

Encoding one word

Eventually we're going to have to read the input text and break it into words as mentioned in the intro, but, for now, let's think about how to encode just one word. We've seen that we can use `ord` to get the ordinal value of a character in the ASCII value:

```
>>> ord('g')
103
```

We can use a list comprehension to do this for every character in a word:

```
>>> word = 'gematria'
>>> [ord(char) for char in word]
[103, 101, 109, 97, 116, 114, 105, 97]
```

And then the `sum` function will add those for us:

```
>>> sum([ord(char) for char in word])
842
```

We can do the same thing with `map`. The first argument to `map` is a function which is applied to every element in the second argument which must be a sequence of items, something we call an *iterable* like a `list` or a generator. Because the second argument is iterable, we don't have to spell out `for char in`. Remember that a `str` can be interchangably treated as a `list` of characters, so strings are iterable!

```
>>> map(ord, word)
<map object at 0x105c3c550>
```

We see this because `map` is a *lazy* function that doesn't actually produce results until they are actually required. For purposes of viewing in the REPL, we can use `list` to see the values. (You do not have to use `list` in your actual code.)

```
>>> list(map(ord, word))
[103, 101, 109, 97, 116, 114, 105, 97]
```

And now we can `sum` that. Note that `sum` will consume the `map` object, so we don't have to use `list`. To

me, this is an extremely clean bit of code:

```
>>> sum(map(ord, word))
842
```

Since ultimately I will be giving these numbers to `print` in a way that will expect strings, I will additionally coerce the number using `str`. Let's define a function called `word2num` to hold this idea:

```
>>> def word2num(word):
...     return str(sum(map(ord, word)))
...
>>> word2num('gematria')
'842'
```

That's a start, but it won't pass our test. If I run `pytest -v gematria.py` with this function, I'll see it fails when a word has punctuation:

```
gematria.py::test_word2num FAILED [100%]

===== FAILURES =====
----- test_word2num -----
def test_word2num():
    """Test word2num"""

    assert word2num("") == "0"
    assert word2num("a") == "97"
    assert word2num("abc") == "294"
>   assert word2num("ab'c") == "294"
E       AssertionError: assert '333' == '294'
E           - 333
E           + 294

gematria.py:41: AssertionError
```

I suggested you use the `re.sub` method that takes a regular expression to find along with a value to substitute for matches in a given string. If we add change `word2num` like so, it will pass our tests:

```
def word2num(word):
    return str(sum(map(ord, re.sub('[^a-zA-Z0-9]', '', word))))
```

We've written a function that encodes a single word. Ne need to find all the words!

Finding the words

As mentioned in the description of the program, the test suite is looking for the lines of input text to be maintained in the output. It's straightforward to `read` an `open` file line-by-line:

```
>>> file = '../inputs/spiders.txt'  
>>> for line in open(file):  
...     print(line, end='')  
...  
Don't worry, spiders,  
I keep house  
casually.
```

If I've taken the `text` from a file, then it's all just now just one string and a `for` loop on a string will iterate over each *character* not each line.

```
>>> text = open(file).read()  
>>> text  
'Don't worry, spiders,\nI keep house\nCasually.\n'  
>>> for char in text:  
...     print(char, end=' - ')  
...  
D-o-n-t- -w-o-r-r-y-, -s-p-i-d-e-r-s-, -  
-I- -k-e-e-p- -h-o-u-s-e-  
-c-a-s-u-a-l-l-y- -
```

So instead we can use `str.splitlines()` to break `text` on the newlines:

```
>>> for line in text.splitlines():  
...     print(line)  
...  
Don't worry, spiders,  
I keep house  
casually.
```

We can use `str.split()` to break each `line` into `words`:

```
>>> for line in text.splitlines():  
...     words = line.split()  
...     print(words)  
...  
['Don't', 'worry,', 'spiders,']  
['I', 'keep', 'house']  
['casually.']
```

Putting it all together

So we're finally to the point where we have lines of text and lists of words to encode. As we've seen, a list comprehension works adequately:

```
>>> for line in text.splitlines():
...     print([word2num(word) for word in line.split()])
...
['405', '579', '762']
['73', '421', '548']
['862']
```

But, to my eye, `map` is cleaner:

```
>>> for line in text.splitlines():
...     print(list(map(word2num, line.split())))
...
['405', '579', '762']
['73', '421', '548']
['862']
```

Now we need to `join` the values on spaces:

```
>>> for line in text.splitlines():
...     print(' '.join(map(word2num, line.split())))
...
405 579 762
73 421 548
862
```

Going Further

- Some scholars believe that the "666" in the above verse from Revelations represented Nero Caesar (written in Aramaic and then encoded via Hebrew gematria) and was used to avoid mentioning him by name. Write a program to find all words where your algorithm calculates the value as "666". Are these particularly scary words?

