

Playful Python: Learning the language through games and puzzles

Ken Youens-Clark



Figure 1: The Playful Python

Contents

Introduction	6
Forking GitHub repo	6
new.py	7
\$PATH	7
Testing your programs	8
Author	8
Copyright	8
Outline	9
Programs	9
Chapter 1: Article Selector	12
Solution	14
Discussion	15
Chapter 2: Jump the Five	17
Solution	19
Discussion	20
Chapter 3: Picnic	21
Solution	22
Discussion	23
Chapter 4: Howler	24
Solution	25
Discussion	27
Chapter 5: Apples and Bananas	28
Solution	30
Discussion	33
Method 1: Iterate every character	33
Method 2: str.replace	34
Method 3: str.translate	34
Method 4: List comprehension	35
Method 5: List comprehension with function	37
Method 6: map with a lambda	38
Method 7: map with new_char	39
Method 8: Regular expressions	39
Chapter 6: Telephone	41
Solution	43
Discussion	45
Mutations in DNA	46
Chapter 7: Bottles of Beer Song	48

Counting down	49
Solution	51
Discussion	53
Chapter 8: Gashlycrumb	54
Solution	56
Discussion	58
Edward Gorey	60
Alternate text	60
Interactive version	60
Chapter 9: Movie Reader	61
Solution	62
Chapter 10: Palindromes	63
Solution	64
Chapter 11: Ransom	66
Solution	67
Discussion	69
Chapter 12: Simple Rhymers	72
Solution	74
Discussion	76
Chapter 13: Rock, Paper, Scissors	80
Solution	81
Chapter 14: Abuse	84
Solution	85
Chapter 15: Acronym	87
Solution	88
Chapter 16: Workout Of (the) Day (WOD)	91
Solution	93
Discussion	95
Chapter 17: Blackjack	97
Solution	99
Chapter 18: Family Tree	102
Solution	103
Chapter 19: Gematria: Numeric encoding of text	106
Solution	108
Discussion	110
Reading lines of text	110

List comprehensions vs map	111
Encoding one word	111
Finding the words	112
Encoding all words	113
Chapter 20: Histogram	115
Solution	117
Chapter 21: Guessing Game	119
Solution	121
Chapter 22: Mommy’s Little (Crossword) Helper	124
Hints	125
Solution	126
Discussion	128
Regular Expressions	128
Manual Matching	129
Summary	130
Chapter 23: Kentucky Friar	131
Solution	133
Discussion	135
Chapter 24: Mad Libs	137
Solution	139
Chapter 25: License Plates	141
Solution	143
Chapter 26: Gibberish Generator	145
Solution	147
Chapter 27: Pig Latin	150
Solution	152
Chapter 28: Soundex Rhymer	154
Testing the stemmer	155
Solution	157
Discussion	160
Using Soundex	160
Chapter 29: Substring Guessing Game	162
Solution	163
Chapter 30: Tic-Tac-Toe Outcome	167
Solution	168

Chapter 31: Twelve Days of Christmas	170
Solution	171
Chapter 32: War	173
Solution	176
Chapter 33: Anagram	178
Solution	180
Discussion	183
Logging	183
Reading wordlist	183
defaultdict	184
Identifying anagrams	185
Selecting words to compare	186
Chapter 34: Hangman	189
Solution	191
Chapter 35: First Bank of Change	195
Solution	197
Discussion	199
Chapter 36: Runny Babbit	202
Solution	204
Discussion	206
Chapter 37: Markov Chain	209
Solution	211
Discussion	214
Chapter 38: Hamming Chain	217
Solution	219
Chapter 39: Morse Encoder/Decoder	222
Solution	223
Chapter 40: ROT13 (Rotate 13)	227
Solution	229
Chapter 41: Tranpose ABC Notation	231
Solution	234
Chapter 42: Word Search	236
Solution	239
Discussion	243
Appendix 1: argparse	247
Types of arguments	247

Datatypes of values	247
Number of arguments	250
Choices	252
Automatic help	253
Getting the argument values	253
Appendix 2: Truthiness	255
Appendix 3: File Handles	258
File Modes	258
STDIN, STDOUT, STDERR	259
Appendix 4: N-grams, K-mers, and Markov Chains	260

Introduction

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

I believe you can learn serious things through silly games. I also think you will learn best by *doing*. This is a book of programming exercises. 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.

I won’t necessarily show you beforehand how to write each program. I’ll describe what the program should do and provide some discussion about how to write it. I’ll also create an appendix with short examples of how to do things like how to use `argparse`, how to read/write from/to a file, how to process all the files in a directory, how to extract k-mers from a string, etc. I’ll provide some building blocks, but I want you to figure out how to put the pieces together.

Forking GitHub repo

First use the GitHub interface to “fork” this repository into your own account. Then do `git clone` of *your* repository to get a local copy. Inside that checkout, do:

```
git remote add upstream https://github.com/kyclark/playful_python.git
```

This will allow you to `git pull upstream master` in order to get updates. When you create new files, `git add/commit/push` them to *your* repository. (Please do not create pull requests on *my* repository – unless, of course, you have suggestions for improving my repo!).

new.py

I provide some useful programs in the `bin` directory including one called `new.py` that will help you stub out new Python programs using the `argparse` module to parse the command line arguments and options for your programs. I recommend you start every new program with this program. For example, in the `article` directory the `README.md` wants you to create a program called `article.py`. You should do this:

```
$ cd article
$ new.py article
```

This will create a new file called `article.py` (that has been made executable with `chmod +x`, if your operating system supports that) that has example code for you to start writing your program.

\$PATH

Your `$PATH` is a list of directories where your operating system will look for programs. To see what your `$PATH` looks like, do:

```
$ echo $PATH
```

Probably each directory is separated by a colon (:). *The order of the directories matters!* For instance, it's common to have more than one version of Python installed. When you type `python` on the command line, the directories in your `$PATH` are searched in order, and the first `python` found is the one that is used (and it's probably Python version 2!)

You could execute `new.py` by giving the full path to the program, e.g., `$HOME/work/playful_python/bin/new.py`, but that's really tedious. It's best to put `new.py` into one of the directories that is already in your `$PATH` like maybe `/usr/local/bin`. The problem is that you probably need administrator privileges to write to most of the directories that are in your `$PATH`. If you are working on your laptop, this is probably not a problem, but if you are on a shared system, you probably won't be able to copy the program into your `$PATH` directories.

An alternative is to alter your `$PATH` to include the directory where `new.py` is located. E.g., if `new.py` is in `$HOME/work/playful_python/bin/`, then add this directory to your `$PATH` – probably by editing `.bashrc` or `.bash_profile` located in your `$HOME` directory (if you use `bash`). See the documentation for your shell of choice to understand how to edit and persist your `$PATH`.

For what it's worth, I always create a `$HOME/.local` directory for local installations of software I need, so I add `$HOME/.local/bin` to my `$PATH`. Then I copy programs like `new.py` there and they are available to me anywhere on the system.

Testing your programs

Once you have stubbed out your new program, 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 `main` to accomplish the task described in the README. To run the test suite using `make`, you can type `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 -v 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, e.g., often a test will ask you to alter one bit of text from the command line, and this it will ask you to read and alter the text from a file. I would suggest you solve the tests in order.

Author

Ken Youens-Clark is a Sr. Scientific Programmer in the lab of Dr. Bonnie Hurwitz at the University of Arizona. He started college as a music major at the University of North Texas but changed to English lit for his BA in 1995. He started programming at his first job out of college, working through several languages and companies before landing in bioinformatics in 2001. In 2019 he earned his MS in Biosystems Engineering, and enjoys helping people learn programming. When he's not working, he likes playing music, riding bikes, cooking, and being with his wife and children.

Copyright

© Ken Youens-Clark 2019

Outline

I aim to have 40-50 programs complete with specs, examples, inputs, and test suites. They won't necessarily have a specific order, but they will be grouped into easiest/harder/hardest categories. As many programs use common ideas (e.g., regular expressions, graphs, infinite loops), there will be an appendix section with explanations of how to explore those ideas.

I have in mind a layout where each program gets four pages:

1	2	3	4
+	+	+	+
illus/	specs	solution	notes
info			
+	+	+	+

1. If a short program, perhaps an illustration; if longer, maybe some background or hints.
2. The `README.md` information (specs, example output)
3. The `solution.py` contents
4. Annotation of the solution with comments on lines, sections

Programs

The goal is to get the reader to become a *writer* – to try to solve the problems. One technique in teaching is to first present a problem without showing how to solve it. Once the student engages with the problem, they find they want and need the object of the lesson. Each program is intended to flex some programming technique or idea like playing with lists or contemplating regular expressions or using dictionaries. By using `argparse` for the programs, we also cover validation of user input.

Easiest

- **article**: Select “a” or “an” depending on the given argument
- **howler**: Uppercase input text so they YELL AT YOU LIKE “HOWLER” MESSAGES IN HARRY POTTER. (Could also be called “OWEN MEANY”?)
- **jump_the_five**: Numeric encryption based on “The Wire.”

- **bottles_of_beer**: Produce the “Bottle of Beer on the Wall” song. Explores the basic idea of an algorithm and challenges the programmer to format strings.
- **picnic**: Write the picnic game. Uses input, lists.
- **apples_and_bananas**: Substitute vowels in text, e.g., “bananas” -> “bononos”. While the concept is substitution of characters in a string which is actually trivial, it turns out there are many (at least 7) decent ways to accomplish this task!
- **gashlycrumb**: Create a morbid lookup table from text. Natural use of dictionaries.
- **movie_reader**: Print text character-by-character with pauses like in the movies. How to read text by character, use `STDOUT/flush`, and pause the program.
- **palindromes**: Find palindromes in text. Reading input, manipulation of strings.
- **ransom_note**: Transform input text into “RaNSom cASe”. Manipulation of text.
- **rhymmer**: Produce rhyming “words” from input text.
- **rock_paper_scissors**: Write Rock, Paper, Scissors game. Infinite loops, dictionaries.

Harder

- **abuse**: Generate insults from lists of adjectives and nouns. Use of randomness, sampling, and lists.
- **bacronym**: Retrofit words onto acronyms. Use of randomness and dictionaries.
- **blackjack**: Play Blackjack (card game). Use of randomness, combinations, dictionaries.
- **family_tree**: Use GraphViz to visualize a family tree from text. Parsing text, creating graph structures, creating visual output.
- **gematria**: Calculate numeric values of words from characters. Manipulation of text, use of higher-order functions.
- **guess**: Write a number-guessing game. Use of randomness, validation/coercion of inputs, use of exceptions.
- **kentucky_fryer**: Turn text into Southern American English. Parsing, manipulation of text.
- **mad_libs**: TBD
- **markov_words**: Markov chain to generate words. Use of n-grams/k-mers, graphs, randomness, logging.
- **piggie**: Encode text in Pig Latin. Use of regular expressions, text manipulation.
- **sound**: Use Soundex to find rhyming words from a word list.
- **substring**: Write a game to guess words sharing a common substring. Dictionaries, k-mers/n-grams.

- **tictactoe**: Write a Tic-Tac-Toe game. Randomness, state.
- **twelve_days_of_christmas**: Produce the “12 Days of Christmas” song. Algorithms, loops.
- **war**: Play the War card game. Combinations, randomness.
- **license_plates**: Explore how a regular expression engine works by creating alternate forms of license plates.

Hardest

- **anagram**: Find anagrams of text. Combinations, permutations, dictionaries.
- **hangman**: Write a Hangman (word/letter-guessing game). Randomness, game state, infinite loops, user input, validation.
- **markov_chain**: Markov chain to generate text. N-grams at word level, parsing text, list manipulations.
- **morse**: Write a Morse encoder/decoder. Dictionaries, text manipulation.
- **rot13**: ROT13-encode input text. Lists, encryption.

Chapter 1: Article Selector

Write a Python program called `article.py` that will select `a` or `an` for a given word depending on whether the word starts with a consonant or vowel, respectively.

When run with no arguments or the `-h|--help` flags, it should print a usage statement:

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

Article selector

positional arguments:
 str Word

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

When run with a single positional argument, it should print the correct article and the given argument.

```
$ ./article.py bear
a bear
$ ./article.py Octopus
an Octopus
```

The tests will only give you words that start with an actual alphabetic character, so you won't have to detect numbers or punctuation or other weird stuff. Still, how might you extend the program to ensure that given argument only starts with one of the 26 characters of the English alphabet?

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), 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`.

- For the purposes of `x in y`, a string (`str`) is a **list** of characters, so you could ask if a character is in a string.
- Use the `print` function to print out the article joined to the argument. Put a single space in between.
- 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.

Solution

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

Discussion

As with all the solutions presented, this assumes you have stubbed the program with `new.py` and that you are using the `argparse` module. I suggest putting this logic into a separate function which here is called `get_args` and which I like to define first so that I can see right away when I'm reading the program what the program expects as input. On line 12, I set the `description` for the program that will be displayed with the help documentation. On line 15, I indicate that the program expects just one *positional* argument, no more, no less. Since it is a “word” that I expect, I called the argument `word` which is also how I will access the value on line 25. I use the `metavar` on line 15 to let the user know that this should be a string.

The `get_args` function will **return** the result of parsing the command line arguments which I put into the variable `args` on line 24. I can now access the `word` by call `args.word`. Note the lack of parentheses – it's not `args.word()` – as this is not a function call. Think of it like a slot where the value lives.

On line 26, we need to figure out whether the `article` should be `a` or `an`. We'll use a very simple rule that any word that has a first character that is a vowel should get `an` and otherwise we choose `a`. This obviously misses actual pronunciations like in American English we don't pronounce the “h” in “herb” and so actually say “an herb” whereas the British *do* pronounce the “h” and so would say “an herb”. (Even more bizarre to me is that the British leave off the article entirely for the word “hospital” as in, “The Queen is in hospital!”) Nor will we consider words where the initial `y` acts like a vowel.

We can access the first character of the `word` with `word[0]` which looks the same as how we access the first element of a list. Strings are really list of characters, so this isn't so far-fetched, but we do have to remember that Python, like so many programming languages, starts numbering at 0, so we often talked about the first element of a list as the “zeroth” element.

To decide if the given word starts with a vowel, we ask is `word[0].lower()` in `'aeiou'`. So, to unpack that, `word[0]` returns a one-character-long `str` type which has the method `.lower()` which we call using the parentheses. Without the parens, this would just be the *idea* of the function that returns a lowercased version of the string. Understand that the `word` remains unchanged. The function does not lowercase `word[0]`, it only *returns a lowercase version* of that character.

```
>>> word = 'APPLE'
>>> word
'APPLE'
>>> word[0].lower()
'a'
>>> word
'APPLE'
```

The `x in y` form is a way to ask if element `x` is in the collection `y`:

```
>>> 'a' in 'abc'
True
>>> 'foo' in ['foo', 'bar']
True
>>> 3 in range(5)
True
>>> 10 in range(3)
False
```

The *if expression* (also called a “ternary” expression) is different from an *if statement*. An *expression* returns a value, and a *statement* does not. The *if expression* must have an *else*, but the *if statement* does not have this requirement. The first value is returned if the predicate (the bit after the *if*) evaluates to *True* in a Boolean context (cf. “Truthiness”), otherwise the last value is returned:

```
>>> 'Hooray!' if True else 'Shucks!'
'Hooray!'
```

The longer way to write this would have been:

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

Or more succinctly:

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

Cf. appendices: `argparse`, Truthiness

Chapter 2: Jump the Five



Figure 2: “When I get up, nothing gets me down.” - D. L. Roth

Write a program called `jump.py` that will encode any number using “jump-the-five” algorithm that selects as a replacement for a given number one that is opposite on a US telephone pad if you jump over the 5. The numbers 5 and 0 will exchange with each other. So, “1” jumps the 5 to become “9,” “6” jumps the 5 to become “4,” “5” becomes “0,” etc.

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

Print a usage statement for `-h|--help` or if there are no arguments.

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

Your program should replace numbers *anywhere* in the input string:

```
$ ./jump.py 555-1212
```

```
000-9898
$ ./jump.py 'Call 1-800-329-8044 today!'
Call 9-255-781-2566 today!
```

Hints:

- The numbers can occur anywhere in the text, so I recommend you think of how you can process the input character-by-character.
- To me, the most natural way to represent the substitution table is in a `dict`.
- Read the documentation on Python's `str` class to see what you can do with a string. For instance, there is a `replace` method. Could you use that?

Solution

```
1  #!/usr/bin/env python3
2  """Jump the Five"""
3
4  import argparse
5
6
7  # -----
8  def get_args():
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')
16
17     return parser.parse_args()
18
19
20 # -----
21 def main():
22     """Make a jazz noise here"""
23
24     args = get_args()
25     text = args.text
26     jumper = {'1': '9', '2': '8', '3': '7', '4': '6', '5': '0',
27              '6': '4', '7': '3', '8': '2', '9': '1', '0': '5'}
28
29     for char in text:
30         print(jumper[char] if char in jumper else char, end='')
31
32     print()
33
34
35 # -----
36 if __name__ == '__main__':
37     main()
```

Discussion

On line 15, we indicate the one positional argument our program expects which is some `text` which we can retrieve on line 25. 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. Later problems will require much more complex arguments, so it's good to get used to this now.

I suggested you could represent the substitution table as a `dict` which is what I create on line 26. Each number `key` has its substitute as the `value` in the `dict`. 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. They are being stored as `str` values, not `int`. This is because we will be reading from a `str`. If we stored them as `int` keys and values, we would have to coerce the `str` types using the `int` function:

```
>>> type('4')
<class 'str'>
>>> type(4)
<class 'int'>
>>> type(int('4'))
<class 'int'>
```

To process the `text` by individual character (`char`), we can use a `for` loop on line 29. Like in the `article` solution, I decided to use an *if expression* where I look to see if the `char` is in the `jumper` dictionary. In the `article`, you saw we asked if a character was in the string `'aeiou'` (which can also be thought of as a `list` of characters). Here when we ask if a `char` (which is a string) is in a `dict`, Python looks to see if there is a `key` in the dictionary with that value. So if `char` is `'4'`, then we will print `jumper['4']` which is `'6'`. If the `char` is not in `jumper` (meaning it's not a digit), then we print `char`.

Another way you could have solved this would be to use the `str.translate` method which needs a translation table that you can make with the `str.maketrans` method:

```
>>> s = 'Jenny = 867-5309'
>>> s.translate(str.maketrans(jumper))
'Jenny = 243-0751'
```

Note that you could *not* use `str.replace` to change each number in turn as you would first change 1 to 9 and then you'd get to the 9s that were in the original string and the 9s that you changed from 1s and you'd change them back to 1s!

Chapter 3: Picnic

Write a Python program called `picnic.py` that accepts one or more positional arguments as the items to bring on a picnic. In response, print “You are bringing ...” where “...” should be replaced according to the number of items where:

1. If one item, just state, e.g., if `chips` then “You are bringing chips.”
2. If two items, put “and” in between, e.g., if `chips` `soda` then “You are bringing chips and soda.”
3. If three or more items, place commas between all the items INCLUDING BEFORE THE FINAL “and” BECAUSE WE USE THE OXFORD COMMA, e.g., if `chips` `soda` `cupcakes` then “You are bringing chips, soda, and cupcakes.”

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

Picnic game

positional arguments:
str Item(s) to bring

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

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

Solution

```
1  #!/usr/bin/env python3
2  """Picnic game"""
3
4  import argparse
5
6
7  # -----
8  def get_args():
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',
16                         metavar='str',
17                         nargs='+',
18                         help='Item(s) to bring')
19
20     return parser.parse_args()
21
22
23 # -----
24 def main():
25     """Make a jazz noise here"""
26
27     args = get_args()
28     items = args.item
29     num = len(items)
30
31     bringing = items[0] if num == 1 else ' and '.join(
32         items) if num == 2 else ', '.join(items[:-1] + ['and ' + items[-1]])
33
34     print('You are bringing {}'.format(bringing))
35
36
37 # -----
38 if __name__ == '__main__':
39     main()
```

Discussion

This program can accept a variable number of arguments which are all the same thing, so the most appropriate way to represent this with `argparse` is shown on lines 15-19 where we define an `item` argument with `nargs='+'` where `nargs` is the *number of arguments* and `+` means *one or more*. Remember, even if the user provides only one argument, you will still get a `list` with just one element.

We put the `items` into a variable on line 28. Note that I call it by the plural `items` because it's probably going to be more than one. Also, I call the variable something informative, not just `args` or something too generic. Lastly, I need to decide how to format the items. As in the article selector, I'm using an *if expression* rather than an *if* statement that would look like this:

```
bringing = ''
if num == 1:
    bringing = items[0]
elif num == 2:
    bringing = ' and '.join(items)
else:
    bringing = ', '.join(items[:-1]) + [ 'and ' + items[-1]]
```

But I chose to condense this down into a double *if* expression with the following form:

```
bringing = one_item if num == 1 else two_items if num == 2 else three_items
```

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

```
>>> 'I spy something {}!'.format('blue')
'I spy something blue!'
```

You can also put names inside the `{}` and pass in key/value pairs in any order:

```
>>> 'Give {person} the {thing}!'.format(thing='bread', person='Maggie')
'Give Maggie the bread!'
```

Depending on your version of Python, you may be able to use *f-strings*:

```
>>> color = 'blue'
>>> f'I spy something {color}!'
'I spy something blue!'
```

Chapter 4: Howler

Write a Python program `howler.py` that will uppercase all the text from the command line or from a file. The program should also take a named option of `-o|--outfile` to write the output. The default output should be *standard out* (STDOUT).

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

Howler (upper-case 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:)

```
$ ./howler.py 'One word: Plastics!'
ONE WORD: PLASTICS!
$ ./howler.py ../inputs/fox.txt
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
$ ./howler.py -o out.txt ../inputs/fox.txt
$ cat out.txt
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```


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')
18
19     parser.add_argument('-o',
20                         '--outfile',
21                         help='Output filename',
22                         metavar='str',
23                         type=str,
24                         default='')
25
26     return parser.parse_args()
27
28
29 # -----
30 def main():
31     """Make a jazz noise here"""
32     args = get_args()
33     text = args.text
34     out_file = args.outfile
35
36     if os.path.isfile(text):
37         text = open(text).read().rstrip()
38
39     out_fh = open(out_file, 'wt') if out_file else sys.stdout
40     print(text.upper(), file=out_fh)
41     out_fh.close()
42
43
```

```
44 # -----
45 if __name__ == '__main__':
46     main()
```

Discussion

Cf. Truthiness, File Handles

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 that you should uppercase or it might be the name of a file. This pattern will come up repeatedly in this book, so commit these lines to memory:

```
if os.path.isfile(text):
    text = open(text).read().rstrip()
```

The first line looks on the file system to see 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 bioinformatics we regularly deal with files with sizes in the 10s to 100s of gigabytes!

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. The longer way to write the above would be:

```
if os.path.isfile(text):
    fh = open(text)
    text = fh.read()
    text = text.rstrip()
```

On line 39, we decide where to put the output of our program. The `if` expression will open `out_file` for writing text if `out_file` has been defined. The default value for `out_file` is the empty string which is effectively `False` when evaluated in a Boolean context. Unless the user provides a value, the output file handle `out_fh` will be `sys.stdout`.

To get uppercase, we can use the `text.upper` method. You can either `out_fh.write` this new text or use `print(..., file=...)`, noting which needs a newline and which does not. You can use `fh.close()` to close the file handle, but it's not entirely necessary as the program immediately ends after this. Still, it's good practice to close your file handles.

Chapter 5: Apples and Bananas

Perhaps you remember the children’s song “Apples and Bananas”?

```
I like to eat, eat, eat apples and bananas
I like to eat, eat, eat apples and bananas
```

```
I like to ate, ate, ate ay-ples and ba-nay-nays
I like to ate, ate, ate ay-ples and ba-nay-nays
```

```
I like to eat, eat, eat ee-ples and bee-nee-nees
I like to eat, eat, eat ee-ples and bee-nee-nees
```



Figure 3: Apple and bananas go together like peas and carrots.