Mommy's Little Crossword Helper (Finding patterns of text in files)

[helper string] | *helper-string.png*

Write a Python program called `helper.py` that finds all words matching a given `-p|--pattern` such as one might use to complete a crossword puzzle to find words matching from a given `-w|--wordlist` (default `/usr/share/dict/words`). E.g., all 5-letter words with a "t" as the second character and ending in "ed". I could do this on the command line like so:

```
$ grep '^.{t}' /usr/share/dict/words | grep 'ed$' | awk 'length($0) == 5'  
steed
```

When run with no arguments or a `-h|--help` flag, it should print a usage:

```
$ ./helper.py  
usage: helper.py [-h] [-w str] str  
helper.py: error: the following arguments are required: str  
$ ./helper.py -h  
usage: helper.py [-h] [-w str] str  
  
Crossword helper  
  
positional arguments:  
  str                  The pattern to search  
  
optional arguments:  
  -h, --help            show this help message and exit  
  -w str, --wordlist str  
                        Wordlist to search (default: /usr/share/dict/words)
```

In the "pattern" argument, use an underscore (`_`) to indicate each unknown character and supply any known letters, e.g., the example above would be `_t_ed`:

```
$ ./helper.py _t_ed  
1: steed
```

Or 6-letter words beginning with "ex" and ending in "s":

```
$ ./helper.py ex___s
1: excess
2: excuss
3: exitus
4: exodos
5: exodus
6: exomis
```

Hints

- If you know about regular expressions, that is a natural way to solve this problem. See how elegantly you can solve the problem.
- Even if you do know how to solve use regexes, try solving without them.

Solution

```
1 #!/usr/bin/env python3
2 """Crossword helper"""
3
4 import argparse
5 import io
6 import re
7 from typing import List, TextIO
8
9
10 # -----
11 def get_args() -> argparse.Namespace:
12     """Get command-line arguments"""
13
14     parser = argparse.ArgumentParser(
15         description='Crossword helper',
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18     parser.add_argument('pattern', metavar='str', help='The pattern to
search')
19
20     parser.add_argument('-w',
21                         '--wordlist',
22                         help='Wordlist to search',
23                         metavar='str',
24                         type=argparse.FileType('r'),
25                         default='/usr/share/dict/words')
26
27     return parser.parse_args()
28
29
30 # -----
31 def regex_solution(pattern: str, wordlist: TextIO) -> List[str]:
32     """Using regular expressions"""
33
34     regex = r'\b{}\b'.format(pattern.replace('_', '.'))\n35     return re.findall(regex, wordlist.read())
36
37
38 # -----
39 def test_regex_solution():
40     """Test regex_solution"""
41
42     text = lambda: io.StringIO('apple banana cherry date')
43     assert regex_solution('_ppl_', text()) == ['apple']
44     assert regex_solution('c_e_ry', text()) == ['cherry']
45
46
```

```

47 # -----
48 def manual_solution(pattern: str, wordlist: TextIO) -> List[str]:
49     """Not using regular expressions"""
50
51     letters = filter(lambda t: t[1] != '_', enumerate(pattern))
52     wanted_len = len(pattern)
53
54     return list(
55         filter(
56             lambda word: len(word) == wanted_len and all(
57                 [word[i] == char for i, char in letters]),
58             wordlist.read().split()))
59
60
61 # -----
62 def test_manual_solution():
63     """Test manual_solution"""
64
65     text = lambda: io.StringIO('apple banana cherry date')
66     assert manual_solution('_ppl_', text()) == ['apple']
67     assert manual_solution('c_e_ry', text()) == ['cherry']
68
69
70 # -----
71 def main():
72     """Make a jazz noise here"""
73
74     args = get_args()
75     words = regex_solution(args.pattern, args.wordlist)
76     #words = manual_solution(args.pattern, args.wordlist)
77
78     if words:
79         for i, word in enumerate(words, start=1):
80             print('{:3}: {}'.format(i, word))
81     else:
82         print('Found no words matching "{}".'.format(args.pattern))
83
84
85 # -----
86 if __name__ == '__main__':
87     main()

```

Discussion

I rely on `argparse` so very much, and this example is no different. I define a `pattern` as a positional argument and a the `--wordlist` option as a readable file type that has a reasonable default. With this definition, I can safely `read()` the word list argument to get the entire contents of the file. I decided to show two ways to solve the problem, both of which take the `pattern` (a `str`) and the `wordlist` as an open file handle (`TextIO`).

Also, this is the first example where I start introducing *type hints*, a way to decorate arguments and functions to indicate the expected types. For example:

```
def regex_solution(pattern: str, wordlist: TextIO) -> List[str]:
```

Indicates that the function `regex_solution` takes a `pattern` argument which is a `str` and a `wordlist` which is something that will behave like a file handle (`TextIO`). The function will return `->` a `List` of `str` values, that is the strings in `wordlist` that matched the `pattern`. Python itself ignores these annotations, but a tool like `mypy` can be used to verify that your code is correct.