Write a Python program called `apples.py` that will turn all the vowels in some given text in a single positional argument into just one `-v|--vowel` (default `a`) like this song. It should complain if the `--vowel` argument isn’t a single, lowercase vowel (hint, see `choices` in the `argparse` documentation). If the given text argument is a file, read the text from the file. Replace all vowels with the given vowel, both lower- and uppercase.

```
$ ./apples.py
usage: apples.py [-h] [-v str] str
apples.py: error: the following arguments are required: str
$ ./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 only vowel allowed (default: a)

\$./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',

\$./apples.py foo

faa

\$./apples.py ../inputs/fox.txt

Tha qaack brawn fax jumps avar tha lazy dag.

Solution

```
1  #!/usr/bin/env python3
2  """Apples and Bananas"""
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='Apples and bananas',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text', metavar='str', help='Input text or file')
18
19     parser.add_argument('-v',
20                         '--vowel',
21                         help='The vowel(s) allowed',
22                         metavar='str',
23                         type=str,
24                         default='a',
25                         choices=list('aeiou'))
26
27     return parser.parse_args()
28
29
30 # -----
31 def main():
32     """Make a jazz noise here"""
33
34     args = get_args()
35     text = args.text
36     vowel = args.vowel
37
38     if os.path.isfile(text):
39         text = open(text).read()
40
41     # Method 1: Iterate every character
42     # new_text = []
43     # for char in text:
```

```

44     #     if char in 'aeiou':
45     #         new_text.append(vowel)
46     #     elif char in 'AEIOU':
47     #         new_text.append(vowel.upper())
48     #     else:
49     #         new_text.append(char)
50     # text = ''.join(new_text)
51
52     # Method 2: str.replace
53     # for v in 'aeiou':
54     #     text = text.replace(v, vowel).replace(v.upper(), vowel.upper())
55
56     # Method 3: str.translate
57     # trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5)
58     # text = text.translate(trans)
59
60     # Method 4: Use a list comprehension
61     # new_text = [
62     #     vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c
63     #     for c in text
64     # ]
65     # text = ''.join(new_text)
66
67     # Method 5: Define a function, use list comprehension
68     def new_char(c):
69         return vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c
70
71     # text = ''.join([new_char(c) for c in text])
72
73     # Method 6: Use a `map` to iterate with a `lambda`
74     # text = ''.join(
75     #     map(
76     #         lambda c: vowel if c in 'aeiou' else vowel.upper()
77     #         if c in 'AEIOU' else c, text))
78
79     # Method 7: `map` with the function
80     text = ''.join(map(new_char, text))
81
82     # Method 8: Regular expressions
83     # text = re.sub('[aeiou]', vowel, text)
84     # text = re.sub('[AEIOU]', vowel.upper(), text)
85
86     print(text.rstrip())
87
88
89     # -----

```

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


Discussion

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. Once again, I defer to `argparse` by defining the `text` positional argument and the `-v|--vowel` option with a default value of `'a'`. I additionally use the `choices` option to restrict the values to the `list('aeiou')`. Remember that calling `list` on a string will expand it into a `list` of characters:

```
>>> list('aeiou')
['a', 'e', 'i', 'o', 'u']
```

The next problem is detecting if `text` is the name of a file that should be read for the text or is the text itself. I use `os.path.isfile` to ask the operating system if `text` names a file on disk. If this returns `True`, then I use `open(text).read()` to open the file and read the entire contents of the opened file handle into the `text` variable.

Method 1: Iterate every character

You can use a `for` loop on a string to access each character:

```
>>> text = 'Apples and Bananas!'
>>> vowel = 'o'
>>> new_text = []
>>> for char in text:
...     if char in 'aeiou':
...         new_text.append(vowel)
...     elif char in 'AEIOU':
...         new_text.append(vowel.upper())
...     else:
...         new_text.append(char)
...
>>> text = ''.join(new_text)
>>> text
'Opplos ond Bononos!'
```

So we get each `char` (character) in the `text` and ask if the character is in the string `'aeiou'` to determine if it is a vowel. If it is, we instead use the `vowel` determined by the user. Likewise with checking for membership in `'AEIOU'` to see if it's an uppercase vowel and using the `vowel.upper()`. If neither of those conditions is true, then we stick with the original character. Finally we overwrite `text` by joining the `new_text` on the empty string to make a new string with the vowels replaced.

Method 2: `str.replace`

The `str` class has a `replace` method that will return a new string with all instances of one string replaced by another. Note that the original string remains unchanged:

```
>>> s = 'foo'
>>> s.replace('o', 'a')
'faa'
>>> s.replace('oo', 'x')
'fx'
>>> s
'foo'
```

In this version:

```
>>> text = 'Apples and Bananas!'
>>> for v in 'aeiou':
...     text = text.replace(v, vowel).replace(v.upper(), vowel.upper())
...
>>> text
'Opplos ond Bononos!'
```

We use a `for` loop to iterate over each vowel in `'aeiou'` and then call `text.replace` to change that character to the indicated `vowel` from the user using both lower- and uppercase. If the character is not present, no action is taken.

Method 3: `str.translate`

There is a `str` method called `translate` that is very similar to `replace` that will “replace each character in the string using the given translation table.” To create the translation table, you should call the `str.maketrans` method. I pass it the string of lower- and upper-case vowels (5 of each) and a string that has position-by-position what should be substituted which I create by concatenating the lowercase `vowel` repeated 5 times with the uppercase `vowel` repeated 5 times.

```
>>> vowel * 5
'ooooo'
>>> vowel * 5 + vowel.upper() * 5
'ooooo00000'
>>> trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5)
```

The `trans` table is a `dict` where each character is represented by its ordinal value. You can go back and forth from characters and their ordinal values by using `chr` and `ord`:

```
>>> chr(97)
```

```
'a'
>>> ord('a')
97
```

If you look at the `trans` table:

```
>>> 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}
```

you can see it's mapping all the lowercase vowels to the ordinal value 111 which is 'o' and the uppercase vowels to 79 which is 'O':

```
>>> chr(111)
'o'
>>> chr(79)
'O'
```

And so I hope you can see how this works now. Recall that the original `text` remains unchanged by the `translate` method, so we overwrite `text` with the new version:

```
>>> text = 'Apples and Bananas!'
>>> trans = str.maketrans('aeiouAEIOU', vowel * 5 + vowel.upper() * 5)
>>> text = text.translate(trans)
>>> text
'Opplos ond Bononos!'
```

Method 4: List comprehension

You can stick a modified `for` loop inside brackets `[]` to create what is called a “list comprehension” to create new list from an existing sequence (list/dict/generator/stream) in one line of code. (You can also do likewise with `{}` for a new dict.) For example, here is how you could generate a list of squared numbers:

```
>>> [n ** 2 for n in range(4)]
[0, 1, 4, 9]
```

Additionally, inside the list comprehension we can use an *if expression*. Let's say you wanted **list** of **tuples** with a value and a string declaring if the value is "Even" or "Odd". The typical way to determine even/odd is looking at the remainder after dividing by 2 which we can do with the modulo (%) operator:

```
>>> 4 % 2
0
>>> 5 % 2
1
```

We can use Python's idea of "truthiness" to evaluate 0 as **False** and anything not 0 as **True**:

```
>>> 'Odd' if 4 % 2 else 'Even'
'Even'
>>> 'Odd' if 5 % 2 else 'Even'
'Odd'
```

Then use that inside a list comprehension:

```
>>> [(n, 'Odd' if n % 2 else 'Even') for n in range(4)]
[(0, 'Even'), (1, 'Odd'), (2, 'Even'), (3, 'Odd')]
```

We can chain *if* expressions to handle more than a binary decision. Perhaps you are programming an autonomous vehicle and want to decide how what to do at a traffic signal?

```
>>> color = 'red'
>>> 'STOP' if color == 'red' else 'Slow' if color == 'yellow' else 'Go'
'STOP'
>>> color = 'green'
>>> 'STOP' if color == 'red' else 'Slow' if color == 'yellow' else 'Go'
'Go'
```

In this version:

```
>>> text = 'Apples and Bananas!'
>>> new_text = [
...     vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c
...     for c in text
... ]
>>> text = ''.join(new_text)
>>> text
'Opplos ond Bononos!'
```

You have to find the start of the **for** loop **for c in text** which is "for character in text." We then use our handy compound *if expression* to decide whether to return the chosen **vowel** **if c in 'aeiou'** or the same check with the upper-case version, and finally we default to the character **c** itself if it fails both of those conditions.

Method 5: List comprehension with function

We could define a small function that will decide whether to return the `vowel` or the original character:

```
>>> vowel = 'o'
>>> def new_char(c):
...     return vowel if c in 'aeiou' else vowel.upper() if c in 'AEIOU' else c
...
>>> new_char('a')
'o'
>>> new_char('b')
'b'
```

And then use our list comprehension to call that. To me, this code is far more readable:

```
>>> 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. I could pass it as another argument, but the closure makes all this very compact and readable.

Method 6: map with a lambda

A `map` is essentially another way to write a list comprehension. Functions like `map` and another we'll use later called `filter` are in the class of “higher-order functions” because they take *other functions* as arguments, which is wicked cool. `map` applies another function to every member of a sequence. I like to think of `map` like a paint booth: You load up the booth with, say, blue paint, then unpainted cars go in, blue paint is applied, and blue cars come out.

I tend to think of this left-to-right:

```
car1, car2 -> paint_blue -> blue car1, blue car2
```

But the calling syntax moves right-to-left:

```
>>> paint_blue = lambda car: 'blue ' + car
>>> list(map(paint, ['car1', 'car2']))
['blue car1', 'blue car2']
```

Often you'll see the first argument to `map` starting with `lambda` to create an anonymous function using the `lambda` keyword. Think about regular named functions like `add1` that adds 1 to a value:

```
>>> def add1(n):
...     return n + 1
...
>>> add1(10)
11
>>> add1(11)
12
```

Here is the same idea using a `lambda`. Notice the function pretty much needs to fit on one line, can't really unpack complicated arguments, and doesn't need `return`:

```
>>> add1 = lambda n: n + 1
>>> add1(10)
11
>>> add1(11)
12
```

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 we just define the argument `n`. There is no difference in how you can use them. They are both functions:

```
>>> type(lambda x: x)
<class 'function'>
```

So I could define the `new_char` function using a `lambda` and it works just like the one created with `def new_char`:

```
>>> new_char = lambda c: vowel if c in 'aeiou' else \
...     vowel.upper() if c in 'AEIOU' else c
>>> new_char('a')
'o'
>>> new_char('b')
'b'
```

And here is how I can use it with `map`:

```
>>> text = 'Apples and Bananas!'
>>> text = ''.join(
...     map(
...         lambda c: vowel if c in 'aeiou' else vowel.upper()
...         if c in 'AEIOU' else c, text))
>>>
>>> text
'Opplos ond Bononos!'
```

Method 7: `map` with `new_char`

The previous version is not exactly easy to read, in my opinions, so instead of using `lambda` to make a function *inside* the `map`, I can use the `def new_char` version from above and `map` into that. In my opinion, this is the cleanest and most readable solution:

```
>>> text = 'Apples and Bananas!'
>>> text = ''.join(map(new_char, text))
>>> text
'Opplos ond Bononos!'
```

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'
```

What happens is that `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: Regular expressions

The last method I will introduce uses regular expressions which are a separate domain-specific language (DSL) you can use to describe patterns of text. They

are incredibly powerful and well worth the effort to learn them. To use them in your program, you `import re` and then use methods like `search` to find a pattern in a string or here `sub` to substitute a pattern for a new string. We'll be using brackets `[]` to create a “character class” meaning anything matching one of these characters. The second argument is the string that will replace the found strings, and the third argument is the string on which to work. Note that this string remains unchanged by the operation:

```
>>> import re
>>> text = 'Apples and Bananas!'
>>> vowel = 'o'
>>> re.sub('[aeiou]', vowel, text)
'Applos ond Bononos!'
>>> 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:

```
>>> text = 'Apples and Bananas!'
>>> text = re.sub('[AEIOU]', vowel.upper(), re.sub('[aeiou]', vowel, text))
>>> text
'Opplos ond Bononos!'
```

But I find that fairly hard to read.

Chapter 6: Telephone

Perhaps you remember the game of “Telephone” where a message is secretly passed through a series of intermediaries and then the result at the end of the chain is compared with how it started? This is like that, only we’re going to take some `text` (from the command line or a file) and mutate it by some percentage `-m|--mutations` (a number between 0 and 1, default 0.1 or 10%) and then print out the resulting text.

Each mutation to the text should be chosen using the `random` module, so your program will also need to accept a `-s|--seed` option (default `None`) to pass to the `random.seed` function for testing purposes. Print the resulting text after making the appropriate number of mutations.

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

Telephone

positional arguments:

str	Input text or file
-----	--------------------

optional arguments:

-h, --help	show this help message and exit
-s str, --seed str	Random seed (default: None)
-m float, --mutations float	Percent mutations (default: 0.1)

The program should not accept a bad `--mutations` argument:

```
$ ./telephone.py -m 10 foo
usage: telephone.py [-h] [-s str] [-m float] str
telephone.py: error: --mutations "10.0" must be b/w 0 and 1
```

It can be interesting to watch the accumulation of mutations:

```
$ ./telephone.py -s 1 ../inputs/fox.txt
Tho quick brown foa jumps oWer*the lazy dog.
$ ./telephone.py -s 1 -m .5 ../inputs/fox.txt
Thakqkrck&brow- fo[ jumps#oWe,*L/C lxdy dogos
```

Hints:

- To create a combined error/usage statement for the `--mutations` error, look at `parser.error` in `argparse`.

- To select a character position to change, I suggest using `random.choice` and a `range` from length of the incoming text. With that, you'll need to alter the character at that position, but you'll find that strings in Python are *immutable*. For instance, if I wanted to change “candle” into “handle”:

```
>>> s = 'candle'
>>> s[0] = 'h'
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: 'str' object does not support item assignment
```

- So, I need to create a *new string* that has `h` joined to the rest of the string `s` after the zeroth position. How could you do that?
- For the replacement value, you should use `random.choice` from the union of the `string` class's `ascii_letters` and `punctuation`:

```
>>> import string
>>> string.ascii_letters
'abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ'
>>> string.punctuation
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

Solution

```
1  #!/usr/bin/env python3
2  """Telephone"""
3
4  import argparse
5  import os
6  import random
7  import string
8  import sys
9
10
11  # -----
12  def get_args():
13      """Get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Telephone',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('input',
20                          metavar='str',
21                          help='Input text or file')
22
23      parser.add_argument('-s',
24                          '--seed',
25                          help='Random seed',
26                          metavar='str',
27                          type=str,
28                          default=None)
29
30      parser.add_argument('-m',
31                          '--mutations',
32                          help='Percent mutations',
33                          metavar='float',
34                          type=float,
35                          default=0.1)
36
37      args = parser.parse_args()
38
39      if not 0 < args.mutations <= 1:
40          msg = '--mutations "{}" must be b/w 0 and 1'.format(args.mutations)
41          parser.error(msg)
42
43      return args
```

```

44
45
46 # -----
47 def main():
48     """Make a jazz noise here"""
49
50     args = get_args()
51     text = args.input
52     random.seed(args.seed)
53
54     if os.path.isfile(text):
55         text = open(text).read()
56
57     len_text = len(text)
58     num_mutations = int(args.mutations * len_text)
59     alpha = string.ascii_letters + string.punctuation
60
61     for _ in range(num_mutations):
62         i = random.choice(range(len_text))
63         text = text[:i] + random.choice(alpha) + text[i+1:]
64
65     print(text.rstrip())
66
67 # -----
68 if __name__ == '__main__':
69     main()

```

Discussion



Figure 4: Telephones are for communication.

The number of mutations will be proportional to the length of the text

```
>>> text = 'The quick brown fox jumps over the lazy dog.'
>>> len_text = len(text)
>>> len_text
44
```

Since we chose the `--mutations` to be a `float` between 0 and 1, we can multiply that by the length to get the number of mutations to introduce. Since that number will likely be another `float` and we can introduce a partial number of mutations, we can use `int` to truncate the number to an integer value.

```
>>> mutations = .1
>>> int(mutations * len_text)
4
```

So we can use that number in a `for` loop with `range(4)` to modify four characters. To choose a character in the text to modify, I suggested to use `random.choice`:

```
>>> import random
>>> random.choice(range(len_text))
1
>>> random.choice(range(len_text))
22
```

If you assign that to a value like `i` (for “integer” and/or “index”, it’s pretty common to use `i` for this kind of value), then you could get the character at that position:

```
>>> i = random.choice(range(len_text))
```

```
>>> i
4
>>> text[i]
'q'
```

Now we saw earlier that we can't just change the `text`:

```
>>> text[i] = 'x'
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.

The bit before `i` is:

```
>>> text[:i]
'The '
```

And after `i` (skipping `i` itself, of course):

```
>>> text[i+1:]
'uick brown fox jumps over the lazy dog.'
```

There are many ways to join strings together into new strings, and the `+` operator is perhaps the simplest. So now we need some new character to insert in the middle which we can get with `random.choice` again, this time choosing from all the letters of the alphabet plus punctuation:

```
>>> import string
>>> alpha = string.ascii_letters + string.punctuation
>>> random.choice(alpha)
'n'
```

So to put it together, we overwrite the existing `text` so as to accumulate the changes over the iterations:

```
>>> text = text[:i] + random.choice(alpha) + text[i+1:]
>>> text
'The vuick brown fox jumps over the lazy dog.'
```

Mutations in DNA

For what it's worth, this is (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 them “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! We can revisit the output of this program later by using the Hamming distance to find how many changes we’d need to make to the output to regain the input.

Chapter 7: Bottles of Beer Song

Write a Python program called `bottles.py` that takes a single option `-n|--num` which is an positive integer (default 10) and prints the “bottles of beer on the wall song.” The program should also respond to `-h|--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)
```

If the `--num` argument is not an integer value, print an error message and stop the program:

```
$ ./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'
```

If the `-n` argument is less than 1, die with ‘`-num () must be > 0`’.

```
$ ./bottles.py -n -1
usage: bottles.py [-h] [-n INT]
bottles.py: error: --num (-1) must > 0
```

If the argument is good, then print the appropriate number of verses:

```
$ ./bottles.py -n 1
1 bottle of beer on the wall,
1 bottle of beer,
Take one down, pass it around,
0 bottles of beer on the wall!

$ ./bottles.py | head
10 bottles of beer on the wall,
10 bottles of beer,
Take one down, pass it around,
9 bottles of beer on the wall!

9 bottles of beer on the wall,
9 bottles of beer,
Take one down, pass it around,
```


8 bottles of beer on the wall!

Hints:

- Start with `new.py` and add a named *option* with `-n` for the “short” flag and `--num_bottles` for the “long” flag name. Be sure to choose `int` for the `type`. Note that the `metavar` is just for displaying to the user and has no effect on validation the arguments `type`.
- Look into `parser.error` for how to get `argparse` to printing an error message along with the usage and halt the program.
- Be sure to make the “bottle” into the proper singular or plural depending on the number in the phrase, e.g., “1 bottle” or “0 bottles.”
- Either run your program or do `make test` after *every single change to your program* to ensure that it compiles and is getting closer to passing the tests. Do not change three things and then run it. Make one change, then run or test it.
- If you use `make test`, it runs `pytest -xv test.py` where the `-x` flag tells `pytest` to stop after the first test failure. The tests are written in a order to help you complete the program. For instance, the first test just ensures that the program exists. The next one that you have some sort of handling of `--help` which would probably indicate that you’re using `argparse` and so have defined your arguments.
- Just try to pass each test in order. Focus on just one thing at a time. Create the program. Add the help. Handle bad arguments. Print just one verse. Print two verses. Etc.
- Read the next section on how to count down.

Counting down

You are going to need to count down, so you’ll need to consider how to do that. You can use `range` to get a list of integers from some a “start” (default 0, inclusive) to an “stop” (not inclusive). The `range` function is “lazy” in that it won’t actually generate the list until you ask for the numbers, so I could create a `range` generator for an absurdly large number like `range(10**1000)` and the REPL returns immediately. Try it! To force *see* the list of numbers, I can coerce it into a `list`:

```
>>> list(range(10))
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

OK, so maybe you were expecting the numbers 1-10? Welcome to “computer science” where we often starting counting at 0 and are quite often “off-by-one.” To count 1 to 10, I have to do this:

```
>>> list(range(1, 11))
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

Cool, cool, but we actually need to count *down*. You saw that this function works differently depending on whether you give it one argument (10) or two (1, 11). It also will do something different if you give it a third argument that represents the “step” of the numbers. So, to list every other number:

```
>>> list(range(1, 11, 2))  
[1, 3, 5, 7, 9]
```

And to count *down*, reverse the start and stop and use -1 for the step:

```
>>> list(range(11, 1, -1))  
[11, 10, 9, 8, 7, 6, 5, 4, 3, 2]
```

Wait, what? OK, the start number is inclusive and the stop is not. Try again:

```
>>> list(range(10, 0, -1))  
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
```

There’s a slightly easier way to get that list by using the **reversed** function:

```
>>> list(reversed(range(1, 11)))  
[10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
```

Solution

```
1  #!/usr/bin/env python3
2  """Bottle 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',
16                         '--num',
17                         metavar='INT',
18                         type=int,
19                         default=10,
20                         help='How many bottles')
21
22     args = parser.parse_args()
23
24     if args.num < 1:
25         parser.error('--num ({}) must > 0'.format(args.num))
26
27     return args
28
29
30 # -----
31 def main():
32     """Make a jazz noise here"""
33
34     args = get_args()
35     tmpl = '\n'.join([
36         '{} bottle{} of beer on the wall,',
37         '{} bottle{} of beer,',
38         'Take one down, pass it around,',
39         '{} bottle{} of beer on the wall!',
40     ])
41
42     for bottle in reversed(range(1, args.num + 1)):
43         next_bottle = bottle - 1
```

```
44         s1 = '' if bottle == 1 else 's'
45         s2 = '' if next_bottle == 1 else 's'
46         print(tmpl.format(bottle, s1, bottle, s1, next_bottle, s2))
47         if bottle > 1:
48             print()
49
50
51 # -----
52 if __name__ == '__main__':
53     main()
```

Discussion



Figure 5: “To alcohol! The cause of, and solution to, all of life’s problems.” - H. Simpson

If you used `new.py` and `argparse` to get started, then about 1/4 of the program is done for you. If you define an argument with the appropriate “short” (a dash plus one character) and “long” names (two dashes and a longer bit) with `type=int` and `default=10`, then `argparse` will do loads of hard work to ensure the user provides you with the correct input. We can’t easily tell `argparse` that the number has to be a *positive* integer without defining a new “type”, but it’s fairly painless to add a check and use `parser.error` to both print an error message plus the usage and halt the execution of the program.

Earlier programs have the last line of `get_args` as:

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

But here we capture the arguments inside `get_args` and add a bit of validation. If `args.num_bottles` is less than one, we call `parser.error` with the message we want to tell the user. We don’t have to tell the program to stop executing as `argparse` will exit immediately. Even better is that it will indicate a non-zero exit value to the operating system to indicate there was some sort of error. If you ever start writing command-line programs that chain together to make workflows, this is a way for one program to indicate failure and halt the entire process until the error has been fixed!

Once you get to the line `args = get_args()` in `main`, a great deal of hard work has already occurred to get and validate the input from the user. From here, I decided to create a template for the song putting `{}` in the spots that change from verse to verse. Then I use the `reversed(range(...))` bit we discussed before to count down, with a `for` loop, using the current number `bottle` and `next_bottle` to print out the verse noting the presence or absence of the `s` where appropriate.

Chapter 8: Gashlycrumb

Write a Python program called `gashlycrumb.py` that takes a letter of the alphabet as an argument and looks up the line in a `-f|--file` argument (default `gashlycrumb.txt`) and prints the line starting with that letter. It should generate usage with no arguments or for `-h|--help`:

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

Gashlycrumb

positional arguments:
str Letter

optional arguments:
-h, --help show this help message and exit
-f str, --file str Input file (default: gashlycrumb.txt)

You can see the structure of the default “gashlycrumb.txt” file:

```
$ head -3 gashlycrumb.txt
A is for Amy who fell down the stairs.
B is for Basil assaulted by bears.
C is for Clara who wasted away.
```



Figure 6: D is for Donald, who died from gas.

You will use the first character of the line as a lookup value:

```
$ ./gashlycrumb.py a
```

```
A is for Amy who fell down the stairs.
$ ./gashlycrumb.py z
Z is for Zillah who drank too much gin.
```

If given a value that does not exist in the list of first characters on the lines from the input file (when searched with regard to case), you should print a message:

```
$ ./gashlycrumb.py 3
I do not know "3".
$ ./gashlycrumb.py CH
I do not know "CH".
```

If provided a `--file` argument that does not exist, your program should exit with an error and message:

```
$ ./gashlycrumb.py -f sdf1 b
usage: gashlycrumb.py [-h] [-f str] str
gashlycrumb.py: error: argument -f/--file: can't open 'sdf1': \
[Errno 2] No such file or directory: 'sdf1'
```

Hints:

- To validate that the `--filename` is actually a readable file, look into using `argparse.FileType('r')` to describe the `type` of the `--file` argument so that `argparse` will do the check and create the error.
- 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.”
- Once you have an open file handle to the `--filename` (which is exactly what you get when use `argparse.FileType`), 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`?

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)
16
17     parser.add_argument('-f',
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
33
34      # lookup = {}
35      # for line in args.file:
36      #     lookup[line[0]] = line.rstrip()
37
38      lookup = {line[0]: line.rstrip() for line in args.file}
39
40      if letter.upper() in lookup:
41          print(lookup[letter.upper()])
42      else:
43          print('I do not know "{}".'.format(letter))
```



```
44
45
46 # -----
47 if __name__ == '__main__':
48     main()
```

Discussion

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`:

```
>>> fh = open('gashlycrumb.txt')
```

I can use a `for` loop to read each line of text and get the first letter using `line[0]` and set a dict called `lookup` with the value for the line:

```
>>> lookup = {}
>>> for line in fh:
...     lookup[line[0]] = line.rstrip()
...
>>> from pprint import pprint as pp
>>> pp(lookup)
{'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.',
 'D': 'D is for Desmond thrown out of a sleigh.',
 'E': 'E is for Ernest who choked on a peach.',
 'F': 'F is for Fanny sucked dry by a leech.',
 'G': 'G is for George smothered under a rug.',
 'H': 'H is for Hector done in by a thug.',
 'I': 'I is for Ida who drowned in a lake.',
 'J': 'J is for James who took lye by mistake.',
 'K': 'K is for Kate who was struck with an axe.',
 'L': 'L is for Leo who choked on some tacks.',
 'M': 'M is for Maud who was swept out to sea.',
 'N': 'N is for Neville who died of ennui.',
 'O': 'O is for Olive run through with an awl.',
 'P': 'P is for Prue trampled flat in a brawl.',
 'Q': 'Q is for Quentin who sank on a mire.',
 'R': 'R is for Rhoda consumed by a fire.',
 'S': 'S is for Susan who perished of fits.',
 'T': 'T is for Titus who flew into bits.',
 'U': 'U is for Una who slipped down a drain.',
 'V': 'V is for Victor squashed under a train.'}
```

```
'W': 'W is for Winnie embedded in ice.',
'X': 'X is for Xerxes devoured by mice.',
'Y': 'Y is for Yorick whose head was bashed in.',
'Z': 'Z is for Zillah who drank too much gin.'}
```

We've seen list comprehensions by essentially sticking a `for` inside brackets `[]`, and we can use a dictionary comprehension by doing the same with a `for` loop 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. I need to `open` it again for this next bit:

```
>>> fh = open('gashlycrumb.txt')
>>> lookup = {line[0]: 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).

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 just 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".
```

I don't have to use a dict. I could, for example, use a list of tuple values:

```
>>> fh = open('gashlycrumb.txt')
>>> lookup = [(line[0], line.rstrip()) for line in fh]
>>> pp(lookup[:2])
[('A', 'A is for Amy who fell down the stairs.'),
 ('B', 'B is for Basil assaulted by bears.')]
>>>
```

I can get the letters with a list comprehension:

```
>>> [char for char, line in lookup][:3]
['A', 'B', 'C']
```

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. It's well worth your time to learn dictionaries very well!