Regular Expressions

The `regex_solution` could be one line, but I wrote it in two for readability. The `pattern` argument to the program uses underscores (`_`) to indicate unknown characters. In regular expressions, the `.` is how we represent "one of any character," so we can use `str.replace` to change those:

```
>>> pattern = '_t_ed'  
>>> pattern.replace('_', '.')  
'.t.ed'
```

I could have used `wordlist.read().split()` to get a list of each word (`List[str]`) and then used a pattern that anchors the above to the beginning (`^`) and end (`$`) of each word:

```
>>> regex = '^{}$'.format(pattern.replace('_', '.'))  
>>> regex  
'^\.t.ed$'
```

So that I could apply this to each word individually:

```
>>> import re  
>>> wordlist = open('/usr/share/dict/words')  
>>> [w for w in wordlist.read().split() if re.search(regex, w)]  
['steed']
```

That works just fine, but I chose instead to use the "word boundary" metacharacter `\b` to anchor the pattern to the beginning and end of each word so that I could `read()` the entire file as a stream. Note

that it's important to enclose this pattern in a "raw" string with `r''` so that the `\b` is interpreted correctly. The `re.findall` method will return every match of the given pattern in a body of text.

```
>>> wordlist = open('/usr/share/dict/words')
>>> regex = r'\b{}\b'.format(pattern.replace('_', '.')) 
>>> re.findall(regex, wordlist.read())
['steed']
```

If I needed to get each `match` object, maybe to use the position of the match or whatnot, I would not use `re.findall`, but for this purpose it was exactly the right function.

The whole function looks like this:

```
>>> import io, re
>>> from typing import List, TextIO
>>> def regex_solution(pattern: str, wordlist: TextIO) -> List[str]:
...     regex = r'\b{}\b'.format(pattern.replace('_', '.')) 
...     return re.findall(regex, wordlist.read())
...
```

To test it, I need to pass it something that works like an open file handle. I'm getting tired of having to use `open('/usr/share/dict/words')` repeatedly. I can make a `str` pretend to be an open file handle using `io.StringIO`. This will return a *lazy* object that, like a normal file handle, will not be consumed until I actually use the contents, but then, once consumed, will be exhausted. I won't be able to `read` it a second time. For that reason, I'm going to create `text` as a `lambda` function that will always return a new `io.StringIO` object:

```
>>> text = lambda: io.StringIO('apple banana cherry date')
>>> text().read()
'apple banana cherry date'
>>> text().read()
'apple banana cherry date'
```

This is a handy way to put tests in your code that read files without having to find actual files to pass as arguments. I can manually try my function in the REPL:

```
>>> regex_solution('_ppl_', text())
['apple']
>>> regex_solution('c_e_ry', text())
['cherry']
```

And I can write a test function looks like this:

```
def test_regex_solution():
    text = lambda: io.StringIO('apple banana cherry date')
    assert regex_solution('_ppl_', text()) == ['apple']
    assert regex_solution('c_e_ry', text()) == ['cherry']
```

I can either put the test directly into `helper.py` or possibly into a `unit.py` (for "unit" tests) or add it to `test.py` and then run `pytest` on whatever file contains the test.

Manual Matching

Trying to solve this without regular expressions can give you a real appreciation for exactly how much time regular expressions can save us. For my manual solution, I thought I would use two criteria to find matching words:

1. The length of a word matches the length of the pattern
2. The word has characters matching in the same positions as in the pattern

For the second point, I thought a list of tuples show the position of each character that is not an underscore would be perfect. We can use `enumerate` on any list to give us position and value of each element. Note that I need to use `list` here to force the REPL to evaluate the generator. You don't need this in your actual code (usually, exceptions can occur):

```
>>> pattern = '_t_ed'
>>> list(enumerate(pattern))
[(0, '_'), (1, 't'), (2, '_'), (3, 'e'), (4, 'd')]
```

Note that generators are lazy and won't generate their values unless forced (e.g., with `list`) and can only be iterated once. (They are so, so lazy.)

```
>>> g = enumerate(pattern)
>>> list(g)
[(0, '_'), (1, 't'), (2, '_'), (3, 'e'), (4, 'd')]
>>> list(g)
[]
```

I only care about the positions of the characters that are *not* underscores, so I can `filter` out the underscores. We talked about `filter` a bit in `rhymers.py`. It's another *higher-order function* that takes *another function* as its first argument. That function is often called the "predicate." The second argument to `filter` is some *iterable* like a `list` or a `generator` that will provide a sequence of values. Usually we write the predicate with a `lambda`, which is a Python keyword that allows us to write a small, one-line function without using the `def` keyword.

```
>>> greet = lambda name: print(f'Hi there, {name}!')
>>> greet('Lincoln')
Hi there, Lincoln!
>>> greet('Steve')
Hi there, Steve!
```

I'd like to use a `filter` to remove the elements where the second element in the tuple is an underscore. One limitation to the `lambda` is that it cannot *unpack* the tuples that are returned by `enumerate` into `i, c` (`i` for "integer," `c` for "character"):

```
>>> list(filter(lambda i, c: c != '_', enumerate(pattern)))
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: <lambda>() missing 1 required positional argument: 'c'
```

I have to accept the `tuple` as-is, so I call my argument `t` to remind me of the type. Then I use `t[1]` to indicate the second part of the tuple which is the character. The `filter` will only allow those list elements to pass through for which the predicate (`lambda`) returns something "truthy."

```
>>> list(filter(lambda t: t[1] != '_', enumerate(pattern)))
[(1, 't'), (3, 'e'), (4, 'd')]
```

If you don't care for `filter`, the same idea can be done with a list comprehension:

```
>>> [t for t in enumerate(pattern) if t[1] != '_']
[(1, 't'), (3, 'e'), (4, 'd')]
```

One of the nicer things about this syntax is that you *can* unpack the tuple, but we need to repack the arguments back into `tuple` form:

```
>>> [(i, char) for i, char in enumerate(pattern) if char != '_']
[(1, 't'), (3, 'e'), (4, 'd')]
```

For this solution, I want to look at each word from `wordlist` individually, so I call `wordlist.read().split()`. Remember that `fh.read()` will read the *entire contents* of a file as a `str`, so you should only use this when you are reasonably sure the file is small enough to fit into your computer's memory (RAM). Then `str.split()` will break that string on spaces into "words" where, if there is punctuation and such in your text, that will still be stuck to the words. We will not worry about that in this exercise as we expect to be reading a regular one-word-per-line dictionary-type file. I'll reuse my `text()` function from above.

```
>>> text().read().split()
['apple', 'banana', 'cherry', 'date']
```

I want to find the words in that list where the length matches the length of the given `pattern` and where the letters in my `pattern` are in the same positions in the word. Let's change the `pattern` to one that will find '`apple`':

```
>>> pattern = '_ppl_'
```

What is the length?

```
>>> wanted_len = len(pattern)
>>> wanted_len
5
```

And let's find the positions of all the non-underscore characters in the `pattern` using the same `filter` we saw before:

```
>>> letters = list(filter(lambda t: t[1] != '_', enumerate(pattern)))
>>> letters
[(1, 'p'), (2, 'p'), (3, 'l')]
```

That is, we want a 5-letter word with a "p" at position 1, a "p" at position 2, and an "l" at position 3.

```
>>> word = 'apple'
>>> len(word) == wanted_len
True
```

We can use a list comprehension to check the positions of the `letters` to the `word`:

```
>>> [word[i] == char for i, char in letters]
[True, True, True]
```

If the `word` is "upper," that will fail:

```
>>> word = 'upper'
>>> [word[i] == char for i, char in letters]
[True, True, False]
```

We can use `all` to reduce this to a single Boolean value:

```

>>> word = 'apple'
>>> all([word[i] == char for i, char in letters])
True
>>> word = 'upper'
>>> all([word[i] == char for i, char in letters])
False

```

We need to go through all the words in `wordlist` and find where both conditions are true. A `for` loop is fairly natural. We can create an empty `list` to hold those that pass.

```

>>> words = []
>>> for word in text().read().split():
...     if len(word) == wanted_len and all(
...         [word[i] == char for i, char in letters]):
...         words.append(word)
...
>>> words
['apple']

```

Or we can use `filter` again. That's what we are doing. We only want to return those words for which the predicate (testing word length and character positions) is `True`:

```

>>> words = list(
...     filter(
...         lambda word: len(word) == wanted_len and all(
...             [word[i] == char for i, char in letters]),
...         text().read().split()))
>>>
>>> words
['apple']

```

The whole function looks like this:

```

>>> def manual_solution(pattern: str, wordlist: TextIO) -> List[str]:
...     letters = filter(lambda t: t[1] != '_', enumerate(pattern))
...     wanted_len = len(pattern)
...     return list(
...         filter(
...             lambda word: len(word) == wanted_len and all(
...                 [word[i] == char for i, char in letters]),
...             wordlist.read().split()))
...