Edward Gorey

If you are not familiar with the work of Edward Gorey, please go read about him immediately, e.g. <https://www.brainpickings.org/2011/01/19/edward-gorey-the-gashlycrumb-tinies/>!

Alternate text

Write your own version of Gorey's text and pass in your version as the `--file`. I include my own `alternate.txt` which I used the simple and Soundex rhymers to help me find words.

Interactive version

Write an interactive version that takes input directly from the user.

```
$ ./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
```

Hint: Use `while True` to set up an infinite loop and keep using `input` to get the user's next `letter`.

Chapter 9: Movie Reader



Figure 7: Matt Damon in The Martian (no, really).

Write a Python program called `movie_reader.py` that takes a single positional argument that is a bit of text or the name of an input file. The output will be dynamic, so I cannot write a test for how the program should behave, nor can I include a bit of text that shows you how it should work. Your program should print the input text character-by-character and then pause .5 seconds for ending punctuation like `.`, `!` or `?`, .2 seconds for a pause like `,`, `:`, or `;`, and .05 seconds for anything else.

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

Movie Reader

positional arguments:
str Input text or file

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

```
$ ./movie_reader.py 'Foo, bar!'
Foo, bar!
$ ./movie_reader.py ../inputs/fox.txt
The quick brown fox jumps over the lazy dog.
```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import os
5  import sys
6  import time
7
8
9  # -----
10 def get_args():
11     """Get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='Movie Reader',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text', metavar='str', help='Input text or file')
18
19     return parser.parse_args()
20
21
22 # -----
23 def main():
24     """Make a jazz noise here"""
25
26     args = get_args()
27     text = args.text
28
29     if os.path.isfile(text):
30         text = open(text).read()
31
32     for line in text.splitlines():
33         for char in line:
34             print(char, end='')
35             time.sleep(.5 if char in '!.?\n' else .2 if char in ',,:;' else .05)
36             sys.stdout.flush()
37
38     print()
39
40
41 # -----
42 if __name__ == '__main__':
43     main()
```

Chapter 10: Palindromes

Write a Python program called `palindromic.py` that will find words that are palindromes in positional argument which is either a string or a file name.

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

Find palindromes in text

positional arguments:
str Input text or file

optional arguments:
-h, --help show this help message and exit
-m int, --min int Minimum word length (default: 3)

```
$ ./palindromic.py '"Wow!" said Mom.'
```

wow

mom

```
$ ./palindromic.py input.txt
```

anna

civic

kayak

madam

mom

wow

level

noon

racecar

radar

redder

refer

rotator

rotor

solos

stats

tenet

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import os
5  import re
6
7
8  # -----
9  def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Find palindromes in text',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('text', metavar='str', help='Input text or file')
17
18     parser.add_argument('-m',
19                         '--min',
20                         metavar='int',
21                         type=int,
22                         help='Minimum word length',
23                         default=3)
24
25     return parser.parse_args()
26
27
28 # -----
29 def main():
30     """Make a jazz noise here"""
31
32     args = get_args()
33     text = args.text
34     min_length = args.min
35
36     if os.path.isfile(text):
37         text = open(text).read()
38
39     for line in text.splitlines():
40         for word in re.split(r'(\W+)', line.lower()):
41             if len(word) >= min_length:
42                 rev = ''.join(reversed(word))
43                 if rev == word:
```



```
44             print(word)
45
46
47 # -----
48 if __name__ == '__main__':
49     main()
```

Chapter 11: Ransom

Create a Python program called `ransom.py` that will randomly capitalize the letters in a text. The program should take a `-s|--seed` argument for the `random.seed` to control randomness for the test suite. It should print usage when given no arguments or `-h|--help`.

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

The text can be given on the command line:

```
$ ./ransom.py -s 2 'The quick brown fox jumps over the lazy dog.'
the qUIck BROWN fOX JUmps ovEr ThE LAZY DOg.
```

Or in a file:

```
$ cat ../inputs/fox.txt
The quick brown fox jumps over the lazy dog.
$ ./ransom.py --seed 2 ../inputs/fox.txt
the qUIck BROWN fOX JUmps ovEr ThE LAZY DOg.
```

Hints:

- You can iterate each character in the input string with a `for` loop
- For each character, can use the `random.choice` function to decide whether to force the character to upper or lower case using methods from the `str` class

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')
18
19     parser.add_argument('-s',
20                         '--seed',
21                         help='Random seed',
22                         metavar='int',
23                         type=int,
24                         default=None)
25
26     args = parser.parse_args()
27
28     if os.path.isfile(args.text):
29         args.text = open(args.text).read().rstrip()
30
31     return args
32
33 # -----
34 def choose(c):
35     """Randomly choose an upper or lowercase letter to return"""
36
37     return c.upper() if random.choice([0, 1]) else c.lower()
38
39 # -----
40 def main():
41     """Make a jazz noise here"""
42     args = get_args()
43     text = args.text
```

```

44     random.seed(args.seed)
45
46     # Method 1: Iterate each character, add to list
47     # ransom = []
48     # for char in text:
49     #     ransom.append(char.upper() if random.choice([0, 1]) else char.lower())
50
51     # Method 2: List comprehension
52     #ransom = [c.upper() if random.choice([0, 1]) else c.lower() for c in text]
53
54     # Method 3: List comprehension with function
55     #ransom = [choose(c) for c in text]
56
57     # Method 4: map with lambda
58     # ransom = map(lambda c: c.upper() if random.choice([0, 1]) else c.lower(),
59     #               text)
60
61     # Method 5: map with function
62     ransom = map(choose, text)
63
64     print(''.join(ransom))
65
66
67     # -----
68     if __name__ == '__main__':
69         main()

```

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?

It’s a common pattern in many of these problems that the input can either be given on the command line or in a file, so I have to defined the `text` argument as having `type=str`. In this version of the program, I decided to check in the `get_args` if the `text` is a file (`os.path.isfile(text)`), and, if so, to override the value of `args.text` with the result of reading the contents of the file. That way when I get to the `args = get_args()` line in my program, I’ve already gotten the text from the user, whether given on the command line or in a file.

I set the `--seed` optional `default` to Python’s special `None` value which means nothing at all. As such, I can pass it directly to `random.seed` because setting the seed to `None` is the same as not setting it. Only if the user indicates a `--seed` value (which must be an `int` and which `argparse` will validate) will this affect the behavior of the program.

Assume that we have the following:

```
>>> text = 'The quick brown fox jumps over the lazy dog.'
```

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. Here’s one way to print an uppercase version of the `text`

```
>>> for char in text:
...     print(char.upper(), end='')
...
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```

Let’s use `random.choice` to make a binary selection:

```
>>> import random
>>> random.choice([True, False])
False
>>> random.choice([0, 1])
0
>>> random.choice(['blue', 'green'])
'blue'
```

Now use that to select whether to take the upper- or lowercase character. Note that this version relies on the idea of “truthiness” (cf appendix) where 0 is considered `False` and anything not zero (like 1) is `True`. So if `random.choice([0, 1])` returns a 1 (or `True`) then we take `char.upper()` otherwise we take `char.lower()`:

```
>>> ransom = []
```

```
>>> for char in text:
...     ransom.append(char.upper() if random.choice([0, 1]) else char.lower())
...
>>> ''.join(ransom)
'The quIck brOwn Fox JumpS over ThE lAZy dOG.'
```

We can shorten this to one line of code if we use a list comprehension, essentially putting the `for` loop inside the brackets `[]` that create the `ransom` list:

```
>>> ransom = [c.upper() if random.choice([0, 1]) else c.lower() for c in text]
>>> ''.join(ransom)
'thE quIck bRowN foX JuMpS oVeR tHe lAzY dog.'
```

All the code for deciding which case could go into a very small function which you could either write as a `lambda`:

```
>>> choose = lambda c: c.upper() if random.choice([0, 1]) else c.lower()
>>> choose('t')
'T'
```

Or the more standard `def` version:

```
>>> def choose(c):
...     return c.upper() if random.choice([0, 1]) else c.lower()
...
>>> choose('t')
't'
```

And then use that in your list comprehension. This version reads very well as is perhaps my favorite:

```
>>> ransom = [choose(c) for c in text]
>>> ''.join(ransom)
'thE qUicK broWN fOx JuMpS OVeR the lAZy doG.'
```

But I also quite like the `map` function which takes another function as the first argument which is applied to all the elements of second argument which is an iterable:

```
>>> ransom = map(lambda c: c.upper() if random.choice([0, 1]) else c.lower(), text)
>>> ''.join(ransom)
'ThE qUiCk BRoWn FoX JuMps oVeR ThE lAzY dog.'
```

And that cleans up very nicely if instead we used our named function. This version is the shortest and perhaps cleanest but does require the reader to understand `map`:

```
>>> ransom = map(choose, text)
>>> ''.join(ransom)
'thE qUicK BrOwN FOX jumPs oVeR thE lAZY dOg.'
```

It may seem silly to spend so much time working through five 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, c-like solution while the list comprehensions are very Pythonic and the `map` versions borrow from the world of purely functional languages like Haskell.

Chapter 12: Simple Rhymer

Write a Python program called `rhymer.py` that will create new words by removing the consonant(s) from the beginning (if any) of a given word and then create new words by prefixing the remainder with all the consonants and clusters that were not at the beginning. That is, prefix with all the consonants in the alphabet plus these 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 wh wr sch scr shr sph spl spr squ str thr
```

If given no arguments or the `-h|--help` flags, print a usage statement:

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

Make rhyming "words"

```
positional arguments:
  str                A word
```

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

If the word starts with a vowel, use the word as-is:

```
$ ./rhymer.py apple | head -3
bapple
capple
dapple
```

If the word begins with any consonants, remove them and append all the prefixes above making sure not to include any prefixes that match what you removed:

```
$ ./rhymer.py take | head -3
bake
cake
dake
$ ./rhymer.py take | grep take
stake
```

If the word doesn't match one of the above conditions, e.g., it is entirely consonants, print a message that you cannot rhyme it.

```
$ ./rhymer.py RDNZL
Cannot rhyme "RDNZL"
```


Hints:

The heart of the program for me is the stemming of the word. Do you even need to stemp it? Not if it begins with a vowel, so how can you detect that? I ended up writing a function called `stemmer` and inserted this into my `rhymmer.py`:

```
def test_stemmer():
    """Test the stemmer"""

    assert ('c', 'ake') == stemmer('cake')
    assert ('ch', 'air') == stemmer('chair')
    assert ('', 'apple') == stemmer('apple')
    assert stemmer('bbb') is None
```

If you notice the `make test` target also include `rhymmer.py`:

```
pytest -xv rhymmer.py test.py
```

I wrote my `stemmer(word)` to return a tuple of (`prefix`, `stem`) where `prefix` will be the empty string when the `word` starts with a vowel. If the word starts with a consonant and can be split, I return the two parts of the word e.g., `chair` become (`'ch'`, `'air'`). Otherwise I return `None` to indicate a failure to communicate.

If you choose to do the same, you can add the `test_stemmer` to your program and `pytest` will find any function with a name starting with `test_` to run. You can use this to verify that your `stemmer` does what you expect.

Solution

```
1  #!/usr/bin/env python3
2  """Make rhyming words"""
3
4  import argparse
5  import re
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 stemmer(word):
24     """Return leading consonants (if any), and 'stem' of word"""
25
26
27 def stemmer(word):
28     vowels = 'aeiou'
29     consonants = ''.join(
30         filter(lambda c: c not in vowels, string.ascii_lowercase))
31     match = re.match('^([' + consonants + ']*)([' + vowels + '].*)', word)
32     if match:
33         return match.groups()
34     return None
35
36
37 # -----
38 def test_stemmer():
39     """Test the stemmer"""
40
41     assert ('c', 'ake') == stemmer('cake')
42     assert ('ch', 'air') == stemmer('chair')
43     assert ('', 'apple') == stemmer('apple')
```

```

44     assert stemmer('bbb') is None
45
46
47 # -----
48 def main():
49     """Make a jazz noise here"""
50     args = get_args()
51     word = args.word
52     stemmed = stemmer(word.lower())
53     prefixes = list('bcd fghjklmnpqrstvwxyz') + (
54         'bl br ch cl cr dr fl fr gl gr pl pr sc '
55         'sh sk sl sm sn sp st sw th tr tw wh wr'
56         'sch scr shr sph spl spr squ str thr').split()
57
58     if stemmed:
59         start, rest = stemmed
60         print('\n'.join([p + rest for p in prefixes if p != start]))
61     else:
62         print('Cannot rhyme "{}".format(word))
63
64
65 # -----
66 if __name__ == '__main__':
67     main()

```

Discussion

As stated in the description, I spent most of my time working out how to stem a word. Some other programs in the book require this idea (Soundex rhymers, Runny Babbit), so you might look there, too. I decided to write a function `stemmer(word)` that will return a tuple of (`prefix`, `stem`).

We need to check if the word can be split into one or more consonants followed by at least one vowel and maybe some other stuff, e.g., `'ha'` could be (`'h'`, `'a'`). The easiest way is to write a regular expression using the `re` module. We've already defined the `vowels`, so we can use those to find the complement of `consonants`. I can iterate through the letters of the alphabet by using `string.ascii_lowercase` and find those not in the `vowels`:

```
>>> import string
>>> string.ascii_lowercase
'abcdefghijklmnopqrstuvwxyz'
>>> vowels = 'aeiou'
>>> consonants = ''.join(
...     filter(lambda c: c not in vowels, string.ascii_lowercase))
>>> consonants
'bcdfghjklmnpqrstvwxyz'
```

Here we see the use of `filter` which is a “higher-order” function takes a *another function* as the first argument and an iterable as the second argument. The `lambda c` keyword creates an anonymous function with a single argument I call `c` (for “character”) which can then be referenced in the function body.

The more Pythonic way to write this would be a list comprehension:

```
>>> consonants = ''.join([c for c in string.ascii_lowercase if c not in vowels])
>>> consonants
'bcdfghjklmnpqrstvwxyz'
```

Both ways are fine. It's mostly preference, though true Pythonistas would probably disagree. If nothing else, the `filter` might be slower than a comprehension, especially if the iterable were large, so choose whichever way makes more sense for your style and application.

The regular expression is a bit tricky. We want to find consonants at the beginning, so we can use the caret (`^`) to anchor the regex to the start of the string.

```
>>> r = '^'
>>> r
'^'
```

Then we create a “character class” using `[]` and `enumerate` inside all the characters that are allowed:

```
>>> r = '^[' + consonants + ']'
>>> r
'^[bcdfghjklmnpqrstvwxyz]'
```

We will want to “capture” these so we can extract them later, so we put parentheses () around the character class to group them:

```
>>> r = '^([' + consonants + '])'
>>> r
'^([bcdfghjklmnpqrstvwxyz])'
```

Let’s try that and see what we get:

```
>>> import re
>>> re.search(r, 'chair')
<re.Match object; span=(0, 1), match='c'>
```

Hmm, it didn’t match *ch* because we didn’t tell the regex *how many* to match, so it just matched one. We can add * to indicate “zero or more”:

```
>>> r = '^([' + consonants + ']*)'
>>> r
'^([bcdfghjklmnpqrstvwxyz]*)'
>>> re.search(r, 'chair')
<re.Match object; span=(0, 2), match='ch'>
```

Very nice. Sometimes you’ll see + to mean that a pattern can be repeated, but that one means “one or more.” By using *, I’m relying on the fact that “zero” matches will always be true, so this will also help me find any **word** that begins with a vowel (although it doesn’t seem like it just yet):

```
>>> re.search(r, 'apple')
<re.Match object; span=(0, 0), match=''>
```

Now I want to say that after some optional consonant prefix there must be at least one vowel:

```
>>> r = '^([' + consonants + ']*)' + '^([' + vowels + '])'
>>> r
'^([bcdfghjklmnpqrstvwxyz]*)([aeiou])'
>>> re.search(r, 'chair')
<re.Match object; span=(0, 3), match='cha'>
>>> re.search(r, 'apple')
<re.Match object; span=(0, 1), match='a'>
```

Getting closer, but we need the regular expression to reach the end of the word now, so we add .* where . means “one of anything” and * means “zero or more”:

```
>>> r = '^([' + consonants + ']*)' + '^([' + vowels + '].*)'
>>> r
'^([bcdfghjklmnpqrstvwxyz]*)([aeiou].*)'
```

```
>>> re.search(r, 'chair')
<re.Match object; span=(0, 5), match='chair'>
>>> re.search(r, 'apple')
<re.Match object; span=(0, 5), match='apple'>
```

Great! We're matching the entire word. The true magic comes in when we look at the capture groups:

```
>>> re.search(r, 'chair').groups()
('ch', 'air')
>>> re.search(r, 'apple').groups()
('', 'apple')
```

That is exactly what I wanted to return! For what it's worth, I can get each group individually by referencing their order:

```
>>> re.search(r, 'chair').group(1)
'ch'
>>> re.search(r, 'apple').group(2)
'apple'
```

If I can't match a string:

```
>>> type(re.search(r, 'RDNZL'))
<class 'NoneType'>
```

I return None from my function:

```
>>> def stemmer(word):
...     vowels = 'aeiou'
...     consonants = ''.join(
...         filter(lambda c: c not in vowels, string.ascii_lowercase))
...     match = re.match('^([' + consonants + ']*)([' + vowels + ']*.*)', word)
...     if match:
...         return match.groups()
...     return None
...
>>> stemmer('apple')
('', 'apple')
>>> stemmer('chair')
('ch', 'air')
>>> stemmer('RDNZL')
```

So, given a working `stemmer` I try to stem a given word. If there is no result, I print the message that I cannot rhyme the word. Otherwise I iterate over all the prefixes:

```
>>> prefixes = list('bcd fghjklmnpqrstvwxyz') + (
...     '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()
```

And add them to the stem of the word, being sure to avoid any prefix that was the same as the original word:

```
>>> start, rest = stemmer('chair')
>>> start
'ch'
>>> rest
'air'
>>> [p + rest for p in prefixes if p != start][:3]
['bair', 'cair', 'dair']
```

Chapter 13: Rock, Paper, Scissors

Write a Python program called `rps.py` that will play the ever-popular “Rock, Paper, Scissors” game. As often as possible, insult the player by combining an adjective and a noun from the following lists:

Adjectives = truculent fatuous vainglorious fatuous petulant moribund jejune feckless antiquated rambunctious mundane misshapen glib dreary dopey devoid deleterious degrading clammy brazen indiscreet indecorous imbecilic dysfunctional dubious drunken disreputable dismal dim deficient deceitful damned daft contrary churlish catty banal asinine infantile lurid morbid repugnant unkempt vapid decrepit malevolent impertinent decrepit grotesque puerile

Nouns = abydocomist bedswerver bespawler bobolyne cumberworld dalcop dew-beater dorbel drate-poke driggle-draggle fopdoodle fustylugs fustilarian gillie-wet-foot gnashgab gobermouch gowpenful-o'-anything klazomaniac leasing-monger loiter-sack lubberwort muck-spout mumblecrust quisby rag-gabrash rakefire roiderbanks saddle-goose scobberlotcher skelpie-limmer smell-feast smellfungus snoutband sornier stampcrab stymphalist tallowcatch triptaker wandought whiffle-whaffle yaldson zoilist

The program should accept a `-s|--seed` to pass to `random`.

```
$ ./rps.py
1-2-3-Go! [rps|q] r
You: Rock
Me : Scissors
You win. You are a clammy drate-poke.
1-2-3-Go! [rps|q] t
You dysfunctional dew-beater! Please choose from: p, r, s.
1-2-3-Go! [rps|q] p
You: Paper
Me : Rock
You win. You are a dismal gillie-wet-foot.
1-2-3-Go! [rps|q] q
Bye, you imbecilic fopdoodle!
```


Solution

```
1  #!/usr/bin/env python3
2  """Rock, Paper, Scissors"""
3
4  import argparse
5  import os
6  import random
7  import sys
8
9
10 # -----
11 def get_args():
12     """Get command-line arguments"""
13
14     parser = argparse.ArgumentParser(
15         description='Rock, Paper, Scissors',
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18     parser.add_argument('-s',
19                         '--seed',
20                         help='Random seed',
21                         metavar='int',
22                         type=int,
23                         default=None)
24
25     return parser.parse_args()
26
27
28 # -----
29 def insult():
30     adjective = """
31     truculent fatuous vainglorious fatuous petulant moribund jejune
32     feckless antiquated rambunctious mundane misshapen glib dreary
33     dopey devoid deleterious degrading clammy brazen indiscreet
34     indecorous imbecilic dysfunctional dubious drunken disreputable
35     dismal dim deficient deceitful damned daft contrary churlish
36     catty banal asinine infantile lurid morbid repugnant unkempt
37     vapid decrepit malevolent impertinent decrepit grotesque puerile
38     """.split()
39
40     noun = """
41     abydocomist bedswerver bespawler bobolyne cumberworld dalcop
42     dew-beater dorbel drate-poke driggle-draggle fopdoodle fustylugs
43     fustilarian gillie-wet-foot gnashgab gobermouch
```

```

44     gowpenful-o'-anything klazomaniac leasing-monger loiter-sack
45     lubberwort muck-spout mumblecrust quisby raggabrash rakefire
46     roiderbanks saddle-goose scobberlotcher skelpie-limmer
47     smell-feast smellfungus snoutband sorner stampcrab stymphalist
48     tallowcatch triptaker wandought whiffle-whaffle yaldson zoilist
49     """.split()
50
51     return ' '.join([random.choice(adjective), random.choice(noun)])
52
53
54 # -----
55 def main():
56     """Make a jazz noise here"""
57
58     args = get_args()
59     random.seed(args.seed)
60
61     valid = set('rps')
62     beats = {'r': 's', 's': 'p', 'p': 'r'}
63     display = {'r': 'Rock', 'p': 'Paper', 's': 'Scissors'}
64
65     while True:
66         play = input('1-2-3-Go! [rps|q] ').lower()
67
68         if play.startswith('q'):
69             print('Bye, you {}'.format(insult()))
70             sys.exit(0)
71
72         if play not in valid:
73             print('You {}! Please choose from: {}'.format(
74                 insult(), ', '.join(sorted(valid))))
75             continue
76
77         computer = random.choice(list(valid))
78
79         print('You: {}\nMe : {}'.format(display[play], display[computer]))
80
81         if beats[play] == computer:
82             print('You win. You are a {}'.format(insult()))
83         elif beats[computer] == play:
84             print('You lose, {}'.format(insult()))
85         else:
86             print('Draw, you {}'.format(insult()))
87
88
89 # -----

```

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

Chapter 14: Abuse

Write a Python program called `abuse.py` that generates some `-n|--number` of insults (default 3) by randomly combining some number of `-a|--adjectives` (default 2) with a noun (see below). Be sure your program accepts a `-s|--seed` argument (default `None`) to pass to `random.seed`.

Adjectives:

bankrupt base caterwauling corrupt cullionly detestable dishonest false filth-
some filthy foolish foul gross heedless indistinguishable infected insatiate irk-
some lascivious lecherous loathsome lubbery old peevish rascaly rotten ruinous
scurilous scurvy slanderous sodden-witted thin-faced toad-spotted unmannered
vile wall-eyed

Nouns:

Judas Satan ape ass barbermonger 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

```
$ ./abuse.py -h
usage: abuse.py [-h] [-a int] [-n int] [-s int]
```

Argparse Python script

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)
```

```
$ ./abuse.py
You slanderous, rotten block!
You lubbery, scurilous ratcatcher!
You rotten, foul liar!
$ ./abuse.py -s 1 -n 2 -a 1
You rotten rogue!
You lascivious ape!
$ ./abuse.py -s 2 -n 4 -a 4
You scurilous, foolish, vile, foul milksop!
You cullionly, lubbery, heedless, filthy lunatic!
You foul, lecherous, infected, slanderous degenerate!
You base, ruinous, slanderous, false liar!
```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import random
5  import sys
6
7  adjectives = """
8  bankrupt base caterwauling corrupt cullionly detestable dishonest
9  false filthsome filthy foolish foul gross heedless indistinguishable
10 infected insatiate irksome lascivious lecherous loathsome lubberly old
11 peevish rascally rotten ruinous scurilous scurvy slanderous
12 sodden-witted thin-faced toad-spotted unmannered vile wall-eyed
13 """.strip().split()
14
15 nouns = """
16 Judas Satan ape ass barbermonger beggar block boy braggart butt
17 carbuncle coward coxcomb cur dandy degenerate fiend fishmonger fool
18 gull harpy jack jolthead knave liar lunatic maw milksop minion
19 ratcatcher recreant rogue scold slave swine traitor varlet villain worm
20 """.strip().split()
21
22
23 # -----
24 def get_args():
25     """get command-line arguments"""
26     parser = argparse.ArgumentParser(
27         description='Argparse Python script',
28         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
29
30     parser.add_argument('-a',
31                         '--adjectives',
32                         help='Number of adjectives',
33                         metavar='int',
34                         type=int,
35                         default=2)
36
37     parser.add_argument('-n',
38                         '--number',
39                         help='Number of insults',
40                         metavar='int',
41                         type=int,
42                         default=3)
43
```

```

44     parser.add_argument('-s',
45                          '--seed',
46                          help='Random seed',
47                          metavar='int',
48                          type=int,
49                          default=None)
50
51     return parser.parse_args()
52
53
54 # -----
55 def main():
56     """Make a jazz noise here"""
57     args = get_args()
58     num_adj = args.adjectives
59     num_insults = args.number
60
61     random.seed(args.seed)
62
63     for _ in range(num_insults):
64         adjs = random.sample(adjectives, k=num_adj)
65         noun = random.choice(nouns)
66         print('You {} {}!'.format(', '.join(adjs), noun))
67
68
69 # -----
70 if __name__ == '__main__':
71     main()

```

Chapter 15: Bacronym

Write a Python program called `bacronym.py` that takes a string like “FBI” and retrofits some `-n|--number` (default 5) of acronyms by reading a `-w|--wordlist` argument (default `/usr/share/dict/words`), skipping over words to `-e|--exclude` (default `a, an, the`) and randomly selecting words that start with each of the letters. Be sure to include a `-s|--seed` argument (default `None`) to pass to `random.seed` for the test suite.