```

Again, we can manually test it in the REPL:

```
>>> manual_solution('_ppl_', text())
['apple']
>>> manual_solution('c_e_ry', text())
['cherry']
```

And add a test function to the appropriate file:

```
def test_manual_solution():
    text = lambda: io.StringIO('apple banana cherry date')
    assert manual_solution('_ppl_', text()) == ['apple']
    assert manual_solution('c_e_ry', text()) == ['cherry']
```

To finish, we need to check if any words matched. If so, we print them out, numbered and nicely aligned; otherwise, we let the user know that no matches were found.

```
>>> words = regex_solution('_ppl_', text())
>>> if words:
...     for i, word in enumerate(words, start=1):
...         print('{:3}: {}'.format(i, word))
... else:
...     print('Found no words matching "{}".'.format(args.pattern))
...
1: apple
```

I hope you tried solving this problem with and without regular expressions as there is much to learn by each method.

Going Further

- Add code that will clean up the incoming text so as to remove any character not in the range [a-zA-Z], then use a non-dictionary file like the US Constitution as the --wordlist.
- How would you handle case-insensitive matching?
- What if you were unsure of a letter in a particular position and wanted to indicate a choice, e.g., "_t_e[rd]"? That is almost a valid regex. Could you handle such a pattern in your manual solution?

Workout Of the Day (Parsing CSV file, creating text table output)

[wod string] | *wod-string.png*

Write a Python program called `wod.py` that will create a Workout Of (the) Day (WOD) from a list of exercises provided in CSV format (default `wod.csv`). Accept a `-n|--num_exercises` argument (default 4) to determine the sample size from your exercise list. Also accept a `-e|--easy` flag to indicate that the reps should be cut in half. Finally accept a `-s|--seed` argument to pass to `random.seed` for testing purposes.

When run with `-h|--help`, it should print a usage:

```
$ ./wod.py -h
usage: wod.py [-h] [-f str] [-s int] [-n int] [-e]

Create Workout Of (the) Day (WOD)

optional arguments:
  -h, --help            show this help message and exit
  -f str, --file str    CSV input file of exercises (default: wod.csv)
  -s int, --seed int     Random seed (default: None)
  -n int, --num_exercises int
                        Number of exercises (default: 4)
  -e, --easy             Make it easy (default: False)
```

The input file should be comma-separated values with headers for "exercise" and "reps." See the `wod.csv` file for an example. Here is the structured as shown by `tablify.py` (included in the `bin` directory):

```
$ tablify.py wod.csv
+-----+-----+
| exercise | reps   |
+-----+-----+
| Burpees  | 20-50  |
| Situps   | 40-100 |
| Pushups  | 25-75  |
| Squats   | 20-50  |
| Pullups  | 10-30  |
| HSPU      | 5-20   |
| Lunges   | 20-40  |
| Plank    | 30-60  |
| Jumprope | 50-100 |
| Jumping Jacks | 25-75 |
| Crunches | 20-30  |
| Dips     | 10-30  |
+-----+-----+
```

You should use the range of reps to choose a random integer value in that range. I would suggest you use the `tabulate` module to format the output as shown.

When run with no arguments, it should use all the defaults:

```
$ ./wod.py
Exercise   Reps
-----
Crunches   26
HSPU        9
Squats     43
Pushups    36
```

When run with a seed, the output should be exactly reproducible:

```
$ ./wod.py -s 1
Exercise   Reps
-----
Pushups    32
Jumping Jacks | 56
Situps    88
Pullups    24
```

When run with the `--easy` flag, the upper and lower bounds for the "reps" should halved and then a random number chosen from that range:

```
$ ./wod.py -s 1 -e
Exercise      Reps
-----
Pushups        15
Jumping Jacks 27
Situps         44
Pullups        12
```

The `--file` option tells the program to read a different input file:

```
$ ./wod.py -f wod2.csv -n 5
Exercise      Reps
-----
Erstwhile Lunges 9
Existential Earflaps 32
Rock Squats     21
Squatting Chinups 49
Flapping Leg Raises 17
```

Hints:

- Use the `csv` module's `DictReader` to read the input CSV files
- Break the `reps` field on the `-` character, coerce the low/high values to `int` values, and then use the `random` module to choose a random integer in that range. Also see if the `random` module can help you *sample* some exercises.
- Read the docs on the `tabulate` module to figure out how to print your data.

Solution

```
1 #!/usr/bin/env python3
2 """Create Workout Of (the) Day (WOD)"""
3
4 import argparse
5 import csv
6 import io
7 import random
8 from itertools import starmap
9 from tabulate import tabulate
10
11
12 # -----
13 def get_args():
14     """get command-line arguments"""
15
16     parser = argparse.ArgumentParser(
17         description='Create Workout Of (the) Day (WOD)',
18         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
19
20     parser.add_argument('-f',
21                         '--file',
22                         help='CSV input file of exercises',
23                         metavar='str',
24                         type=argparse.FileType('r'),
25                         default='wod.csv')
26
27     parser.add_argument('-s',
28                         '--seed',
29                         help='Random seed',
30                         metavar='int',
31                         type=int,
32                         default=None)
33
34     parser.add_argument('-n',
35                         '--num_exercises',
36                         help='Number of exercises',
37                         metavar='int',
38                         type=int,
39                         default=4)
40
41     parser.add_argument('-e',
42                         '--easy',
43                         help='Make it easy',
44                         action='store_true')
45
46     return parser.parse_args()
47
```

```

48
49 # -----
50 def read_csv(fh):
51     """Read the CSV input"""
52     exercises = []
53
54     for row in csv.DictReader(fh, delimiter=','):
55         name = row['exercise']
56         low, high = row['reps'].split('-')
57         exercises.append((name, int(low), int(high)))
58
59     return exercises
60
61
62 # -----
63 def test_read_csv():
64     """Test read_csv"""
65
66     text = io.StringIO('exercise,reps\nfoo,10-20\nbar,30-40')
67     assert read_csv(text) == [('foo', 10, 20), ('bar', 30, 40)]
68
69
70 # -----
71 def main():
72     """Make a jazz noise here"""
73
74     args = get_args()
75     random.seed(args.seed)
76     exercises = read_csv(args.file)
77
78     def mk_exercise(name, low, high):
79         if args.easy:
80             low, high = int(low / 2), int(high / 2)
81             return (name, random.randint(low, high))
82
83     wod = starmap(mk_exercise, random.sample(exercises, k=args.num_exercises))
84     print(tabulate(wod, headers=('Exercise', 'Reps')))
85
86
87 # -----
88 if __name__ == '__main__':
89     main()

```

Discussion

Defining the arguments

As usual, I start with my `get_args` first to define what the program expects. Most important is a `--file`. This is not required (it has a `default` value of the `wod.csv` file), so I make it an optional named argument. I use the `type=argparse.FileType('r')` so I can push the validation of the argument to `argparse`. The `--seed` and `--num_exercises` options must to be `type=int`. The `--easy` option is a *flag*—the value of `args.easy` will be `True` when `--easy` is present and will be `False` otherwise

Reading a CSV file

I recommended you use the `csv.DictReader` module to parse the CSV files into a list of dictionaries that represent each rows values merged with the column names in the first row. If the file looks like this:

```
$ head -3 wod.csv
exercise,reps
Burpees,20-50
Situps,40-100
```

The `DictReader` will use the column names in the first row ("exercise" and "reps") as the *key* names for a dictionary it will create for each row:

```
>>> import csv
>>> fh = open('wod.csv')
>>> rows = list(csv.DictReader(fh, delimiter=','))
>>> rows[0]
OrderedDict([('exercise', 'Burpees'), ('reps', '20-50')])
```

Here is the function I wrote to read the CSV file:

```
def read_csv(fh):
    exercises = []
    for row in csv.DictReader(fh, delimiter=','):
        name = row['exercise']
        low, high = row['reps'].split('-')
        exercises.append((name, int(low), int(high)))
    return exercises
```

I set `exercises` to an empty list (`[]`). The call to `csv.DictReader` will return an *iterable* (a sequence of items) of dictionaries representing each row of data. Note that there are 3 lines of `text` but only two rows of data because the first row is a header that names the fields for the dictionaries. Inside the

`for` loop, I get the `name` of the exercise and use `str.split` to break the `reps` field on the `-` character into `low` and `high`. I decided I would represent my `exercises` as a `list` of 3-tuples containing the name of the exercise and the low and high values as integers.

For purposes of showing you:

1. a very small data set
2. how to fake a file handle for testing

I'm going to use the `io.StringIO` module to create a string that will pretend to be an open file handle:

```
>>> import io  
>>> text = io.StringIO('exercise,reps\nfoo,10-20\nbar,30-40')
```

So that we can call our function in the REPL to test it:

```
>>> exercises = read_csv(text)  
>>> exercises  
[('foo', 10, 20), ('bar', 30, 40)]
```

That's what I wanted. Let's put a ring on that test and add it as a function to the `wod.py` file:

```
def test_read_csv():  
    text = io.StringIO('exercise,reps\nfoo,10-20\nbar,30-40')  
    assert read_csv(text) == [('foo', 10, 20), ('bar', 30, 40)]
```

For the purposes of this exercise, I'm not going to give you bad data, but you should think about all the things that *could* go wrong in the real world:

- The file is empty
- The headers are wrong or missing
- The delimiter is not a comma (maybe it's a tab)
- One or more expected fields are missing
- The `reps` field is not in format `x-y`
- The `reps` field is in the correct format, but the `x` and `y` values in `x-y` are not parsable as integers

Choosing the exercises

Before I use the `random` module, I need to be sure to set the `random.seed` with any input from the user. Now I can use the `random.sample` method to choose the number of exercises in `args.num_exercises`. That function will return a new `list` of exercises which I can use a `for` loop to iterate. Remember that `exercise` element in the list is a 3-tuple with `(name, low, high)` which I can *unpack* into three variables. If the `easy` is `True`, then I halve the low/high values and I use