```
$ ./bacronym.py
usage: bacronym.py [-h] [-n NUM] [-w STR] [-x STR] [-s INT] STR
bacronym.py: error: the following arguments are required: STR
$ ./bacronym.py -h
usage: bacronym.py [-h] [-n NUM] [-w STR] [-x STR] [-s INT] STR
```

Explain acronyms

positional arguments:

STR	Acronym
-----	---------

optional arguments:

<code>-h, --help</code>	show this help message and exit
<code>-n NUM, --num NUM</code>	Maximum number of definitions (default: 5)
<code>-w STR, --wordlist STR</code>	Dictionary/word file (default: <code>/usr/share/dict/words</code>)
<code>-x STR, --exclude STR</code>	List of words to exclude (default: <code>a,an,the</code>)
<code>-s INT, --seed INT</code>	Random seed (default: <code>None</code>)

```
$ ./bacronym.py FBI -s 1
```

FBI =

- Fecundity Brokage Imitant
- Figureless Basketmaking Ismailite
- Frumpery Bonedog Irregardless
- Foxily Blastomyces Inedited
- Fastland Bouncingly Idiospasm

Solution

```
1  #!/usr/bin/env python3
2  """Make guesses about acronyms"""
3
4  import argparse
5  import sys
6  import os
7  import random
8  import re
9  from collections import defaultdict
10
11
12  # -----
13  def get_args():
14      """get arguments"""
15      parser = argparse.ArgumentParser(
16          description='Explain acronyms',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('acronym', help='Acronym', type=str, metavar='STR')
20
21      parser.add_argument('-n',
22                          '--num',
23                          help='Maximum number of definitions',
24                          type=int,
25                          metavar='NUM',
26                          default=5)
27
28      parser.add_argument('-w',
29                          '--wordlist',
30                          help='Dictionary/word file',
31                          type=str,
32                          metavar='STR',
33                          default='/usr/share/dict/words')
34
35      parser.add_argument('-x',
36                          '--exclude',
37                          help='List of words to exclude',
38                          type=str,
39                          metavar='STR',
40                          default='a,an,the')
41
42      parser.add_argument('-s',
43                          '--seed',
```



```

44             help='Random seed',
45             type=int,
46             metavar='INT',
47             default=None)
48
49     return parser.parse_args()
50
51
52 # -----
53 def main():
54     """main"""
55
56     args = get_args()
57     acronym = args.acronym
58     wordlist = args.wordlist
59     limit = args.num
60     goodword = r'^[a-z]{2,}$'
61     badwords = set(re.split(r'\s*,\s*', args.exclude.lower()))
62
63     random.seed(args.seed)
64
65     if not re.match(goodword, acronym.lower()):
66         print('"{}" must be >1 in length, only use letters'.format(acronym))
67         sys.exit(1)
68
69     if not os.path.isfile(wordlist):
70         print('"{}" is not a file.'.format(wordlist))
71         sys.exit(1)
72
73     seen = set()
74     words_by_letter = defaultdict(list)
75     for word in open(wordlist).read().lower().split():
76         clean = re.sub('[^a-z]', '', word)
77         if not clean: # nothing left?
78             continue
79
80         if re.match(goodword,
81                    clean) and clean not in seen and clean not in badwords:
82             seen.add(clean)
83             words_by_letter[clean[0]].append(clean)
84
85     len_acronym = len(acronym)
86     definitions = []
87     for i in range(0, limit):
88         definition = []
89         for letter in acronym.lower():

```

```

90         possible = words_by_letter.get(letter, [])
91         if len(possible) > 0:
92             definition.append(
93                 random.choice(possible).title() if possible else '?')
94
95         if len(definition) == len_acronym:
96             definitions.append(' '.join(definition))
97
98     if len(definitions) > 0:
99         print(acronym.upper() + ' =')
100         for definition in definitions:
101             print(' - ' + definition)
102     else:
103         print('Sorry I could not find any good definitions')
104
105
106 # -----
107 if __name__ == '__main__':
108     main()

```

Chapter 16: Workout Of (the) Day (WOD)

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. You should use the `tabulate` module to format the output as expected.

The input file should be comma-separated values with headers for “exercise” and “reps,” e.g.:

```
$ 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.

```
$ ./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)
$ ./wod.py
Exercise      Reps
```

```

-----
Crunches      26
HSPU          9
Squats       43
Pushups      36
$ ./wod.py -s 1
Exercise      Reps
-----
Pushups      32
Jumping Jacks 56
Situps       88
Pullups      24
$ ./wod.py -s 1 -e
Exercise      Reps
-----
Pushups      15
Jumping Jacks 27
Situps       44
Pullups      12
$ ./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 to get it 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 os
7  import random
8  from tabulate import tabulate
9  from dire import die
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
53     exercises = []
54
55     for row in csv.DictReader(fh, delimiter=','):
56         name = row['exercise']
57         low, high = row['reps'].split('-')
58         exercises.append((name, int(low), int(high)))
59
60     return exercises
61
62
63 # -----
64 def main():
65     """Make a jazz noise here"""
66
67     args = get_args()
68     random.seed(args.seed)
69     exercises = read_csv(args.file)
70     table = []
71
72     for name, low, high in random.sample(exercises, k=args.num_exercises):
73         if args.easy:
74             low = int(low / 2)
75             high = int(high / 2)
76
77         table.append((name, random.randint(low, high)))
78
79     print(tabulate(table, headers=('Exercise', 'Reps')))
80
81
82 # -----
83 if __name__ == '__main__':
84     main()

```

Discussion

As usual, I start with my `get_args` first to define what the program expects. Most important is a `file` which is not required since 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 offload the validation of the argument to `argparse`. The `--seed` and `--num_exercises` options must be `type=int`, and the `--easy` option is a `True/False` flag.

Reading the WOD file

Since I know I will return a `list` of exercises and low/high ranges, I first set `exercises = []`. I recommended you use the `csv.DictReader` module to parse the CSV files into a list of dictionaries that represent each row's 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
```

You can read it like so:

```
>>> import csv
>>> fh = open('wod.csv')
>>> rows = list(csv.DictReader(fh, delimiter=','))
>>> rows[0]
OrderedDict([('exercise', 'Burpees'), ('reps', '20-50')])
```

On line 55-58, I iterate the rows, `split` the `reps` values like 20-50 into a low and high values, coerce them into `int` values. I want to `return` a `list` of tuples containing the exercise name along with the minimum and maximum reps.

For the purposes of this exercise, you can assume the CSV files you are given will have the correct headers and the reps can be safely converted.

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. The output will be formatted using the `tabulate` module which wants the data as a single `list` of rows to format, so I first create a `table` to hold the chosen exercises and reps. Then I get the workout options and reps from the file (line 69) which looks like this:

```
>>> from pprint import pprint as pp
>>> pp(exercises)
[('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)]

```

and can then use `random.sample` to select some `k` number given by the user from the `exercises`:

```

>>> import random
>>> random.sample(exercises, 3)
[('Dips', 10, 30), ('Jumprope', 50, 100), ('Lunges', 20, 40)]

```

The sampling returns a `list` from `exercises` which holds tuples with three values each, so I can iterate over those tuples and unpack them all on line 72. If `args.easy` is `True`, then I halve the `low` and `high` values.

```

>>> random.randint(5, 10)
6
>>> random.randint(5, 10)
8

```

Printing the table

Then I can `append` to the `table` a new tuple containing the `name` of the exercise and a `randint` (random integer) selected from the range given by `low` and `high`. Finally I can `print` the result of having the `tabulate` module create a text table using the given `headers`. You can explore the documentation of the `tabulate` module to discover the many options the module has.

Chapter 17: Blackjack

Write a Python program called `blackjack.py` that plays an abbreviated game of Blackjack. You will need to `import random` to get random cards from a deck you will construct, and so your program will need to accept a `-s|--seed` that will set `random.seed()` with the value that is passed in so that the test suite will work. The other arguments you will accept are two flags (Boolean values) of `-p|--player_hits` and `-d|--dealer_hits`. As usual, you will also have a `-h|--help` option for usage statement.

To play the game, the user will run the program and will see a display of what cards the dealer has (noted “D”) and what cards the player has (noted “P”) along with a sum of the values of the cards. In Blackjack, number cards are worth their value, face cards are worth 10, and the Ace will be worth 1 for our game (though in the real game it can alternate between 1 and 11).

To create your deck of cards, you will need to use the Unicode symbols for the suites () [which won’t display in the PDF, so consult the Markdown file].

Combine these with the numbers 2-10 and the letters “A”, “J”, “Q,” and “K” (hint: look at `itertools.product`). Because your game will use randomness, you will need to sort your deck and then use the `random.shuffle` method so that your cards will be in the correct order to pass the tests.

When you make the initial deal, keep in mind how cards are actually dealt – first one card to each of the players, then one to the dealer, then the players, then the dealer, etc. You might be tempted to use `random.choice` or something like that to select your cards, but you need to keep in mind that you are modeling an actual deck and so selected cards should no longer be present in the deck. If the `-p|--player_hits` flag is present, deal an additional card to the player; likewise with the `-d|--dealer_hits` flag.

After displaying the hands, the code should:

1. Check if the player has more than 21; if so, print ‘Player busts! You lose, loser!’ and `exit(0)`
2. Check if the dealer has more than 21; if so, print ‘Dealer busts.’ and `exit(0)`
3. Check if the player has exactly 21; if so, print ‘Player wins. You probably cheated.’ and `exit(0)`
4. Check if the dealer has exactly 21; if so, print ‘Dealer wins!’ and `exit(0)`
5. If either the dealer or the player has less than 18, you should indicate “X should hit.”

NB: Look at the Markdown format to see the actual output as the suites won’t display in the PDF version!

```
$ ./blackjack.py
D [11]: J A
```

```
P [18]: 8 10
Dealer should hit.
$ ./blackjack.py
D [13]: 3 J
P [16]: 6 10
Dealer should hit.
Player should hit.
$ ./blackjack.py -s 5
D [ 5]: 4 A
P [19]: 10 9
Dealer should hit.
$ ./blackjack.py -s 3 -p
D [19]: K 9
P [22]: 3 9 J
Player busts! You lose, loser!
$ ./blackjack.py -s 15 -p
D [19]: 10 9
P [21]: 10 8 3
Player wins. You probably cheated.
```

Solution

```
1  #!/usr/bin/env python3
2  """Blackjack"""
3
4  import argparse
5  import random
6  import sys
7  from itertools import product
8  from dire import die
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14      parser = argparse.ArgumentParser(
15          description='Argparse Python script',
16          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18      parser.add_argument('-s',
19                          '--seed',
20                          help='Random seed',
21                          metavar='int',
22                          type=int,
23                          default=None)
24
25      parser.add_argument('-d',
26                          '--dealer_hits',
27                          help='Dealer hits',
28                          action='store_true')
29
30      parser.add_argument('-p',
31                          '--player_hits',
32                          help='Player hits',
33                          action='store_true')
34
35      return parser.parse_args()
36
37
38  # -----
39  def bail(msg):
40      """print() and exit(0)"""
41      print(msg)
42      sys.exit(0)
43
```

```

44
45 # -----
46 def card_value(card):
47     """card to numeric value"""
48     val = card[1:]
49     faces = {'A': 1, 'J': 10, 'Q': 10, 'K': 10}
50     if val.isdigit():
51         return int(val)
52     elif val in faces:
53         return faces[val]
54     else:
55         die('Unknown card value for "{}".format(card))
56
57
58 # -----
59 def main():
60     """Make a jazz noise here"""
61     args = get_args()
62     random.seed(args.seed)
63     suites = list(' ')
64     values = list(range(2, 11)) + list('AJQK')
65     cards = sorted(map(lambda t: '{}{}'.format(*t), product(suites, values)))
66     random.shuffle(cards)
67
68     p1, d1, p2, d2 = cards.pop(), cards.pop(), cards.pop(), cards.pop()
69     player = [p1, p2]
70     dealer = [d1, d2]
71
72     if args.player_hits:
73         player.append(cards.pop())
74     if args.dealer_hits:
75         dealer.append(cards.pop())
76
77     player_hand = sum(map(card_value, player))
78     dealer_hand = sum(map(card_value, dealer))
79
80     print('D [{:2}]: {}'.format(dealer_hand, ' '.join(dealer)))
81     print('P [{:2}]: {}'.format(player_hand, ' '.join(player)))
82
83     if player_hand > 21:
84         bail('Player busts! You lose, loser!')
85     elif dealer_hand > 21:
86         bail('Dealer busts.')
87     elif player_hand == 21:
88         bail('Player wins. You probably cheated.')
89     elif dealer_hand == 21:

```

```
90         bail('Dealer wins!')
91
92     if dealer_hand < 18: print('Dealer should hit.')
93     if player_hand < 18: print('Player should hit.')
94
95
96 # -----
97 if __name__ == '__main__':
98     main()
```

Chapter 18: Family Tree

Write a program called `tree.py` that will take an input file as a single positional argument and produce a graph of the family tree described therein. The file can have only three kinds of statements:

1. `INITIALS = Full Name`
2. `person1 married person2`
3. `person1 and person2 begat child1[, child2...]`

Use the `graphviz` module to generate a graph like the `kyc.gv.pdf` included here that was generated from the following input:

```
$ cat tudor.txt
H7 = Henry VII
EOY = Elizabeth of York
H8 = Henry VIII
COA = Catherine of Aragon
AB = Anne Boleyn
JS = Jane Seymour
AOC = Anne of Cleves
CH = Catherine Howard
CP = Catherine Parr
HDC = Henry, Duke of Cornwall
M1 = Mary I
E1 = Elizabeth I
E6 = Edward VI

H7 married EOY
H7 and EOY begat H8
H8 married COA
H8 married AB
H8 married JS
H8 married AOC
H8 married CH
H8 married CP
H8 and COA begat HDC, M1
H8 and AB begat E1
H8 and JS begat E6
$ ./tree.py tudor.txt
Done, see output in "tudor.txt.gv".
```

Solution

```
1  #!/usr/bin/env python3
2  """
3  Author : kycklark
4  Date   : 2019-05-24
5  Purpose: Display a family tree
6  """
7
8  import argparse
9  import os
10 import re
11 import sys
12 from graphviz import Digraph
13
14
15 # -----
16 def get_args():
17     """Get command-line arguments"""
18
19     parser = argparse.ArgumentParser(
20         description='Display a family tree',
21         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
22
23     parser.add_argument('file',
24                         metavar='FILE',
25                         type=argparse.FileType('r'),
26                         help='File input')
27
28     parser.add_argument('-o',
29                         '--outfile',
30                         help='Output filename',
31                         metavar='str',
32                         type=str,
33                         default='')
34
35     return parser.parse_args()
36
37
38 # -----
39 def main():
40     """Make a jazz noise here"""
41
42     args = get_args()
43     fh = args.file
```

```

44     out_file = args.outfile or os.path.basename(fh.name) + '.gv'
45
46     nodes, edges = parse_tree(fh)
47     dot = Digraph(comment='Tree')
48     for initials, name in nodes.items():
49         dot.node(name)
50
51     for n1, n2 in edges:
52         if n1 in nodes:
53             n1 = nodes[n1]
54         if n2 in nodes:
55             n2 = nodes[n2]
56
57         dot.edge(n1, n2)
58
59     dot.render(out_file, view=True)
60
61     print('Done, see output in "{}".'.format(out_file))
62
63 # -----
64 def parse_tree(fh):
65     """parse input file"""
66
67     ini_patt = '([A-Za-z0-9]+)'
68     name_patt = ini_patt + '\s*=\s*(.+)'
69     begat_patt = ini_patt + '\s+and\s+' + ini_patt + '\s+begat\s*(.+)'
70     married_patt = ini_patt + '\s+married\s+' + ini_patt
71     edges = set()
72     nodes = {}
73
74     for line in fh:
75         name_match = re.match(name_patt, line)
76         begat_match = re.match(begat_patt, line)
77         married_match = re.match(married_patt, line)
78
79         if name_match:
80             initials, name = name_match.groups()
81             nodes[initials] = name
82         elif married_match:
83             p1, p2 = married_match.groups()
84             edges.add((p1, p2))
85         elif begat_match:
86             p1, p2, begat = begat_match.groups()
87             children = re.split('\s*,\s*', begat)
88             for parent in p1, p2:
89                 for child in children:

```



```
90             edges.add((parent, child))
91
92     return nodes, edges
93
94
95 # -----
96 if __name__ == '__main__':
97     main()
```

Chapter 19: Gematria: Numeric encoding of text

Write a Python program called `gematria.py` that will numerically encode each word in a given text. 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 'python':
...     print(ord(char))
...
112
121
116
104
111
110
```

We’ve seen before how you can put a `for` loop inside brackets `[]` for a list comprehension. Do that and then `sum` the list.

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
```

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.
- Can you write a function that can encode just one word? E.g., "gematria" = 842.
- Be sure you only encode the words themselves and not any punctuation that might be next to a word. E.g., if you use `str.split` to break text on whitespaces, quotes/commas/periods and such will still be attached to the words. Additionally, you should remove any internal punctuation like apostrophes. Maybe look into the `re` module to use regular expressions.
- Now can you apply that function to each word in a line of text?

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     word = re.sub('[^a-zA-Z0-9]', '', word)
32     return str(sum(map(ord, word)))
33
34
35 # -----
36 def main():
37     """Make a jazz noise here"""
38
39     args = get_args()
40     text = args.text
41
42     for line in text.splitlines():
43         print(' '.join(map(word2num, line.split())))
```

```
44
45
46 # -----
47 if __name__ == '__main__':
48     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.

Reading lines of text

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.
```

List comprehensions vs map

Like several other programs, we now are left with applying some function to each member of a list. That is, we want to turn each word into a number, and we've seen there are several ways to go about this. I will focus on two methods which I use twice each: list comprehensions and the `map` function.

Encoding one word

First let's take just one word. We've seen how we can use `ord` to turn one character into a number:

```
>>> 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 something *iterable* like a `list` or a generator. Because the second argument is iterable, we don't have to spell out `for char in`.

```
>>> 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 only, 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`. We can put this into a function either writing it with `lambda` on one line:

```
>>> word2num = lambda word: str(sum(map(ord, word)))
>>> word2num('gematria')
'842'
```

Or using `def`:

```
>>> def word2num(word):
...     return str(sum(map(ord, word)))
...
>>> word2num('gematria')
'842'
```

Finding the words

Just above we were applying the `ord` function to every character in a word. Now we want to apply our new `word2num` function to every word in a line. I hope you see it's the exact same problem, and both list comprehensions and `map` will serve equally.

So how to find “words” in a line? We know that 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.']
```

There's a small problem, though. Notice that we get `worry,` and not `worry` and `spiders,` instead of `spiders`. We don't want to encode the punctuation that is still attached to the words. Also, let's just say we also don't want to encode the apostrophe in `Don't`. So how can we remove these offending characters? The first step is in identifying what they are. If we say “remove anything that is not the a letter in the set A-Z or a number in the list 0-9”, that helps. We can use regular expressions to describe that exactly using `[]` to create a “character class” and putting the allowed characters in there. Notice this filters out the unwanted characters:

```
>>> import re
>>> re.findall('[a-zA-Z0-9]', "Don't")
['D', 'o', 'n', 't']
>>> re.findall('[a-zA-Z0-9]', "spiders,")
['s', 'p', 'i', 'd', 'e', 'r', 's']
```

Or we could use the `re.sub` function to “substitute” any matches. We can negate our character class by putting a caret (`^`) just *inside* the start of the

brackets to indicate we want to find anything that's *not* an English alphabet character or an Arabic number and replace it with the empty string:

```
>>> re.sub('[^a-zA-Z0-9]', '', "Don't")
'Dont'
>>> re.sub('[^a-zA-Z0-9]', '', "spiders,")
'spiders'
```

Let's put that into a function:

```
>>> def clean(word):
...     return re.sub('[^a-zA-Z0-9]', '', word)
...
>>> clean("Don't")
'Dont'
>>> clean("spiders,")
'spiders'
```

Compare this with the earlier version to see that we now have “clean” words to encode:

```
>>> for line in text.splitlines():
...     words = map(clean, line.split())
...     print(list(words))
...
['Dont', 'worry', 'spiders']
['I', 'keep', 'house']
['casually']
```

For convenience, let's update the `word2num` function to use that:

```
def word2num(word):
    word = re.sub('[^a-zA-Z0-9]', '', word)
    return str(sum(map(ord, word)))
>>> word2num('spiders,')
'762'
>>> word2num('spiders')
'762'
```

Encoding all words

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:

```
>>> words = ['Dont', 'worry', 'spiders']
>>> [word2num(word) for word in words]
['405', '579', '762']
```

But `map` is cleaner:

```
>>> list(map(word2num, words))  
['405', '579', '762']
```

All that is left is to print the encoded words back out:

```
>>> for line in text.splitlines():  
...     print(' '.join(map(word2num, line.split())))  
...  
405 579 762  
73 421 548  
862
```

Chapter 20: Histogram

Write a Python program called `histy.py` that takes a single positional argument that may be plain text or the name of a file to read for the text. Count the frequency of each character (not spaces) and print a histogram of the data. By default, you should order the histogram by the characters but include `-f|--frequency_sort` option to sort by the frequency (in descending order). Also include a `-c|--character` option (default `|`) to represent a mark in the histogram, a `-m|--minimum` option (default `1`) to include a character in the output, a `-w|--width` option (default `70`) to limit the size of the histogram, and a `-i|--case_insensitive` flag to force all input to uppercase.