`random.randint` to select a number from them:

```
>>> easy = True
>>> wod = []
>>> for name, low, high in random.sample(exercises, k=2):
...     if easy:
...         low, high = int(low / 2), int(high / 2)
...     wod.append((name, random.randint(low, high)))
...
...
```

This is a very standard, imperative approach where we create an empty variable (here `wod = []`), then use a loop (`for` or `while`) to add elements to the variable. In the end, the `wod` (workout of the day) has these exercises:

```
>>> wod
[('bar', 17), ('foo', 10)]
```

Let's see if we get rid of the loop and create the `wod` in a *functional* manner. To do that, we need to create a function that will take just *one* exercise and transform it:

```
>>> def mk_exercise(name, low, high):
...     if easy:
...         low, high = int(low / 2), int(high / 2)
...     return (name, random.randint(low, high))
...
...
```

Could you write that as a `lambda` on one line? (*Should* you do that?)

Then we can `map` our `random.sample` into that function:

```
>>> wod = list(map(mk_exercise, random.sample(exercises, k=2)))
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: mk_exercise() missing 2 required positional arguments: 'low' and 'high'
```

Ooops. Not what we expected. It can be difficult to understand this error message, but the problem lies in the fact that we wrote `mk_exercise (name, low, high)` to accept 3 arguments but the `map` is passing in a single `tuple`. We can either use a `lambda` to get access to the tuple (which I'll call `t`) and then use the "splat" (`*t`) to expand it into a `list`:

```
>>> wod = list(map(lambda t: mk_exercise(*t), random.sample(exercises, k=2)))
>>> wod
[('foo', 10), ('bar', 15)]
```

Or use the `itertools.starmap` to do this (where `*` is the "star"):

```
>>> from itertools import starmap
>>> wod = list(starmap(mk_exercise, random.sample(exercises, k=2)))
>>> wod
[('foo', 5), ('bar', 16)]
```

Could you rewrite the `read_csv` function to use a `map` or a list comprehension rather than a `for` loop? Could you combine the logic of these two ideas into one? *Should* you combine the logic of these two ideas into one? Is it better to have a function do just one thing or all the things? If you choose to write a single function to do both, be sure to write a test that checks your work.

Formatting text tables with `tabulate`

I suggested you use the `tabulate` module for format the output. I chose the default table style, but I hope you read the docs and see all the other formats you could use.

```
>>> from tabulate import tabulate
>>> print(tabulate(wod, headers=('Exercise', 'Reps')))
Exercise      Reps
-----  -----
foo          5
bar         16
```

Going Further

- Create your own input file. Add tests to verify it is processed correctly.
- Add an option to choose a different `tabulate` format.

Password Strength

Creating passwords that are both difficult to guess and easy to remember is not easy. An XKCD comic (<https://xkcd.com/936/>) describes an algorithm that provides both security and recall by suggesting that a password be composed of "four random common words." For instance, the comic suggests that the password composed of the words "correct," "horse," "battery," and "staple" would provide "~44 bits of entropy" which would require around 550 years for a computer to guess given 1,000 guess per second.

In this exercise, we will write a program called `password.py` that will create some `--num` number of passwords (default 3) each created by randomly choosing some `--num_words` (default 4) from a unique set of words from one or more input files (default `/usr/share/dict/words`). As it will use the `random` module, the program will also accept a random `--seed` argument. The words from the input files will need to be a minimum length of some `--min_word_len` (default 3) after removing any non-characters.

The words from the input files should be lowercase except for the first letter which should be uppercase (see the `str.title()` method). This is both to introduce both cases of letters and to make it a bit easier to see and remember the individual words:

```
$ ./password.py -s 1
ChromePorometerUnwastableUnconditionated
ThujaAwesomelyEyeglaceCatabolin
OptiveThicketMoratoriaNoncompetent
```

Or we could use the US Constitution as the input:

```
$ ./password.py -s 1 ../inputs/const.txt
DetermineBusinessJacobDaniel
UponSubjectsThirdQuorum
GilmanConfessionTwentyApportioned
```

Additionally the program should have an option to perform common character substitutions to make it `133+` ("leet" or "elite hacker"-style):

```
$ ./password.py -s 1 ../inputs/const.txt --133t
D3T3rm1n3Bu51ne5sj@c0bD@n13l
P05+hUnDReDGUNn1nGT@X3s
51X+Hw45hing+0NC0NViC+eDeNG@G3
```

The input (or training) texts can have a huge effect on the memorability of the generated passwords. Rather than select from all known English words, it would be better to select, say, just from some interesting nouns, verbs, and adjectives. To produce such lists, I have included a program called `harvest.py` which you can use to extract these sorts of words from texts you find interesting. For instance, I have used this to extract just those parts of speech from the US