```
$ ./histy.py
usage: histy.py [-h] [-c str] [-m int] [-w int] [-i] [-f] str
histy.py: error: the following arguments are required: str
$ ./histy.py -h
usage: histy.py [-h] [-c str] [-m int] [-w int] [-i] [-f] str
```

Histogrammer

positional arguments:

str	Input text or file
-----	--------------------

optional arguments:

-h, --help	show this help message and exit
-c str, --character str	Character for marks (default:)
-m int, --minimum int	Minimum frequency to print (default: 1)
-w int, --width int	Maximum width of output (default: 70)
-i, --case_insensitive	Case insensitive search (default: False)
-f, --frequency_sort	Sort by frequency (default: False)

```
$ ./histy.py ../inputs/fox.txt
```

```
T      1 |
a      1 |
b      1 |
c      1 |
d      1 |
e      3 |||
f      1 |
g      1 |
h      2 ||
i      1 |
j      1 |
k      1 |
```

```

l      1 |
m      1 |
n      1 |
o      4 ||||
p      1 |
q      1 |
r      2 ||
s      1 |
t      1 |
u      2 ||
v      1 |
w      1 |
x      1 |
y      1 |
z      1 |
$ ./histy.py ../inputs/const.txt -fim 100 -w 50 -c '#'
E 5107 #####
T 3751 #####
O 2729 #####
S 2676 #####
A 2675 #####
N 2630 #####
I 2433 #####
R 2206 #####
H 2029 #####
L 1490 #####
D 1230 #####
C 1164 #####
F 1021 #####
U 848 #####
P 767 #####
M 730 #####
B 612 #####
Y 504 #####
V 460 #####
G 444 #####
W 375 ###

```

Solution

```
1  #!/usr/bin/env python3
2  """Histogrammer"""
3
4  import argparse
5  import os
6  import re
7  from collections import Counter
8  from dire import die
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14      parser = argparse.ArgumentParser(
15          description='Histogrammer',
16          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18      parser.add_argument('text', metavar='str', help='Input text or file')
19
20      parser.add_argument('-c',
21                          '--character',
22                          help='Character for marks',
23                          metavar='str',
24                          type=str,
25                          default='|')
26
27      parser.add_argument('-m',
28                          '--minimum',
29                          help='Minimum frequency to print',
30                          metavar='int',
31                          type=int,
32                          default=1)
33
34      parser.add_argument('-w',
35                          '--width',
36                          help='Maximum width of output',
37                          metavar='int',
38                          type=int,
39                          default=70)
40
41      parser.add_argument('-i',
42                          '--case_insensitive',
43                          help='Case insensitive search',
```

```

44             action='store_true')
45
46     parser.add_argument('-f',
47                         '--frequency_sort',
48                         help='Sort by frequency',
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     text = args.text
60     char = args.character
61     width = args.width
62     min_val = args.minimum
63
64     if len(char) != 1:
65         die('--character "{}" must be one character'.format(char))
66
67     if os.path.isfile(text):
68         text = open(text).read()
69     if args.case_insensitive:
70         text = text.upper()
71
72     freqs = Counter(filter(lambda c: re.match(r'\w', c), list(text)))
73     high = max(freqs.values())
74     scale = high / width if high > width else 1
75     items = map(lambda t: (t[1], t[0]),
76                 sorted([(v, k) for k, v in freqs.items()],
77                        reverse=True)) if args.frequency_sort else sorted(
78                     freqs.items()))
79
80     for c, num in items:
81         if num < min_val:
82             continue
83         print('{} {:6} {}'.format(c, num, char * int(num / scale)))
84
85
86 # -----
87 if __name__ == '__main__':
88     main()

```

Chapter 21: Guessing Game

Write a Python program called `guess.py` that plays a guessing game for a number between a `-m|--min` and `-x|--max` value (default 1 and 50, respectively) with a limited number of `-g|--guesses` (default 5). Complain if either `--min` or `--guesses` is less than 1. Accept a `-s|--seed` for `random.seed`. If the user guesses something that is not a number, complain about it.

The game is intended to actually be interactive, which makes it difficult to test. Here is how it should look in interactive mode:

```
$ ./guess.py -s 1
Guess a number between 1 and 50 (q to quit): 25
"25" is too high.
Guess a number between 1 and 50 (q to quit): foo
"foo" is not a number.
Guess a number between 1 and 50 (q to quit): 12
"12" is too high.
Guess a number between 1 and 50 (q to quit): 6
"6" is too low.
Guess a number between 1 and 50 (q to quit): 9
"9" is correct. You win!
```



Figure 8: What does the future hold?

Because I want to be able to write a test for this, I also want the program to accept an `-i|--inputs` option so that the game can also be played exactly the same but without the prompts for input:

```
$ ./guess.py -s 1 -i 25 foo 12 6 9
"25" is too high.
"foo" is not a number.
"12" is too high.
```

"6" is too low.

"9" is correct. You win!

You should be able to handle this in your infinite game loop.

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import random
5  import re
6  import sys
7  from dire import die
8
9
10 # -----
11 def get_args():
12     """get args"""
13     parser = argparse.ArgumentParser(
14         description='Guessing game',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('-m',
18                         '--min',
19                         help='Minimum value',
20                         metavar='int',
21                         type=int,
22                         default=1)
23
24     parser.add_argument('-x',
25                         '--max',
26                         help='Maximum value',
27                         metavar='int',
28                         type=int,
29                         default=50)
30
31     parser.add_argument('-g',
32                         '--guesses',
33                         help='Number of guesses',
34                         metavar='int',
35                         type=int,
36                         default=5)
37
38     parser.add_argument('-s',
39                         '--seed',
40                         help='Random seed',
41                         metavar='int',
42                         type=int,
43                         default=None)
```

```

44
45     parser.add_argument('-i',
46                         '--inputs',
47                         help='Inputs',
48                         metavar='str',
49                         type=str,
50                         nargs='+',
51                         default=[])
52
53     return parser.parse_args()
54
55
56 # -----
57 def main():
58     """main"""
59     args = get_args()
60     low = args.min
61     high = args.max
62     guesses_allowed = args.guesses
63     inputs = args.inputs
64     random.seed(args.seed)
65
66     if low < 1:
67         die('--min "{}" cannot be lower than 1'.format(low))
68
69     if guesses_allowed < 1:
70         die('--guesses "{}" cannot be lower than 1'.format(guesses_allowed))
71
72     if low > high:
73         die('--min "{}" is higher than --max "{}"'.format(low, high))
74
75     secret = random.randint(low, high)
76     prompt = 'Guess a number between {} and {} (q to quit): '.format(low, high)
77     num_guesses = 0
78
79     while True:
80         guess = inputs.pop(0) if inputs else input(prompt)
81         num_guesses += 1
82
83         if re.match('q(uit)?', guess.lower()):
84             print('Now you will never know the answer.')
85             sys.exit()
86
87         # Method 1: test if the guess is a digit
88         if not guess.isdigit():
89             print("{} is not a number.".format(guess))

```

```

90         continue
91     num = int(guess)
92
93     # Method 2: try/except
94     num = 0
95     try:
96         num = int(guess)
97     except:
98         warn("{} is not an integer".format(guess))
99         continue
100
101     if not low <= num <= high:
102         print('Number "{}" is not in the allowed range'.format(num))
103     elif num == secret:
104         print("{} is correct. You win!".format(num))
105         break
106     else:
107         print("{} is too {}.{}".format(num,
108                                         'low' if num < secret else 'high'))
109
110     if num_guesses >= guesses_allowed:
111         print(
112             'Too many guesses, loser! The number was "{}.".format(secret))
113         sys.exit(1)
114
115
116 # -----
117 if __name__ == '__main__':
118     main()

```

Chapter 22: Mommy’s Little (Crossword) Helper

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
```

Here is how a program could look:

```
$ ./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)

We’ll use an underscore (`_`) to indicate a blank 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 os
6  import re
7  import sys
8  from typing import List, TextIO
9
10
11  # -----
12  def get_args() -> argparse.Namespace:
13      """Get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Crossword helper',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('pattern', metavar='str', help='The pattern to search')
20
21      parser.add_argument('-w',
22                          '--wordlist',
23                          help='Wordlist to search',
24                          metavar='str',
25                          type=argparse.FileType('r'),
26                          default='/usr/share/dict/words')
27
28      return parser.parse_args()
29
30
31  # -----
32  def regex_solution(pattern: str, wordlist: TextIO) -> List[str]:
33      """Using regular expressions"""
34
35      regex = r'\b{}\b'.format(pattern.replace('_', '.'))
36      return re.findall(regex, wordlist.read())
37
38
39  # -----
40  def manual_solution(pattern: str, wordlist: TextIO) -> List[str]:
41      """Not using regular expressions"""
42
43      letters = [t for t in enumerate(pattern) if t[1] != '_']
```

```

44     #letters = filter(lambda t: t[1] != '_', enumerate(pattern))
45     wanted_len = len(pattern)
46     words = []
47
48     for word in wordlist.read().split():
49         if len(word) == wanted_len and all(
50             [word[i] == char for i, char in letters]):
51             words.append(word)
52
53     return words
54
55
56 # -----
57 def main():
58     """Make a jazz noise here"""
59
60     args = get_args()
61     words = regex_solution(args.pattern, args.wordlist)
62     #words = manual_solution(args.pattern, args.wordlist)
63
64     if words:
65         for i, word in enumerate(words, start=1):
66             print('{:3}: {}'.format(i, word))
67     else:
68         print('Found no words matching "{}".'.format(args.pattern))
69
70
71 # -----
72 if __name__ == '__main__':
73     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`).

Regular Expressions

The `regex_solution` could be one line, but I wrote it in two for readability. The `pattern` uses underscores (`_`) to indicate a character. 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 chosen to use `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.

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 only need to use `list` here to force the REPL to evaluate the generator.

```
>>> pattern = '_t_ed'
>>> list(enumerate(pattern))
[(0, '_'), (1, 't'), (2, '_'), (3, 'e'), (4, 'd')]
```

You don't need to use `list` in your code unless you will need to iterate the generated list more than once. This is because generators are lazy, hence they won't generate their values unless forced, and they can only be iterated once:

```
>>> 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. One limitation to the `lambda` is that it cannot unpack the tuple, so I use `t` to remind me of the type and use `[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 return the tuple all the same):

```
>>> [(i, char) for i, char in enumerate(pattern) if char != '_']
[(1, 't'), (3, 'e'), (4, 'd')]
```

For this solution, I do want to look at each word individually, so I call `for word in wordlist.read().split()` and then check first for the length. The second condition is a little trickier and worth exploring. I decided to use the `all` function to find if *all* the characters in the `pattern` are the same in the `word`. Here I use the list comprehension syntax to unpack the list of tuples in `letters` to get their positions (`i`) and characters (`char`) and check if the `word` at that position matches the character (`word[i] == char`):

```
>>> word = 'steed'
>>> [word[i] == char for i, char in letters]
[True, True, True]
>>> word = 'steer'
>>> [word[i] == char for i, char in letters]
[True, True, False]
```

And then `all` will reduce it to a single value:

```
>>> word = 'steed'
>>> all([word[i] == char for i, char in letters])
True
>>> word = 'steer'
>>> all([word[i] == char for i, char in letters])
False
```

If both conditions are `True` (same length, all characters the same), then I `append` the `word` to the list of `words` I finally `return` from the function.

Summary

All that is left is 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. I hope you tried solving this problem with and without regular expressions as there is much to learn by each method.

Chapter 23: Kentucky Friar

Write a Python program called `friar.py` that reads some input text from a single positional argument on the command line (which could be a file to read) and transforms the text by dropping the “g” from words two-syllable words ending in “-ing” and also changes “you” to “y’all”. Be mindful to keep the case the same on the first letter, e.g, “You” should become “Y’all,” “Hunting” should become “Huntin”’.



Figure 9: The friar is fixin’ ta do some cookin’!

```
$ ./friar.py
usage: friar.py [-h] str
friar.py: error: the following arguments are required: str
$ ./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
$ ./friar.py you
y'all
$ ./friar.py Fishing
Fishin'
```

```
$ ./friar.py string
string
$ cat tests/input1.txt
So I was fixing to ask him, "Do you want to go fishing?" I was dying
to go for a swing and maybe do some swimming, too.
$ ./friar.py tests/input1.txt
So I was fixin' to ask him, "Do y'all want to go fishin'?" I was dyin'
to go for a swing and maybe do some swimmin', too.
```

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     return parser.parse_args()
19
20
21 # -----
22 def fry(word):
23     """
24     Drop the 'g' from '-ing' words, change "you" to "y'all"
25     """
26
27     ing_word = re.search('(.+)ing$', word)
28     you = re.match('([Yy])ou$', word)
29
30     if ing_word:
31         prefix = ing_word.group(1)
32         if re.search('[aeiouy]', prefix):
33             return prefix + "in'"
34     elif you:
35         return you.group(1) + "'all"
36
37     return word
38
39
40 # -----
41 def main():
42     """Make a jazz noise here"""
43
```

```

44     args = get_args()
45     text = args.text
46
47     if os.path.isfile(text):
48         text = open(text).read()
49
50     for line in text.splitlines():
51         print(''.join(map(fry, re.split(r'(\W+)', line.rstrip()))))
52
53
54 # -----
55 if __name__ == '__main__':
56     main()

```

Discussion

The heart of this program for me is the `fry` function. The `main` and `get_args` should look pretty standard by now. We get some argument that is either the text or the name of a file with the text. I chose to handle the input line-by-line because of the need to `print` the output. I don't want to worry about messing up the existing new lines, so I decided to read a line, strip off the newline, process the words, and `print` it all back out.

I wouldn't want to try to solve this problem without regular expressions, so I didn't really bother exploring a way that doesn't use them. For one thing, I use `re.split` to split the text on things that do and do not look like words. The first argument to this function is the regex that matches the thing you want to split on. Normally this is thrown away, for instance, if I `split` on any amount of whitespace, then the whitespace is not included:

```
>>> import re
>>> s = 'I said, "How do you do?"'
>>> re.split('\s+', s)
['I', 'said,', '"How', 'do', 'you', 'do?"]
```

It's a funny trick with this method that if you put the regex in capturing parens, it will return both the splitting text and the bits in between. The expression `\w` is any “word”-like character, so `\W` is the complement (non-word characters). The plus sign means “one or more”, and so it finds all the non-word characters between the words. This is important because I don't want to lose them!

```
>>> re.split(r'(\W+)', s)
['I', ' ', 'said', ', ', '"', 'How', ' ', ' ', 'do', ' ', ' ', 'you', ' ', ' ', 'do', '?", '']
```

Now I need to process any string that ends in “ing”:

```
>>> re.search('(.)ing$', 'spam')
>>> re.search('(.)ing$', 'fishing')
<re.Match object; span=(0, 7), match='fishing'>
```

I only want to remove the “g” from two-syllable words, though. A rough guess is to look for a vowel in the part of the word before the “ing”, so I wrote the regex to capture the first part:

```
>>> match = re.search('(.)ing$', 'fishing')
>>> prefix = match.group(1)
>>> prefix
'fish'
>>> re.search('[aeiouy]', prefix)
<re.Match object; span=(1, 2), match='i'>
```

But a word like “swing” would not work:

```
>>> match = re.search('(.)ing$', 'swing')
```

```
>>> prefix = match.group(1)
>>> prefix
'sw'
>>> re.search('[aeiouy]', prefix)
```

If all the conditions are true, I return the **prefix** of the word with “in”.

The other word to match is “you” either with an upper- or lowercase “y” which I can represent with a character class `[Yy]` for “either ‘Y’ or ‘y’” which I additionally capture so as to reuse it and maintain the proper case:

```
>>> match = re.match('([Yy])ou$', 'You')
>>> match.group(1) + "'all"
"Y'all"
```

Finally we need to apply our `fry` function to all the pieces we got from splitting the input text. I know that a list comprehension is more “Pythonic,” but I just prefer how `map` reads. I also understand that `map` is a bit slower due to the overhead of calling another function, but I don’t usually choose Python for performance.

```
>>> def fry(word):
...     ing_word = re.search('(.+)ing$', word)
...     you = re.match('([Yy])ou$', word)
...     if ing_word:
...         prefix = ing_word.group(1)
...         if re.search('[aeiouy]', prefix):
...             return prefix + "in"
...     elif you:
...         return you.group(1) + "'all"
...     return word
...
>>> s = "Hunting and fishing all you care about."
>>> ''.join([fry(w) for w in re.split(r'(\W+)', s)])
"Huntin' and fishin' all y'all care about."
>>> ''.join(map(fry, re.split(r'(\W+)', s)))
"Huntin' and fishin' all y'all care about."
```


Chapter 24: Mad Libs

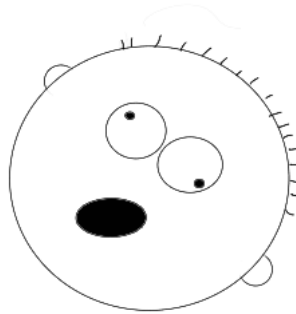


Figure 10: This definitely not a copyright infringement.

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, please add a `-i|--inputs` option so the test suite can pass in all the answers and bypass the `input` calls.

```
$ ./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)

```
$ cat help.txt
<exclamation>! I need <noun>!
<exclamation>! Not just <noun>!
<exclamation>! You know I need <noun>!
```

```

<exclamation>!
$ ./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!
$ ./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.

```

Solution

```
1  #!/usr/bin/env python3
2  """Mad Libs"""
3
4  import argparse
5  import os
6  import re
7  import sys
8  from dire import die
9
10
11  # -----
12  def get_args():
13      """Get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Mad Libs',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('file',
20                          metavar='FILE',
21                          type=argparse.FileType('r'),
22                          help='Input file')
23
24      parser.add_argument('-i',
25                          '--inputs',
26                          help='Inputs (for testing)',
27                          metavar='str',
28                          type=str,
29                          nargs='+',
30                          required=False)
31
32      return parser.parse_args()
33
34
35  # -----
36  def main():
37      """Make a jazz noise here"""
38
39      args = get_args()
40      inputs = args.inputs
41      regex = re.compile('(<[^\>]+\>)' )
42      text = args.file.read().rstrip()
43      blanks = list(regex.finditer(text))
```

```

44
45     if not blanks: die('File "{}" has no placeholders'.format(args.file.name))
46
47     for blank in blanks:
48         name = blank.group(1)
49         answer = inputs.pop(0) if inputs else input('{}: '.format(
50             name.replace('<', '').replace('>', '')))
51         text = re.sub(name, answer, text, count=1)
52
53     print(text)
54
55
56 # -----
57 if __name__ == '__main__':
58     main()

```

Chapter 25: License Plates

Write a Python program called `license.py` that will create a regular expression for a license plate that accounts for characters and numbers which might be confused according to the following list:

- 5 S
- X K Y
- 1 I
- 3 E
- D 0 O Q
- M N
- U V W
- 2 8

Print the plate, the regular expression that would match that plate with all possible ambiguities, and then print all possible combinations of plates that includes the options along with the result of comparing the regular expression you created to the generated plate.

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

License plate regular expression

positional arguments:

PLATE License plate

optional arguments:

-h, --help show this help message and exit

```
$ ./license.py ABC1234
plate = "ABC1234"
regex = "^ABC[1I] [27] [3E] 4$"
ABC1234 OK
ABC12E4 OK
ABC1734 OK
ABC17E4 OK
ABCI234 OK
ABCI2E4 OK
ABCI734 OK
ABCI7E4 OK
$ ./license.py 123456
plate = "123456"
regex = "^ [1I] [27] [3E] 4 [5S] 6$"
```

```

123456 OK
1234S6 OK
12E456 OK
12E4S6 OK
173456 OK
1734S6 OK
17E456 OK
17E4S6 OK
I23456 OK
I234S6 OK
I2E456 OK
I2E4S6 OK
I73456 OK
I734S6 OK
I7E456 OK
I7E4S6 OK

```

Owing to the vagaries of the typefaces chosen by different states as well as the wear of the plates themselves, it would seem to me that people might easily confuse certain letters and numbers on plates. In the above example, **ABC1234**, the number 1 might look like the letter I, so the plate could be **ABD1234** or **ABCI234**. Granted, most license plates follow a pattern of using only letters in some spots and numbers in others, e.g., 3 letters plus 4 numbers, but I want to focus on all possibilities in this problem both because it makes the problem a bit easier and also because it doesn't have to worry about how each state formats their plates. Additionally, I want to account for customized plates that do not follow any pattern and might use any combination of characters.

I represented the above confusion table as a list of tuples. At first I thought I might use a dictionary, but there is a problem when three characters are involved, e.g., 0, O, and Q. I iterate through each character in the provided plate and decide if the character exists in any of the tuples. If so, I represent that position in the regular expression as a choice; if not, it is just the character.

If you think about a regular expression as a graph, it starts with the first character, e.g., **A** which must be followed by **B** which must be followed by **C** which must be followed by either a **1** or an **I** which must be followed by a **2** or a **7**, etc.

```

          1          2          3
A -> B -> C -> <   > -> <   > -> <   > -> 4
          I          7          E

```

In creating all the possible plates from your regular expression, you are making concrete what the regular expression is, well, ... expressing. I find `itertools.product` to be just the ticket for creating all those possibilities, which must be sorted for the sake of the test.

Solution

```
1  #!/usr/bin/env python3
2  """License plate regular expression"""
3
4  import argparse
5  import re
6  import sys
7  from itertools import product
8
9
10 # -----
11 def get_args():
12     """get command-line arguments"""
13     parser = argparse.ArgumentParser(
14         description='License plate regular expression',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('plate', metavar='PLATE', help='License plate')
18
19     return parser.parse_args()
20
21
22 # -----
23 def main():
24     """Make a jazz noise here"""
25     args = get_args()
26     plate = args.plate
27     mixups = [('5', 'S'), ('X', 'K', 'Y'), ('1', 'I'), ('3', 'E'),
28              ('D', 'O', 'O', 'Q'), ('M', 'N'), ('U', 'V', 'W'), ('2', '7')]
29
30     chars = []
31     for char in plate:
32         group = list(filter(lambda t: char in t, mixups))
33         if group:
34             chars.append(group[0])
35         else:
36             chars.append((char, ))
37
38     regex = '^{}$'.format(''.join(
39         map(lambda t: '[' + ''.join(t) + ']' if len(t) > 1 else t[0], chars)))
40
41     print('plate = {}'.format(plate))
42     print('regex = {}'.format(regex))
43
```

```
44     for possible in sorted(product(*chars)):
45         s = ''.join(possible)
46         print(s, 'OK' if re.search(regex, s) else 'NO')
47
48
49 # -----
50 if __name__ == '__main__':
51     main()
```


Chapter 26: Gibberish Generator

Write a Python program called `gibberish.py` that uses the Markov chain algorithm to generate new words from a set of training files. The program should take one or more positional arguments which are files that you read, word-by-word, and note the options of letters after a given `-k|--kmer_size` (default 2) grouping of letters. E.g., in the word “alabama” with `k=1`, the frequency table will look like:

```
a = 1, b, m
l = a
b = a
m = a
```

That is, given this training set, if you started with `l` you could only choose an `a`, but if you have `a` then you could choose `l`, `b`, or `m`.

The program should generate `-n|--num_words` words (default 10), each a random size between `k + 2` and a `-m|--max_word` size (default 12). Be sure to accept `-s|--seed` to pass to `random.seed`. My solution also takes a `-d|--debug` flag that will emit debug messages to `.log` for you to inspect.

Chose the best words and create definitions for them:

- yulcogicism: the study of Christmas gnostics
- umjump: skateboarding trick
- callots: insignia of officers in Greek army
- urchenev: fungal growth found under cobblestones

```
$ ./gibberish.py
usage: gibberish.py [-h] [-n int] [-k int] [-m int] [-s int] [-d] FILE [FILE ...]
gibberish.py: error: the following arguments are required: FILE
$ ./gibberish.py -h
usage: gibberish.py [-h] [-n int] [-k int] [-m int] [-s int] [-d] FILE [FILE ...]
```

Markov chain for characters/words

positional arguments:

FILE	Training file(s)
------	------------------

optional arguments:

<code>-h, --help</code>	show this help message and exit
<code>-n int, --num_words int</code>	Number of words to generate (default: 10)
<code>-k int, --kmer_size int</code>	Kmer size (default: 2)
<code>-m int, --max_word int</code>	Max word length (default: 12)

```

-s int, --seed int    Random seed (default: None)
-d, --debug          Debug to ".log" (default: False)
$ ./gibberish.py /usr/share/dict/words -s 1
1: oveli
2: uming
3: uylatiteda
4: owsh
5: uuse
6: ismandl
7: efortai
8: eyhopy
9: auretrab
10: ozogralach
$ ./gibberish.py ../inputs/const.txt -s 2 -k 3
1: romot
2: leasonsusp
3: gdoned
4: bunablshed
5: neithere
6: achmen
7: reason
8: nmentyone
9: effereof
10: eipts
$ ./gibberish.py -k 2 ../inputs/1945-boys.txt
1: baronaler
2: lip
3: oselli
4: ard
5: vicharley
6: melli
7: denry
8: jerictomank
9: rick
10: larvichaell

```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import logging
5  import os
6  import random
7  import re
8  import sys
9  from collections import defaultdict
10
11
12  # -----
13  def get_args():
14      """Get command-line arguments"""
15
16      parser = argparse.ArgumentParser(
17          description='Markov chain for characters/words',
18          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
19
20      parser.add_argument('file',
21                          metavar='FILE',
22                          nargs='+',
23                          help='Training file(s)')
24
25      parser.add_argument('-n',
26                          '--num_words',
27                          help='Number of words to generate',
28                          metavar='int',
29                          type=int,
30                          default=10)
31
32      parser.add_argument('-k',
33                          '--kmer_size',
34                          help='Kmer size',
35                          metavar='int',
36                          type=int,
37                          default=2)
38
39      parser.add_argument('-m',
40                          '--max_word',
41                          help='Max word length',
42                          metavar='int',
43                          type=int,
```

```

44             default=12)
45
46     parser.add_argument('-s',
47                         '--seed',
48                         help='Random seed',
49                         metavar='int',
50                         type=int,
51                         default=None)
52
53     parser.add_argument('-d',
54                         '--debug',
55                         help='Debug to ".log"',
56                         action='store_true')
57
58     return parser.parse_args()
59
60
61 # -----
62 def main():
63     """Make a jazz noise here"""
64
65     args = get_args()
66     k = args.kmer_size
67     random.seed(args.seed)
68
69     logging.basicConfig(
70         filename='.log',
71         filemode='w',
72         level=logging.DEBUG if args.debug else logging.CRITICAL)
73
74     # debate use of set/list in terms of letter frequencies
75     chains = defaultdict(list)
76     for file in args.file:
77         for line in open(file):
78             for word in line.lower().split():
79                 word = re.sub('[^a-z]', '', word)
80                 for i in range(0, len(word) - k):
81                     kmer = word[i:i + k + 1]
82                     chains[kmer[:-1]].append(kmer[-1])
83
84     logging.debug(chains)
85
86     kmers = list(chains.keys())
87     starts = set()
88
89     for i in range(1, args.num_words + 1):

```

```

90     word = ''
91     while not word:
92         kmer = random.choice(kmers)
93         if not kmer in starts and chains[kmer] and re.search(
94             '[aeiou]', kmer):
95             if k > 1:
96                 starts.add(kmer)
97             word = kmer
98
99     length = random.choice(range(k + 2, args.max_word))
100    logging.debug('Make a word {} long starting with "{}".format(
101        length, word))
102    while len(word) < length:
103        if not chains[kmer]: break
104        char = random.choice(list(chains[kmer]))
105        logging.debug('char = "{}".format(char))
106        word += char
107        kmer = kmer[1:] + char
108
109    logging.debug('word = "{}".format(word))
110    print('{:3}: {}'.format(i, word))
111
112
113    # -----
114    if __name__ == '__main__':
115        main()

```

Chapter 27: Pig Latin

Write a Python program named `piggie.py` that takes one or more file names as positional arguments and converts all the words in them into “Pig Latin” (see rules below). Write the output to a directory given with the flags `-o|--outdir` (default `out-yay`) using the same basename as the input file, e.g., `input/foo.txt` would be written to `out-yay/foo.txt`.

if a file argument names a non-existent file, print a warning to `STDERR` and skip that file. If the output directory does not exist, create it.

To create “Pig Latin”:

1. If the word begins with consonants, e.g., “k” or “ch”, move them to the end of the word and append “ay” so that “mouse” becomes “ouse-may” and “chair” becomes “air-chay.”
2. If the word begins with a vowel, simple append “-yay” to the end, so “apple” is “apple-yay.”

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

Convert to Pig Latin

positional arguments:

FILE	Input file(s)
------	---------------

optional arguments:

-h, --help	show this help message and exit
-o str, --outdir str	Output directory (default: out-yay)

```
[cholla@~/work/python/playful_python/piggie]$ ./piggie.py
usage: piggie.py [-h] [-o str] FILE [FILE ...]
piggie.py: error: the following arguments are required: FILE
[cholla@~/work/python/playful_python/piggie]$ ./piggie.py -h
usage: piggie.py [-h] [-o str] FILE [FILE ...]
```

Convert to Pig Latin

positional arguments:

FILE	Input file(s)
------	---------------

optional arguments:

-h, --help	show this help message and exit
-o str, --outdir str	Output directory (default: out-yay)

```
$ ./piggie.py ../inputs/sonnet-29.txt
1: sonnet-29.txt
Done, wrote 1 file to "out-yay".
$ head out-yay/sonnet-29.txt
onnet-Say 29-yay
illiam-Way akespeare-Shay
```

```
en-Whay, in-yay isgrace-day ith-way ortune-fay and-yay en-may's-yay eyes-yay,
I-yay all-yay alone-yay eweep-bay y-may outcast-yay ate-stay,
And-yay ouble-tray eaf-day eaven-hay ith-way y-may ootless-bay ies-cray,
And-yay ook-lay upon-yay elf-mysay and-yay urse-cay y-may ate-fay,
ishing-Way e-may ike-lay o-tay one-yay ore-may ich-ray in-yay ope-hay,
eatured-Fay ike-lay im-hay, ike-lay im-hay ith-way iends-fray ossessed-pay,
esiring-Day is-thay an-may's-yay art-yay and-yay at-thay an-may's-yay ope-s cay,
```

Solution

```
1  #!/usr/bin/env python3
2  """Convert text to Pig Latin"""
3
4  import argparse
5  import os
6  import re
7  import string
8  from dire import warn
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Convert to Pig Latin',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('file',
20                          metavar='FILE',
21                          nargs='+',
22                          help='Input file(s)')
23
24      parser.add_argument('-o',
25                          '--outdir',
26                          help='Output directory',
27                          metavar='str',
28                          type=str,
29                          default='out-yay')
30
31      return parser.parse_args()
32
33
34  # -----
35  def main():
36      """Make a jazz noise here"""
37
38      args = get_args()
39      out_dir = args.outdir
40
41      if not os.path.isdir(out_dir):
42          os.makedirs(out_dir)
43
```



```

44     num_files = 0
45     for i, file in enumerate(args.file, start=1):
46         basename = os.path.basename(file)
47         out_file = os.path.join(out_dir, basename)
48         out_fh = open(out_file, 'wt')
49         print('{:3}: {}'.format(i, basename))
50
51         if not os.path.isfile(file):
52             warn("{} is not a file.".format(file))
53             continue
54
55         num_files += 1
56         for line in open(file):
57             for bit in re.split(r"([\w']+)", line):
58                 out_fh.write(pig(bit))
59
60         out_fh.close()
61
62     print('Done, wrote {} file{} to "{}".'.format(
63         num_files, ' ' if num_files == 1 else 's', out_dir))
64
65
66 # -----
67 def pig(word):
68     """Create Pig Latin version of a word"""
69
70     if re.match(r"^[^\w']+$", word):
71         consonants = re.sub('[aeiouAEIOU]', '', string.ascii_letters)
72         match = re.match('^([' + consonants + ']+)(.+)', word)
73         if match:
74             word = '-'.join([match.group(2), match.group(1) + 'ay'])
75         else:
76             word = word + '-yay'
77
78     return word
79
80
81 # -----
82 if __name__ == '__main__':
83     main()

```

Chapter 28: Soundex Rhymer

Write a Python program called `rhymer.py` that uses the Soundex algorithm/module to find words that rhyme with a given input word. When comparing words, you sometimes want to discount any leading consonants, e.g., the words “listen” and “glisten” rhyme but only if you compare the “isten” part, so the program should have an optional flag `-s|--stem` to indicate that the given word and the words you compare should both be trimmed to the “stem”. The program should take an optional `-w|--wordlist` argument (default `/usr/share/dict/words`) for the comparisons and should respond, as always, to `-h|--help` for usage.

For more background on the Soundex algorithm, I recommend the Wikipedia page and the PyPi module documentation for `soundex`.

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

Find rhyming words using the Soundex

positional arguments:

str	Word
-----	------

optional arguments:

-h, --help	show this help message and exit
-w str, --wordlist str	Wordlist (default: /usr/share/dict/words)
-s, --stem	Stem the word (remove starting consonants (default: False))

With my words list, I can find 37 words that rhyme with “listen” and 161 words that rhyme with the “isten” part:

```
$ ./rhymer.py listen | wc -l
37
$ ./rhymer.py -s listen | wc -l
161
```

I can verify that “glisten” only turns up when stemming is on:

```
$ ./rhymer.py listen | grep glisten
$ ./rhymer.py -s listen | grep glisten
glisten
```

Here is a sample of the words that my version finds:

```
$ ./rhymer.py listen | head -3
lackeydom
lactam
```

`lactation`

This program could be useful in creating custom input for the Gashlycrumb program.

Hints:

- You need to be sure that the given `word` actually has a vowel.
- If you are going to remove consonants from the beginning of a string, it might be easiest to find a regular expression to find things that are not vowels (because there are fewer of them to list).
- Another way to remove leading consonants would be to manually find the position of the first vowel in the string and then use a list slice on the given word to take the substring from that position to the end
- I suggest you use the `soundex` module

Testing the stemmer

I found the stemming part somewhat challenging, especially as I explored three different methods. I added the following test inside my `rhymers.py`:

```
def test_stemmer():
    """test stemmer"""

    assert stemmer('listen', True) == 'isten'
    assert stemmer('listen', False) == 'listen'
    assert stemmer('chair', True) == 'air'
    assert stemmer('chair', False) == 'chair'
    assert stemmer('apple', True) == 'apple'
    assert stemmer('apple', False) == 'apple'
    assert stemmer('xxxxxx', True) == 'xxxxxx'
    assert stemmer('xxxxxx', False) == 'xxxxxx'

    assert stemmer('LISTEN', True) == 'ISTEN'
    assert stemmer('LISTEN', False) == 'LISTEN'
    assert stemmer('CHAIR', True) == 'AIR'
    assert stemmer('CHAIR', False) == 'CHAIR'
    assert stemmer('APPLE', True) == 'APPLE'
    assert stemmer('APPLE', False) == 'APPLE'
    assert stemmer('XXXXXX', True) == 'XXXXXX'
    assert stemmer('XXXXXX', False) == 'XXXXXX'
```

And then I modified `make_test` to include `rhymers.py` in the list of files to test. The `pytest` module looks for any function name that starts with `test_` and runs them. The `assert` will halt execution of the program if the test fails.

Some of the words in my system dictionary don't have vowels, so some of methods that assumed the presence of a vowel failed. Writing a test just for this one

function really helped me find errors in my code.

Solution

```
1  #!/usr/bin/env python3
2  """Find rhyming words using the Soundex"""
3
4  import argparse
5  import re
6  import string
7  from soundex import Soundex
8
9
10 # -----
11 def get_args():
12     """get command-line arguments"""
13
14     parser = argparse.ArgumentParser(
15         description='Find rhyming words using the Soundex',
16         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18     parser.add_argument('word', metavar='str', help='Word')
19
20     parser.add_argument('-w',
21                         '--wordlist',
22                         metavar='str',
23                         help='Wordlist',
24                         type=argparse.FileType('r'),
25                         default='/usr/share/dict/words')
26
27     parser.add_argument('-s',
28                         '--stem',
29                         help='Stem the word (remove starting consonants',
30                         action='store_true')
31
32     args = parser.parse_args()
33
34     #if not any([c in 'aeiouy' for c in args.word.lower()]):
35     if not re.search('[aeiouy]', args.word, re.IGNORECASE):
36         msg = 'word "{}" must contain at least one vowel'
37         parser.error(msg.format(args.word))
38
39     return args
40
41
42 # -----
43 def stemmer(s: str, stem: bool) -> str:
```

```

44     """Use regular expressions"""
45
46     if stem:
47         match = re.search(r'^[aeiou]+([aeiou].*)', s, re.IGNORECASE)
48         return match.group(1) if match else s
49     return s
50
51
52 # -----
53 # def stemmer(s: str, stem: bool) -> str:
54 #     """Manually `find` first vowel"""
55
56 #     if stem:
57 #         positions = list(
58 #             filter(lambda p: p >= 0, [s.lower().find(v) for v in 'aeiou']))
59 #         if positions:
60 #             first = min(positions)
61 #             return s[first:] if first else s
62 #     return s
63
64 # -----
65 # def stemmer(s: str, stem: bool) -> str:
66 #     """Manually find first vowel with generator/next"""
67
68 #     if stem:
69 #         first = next(
70 #             (t[0] for t in enumerate(s) if t[1].lower() in 'aeiou'), False)
71 #         return s[first:] if first else s
72 #     return s
73
74
75 # -----
76 def test_stemmer():
77     """test stemmer"""
78
79     assert stemmer('listen', True) == 'isten'
80     assert stemmer('listen', False) == 'listen'
81     assert stemmer('chair', True) == 'air'
82     assert stemmer('chair', False) == 'chair'
83     assert stemmer('apple', True) == 'apple'
84     assert stemmer('apple', False) == 'apple'
85     assert stemmer('xxxxxx', True) == 'xxxxxx'
86     assert stemmer('xxxxxx', False) == 'xxxxxx'
87
88     assert stemmer('LISTEN', True) == 'ISTEN'
89     assert stemmer('LISTEN', False) == 'LISTEN'

```

```

90     assert stemmer('CHAIR', True) == 'AIR'
91     assert stemmer('CHAIR', False) == 'CHAIR'
92     assert stemmer('APPLE', True) == 'APPLE'
93     assert stemmer('APPLE', False) == 'APPLE'
94     assert stemmer('XXXXXX', True) == 'XXXXXX'
95     assert stemmer('XXXXXX', False) == 'XXXXXX'
96
97
98     # -----
99     def main():
100         """Make a jazz noise here"""
101
102         args = get_args()
103         given = args.word
104         words = args.wordlist.read().split()
105
106         def sndx(s):
107             return Soundex().soundex(stemmer(s, args.stem))
108
109         wanted = sndx(given)
110
111         for word in words:
112             if given != word and sndx(word) == wanted:
113                 print(word)
114
115         # print('\n'.join(
116         #     filter(lambda word: given != word and sndx(word) == wanted, words)))
117
118         # print('\n'.join([
119         #     word for word in words if given != word and sndx(word) == wanted
120         # ]))
121
122
123     # -----
124     if __name__ == '__main__':
125         main()

```

Discussion

The first thing to check is that the given word contains a vowel which is simple enough if you use regular expressions. We'll include “y” for this purpose:

```
>>> re.search('[aeiouy]', 'YYZ', re.IGNORECASE) or 'Fail'
<re.Match object; span=(0, 1), match='Y'>
>>> re.search('[aeiouy]', 'bbbb', re.IGNORECASE) or 'Fail'
'Fail'
```

Another way that doesn't use a regex could use a list comprehension to iterate over character in the given word to see if it is in the list of vowels 'aeiouy':

```
>>> [c in 'aeiouy' for c in 'CAT'.lower()]
[False, True, False]
```

You can then ask if any of these tests are true:

```
>>> any([c in 'aeiouy' for c in 'CAT'.lower()])
True
>>> any([c in 'aeiouy' for c in 'BCD'.lower()])
False
```

By far the regex version is simpler, but it's always interesting to think about other ways to accomplish a task. Anyway, if the given word does not have a vowel, I throw a `parser.error`.

Using Soundex

The `soundex` module has you create a `Soundex` object and then call a `soundex` function, which all seems a bit repetitive. Still, it gives us a way to get a Soundex value for a given word:

```
>>> from soundex import Soundex
>>> sndx = Soundex()
>>> sndx.soundex('paper')
'p16'
```

The problem is that sometimes we want the stemmed version of the word:

```
>>> sndx.soundex('aper')
'a16'
```

So I wrote a `stemmer` function that does (or does not) stem the word using the value of the `--stem` option which I defined in `argparse` as a Boolean value. I tried to find a way to remove leading consonants both with and without regular expressions. The regex version builds a somewhat complicated regex. Let's start with how to match something at the start of a string that is *not* a vowel (again, because there are only 5 to list):


```
>>> import re
>>> re.search(r'^[~aeiou]+', 'chair')
<re.Match object; span=(0, 2), match='ch'>
```

So we saw earlier that `[aeiou]` is the character class that matches vowels, so we can *negate* the class with `^` **inside** the character class. It's a bit confusing because there is also a `^` at the beginning of the `r''` (raw) string that anchors the expression to the beginning of the string.

OK, so that find the non-vowels leading the word, but we want the bit afterwards. It seems like we could just write something like this:

```
>>> re.search(r'^[~aeiou]+(.+)$', 'chr')
<re.Match object; span=(0, 3), match='chr'>
```

Which seems to say “one or more non-vowels followed by one or more of anything” and it looks to work, but look further:

```
>>> re.search(r'^[~aeiou]+(.+)$', 'chr').groups()
('r',)
```

It finds the last `r`. We need to specify that after the non-vowels there needs to be at least one vowel:

```
>>> re.search(r'^[~aeiou]+([aeiou].*)', 'chr')
```

And now it works:

```
>>> re.search(r'^[~aeiou]+([aeiou].*)', 'chr')
>>> re.search(r'^[~aeiou]+([aeiou].*)', 'car')
<re.Match object; span=(0, 3), match='car'>
>>> re.search(r'^[~aeiou]+([aeiou].*)', 'car').groups()
('ar',)
```

So the `stemmer` works by first looking to see if we should even attempt to `stem`. If so, it attempts to match the regular expression. If that succeeds, then it returns the match. The `else` for everything is to return the original string `s`.

The two other versions of `stemmer` rely on some things I'll discuss later.

As stated in the intro, it was most helpful to me to add the `test_stemmer` function to ensure that all my versions of the `stemmer` function actually had the same behavior.

Once I have the `stemmer` function, I can apply it to the given `word` and every word in the `--wordlist` and then call the “

Chapter 29: Substring Guessing Game

Write a Python program called `sub.py` that plays a guessing game where you read a `-f|--file` input (default `/usr/share/dict/words`) and use a given `-k|--ksize` to find all the words grouped by their shared kmers. Remove any kmers where the number of words is fewer than `-m|--min_words`. Also accept a `-s|--seed` for `random.seed` for testing purposes. Prompt the user to guess a word for a randomly chosen kmer. If their guess is not present in the shared list, taunt them mercilessly. If their guess is present, affirm their worth and prompt to guess again. Allow them to use `!` to quit and `?` to be provided a hint (a word from the list). For both successful guesses and hints, remove the word from the shared list. When they have quit or exhausted the list, quit play. At the end of the game, report the number of found words.

```
$ ./sub.py -h
```

```
usage: sub.py [-h] [-f str] [-s int] [-m int] [-k int]
```

Find words sharing a substring

optional arguments:

```
-h, --help            show this help message and exit
-f str, --file str    Input file (default: /usr/share/dict/words)
-s int, --seed int    Random seed (default: None)
-m int, --min_words int
                        Minimum number of words for a given kmer (default: 3)
-k int, --ksize int   Size of k (default: 4)
```

```
$ ./sub.py
```

```
Name a word that contains "slak" [!=quit, ?=hint] (10 left) slake
```

```
Totes! "slake" is found!
```

```
Name a word that contains "slak" [!=quit, ?=hint] (9 left) ?
```

```
For instance, "breislakite"...
```

```
Name a word that contains "slak" [!=quit, ?=hint] (8 left) unslakable
```

```
Totes! "unslakable" is found!
```

```
Name a word that contains "slak" [!=quit, ?=hint] (7 left) q
```

```
What is wrong with you?
```

```
Name a word that contains "slak" [!=quit, ?=hint] (7 left) !
```

```
Quitter!
```

```
Hey, you found 2 words! Not bad.
```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import os
5  import random
6  import re
7  import sys
8  from collections import defaultdict
9  from dire import die
10
11
12  # -----
13  def get_args():
14      """get command-line arguments"""
15      parser = argparse.ArgumentParser(
16          description='Find words sharing a substring',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('-f',
20                          '--file',
21                          metavar='str',
22                          help='Input file',
23                          type=str,
24                          default='/usr/share/dict/words')
25
26      parser.add_argument('-s',
27                          '--seed',
28                          help='Random seed',
29                          metavar='int',
30                          type=int,
31                          default=None)
32
33      parser.add_argument('-m',
34                          '--min_words',
35                          help='Minimum number of words for a given kmer',
36                          metavar='int',
37                          type=int,
38                          default=3)
39
40      parser.add_argument('-k',
41                          '--ksize',
42                          help='Size of k',
43                          metavar='int',
```

```

44                                     type=int,
45                                     default=4)
46
47     return parser.parse_args()
48
49
50 # -----
51 def get_words(file):
52     """Get words from input file"""
53
54     if not os.path.isfile(file):
55         die("{} is not a file")
56
57     words = set()
58     for line in open(file):
59         for word in line.split():
60             words.add(re.sub('[^a-zA-Z0-9]', '', word.lower()))
61
62     if not words:
63         die('No usable words in "{}".format(file))
64
65     return words
66
67 # -----
68 def get_kmers(words, k, min_words):
69     """ Find all words sharing kmers"""
70
71
72     if k <= 1:
73         die('-k "{}" must be greater than 1'.format(k))
74
75     shared = defaultdict(list)
76     for word in words:
77         for kmer in [word[i:i + k] for i in range(len(word) - k + 1)]:
78             shared[kmer].append(word)
79
80     # Select kmers having enough words (can't use `pop`!)
81
82     # Method 1: for loop
83     ok = dict()
84     for kmer in shared:
85         if len(shared[kmer]) >= min_words:
86             ok[kmer] = shared[kmer]
87
88     # Method 2: list comprehension
89     # ok = dict([(kmer, shared[kmer]) for kmer in shared

```

```

90         #             if len(shared[kmer]) >= min_words])
91
92     # Method 3: map/filter
93     # ok = dict(
94     #         map(lambda kmer: (kmer, shared[kmer]),
95     #             filter(lambda kmer: len(shared[kmer]) >= min_words,
96     #                 shared.keys()))))
97
98     return ok
99
100
101 # -----
102 def main():
103     """Make a jazz noise here"""
104
105     args = get_args()
106
107     random.seed(args.seed)
108
109     shared = get_kmers(get_words(args.file), args.ksize, args.min_words)
110
111     # Choose a kmer, setup game state
112     kmer = random.choice(list(shared.keys()))
113     guessed = set()
114     found = []
115     prompt = 'Name a word that contains "{}" [!=quit, ?=hint] '.format(kmer)
116     compliments = ['Nice', 'Rock on', 'Totes', 'Fantastic', 'Excellent']
117     taunts = [
118         'Surely you jest!', 'Are you kidding me?',
119         'You must have rocks for brains.', 'What is wrong with you?'
120     ]
121
122     #print(kmer, shared[kmer])
123
124     while True:
125         num_left = len(shared[kmer])
126         if num_left == 0:
127             print('No more words!')
128             break
129
130         guess = input(prompt + '({} left) '.format(num_left)).lower()
131
132         if guess == '?':
133             # Provide a hint
134             pos = random.choice(range(len(shared[kmer])))
135             word = shared[kmer].pop(pos)

```

```

136         print('For instance, "{}"...'.format(word))
137
138     elif guess == '!':
139         # Bail
140         print('Quitter!')
141         break
142
143     elif guess in guessed:
144         # Chastise
145         print('You have already guessed "{}"'.format(guess))
146
147     elif guess in shared[kmer]:
148         # Remove the word, feedback with compliment
149         pos = shared[kmer].index(guess)
150         word = shared[kmer].pop(pos)
151         print('{}! "{}" is found!'.format(random.choice(compliments),
152                                           word))
153         found.append(word)
154         guessed.add(guess)
155
156     else:
157         # Taunt
158         print(random.choice(taunts))
159
160     # Game over, man!
161     if found:
162         n = len(found)
163         print('Hey, you found {} word{}! Not bad.'.format(
164             n, '' if n == 1 else 's'))
165     else:
166         print('Wow, you found no words. You suck!')
167
168
169     # -----
170 if __name__ == '__main__':
171     main()

```

Chapter 30: Tic-Tac-Toe Outcome

Create a Python program called `outcome.py` that takes a given Tic-Tac-Toe state as its only (positional) argument and reports if X or O has won or if there is no winner. The state should only contain the characters “.”, “O”, and “X”, and must be exactly 9 characters long. If there is not exactly one argument, print a “usage” statement.

```
$ ./outcome.py
Usage: outcome.py STATE
$ ./outcome.py ..X.OA..X
State "..X.OA..X" must be 9 characters of only ., X, O
$ ./outcome.py ..X.OX...
No winner
$ ./outcome.py ..X.OX..X
X has won
```

Solution

```
1  #!/usr/bin/env python3
2
3  import os
4  import re
5  import sys
6
7
8  # -----
9  def main():
10     args = sys.argv[1:]
11
12     if len(args) != 1:
13         print('Usage: {} STATE'.format(os.path.basename(sys.argv[0])))
14         sys.exit(1)
15
16     state = args[0]
17
18     if not re.search('^.[XO]{9}$', state):
19         print('State "{}" must be 9 characters of only ., X, O'.format(state),
20               file=sys.stderr)
21         sys.exit(1)
22
23     winning = [[0, 1, 2], [3, 4, 5], [6, 7, 8], [0, 3, 6], [1, 4, 7],
24               [2, 5, 8], [0, 4, 8], [2, 4, 6]]
25
26     winner = 'No winner'
27
28     # for player in ['X', 'O']:
29     #     for combo in winning:
30     #         i, j, k = combo
31     #         if state[i] == player and state[j] == player and state[k] == player:
32     #             winner = player
33     #             break
34
35     # for player in ['X', 'O']:
36     #     for combo in winning:
37     #         chars = []
38     #         for i in combo:
39     #             chars.append(state[i])
40
41     #         if ''.join(chars) == player * 3:
42     #             winner = player
43     #             break
```



```

44
45     # for player in ['X', 'O']:
46     #     for i, j, k in winning:
47     #         chars = ''.join([state[i], state[j], state[k]])
48     #         if ''.join(chars) == '{}{}{}'.format(player, player, player):
49     #             winner = player
50     #             break
51
52     for player in ['X', 'O']:
53         for i, j, k in winning:
54             combo = [state[i], state[j], state[k]]
55             if combo == [player, player, player]:
56                 winner = '{} has won'.format(player)
57                 break
58
59     # for combo in winning:
60     #     group = list(map(lambda i: state[i], combo))
61     #     for player in ['X', 'O']:
62     #         if all(x == player for x in group):
63     #             winner = player
64     #             break
65
66     print(winner)
67
68
69     # -----
70 if __name__ == '__main__':
71     main()

```

Chapter 31: Twelve Days of Christmas

Write a Python program called `twelve_days.py` that will generate the “Twelve Days of Christmas” song up to the `-n|--number_days` argument (default 12), writing the resulting text to the `-o|--outfile` argument (default STDOUT).

```
$ ./twelve_days.py -h
usage: twelve_days.py [-h] [-o str] [-n int]
```

Twelve Days of Christmas

optional arguments:

```
-h, --help            show this help message and exit
-o str, --outfile str  Outfile (STDOUT) (default: )
-n int, --number_days int
                        Number of days to sing (default: 12)
```

```
$ ./twelve_days.py -n 1
On the first day of Christmas,
My true love gave to me,
A partridge in a pear tree.
```

```
$ ./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.
```

```
$ ./twelve_days.py -o out
$ wc -l out
    113 out
```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import sys
5  from dire import die
6
7
8  # -----
9  def get_args():
10     """get command-line arguments"""
11     parser = argparse.ArgumentParser(
12         description='Twelve Days of Christmas',
13         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
14
15     parser.add_argument('-o',
16                         '--outfile',
17                         help='Outfile (STDOUT)',
18                         metavar='str',
19                         type=str,
20                         default='')
21
22     parser.add_argument('-n',
23                         '--number_days',
24                         help='Number of days to sing',
25                         metavar='int',
26                         type=int,
27                         default=12)
28
29     return parser.parse_args()
30
31
32 # -----
33 def main():
34     """Make a jazz noise here"""
35
36     args = get_args()
37     out_file = args.outfile
38     num_days = args.number_days
39     out_fh = open(out_file, 'wt') if out_file else sys.stdout
40
41     days = {
42         12: 'Twelve drummers drumming',
43         11: 'Eleven pipers piping',
```

```

44         10: 'Ten lords a leaping',
45         9: 'Nine ladies dancing',
46         8: 'Eight maids a milking',
47         7: 'Seven swans a swimming',
48         6: 'Six geese a laying',
49         5: 'Five gold rings',
50         4: 'Four calling birds',
51         3: 'Three French hens',
52         2: 'Two turtle doves',
53         1: 'a partridge in a pear tree',
54     }
55
56     ordinal = {
57         12: 'twelfth', 11: 'eleven', 10: 'tenth',
58         9: 'ninth', 8: 'eighth', 7: 'seventh',
59         6: 'sixth', 5: 'fifth', 4: 'fourth',
60         3: 'third', 2: 'second', 1: 'first',
61     }
62
63     if not num_days in days:
64         die('Cannot sing "{}" days'.format(num_days))
65
66     for i in range(1, num_days + 1):
67         first = 'On the {} day of Christmas,\nMy true love gave to me,'
68         out_fh.write(first.format(ordinal[i]) + '\n')
69         for j in reversed(range(1, i + 1)):
70             if j == 1:
71                 if i == 1:
72                     out_fh.write('{}.\n'.format(days[j].title()))
73                 else:
74                     out_fh.write('And {}.\n'.format(days[j]))
75             else:
76                 out_fh.write('{}.\n'.format(days[j]))
77
78         if i < max(days.keys()):
79             out_fh.write('\n')
80
81
82     # -----
83     if __name__ == '__main__':
84         main()

```

Chapter 32: War

The generation of random numbers is too important to be left to chance. – Robert R. Coveyou

Create a Python program called `war.py` that plays the card game “War.” The program will use the `random` module to shuffle a deck of cards, so your program will need to accept a `-s|--seed` argument (default: `None`) which you will use to call `random.seed`, if present.