Constitution and which I have included in the `password/const` directory. Notice the difference:

```
$ ./password.py -s 1 const/nouns.txt  
DeclarationResignationChoiceImportation  
DeathPunishmentPowerPrivilege  
ThingObjectionEnumerationContinuance
```

Or you could include the verbs and adjectives, too:

```
$ ./password.py -s 1 const/*  
DollarCitizenLandDevolve  
VacancySubjectThingQuestion  
GreatCrimeUniformApportion
```

Here is a sample of passwords generated from the text of *The Scarlet Letter* by Nathaniel Hawthorne:

```
$ ./password.py -s 2 scarlet/*  
ChurchDrearyDiscernScarlet  
HighPeopleMinisterLess  
BringHeadTalkSmile
```

You will learn:

- How to use random
- How clean text
- How to use sets to unique input
- How to randomly sample from a collection
- The effect of sorting on selections and randomness
- How to transform text

l33t-ify

To convert the text into l33t-speak, you should first use your ransom note algorithm to randomly capitalize letters in the text. Here is the test for your `ransom` function:

```
def test_ransom():
    """Test ransom"""

    random.seed(1)
    assert(ransom('Money') == 'moNeY')
    assert(ransom('Dollars') == 'DOLLaRs')
    random.seed(None)
```

Then use this substitution table to transform characters:

'a': '@'
'A': '4'
'o': '0'
'O': '0'
't': '+'
'e': '3'
'E': '3'
'I': '1'
'S': '5'

Here is the test for the `l33t` function:

```
def test_l33t():
    """Test l33t"""

    random.seed(1)
    assert(l33t('Money') == 'm0N3Y')
    assert(l33t('Dollars') == 'D0LL@Rs')
    random.seed(None)
```

They have been included in the `unit.py` that you could use like `pytest -xv unit.py`, or you could include them directly into your `password.py` program and run `pytest -xv password.py`.

Solution

```
1 #!/usr/bin/env python3
2 """Password maker"""
3
4 import argparse
5 import os
6 import sys
7 import re
8 import random
9
10
11 # -----
12 def get_args():
13     """Get command-line arguments"""
14
15     parser = argparse.ArgumentParser(
16         description='Password maker',
17         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19     parser.add_argument('file',
20                         metavar='FILE',
21                         type=argparse.FileType('r'),
22                         nargs='*',
23                         help='Input file(s)',
24                         default=[open('/usr/share/dict/words')])
25
26     parser.add_argument('-n',
27                         '--num',
28                         metavar='INT',
29                         type=int,
30                         default=3,
31                         help='Number of passwords to generate')
32
33     parser.add_argument('-w',
34                         '--num_words',
35                         metavar='INT',
36                         type=int,
37                         default=4,
38                         help='Number of words to use for password')
39
40     parser.add_argument('-m',
41                         '--min_word_len',
42                         metavar='INT',
43                         type=int,
44                         default=3,
45                         help='Minimum word length')
46
47     parser.add_argument('-s',
```

```

48             '--seed',
49             metavar='INT',
50             type=int,
51             help='Random seed')
52
53     parser.add_argument('-l',
54                         '--l33t',
55                         action='store_true',
56                         help='Obsfuscate letters')
57
58     return parser.parse_args()
59
60
61 # -----
62 def main():
63     """Make a jazz noise here"""
64
65     args = get_args()
66     random.seed(args.seed)
67     words = set()
68
69     clean = lambda word: re.sub('[^a-zA-Z]', '', word)
70     for fh in args.file:
71         for word in filter(lambda w: len(w) > args.min_word_len,
72                             map(clean,
73                                 fh.read().lower().split())):
74             words.add(word.title())
75
76     words = sorted(list(words))
77
78     for _ in range(args.num):
79         password = ''.join(random.sample(words, args.num_words))
80         print(l33t(password) if args.l33t else password)
81
82
83 # -----
84 def l33t(text):
85     """l33t"""
86
87     text = ransom(text)
88     xform = {
89         'a': '@',
90         'A': '4',
91         'o': '0',
92         'O': '0',
93         't': '+',
94         'e': '3',
95         'E': '3',
96         'I': '1',
97         'S': '5'
98     }

```

```
99      for x, y in xform.items():
100          text = text.replace(x, y)
101
102      return text
103
104
105 # -----
106 def ransom(text):
107     """Randomly choose an upper or lowercase letter to return"""
108
109     return ''.join(
110         map(lambda c: c.upper() if random.choice([0, 1]) else c.lower(),
111             text))
112
113 # -----
114 if __name__ == '__main__':
115     main()
```

Discussion

Stuff

Review

What we did.

Going Further

What else?