First your program will need to create a deck of cards. You will need to use the Unicode symbols for the suites () [which won’t display in the PDF, so consult the Markdown file] and combine those with the numbers 2-10 and the letters “J”, “Q”, “K”, and “A.” (hint: look at `itertools.product`).

```
>>> from itertools import product
>>> a = list('AB')
>>> b = range(2)
>>> list(product(a, b))
[('A', 0), ('A', 1), ('B', 0), ('B', 1)]
```

NB: You must sort your deck and then use the `random.shuffle` method so that your cards will be in the correct order to pass the tests!

In the real game of War, the cards are shuffled and then dealt one card each first to the non-dealer, then to the dealer, until all cards are dealt and each player has 26 cards. We will not be modeling this behavior. When writing your version of the game, simply `pop` two cards off the deck as the cards for player 1 and player 2, respectively. Compare the two cards by ignoring the suite and evaluating the value where 2 is the lowest and Aces are the highest. When two cards have the same values (e.g., two 5s or two Jacks), print “WAR!” In the real game, this initiates a sub-game of War which is a “recursive” algorithm which we will not bother modeling. Keep track of which player wins each round where no points are awarded in a tie. At the end, report the points for each player and state the winner. In the event of a tie, print “DRAW.”

```
$ ./war.py -h
usage: war.py [-h] [-s int]
```

“War” cardgame

optional arguments:

```
  -h, --help            show this help message and exit
  -s int, --seed int    Random seed (default: None)
$ ./war.py -s 1
9   J P2
A   5 P1
4   8 P2
```

```

6 3 P1
5 3 P1
K 10 P1
7 7 WAR!
2 4 P2
2 10 P2
6 5 P1
2 6 P2
4 8 P2
J 9 P1
10 Q P2
8 7 P1
K Q P1
10 2 P1
9 9 WAR!
8 J P2
3 5 P2
Q 4 P1
6 A P2
K 7 P1
Q 3 P1
A K P1
A J P1
P1 14 P2 10: Player 1 wins
$ ./war.py -s 2
4 6 P2
K J P1
J 4 P1
7 4 P1
Q 10 P1
5 3 P1
K 9 P1
2 Q P2
7 A P2
3 A P2
5 8 P2
2 10 P2
10 K P2
2 3 P2
Q 8 P1
6 J P2
6 8 P2
8 7 P1
5 2 P1
6 J P2
9 9 WAR!

```

```

K   A P2
10  Q P2
7   5 P1
9   A P2
4   3 P1
P1 11 P2 14: Player 2 wins
$ ./war.py -s 10
J   3 P1
2   5 P2
Q  10 P1
10  4 P1
6   5 P1
3   J P2
K   8 P1
5   8 P2
5   3 P1
J  10 P1
10  J P2
A   7 P1
K   Q P1
7   A P2
9   9 WAR!
2   6 P2
K   A P2
6   Q P2
8   9 P2
3   7 P2
8   Q P2
6   4 P1
7   2 P1
4   4 WAR!
9   2 P1
K   A P2
P1 12 P2 12: DRAW

```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import random
5  import sys
6  from itertools import product
7
8
9  # -----
10 def get_args():
11     """get command-line arguments"""
12     parser = argparse.ArgumentParser(
13         description='War cardgame',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('-s',
17                         '--seed',
18                         help='Random seed',
19                         metavar='int',
20                         type=int,
21                         default=None)
22
23     return parser.parse_args()
24
25
26 # -----
27 def main():
28     """Make a jazz noise here"""
29     args = get_args()
30     seed = args.seed
31
32     if seed is not None:
33         random.seed(seed)
34
35     suits = list(' ')
36     values = list(map(str, range(2, 11))) + list('JQKA')
37     cards = sorted(map(lambda t: '{}{}'.format(*t), product(suits, values)))
38     random.shuffle(cards)
39
40     p1_wins = 0
41     p2_wins = 0
42
43     card_value = dict(
```



```

44         list(map(lambda t: list(reversed(t)), enumerate(list(values))))
45
46     while cards:
47         p1, p2 = cards.pop(), cards.pop()
48         v1, v2 = card_value[p1[1:]], card_value[p2[1:]]
49         res = ''
50
51         if v1 > v2:
52             p1_wins += 1
53             res = 'P1'
54         elif v2 > v1:
55             p2_wins += 1
56             res = 'P2'
57         else:
58             res = 'WAR!'
59
60         print('{:>3} {:>3} {}'.format(p1, p2, res))
61
62     print('P1 {} P2 {}: {}'.format(
63         p1_wins, p2_wins, 'Player 1 wins' if p1_wins > p2_wins else
64         'Player 2 wins' if p2_wins > p1_wins else 'DRAW'))
65
66
67 # -----
68 if __name__ == '__main__':
69     main()

```

Chapter 33: Anagram

Write a program called `presto.py` that will find anagrams of a given positional argument. The program should take an optional `-w|--wordlist` (default `/usr/share/dict/words`) and produce output that includes combinations of `-n|num_combos` words (default 1) that are anagrams of the given input.

It should provide a usage with no input or the `-h|--help` flags:

```
$ ./presto.py
usage: presto.py [-h] [-w str] [-n int] [-d] str
presto.py: error: the following arguments are required: str
$ ./presto.py -h
usage: presto.py [-h] [-w str] [-n int] [-d] str
```

Find anagrams

positional arguments:

str	Input text
-----	------------

optional arguments:

-h, --help	show this help message and exit
-w str, --wordlist str	Wordlist (default: /usr/share/dict/words)
-n int, --num_combos int	Number of words combination to test (default: 1)
-d, --debug	Debug (default: False)

Be default, it should search the `--wordlist` file for other words of the same length as the input that have the same letters in the same frequency:

```
$ ./presto.py presto
presto =
  1. poster
  2. repost
  3. respot
  4. stoper
$ ./presto.py listen
listen =
  1. enlist
  2. silent
  3. tinsel
```

If `-n` is greater than 1 (the default), then the program should additionally find all combinations of two words that together create the original word.

```
$ ./presto.py listen -n 2 | tail
82. sten li
```

```
83. te nils
84. ten lis
85. ten sil
86. ti lens
87. til ens
88. til sen
89. tin els
90. tin les
91. tinsel
```

Hints:

- How will you determine that two strings are anagrams? That is, what is the code you will use to compare two strings and return **True** or **False** that they are anagrams? Start there.
- You can assume a strict dictionary-type input file like the default, but you might also consider mining some other text as the source for anagrams, one that might have punctuation and mixed-case letters.
- When you move to `-n > 1`, you may find you quickly have an overwhelming number of combinations to consider. My `/usr/share/dict/words` has 235886 words. At `n=2`, that could produce over 55 *billion* combinations of words. Obviously I don't need to consider the entire Cartesian product of the list, only those whose lengths sum to equal the length of the input word. How can you find all the combinations of numbers that sum to that length? E.g, for 5, you can add $0 + 5$, $1 + 4$, and $2 + 3$. How can you segregate all the input words by their lengths?

Solution

```
1  #!/usr/bin/env python3
2  """Find anagrams"""
3
4  import argparse
5  import logging
6  import re
7  from collections import defaultdict, Counter
8  from itertools import combinations, permutations, product, chain
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Find anagrams',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('text', metavar='str', help='Input text')
20
21      parser.add_argument('-w',
22                          '--wordlist',
23                          help='Wordlist',
24                          metavar='str',
25                          type=argparse.FileType('r'),
26                          default='/usr/share/dict/words')
27
28      parser.add_argument('-n',
29                          '--num_combos',
30                          help='Number of words combination to test',
31                          metavar='int',
32                          type=int,
33                          default=1)
34
35      parser.add_argument('-d', '--debug', help='Debug', action='store_true')
36
37      return parser.parse_args()
38
39  # -----
40
41  def main():
42      """Make a jazz noise here"""
43
```

```

44     args = get_args()
45     text = args.text
46
47     logging.basicConfig(
48         filename='.log',
49         filemode='w',
50         level=logging.DEBUG if args.debug else logging.CRITICAL)
51
52     words = defaultdict(set)
53     regex = re.compile('[^a-z0-9]')
54     for line in args.wordlist:
55         for word in line.split():
56             clean = regex.sub('', word.lower())
57             if len(clean) == 1 and clean not in 'ai':
58                 continue
59             words[len(clean)].add(clean)
60
61     text_len = len(text)
62     counts = Counter(text)
63     anagrams = set()
64     lengths = list(words.keys())
65     for n in range(1, args.num_combos + 1):
66         key_combos = list(
67             filter(
68                 lambda t: sum(t) == text_len,
69                 set(
70                     map(lambda t: tuple(sorted(t)),
71                       combinations(chain(lengths, lengths), n))))
72         logging.debug('key combos = %s', key_combos)
73
74         for keys in key_combos:
75             logging.debug('Searching keys %s', keys)
76             word_combos = list(product(*list(map(lambda k: words[k], keys))))
77
78             for t in word_combos:
79                 if Counter(''.join(t)) == counts:
80                     logging.debug('combo = %s', t)
81                     for s in [' '.join(l) for l in permutations(t)]:
82                         if s != text:
83                             anagrams.add(s)
84
85             logging.debug('# anagrams = %s', len(anagrams))
86
87     logging.debug('Finished searching')
88
89     if anagrams:

```

```

90         print('{} ='.format(text))
91         for i, t in enumerate(sorted(anagrams), 1):
92             print('{:4}. {}'.format(i, t))
93     else:
94         print('No anagrams for "{}".'.format(text))
95
96
97 # -----
98 if __name__ == '__main__':
99     main()

```

Discussion

I rely on `type=argparse.FileType('r')` for any “file” argument, so my `get_args` once again uses that to define the input `--wordlist`. Likewise, I defined `--num_combos` as an `int` and let `argparse` handle argument validation for me.

Logging

My solution also incorporates the `logging` I used while solving this problem for myself. I tend copy and paste this block all the time:

```
logging.basicConfig(
    filename='.log',
    filemode='w',
    level=logging.DEBUG if args.debug else logging.CRITICAL)
```

If I define `args.debug` as a Boolean, then I can effectively turn `logging` on and off because I tend not to write `CRITICAL` messages. Since I use `filemode='w'` to overwrite the `.log` file, then that file will be empty after every run that `--debug` isn't on (and the default is that it is not). Also, I like to use a filename starting with a `.` (e.g., `.log`) as it will be hidden in most Unix-style `ls` commands. This makes logging as transparent and easy as I can think.

Reading wordlist

First I handle getting a wordlist. I wrote a rather verbose way to process what could be a large input file. Rather than called `args.wordlist.read().split()` which reads the *entire file into memory*, I chose to read each line one-by-one into memory and call `line.split()` on that. If you have to deal with large input file (e.g., I regularly deal with files in the gigabytes in biology!), it's best to read line-by-line.

I iterate **for** each **word** in the line and clean it up with a regular expression that defines a character class of all the characters `a-z` and `0-9` with `[a-z0-9]` and then uses a caret **inside** the character class to negate it. Then I use the `sub` (substitute) function to replace characters that match with the empty string:

```
>>> import re
>>> regex = re.compile('[^a-z0-9]')
>>> regex.sub('', '"hey!"')
'hey'
```

If the remaining word is only 1 character long, I only accept it if it is “a” or “i”.

To store the words, I decided to use a dictionary where the key is the length of a given word and the value is a `set` of the words of that length. I chose a

`set` in case I was reading a file other than a standard dictionary-type file where words might be repeated. I use the length of the each word as the key so that I can select the combinations of words whose lengths sum to the desired lengths. That is, if my input word is 5 characters long, there is no reason to look at words longer than 5 characters.

defaultdict

To define this data structure, I used `words = defaultdict(set)` where `defaultdict` takes a datatype like `str` or `list` as the default *value* to initiate when a given key does not exist. For instance, using `int` will create a new entry in the dictionary with a value of 0:

```
>>> from collections import defaultdict
>>> d = defaultdict(int)
>>> d
defaultdict(<class 'int'>, {})
>>> d['foo'] += 1
>>> d
defaultdict(<class 'int'>, {'foo': 1})
```

If you use `str`, the empty string will be used:

```
>>> d = defaultdict(str)
>>> d
defaultdict(<class 'str'>, {})
>>> d['foo'] += 'a'
>>> d
defaultdict(<class 'str'>, {'foo': 'a'})
>>> d['foo'] += 'b'
>>> d
defaultdict(<class 'str'>, {'foo': 'ab'})
```

Likewise with a `list`, you get an empty list:

```
>>> d = defaultdict(list)
>>> d
defaultdict(<class 'list'>, {})
>>> d['foo'] += 'a'
>>> d['foo'] += 'b'
>>> d
defaultdict(<class 'list'>, {'foo': ['a', 'b']})
```

And so, with `set` you get an empty set to which you can add:

```
>>> d = defaultdict(set)
>>> d
defaultdict(<class 'set'>, {})
```



```
>>> d['foo'].add('a')
>>> d['foo'].add('b')
```

Note that the argument to `defaultdict` is *not* in quotes. You are passing the class `set` not the string `'set'`!

So, with all that, I end up adding words like so:

```
>>> words = defaultdict(set)
>>> word = 'apple'
>>> words[len(word)].add(word)
>>> word = 'bear'
>>> words[len(word)].add(word)
>>> words
defaultdict(<class 'set'>, {5: {'apple'}, 4: {'bear'}})
```

Identifying anagrams

In the intro to the problem, I mentioned my algorithm for finding an anagram:

1. Same length as the given word
2. Same frequency of characters as the given word

The first one is easy enough to find using `len`. If our given word is “listen,” then we only need to look at words of length 6 or less:

```
>>> given = 'listen'
>>> len(given)
6
```

How about the character frequency? There are many ways to find this, but I know of no easier method than to use the `Counter` from the `collections` module:

```
>>> from collections import Counter
>>> Counter('listen')
Counter({'l': 1, 'i': 1, 's': 1, 't': 1, 'e': 1, 'n': 1})
```

If we are looking at the word “tinsel,” we see that we have found an anagram:

```
>>> word = 'tinsel'
>>> len(given) == len(word)
True
>>> Counter(given) == Counter(word)
True
```

Selecting words to compare

My first implementation of this program was quite naive and yet worked fine for find all other single words that were anagrams. Everything came crashing down when I attempted to find combinations. I suddenly realized the number of 2-word combinations I needed to check (that 55 *billion* I mentioned before). As it happened, I then rewatched “The Imitation Game” about Alan Turing and the creation of his machine (“Christopher”) to crack the Enigma code which has a possible 150 million million possible states. He was unable to build a machine that could churn through that many possibilities in the 18 hours or so per day they had to find the right combination, so they had to find a way to cut down the number of combinations they attempted. Similarly, I realized I only needed to look at combinations of words whose lengths sum to the length of the given word; hence my decision to store **words** using the word length as the key and then as a **set** of words that length.

Next I needed to find all combinations of numbers that add up to that number. Let’s consider we are using “listen” as the **text**:

```
>>> text = 'listen'
>>> text_len = len(text)
>>> text_len
6
```

I need to do quite a few complex operations for which the **itertools** module provides very handy functions:

```
>>> from itertools import combinations, permutations, product, chain
```

First assume that we had words ranging from 1 to 10 characters in our word list file:

```
>>> lengths = list(range(1, 11))
>>> lengths
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

Now find all **combinations** of all the different lengths, I first need to **chain** the **lengths** to add it to itself:

```
>>> list(chain(lengths, lengths))
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

And then find the combinations of which there are many:

```
>>> len(list(combinations(chain(lengths, lengths), 2)))
190
>>> list(combinations(chain(lengths, lengths), 2))[:5]
[(1, 2), (1, 3), (1, 4), (1, 5), (1, 6)]
```

It turns out the list is longer than necessary because the tuples are unique, so we can fix that with a **set**:

```
>>> combos = combinations(chain(lengths, lengths), 2)
>>> uniq_combos = set(map(lambda t: tuple(sorted(t)), combos))
>>> len(uniq_combos)
55
>>> list(uniq_combos)[:3]
[(5, 9), (4, 7), (1, 3)]
```

And then find those where the sum is 6:

```
>>> list(filter(lambda t: sum(t) == 6, uniq_combos))
[(3, 3), (1, 5), (2, 4)]
```

If we put it all together and look for combinations of 3 numbers that sum to 6:

```
>>> n = 3
>>> text_len = 6
>>> key_combos = list(
...     filter(
...         lambda t: sum(t) == text_len,
...         set(
...             map(lambda t: tuple(sorted(t)),
...                 combinations(chain(lengths, lengths), n))))))
>>>
>>> key_combos
[(1, 1, 4), (1, 2, 3)]
```

Now I have the *keys* for the *words* to look to check for word combinations!

```
>>> key_combos
[(3, 3), (1, 5), (2, 4)]
```

Let's take the first combo:

```
>>> keys = key_combos[0]
>>> keys
(3, 3)
```

And pretend we have a very small *words* list:

```
>>> words = defaultdict(set)
>>> words[3].add('les')
>>> words[3].add('tin')
>>> words[4].add('lest')
>>> words[4].add('list')
>>> words[3].add('len')
>>> words[3].add('its')
>>> words
defaultdict(<class 'set'>, {3: {'len', 'its', 'tin', 'les'}, 4: {'list', 'lest'}})
```

I can map to find to find all the words for those lengths.

```
>>> list(map(lambda k: words[k], keys))
```

```
[{'len', 'its', 'tin', 'les'}, {'len', 'its', 'tin', 'les'}]
```

And then use `product` to get the Cartesian combination:

```
>>> word_combos = list(product(*list(map(lambda k: words[k], keys))))
>>> word_combos[:3]
[('len', 'len'), ('len', 'its'), ('len', 'tin')]
```

Which I can then iterate and apply my algorithm described above to decide if there are any anagrams:

```
>>> counts = Counter('listen')
>>> for t in word_combos:
...     if Counter(''.join(t)) == counts:
...         for s in [' '.join(l) for l in permutations(t)]:
...             if s != text:
...                 print(s)
...
len its
its len
its len
len its
tin les
les tin
les tin
tin les
```

Some are repeated which is why I chose to create my `anagrams` holder as a `set` to make them unique.

In the end, I look to see how many `anagrams` I found using `len(anagrams)`. If there are some, I report how many and what they are in `sorted` order; otherwise I let the user know that none were found.

Chapter 34: Hangman

Write a Python program called `hangman.py` that will play a game of Hangman which is a bit like “Wheel of Fortune” where you present the user with a number of elements indicating the length of a word. For our game, use the underscore `_` to indicate a letter that has not been guessed. The program should take `-n|--minlen` minimum length (default 5) and `-l|--maxlen` maximum length options (default 10) to indicate the minimum and maximum lengths of the randomly chosen word taken from the `-w|--wordlist` option (default `/usr/share/dict/words`). It also needs to take `-s|--seed` to for the random seed and the `-m|--misses` number of misses to allow the player.

To play, you will initiate an infinite loop and keep track of the game state, e.g., the word to guess, the letters already guessed, the letters found, the number of misses. As this is an interactive game, I cannot write an test suite, so you can play my version and then try to write one like it. If the user guesses a letter that is in the word, replace the `_` characters with the letter. If the user guesses the same letter twice, admonish them. If the user guesses a letter that is not in the word, increment the misses and let them know they missed. If the user guesses too many times, exit the game and insult them. If they correctly guess the word, let them know and exit the game.

```
$ ./hangman.py -h
usage: hangman.py [-h] [-l MAXLEN] [-n MINLEN] [-m MISSES] [-s SEED]
                  [-w WORDLIST]
```

Hangman

optional arguments:

```
-h, --help            show this help message and exit
-l MAXLEN, --maxlen MAXLEN
                        Max word length (default: 10)
-n MINLEN, --minlen MINLEN
                        Min word length (default: 5)
-m MISSES, --misses MISSES
                        Max number of misses (default: 10)
-s SEED, --seed SEED  Random seed (default: None)
-w WORDLIST, --wordlist WORDLIST
                        Word list (default: /usr/share/dict/words)
```

```
$ ./hangman.py
- - - - - (Misses: 0)
Your guess? ("?" for hint, "!" to quit) a
- - - - - (Misses: 1)
Your guess? ("?" for hint, "!" to quit) i
- - - - - i _ (Misses: 1)
Your guess? ("?" for hint, "!" to quit) e
```

```

_ _ _ _ _ i _ (Misses: 2)
Your guess? ("?" for hint, "!" to quit) o
_ o _ _ _ i _ (Misses: 2)
Your guess? ("?" for hint, "!" to quit) u
_ o _ _ _ i _ (Misses: 3)
Your guess? ("?" for hint, "!" to quit) y
_ o _ _ _ i _ (Misses: 4)
Your guess? ("?" for hint, "!" to quit) c
_ o _ _ _ i _ (Misses: 5)
Your guess? ("?" for hint, "!" to quit) d
_ o _ _ _ i _ (Misses: 6)
Your guess? ("?" for hint, "!" to quit) p
_ o _ _ _ i p (Misses: 6)
Your guess? ("?" for hint, "!" to quit) m
_ o _ _ _ i p (Misses: 7)
Your guess? ("?" for hint, "!" to quit) n
_ o _ _ _ i p (Misses: 8)
Your guess? ("?" for hint, "!" to quit) s
_ o s _ s _ i p (Misses: 8)
Your guess? ("?" for hint, "!" to quit) t
_ o s t s _ i p (Misses: 8)
Your guess? ("?" for hint, "!" to quit) h
You win. You guessed "hostship" with "8" misses!
$ ./hangman.py -m 2
_ _ _ _ _ _ _ _ _ (Misses: 0)
Your guess? ("?" for hint, "!" to quit) a
_ _ _ _ _ a _ _ a (Misses: 0)
Your guess? ("?" for hint, "!" to quit) b
_ _ _ _ _ a _ _ a (Misses: 1)
Your guess? ("?" for hint, "!" to quit) c
You lose, loser! The word was "metromania."

```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import os
5  import random
6  import re
7  import sys
8  from dire import die
9
10
11  # -----
12  def get_args():
13      """parse arguments"""
14      parser = argparse.ArgumentParser(
15          description='Hangman',
16          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18      parser.add_argument('-l',
19                          '--maxlen',
20                          help='Max word length',
21                          type=int,
22                          default=10)
23
24      parser.add_argument('-n',
25                          '--minlen',
26                          help='Min word length',
27                          type=int,
28                          default=5)
29
30      parser.add_argument('-m',
31                          '--misses',
32                          help='Max number of misses',
33                          type=int,
34                          default=10)
35
36      parser.add_argument('-s',
37                          '--seed',
38                          help='Random seed',
39                          type=str,
40                          default=None)
41
42      parser.add_argument('-w',
43                          '--wordlist',
```

```

44             help='Word list',
45             type=str,
46             default='/usr/share/dict/words')
47
48     return parser.parse_args()
49
50
51 # -----
52 def bail(msg):
53     """Print a message to STDOUT and quit with no error"""
54     print(msg)
55     sys.exit(0)
56
57 # -----
58
59 def main():
60     """main"""
61     args = get_args()
62     max_len = args.maxlen
63     min_len = args.minlen
64     max_misses = args.misses
65     wordlist = args.wordlist
66
67     random.seed(args.seed)
68
69     if not os.path.isfile(wordlist):
70         die('--wordlist "{}" is not a file.'.format(wordlist))
71
72     if min_len < 1:
73         die('--minlen must be positive')
74
75     if not 3 <= max_len <= 20:
76         die('--maxlen should be between 3 and 20')
77
78     if min_len > max_len:
79         die('--minlen ({}) is greater than --maxlen ({})'.format(
80             min_len, max_len))
81
82     good_word = re.compile('^([a-z]{' + str(min_len) + ',' + str(max_len) +
83                             '})$')
84     words = [w for w in open(wordlist).read().split() if good_word.match(w)]
85
86     word = random.choice(words)
87     play({'word': word, 'max_misses': max_misses})
88
89

```



```

90 # -----
91 def play(state):
92     """Loop to play the game"""
93     word = state.get('word') or ''
94
95     if not word: die('No word!')
96
97     guessed = state.get('guessed') or list('_' * len(word))
98     prev_guesses = state.get('prev_guesses') or set()
99     num_misses = state.get('num_misses') or 0
100    max_misses = state.get('max_misses') or 0
101
102    if ''.join(guessed) == word:
103        msg = 'You win. You guessed "{}" with "{}" miss{}!'
104        bail(msg.format(word, num_misses, '' if num_misses == 1 else 'es'))
105
106    if num_misses >= max_misses:
107        bail('You lose, loser! The word was "{}".'.format(word))
108
109    print('{} (Misses: {})'.format(' '.join(guessed), num_misses))
110    new_guess = input('Your guess? ("?" for hint, "!" to quit) ').lower()
111
112    if new_guess == '!':
113        bail('Better luck next time, loser.')
114    elif new_guess == '?':
115        new_guess = random.choice([x for x in word if x not in guessed])
116        num_misses += 1
117
118    if not re.match('[a-z]$', new_guess):
119        print('"{}" is not a letter'.format(new_guess))
120        num_misses += 1
121    elif new_guess in prev_guesses:
122        print('You already guessed that')
123    elif new_guess in word:
124        prev_guesses.add(new_guess)
125        last_pos = 0
126        while True:
127            pos = word.find(new_guess, last_pos)
128            if pos < 0:
129                break
130            elif pos >= 0:
131                guessed[pos] = new_guess
132                last_pos = pos + 1
133    else:
134        num_misses += 1
135

```

```
136     play({
137         'word': word,
138         'guessed': guessed,
139         'num_misses': num_misses,
140         'prev_guesses': prev_guesses,
141         'max_misses': max_misses
142     })
143
144
145 # -----
146 if __name__ == '__main__':
147     main()
```

Chapter 35: First Bank of Change

Write a Python program called `fboc.py` that will figure out all the different combinations of pennies, nickels, dimes, and quarters in a given **value** provided as a single positional argument. The value must be greater than 0 and less than or equal to 100.

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

First Bank of Change

positional arguments:
int Sum

```
optional arguments:
  -h, --help  show this help message and exit
$ ./fboc.py 0
usage: fboc.py [-h] int
fboc.py: error: value "0" must be > 0 and <= 100
$ ./fboc.py 124
usage: fboc.py [-h] int
fboc.py: error: value "124" must be > 0 and <= 100
$ ./fboc.py 1
If you give me 1 cent, I can give you:
  1: 1 penny
$ ./fboc.py 4
If you give me 4 cents, I can give you:
  1: 4 pennies
$ ./fboc.py 6
If you give me 6 cents, I can give you:
  1: 6 pennies
  2: 1 nickel, 1 penny
$ ./fboc.py 13
If you give me 13 cents, I can give you:
  1: 13 pennies
  2: 1 dime, 3 pennies
  3: 1 nickel, 8 pennies
  4: 2 nickels, 3 pennies
$ ./fboc.py 27
If you give me 27 cents, I can give you:
  1: 27 pennies
  2: 1 quarter, 2 pennies
```

3: 1 dime, 17 pennies
4: 2 dimes, 7 pennies
5: 1 nickel, 22 pennies
6: 1 dime, 1 nickel, 12 pennies
7: 2 dimes, 1 nickel, 2 pennies
8: 2 nickels, 17 pennies
9: 1 dime, 2 nickels, 7 pennies
10: 3 nickels, 12 pennies
11: 1 dime, 3 nickels, 2 pennies
12: 4 nickels, 7 pennies
13: 5 nickels, 2 pennies

Solution

```
1  #!/usr/bin/env python3
2  """Coin combos for value"""
3
4  import argparse
5  from itertools import product
6  from functools import partial
7
8
9  # -----
10 def get_args():
11     """Get command-line arguments"""
12
13     parser = argparse.ArgumentParser(
14         description='First Bank of Change',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('value', metavar='int', type=int, help='Sum')
18
19     args = parser.parse_args()
20
21     if not 0 < args.value <= 100:
22         parser.error('value "{}" must be > 0 and <= 100'.format(args.value))
23
24     return args
25
26
27 # -----
28 def main():
29     """Make a jazz noise here"""
30
31     args = get_args()
32     value = args.value
33     nickels = range((value // 5) + 1)
34     dimes = range((value // 10) + 1)
35     quarters = range((value // 25) + 1)
36     fig = partial(figure, value)
37     combos = [c for c in map(fig, product(nickels, dimes, quarters)) if c]
38
39     print('If you give me {} cent{}, I can give you:'.format(
40         value, '' if value == 1 else 's'))
41
42     for i, combo in enumerate(combos, 1):
43         print('{:3}: {}'.format(i, fmt_combo(combo)))
```

```

44
45
46 # -----
47 def fmt_combo(combo):
48     """English version of combo"""
49
50     out = []
51     for coin, val in zip(('quarter', 'dime', 'nickel', 'penny'), combo):
52         if val:
53             plural = 'pennies' if coin == 'penny' else coin + 's'
54             out.append('{} {}'.format(val, coin if val == 1 else plural))
55
56     return ', '.join(out)
57
58
59 # -----
60 def figure(value, coins):
61     """
62     If there is a valid combo of 'coins' in 'value',
63     return a tuple of ints for (quarters, dimes, nickels, pennies)
64     """
65
66     nickels, dimes, quarters = coins
67     big_coins = sum([5 * nickels, 10 * dimes, 25 * quarters])
68
69     if big_coins <= value:
70         return (quarters, dimes, nickels, value - big_coins)
71
72
73 # -----
74 if __name__ == '__main__':
75     main()

```

Discussion

Let's start with a short look at `get_args` where I've decided to move the validation of the single `value` argument into this function rather than getting the arguments in `main` and checking there. We can use `argparse` to ensure the user provides an `int` value, but there's no `type` to say that it must be in our desired range; however, I can use the `parser.error` function on line 22 to trigger the normal fail-with-usage behaviour we normally get from `argparse`. From the standpoint of the calling code on line 32, all the work to coerce and validate the user happens in `get_args`. If we make it past line 32, then all must have been good and we can just focus on the task at hand.

I'd like to mention that I worked for a couple of days on this solution. I tried many different approaches before settling on the way I solved this problem, so what I do next may not be at all how you solved the problem. My idea was to find how many possible nickels, dimes, and quarters are in the given `value` and then find every combination of those values to see which ones sum to the `value` or less. To do this, I can use the `//` operator to find the integer division of the `value` by each of 5, 10, and 25 for nickels, dimes, and quarters, e.g.:

```
>>> value = 13
>>> value // 5
2
```

Finds there are two nickels in 13 cents. I construct a range that includes 0, 1, and 2 like so:

```
>>> nickels = range((value // 5) + 1)
>>> nickels
range(0, 3)
>>> list(nickels)
[0, 1, 2]
```

I used the `itertools.product` function and three ranges for nickels, dimes, and quarters to find every possible combination of every number of coins

```
>>> dimes = range((value // 10) + 1)
>>> quarters = range((value // 25) + 1)
>>> from itertools import product
>>> list(product(nickels, dimes, quarters))
[(0, 0, 0), (0, 1, 0), (1, 0, 0), (1, 1, 0), (2, 0, 0), (2, 1, 0)]
```

I want to include 0 of every coin so that I can make up the remainder in pennies. Let's jump ahead to the `figure` function to see how I wanted to use these values. Because `product` gives me a list of 3-tuples, I decided to pass `figure` the `value` and then a `coins` tuple that I unpack on line 66. I `sum` the values of the `nickels`, `dimes`, and `quarters` on line 67 and see if that is less than or equal to the `value`. If so, I get the number of pennies by subtracting the sum of the larger coins and return a 4-tuple with the number of each coin. If the previous sum was

larger than the `value`, we don't bother defining the `return` of the function and so `None` is used.

Going back to line 37 where I want to call `figure` for each of the combinations returned by `product`, I use a list comprehension combined with a `map` which may seem rather dense but works quite well. The `map` wants a function and a list of items to apply the function. There's a slight problem in that the `figure` function wants 2 arguments – the `value` and the 3-tuple. I could have written the `map` using a `lambda`:

```
>>> def figure(value, coins):
...     nickels, dimes, quarters = coins
...     big_coins = sum([5 * nickels, 10 * dimes, 25 * quarters])
...     if big_coins <= value:
...         return (quarters, dimes, nickels, value - big_coins)
...
>>> list(map(lambda c: figure(value, c), product(nickels, dimes, quarters)))
[(0, 0, 0, 13), (0, 1, 0, 3), (0, 0, 1, 8), None, (0, 0, 2, 3), None]
```

But I thought it would be cleaner to create a partial application of the `figure` function with the `value` already bound. The `functools.partial` is exactly the tool we need and then we only need to pass in the 3-tuple of the coins:

```
>>> from functools import partial
>>> fig = partial(figure, value)
>>> fig((1,0,0))
(0, 0, 1, 8)
```

And so now I can use this `partial` function in my `map`:

```
>>> list(map(fig, product(nickels, dimes, quarters)))
[(0, 0, 0, 13), (0, 1, 0, 3), (0, 0, 1, 8), None, (0, 0, 2, 3), None]
```

Notice how we get some `None` values returned. Remember, this is because some of the combinations we are trying are too large, e.g., the maximum number of all the coins will be too large. So, to filter out those value, I can use a list comprehension with a guard at the end:

```
>>> combos = [c for c in map(fig, product(nickels, dimes, quarters)) if c]
>>> combos
[(0, 0, 0, 13), (0, 1, 0, 3), (0, 0, 1, 8), (0, 0, 2, 3)]
```

I could have used a `filter` for this, but it just doesn't seem to read as well:

```
>>> list(filter(lambda c: c, map(fig, product(nickels, dimes, quarters))))
[(0, 0, 0, 13), (0, 1, 0, 3), (0, 0, 1, 8), (0, 0, 2, 3)]
```

This is a list of 4-tuples representing the number of quarters, dimes, nickels, and pennies that will sum to 13. We still need to report back to the user, so that is the purpose of the `fmt_combo` function. Given that 4-tuple, I want to report, e.g., “1 quarter” or “3 dimes”, so I need to know the value of the denomination

and the singular/plural versions of name of the denomination. I use the `zip` function to pair the coin denominations with their values:

```
>>> combo = (0, 0, 0, 13)
>>> list(zip(('quarter', 'dime', 'nickel', 'penny'), combo))
[('quarter', 0), ('dime', 0), ('nickel', 0), ('penny', 13)]
```

The `plural` version of each name is made by adding `s` except for `penny`, so line 53 handles that. If the denomination is not in the `combo` (e.g., here we have only pennies), then we skip those by using `if val` where `val` will be the number of coins. The integer value 0 will evaluate to `False` in a Boolean context, so only those with a non-zero value will be included. I decided to create a `list` of the strings for each denomination, so I `append` to that list the `val` plus the correct singular or plural version of the name, finally returning that list joined on comma-space (`', '`).

Finally lines 39-43 are left to formatting the report to the user, being sure to provide feedback that includes the original `value` (“If you give me ...”) and an enumerated list of all the possible ways we could make change. The test suite does not bother to check the order in which you return the combinations, only that the correct number are present and they are in the correct format.

Chapter 36: Runny Babbit

Are you familiar with Spoonerisms where the initial consonant sounds of two words are switched? According to Wikipedia, they get their name from William Archibald Spooner who did this often. The author Shel Silverstein wrote a wonderful book called *Runny Babbit* (“bunny rabbit”) based on this. So, let’s write a Python program called `runny_babbit.py` that will read some text or an input file given as a single positional argument and finds neighboring words with initial consonant sounds to swap. As we’ll need to look at pairs of words and in such a way that it will make it difficult to remember the original formatting of the text, let’s also take a `-w|--width` (default 70) to format the output text to a maximum width.

As usual, the program should show usage with no arguments or for `-h|--help`:

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

Introduce Spoonerisms

positional arguments:

str	Input text or file
-----	--------------------

optional arguments:

-h, --help	show this help message and exit
-w int, --width int	Output text width (default: 70)

It should handle text from the command line:

```
$ ./runny_babbit.py 'the bunny rabbit'
the runny babbitt
```

Or a named file:

```
$ cat input1.txt
The bunny rabbit is cute.
$ ./runny_babbit.py input1.txt
The runny babbitt is cute.
```

We’ll use a set of “stop” words to prevent the switching of sounds when one of the words is in the following list:

before behind between beyond but by concerning despite down during following
for from into like near plus since that the through throughout to towards which
with within without

The results are endlessly entertaining:

```
$ ./runny_babbit.py ../inputs/preamble.txt
```

When, in the course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the laws of nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

Hints:

- You'll need to consider all the words in the input as pairs, like [(0, 1), (1, 2)] up to n (number of words) etc. How can you create such a list where instead of 0 and 1 you have the actual words, e.g., [('The', 'bunny'), ('bunny', 'rabbit')]?
- There are several exercises where we try to break words into initial consonant sounds and whatever else that follows. Can you reuse code from elsewhere? I'd recommend using regular expressions!
- Be sure you don't use a word more than once in a swap. E.g., in the phrase "the brown, wooden box", we'd skip "the" and consider the other two pairs of words ('brown', 'wooden') and ('wooden', 'box'). If we swap the first pair to make ('wown', 'brooden'), we would not want to consider the next pair because 'wooden' has already been used.
- Use the `textwrap` module to handle the formatting of the output text to a maximum `--width`

Solution

```
1  #!/usr/bin/env python3
2  """Spoonerisms"""
3
4  import argparse
5  import os
6  import re
7  import string
8  import textwrap
9
10
11  # -----
12  def get_args():
13      """Get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Introduce Spoonerisms',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('text',
20                          metavar='str',
21                          help='Input text or file')
22
23      parser.add_argument('-w',
24                          '--width',
25                          help='Output text width',
26                          metavar='int',
27                          type=int,
28                          default=70)
29
30      args = parser.parse_args()
31
32      if os.path.isfile(args.text):
33          args.text = open(args.text).read()
34
35      return args
36
37
38  # -----
39  def main():
40      """Make a jazz noise here"""
41
42      args = get_args()
43      text = args.text
```

```

44     words = text.split()
45     pairs = []
46
47     for k in range(len(words) - 1):
48         pairs.append((words[k], words[k+1]))
49
50     vowels = 'aeiouAEIOU'
51     consonants = ''.join([c for c in string.ascii_letters if c not in vowels])
52     regex = re.compile('^([' + consonants + ']+)([' + vowels + ']*\w)')
53     stop = set('before behind between beyond but by concerning'
54               'despite down during following for from into like near'
55               'plus since that the through throughout to towards'
56               'which with within without'.split())
57     skip = set()
58
59     for i, pair in enumerate(pairs):
60         w1, w2 = pair
61         if set([w1.lower(), w2.lower()]).intersection(stop):
62             continue
63
64         i1, i2 = i, i + 1
65         if i1 in skip or i2 in skip:
66             continue
67
68         m1 = regex.search(w1)
69         m2 = regex.search(w2)
70         if m1 and m2:
71             prefix1, suffix1 = m1.groups()
72             prefix2, suffix2 = m2.groups()
73             words[i1] = prefix2 + suffix1
74             words[i2] = prefix1 + suffix2
75             skip.add(i1)
76             skip.add(i2)
77
78     print('\n'.join(textwrap.wrap(' '.join(words), width=args.width)))
79
80     # -----
81     if __name__ == '__main__':
82         main()

```

Discussion



Figure 11: Also definitely not copyright infringement.

For this exercise, I thought I might move the logic to read an optionally named input *file* into the `get_args` function so that by the time I call `args = get_args()` the `args.text` really is just whatever “text” I need to consider, regardless if the source was the command line or a file. If I’m using `input1.txt`, then I essentially have this:

```
>>> text = open('input1.txt').read()
>>> text
'The bunny rabbit is cute.\n'
```

I need all the pairs of words, so that means I first need all the “words” which I’ll get by naively using `str.split` (that is, I won’t worry about punctuation and such):

```
>>> words = text.split()
>>> words
['The', 'bunny', 'rabbit', 'is', 'cute.']
```

Now I need all *pairs* of words which I can get by going from the zeroth word to the second to last word:

```
>>> pairs = []
>>> for k in range(len(words) - 1):
...     pairs.append((words[k], words[k+1]))
...
>>> pairs
[('The', 'bunny'), ('bunny', 'rabbit'), ('rabbit', 'is'), ('is', 'cute.')]

```

I need to find all the pairs where both words start with some consonant sounds and where neither of them is in my stop list, which I’ll create like so:

```
>>> stop = set('before behind between beyond but by concerning'
...           'despite down during following for from into like near'
...           'plus since that the through throughout to towards'
...           'which with within without'.split())
```

How will I find words that start with consonants? I can easily list all the vowels:

```
>>> vowels = 'aeiouAEIOU'
```

And then create the complement from `string.ascii_lowercase`:

```
>>> import string
>>> consonants = ''.join([c for c in string.ascii_letters if c not in vowels])
>>> consonants
'bcdfghjklmnpqrstvwxyzBCDFGHJKLMNPQRSTVWXYZ'
```

And then build a regular expression that looks for the start of a string `~` followed by a character class of all the `consonants` followed by the character class of `vowels` maybe followed by something else. I'll use parentheses `()` to capture both parts:

```
>>> import re
>>> regex = re.compile('^([' + consonants + ']+)([' + vowels + '].*)')
>>> regex.search('chair')
<re.Match object; span=(0, 5), match='chair'>
>>> regex.search('chair').groups()
('ch', 'air')
```

Now I can iterate over the `pairs`. First I check if either of the words is in the `stop` set by using the `set.intersection` function. For the first pair ('The', 'bunny') we see there is an intersection:

```
>>> w1 = 'The'
>>> w2 = 'bunny'
>>> set([w1.lower(), w2.lower()]).intersection(stop)
{'the'}
```

For the next pair, there is not:

```
>>> w1 = 'bunny'
>>> w2 = 'rabbit'
>>> set([w1.lower(), w2.lower()]).intersection(stop)
set()
```

The next check in my code is whether I've previously determined that I need to skip these words, so I have to know their positions in the original list. I decided to use `enumerate` over the `words` to get the number of the pair which will equal the position of the first word of each tuple in the original list of `words`.

Next I need to see if *both* words match my regular expression:

```
>>> m1 = regex.search(w1)
```

```
>>> m2 = regex.search(w2)
>>> m1
<re.Match object; span=(0, 5), match='bunny'>
>>> m2
<re.Match object; span=(0, 6), match='rabbit'>
```

They do! So I can use their `groups` to get the parts of each word to swap:

```
>>> m1.groups()
('b', 'unny')
>>> m2.groups()
('r', 'abbit')
>>> prefix1, suffix1 = m1.groups()
>>> prefix2, suffix2 = m2.groups()
```

This is the 2nd pair, so `i` would be equal to 1 in the actual code. I can use this to go mutate the words at positions `i` and `i + 1`:

```
>>> i = 1
>>> words[i] = prefix2 + suffix1
>>> words[i + 1] = prefix1 + suffix2
>>> words
['The', 'runny', 'babbit', 'is', 'cute.']
```

I need to be sure to add those positions to the `skip` set I created for the check that I discussed just above.

Finally we need to `print` the words back out, joining them on a blank and using `textwrap.wrap` with the `--width` argument to make it pretty:

```
>>> import textwrap
>>> print('\n'.join(textwrap.wrap(' '.join(words), width=70)))
The runny babbit is cute.
```


Chapter 37: Markov Chain

Write a Python program called `markov.py` that takes one or more text files as positional arguments for training. Use the `-n|--num_words` argument (default 2) to find clusters of words and the words that follow them, e.g., in “The Bustle” by Emily Dickinson:

```
The bustle in a house
The morning after death
Is solemnest of industries
Enacted upon earth,-
```

```
The sweeping up the heart,
And putting love away
We shall not want to use again
Until eternity.
```

If `n=1`, then we find that “The” can be followed by “bustle,” “morning,” and “sweeping.” There is a “the” followed by “heart,” but we’re not going to alter the text in any way, including removing punctuation, so just use `str.split` on the text to break up the words.

To begin your text, choose a random word (or words) that begin with an uppercase letter. Then randomly select the next word in the chain, keep track of the floating window of the `-n` words, and keep selecting the next words until you have matched or exceeded the `-l|--length` argument of the number of characters (default 500) to emit at which point you should stop when you find a word that terminates with `.`, `!`, or `?`.

If you use `str.split` to get the words from the training text, you’ll be removing any newlines from the text, so use a `-w|--text_width` argument (default 70) to introduce newlines in the output before the text exceeds that number of characters on the line. I recommend you use the `textwrap` module for this.

Because of the use of randomness, you should include a `-s|--seed` argument (default `None`) to pass to `random.seed`.

Occasionally you may chose a path that terminates. That is, in selecting the next word, you may find there is no next-next word. In that case, just exit the program.

My implementation includes a `-d|--debug` option that will write a `.log` file so you can inspect my data structures and logic as you write your own version.

You should find many diverse texts and use them all as training files with varying numbers for `-n` to see how the texts will be mixed. The results are endlessly entertaining.

```
$ ./markov.py
usage: markov.py [-h] [-l int] [-n int] [-s int] [-w int] [-d] FILE [FILE ...]
```

```
markov.py: error: the following arguments are required: FILE
$ ./markov.py -h
usage: markov.py [-h] [-l int] [-n int] [-s int] [-w int] [-d] FILE [FILE ...]
```

Markov Chain

positional arguments:

```
FILE           Training file(s)
```

optional arguments:

```
-h, --help           show this help message and exit
-l int, --length int Output length (characters) (default: 500)
-n int, --num_words int
                        Number of words (default: 2)
-s int, --seed int    Random seed (default: None)
-w int, --text_width int
                        Max number of characters per line (default: 70)
-d, --debug           Debug to ".log" (default: False)
```

```
$ ./markov.py ../inputs/const.txt -s 1
```

```
States, shall have no Vote, unless they shall meet in their respective
Numbers, which shall abridge the privileges or immunities of citizens
of the Militia to execute the Laws thereof, escaping into another,
shall, in the land and naval Forces; To provide for the loss or
emancipation of any slave; but all such Laws shall be bound thereby,
any Thing in the case wherein neither a President or Vice President
and Vice-President, or hold any office, civil or military, under the
United States; he may adjourn them to such Time as he shall have
failed to qualify, then the Vice-President chosen for the purpose
shall consist of a term to which the United States of America.
```

```
$ ./markov.py -s 2 ../inputs/dickinson.txt -w 30 -l 100
```

```
His knowledge to unfold On
what concerns our mutual mind,
The literature of old; What
interested scholars most, What
competitions ran When Plato
was a living girl, And
Beatrice wore The gown that
Dante deified.
```

Solution

```
1  #!/usr/bin/env python3
2  """Markov Chain"""
3
4  import argparse
5  import logging
6  import random
7  import textwrap
8  from pprint import pformat as pf
9  from collections import defaultdict
10
11
12  # -----
13  def get_args():
14      """Get command-line arguments"""
15
16      parser = argparse.ArgumentParser(
17          description='Markov Chain',
18          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
19
20      parser.add_argument('training',
21                          metavar='FILE',
22                          nargs='+',
23                          type=argparse.FileType('r'),
24                          help='Training file(s)')
25
26      parser.add_argument('-l',
27                          '--length',
28                          help='Output length (characters)',
29                          metavar='int',
30                          type=int,
31                          default=500)
32
33      parser.add_argument('-n',
34                          '--num_words',
35                          help='Number of words',
36                          metavar='int',
37                          type=int,
38                          default=2)
39
40      parser.add_argument('-s',
41                          '--seed',
42                          help='Random seed',
43                          metavar='int',
```

```

44             type=int,
45             default=None)
46
47     parser.add_argument('-w',
48                         '--text_width',
49                         help='Max number of characters per line',
50                         metavar='int',
51                         type=int,
52                         default=70)
53
54     parser.add_argument('-d',
55                         '--debug',
56                         help='Debug to ".log"',
57                         action='store_true')
58
59     return parser.parse_args()
60
61
62 # -----
63 def main():
64     """Make a jazz noise here"""
65
66     args = get_args()
67     char_max = args.length
68     random.seed(args.seed)
69     num_words = args.num_words
70
71     logging.basicConfig(
72         filename='.log',
73         filemode='w',
74         level=logging.DEBUG if args.debug else logging.CRITICAL)
75
76     training = read_training(args.training, num_words)
77     logging.debug('training = %s', pf(training))
78
79     # Find a word starting with a capital letter
80     words = list(
81         random.choice(
82             list(filter(lambda t: t[0][0].isupper(), training.keys()))))
83
84     logging.debug('starting with "%s"', words)
85     logging.debug(training[tuple(words)])
86
87     while True:
88         # get last two words
89         prev = tuple(words[-1 * num_words:])

```

```

90
91     # bail if dead end
92     if not prev in training:
93         break
94
95     new_word = random.choice(training[prev])
96     logging.debug('chose "{}" from {}'.format(new_word, training[prev]))
97     words.append(new_word)
98
99     # try to find ending punctuation if we've hit wanted char count
100    char_count = sum(map(len, words)) + len(words)
101    if char_count >= char_max and new_word[-1] in '!.?':
102        break
103
104    print('\n'.join(textwrap.wrap(' '.join(words), width=args.text_width)))
105    logging.debug('Finished')
106
107
108    # -----
109    def read_training(fhs, num_words):
110        """Read training files, return dict of chains"""
111
112        all_words = defaultdict(list)
113        for fh in fhs:
114            words = fh.read().split()
115
116            for i in range(0, len(words) - num_words):
117                l = words[i:i + num_words + 1]
118                all_words[tuple(l[:-1])].append(l[-1])
119
120        return all_words
121
122
123    # -----
124    if __name__ == '__main__':
125        main()

```

Discussion

As usual, I like to start my program by defining the options to my program with `get_args`. There will be one or more positional arguments which are `training` files, so I defined this argument with `narg='+'` and the `type=argparse.FileType('r')` so that `argparse` will validate the user input. Per the README, I define four other `int` arguments for the `--length` of the output, the `--num_words` in the patterns, the random `--seed`, and the `--text_width` of each line of output.

I also define a `--debug` option that will turn on `logging` to a `.log` file. Lines 71-74 initialize the `logging` module with `filemode='w'` so that it will overwrite an existing file and only emitting `DEBUG`-level messages if `--debug` is present; otherwise, only `CRITICAL` messages are shown and, since I have no calls to `logging.critical`, nothing will go into the logfile.

On line 76, I call a function to read the training files which I pass as a `list` as the first argument and the `args.num_words` as the second. While I could have put these few lines of code in the `main`, I prefer having short functions that do one thing. One of the hardest things to figure out for this program was the data structure I needed to represent a Markov chain. I settled on using a `dict` that would have as keys a tuple of word pairs and as values a list of words that follow that word pair. I call this `all_words` on line 114 and create it using the `collections.defaultdict(list)`. The advantage to `defaultdict` is that keys are created automatically using a default value for the indicated data type – an empty string for `str`, the value 0 for `int`, and the empty list `[]` for `list`. (If you're into category theory, these are the “empty” values for the monoids of strings, integers, and lists.)

On line 115, I iterate over the file handles that `argparse` opened for me. Note I call each file handle `fh` and the list of file handles `fhs` (the “plural” of `fh`). On line 116, the call to `fh.read().split()` will read the *entire* file and split it into “words” which I quote because I'm specifically not removing any sort of punctuation as I decided to follow the example in the Kernighan/Pike book where quotes and punctuation from the original text will determine the same kinds of patterns in the resulting text. Of course, this can lead to mismatched quotes and randomly distributed punctuation, but *c'est la vie*.

To create the chains, I want to select continuous sequences of words of length `--num_words` plus the word that follows. So, if `--num_words` is 1, I will use the first word as my key, then look ahead at the next word as a choice of what can come next. Given a phrase like “The Lion, The Witch and The Wardrobe”, we can see that “The” may be followed by either “Lion,” “Witch,” or “Wardrobe.” To write this in code, we use the same idea as extracting k-mers from a word but instead think of “mers” as words and select “k-words” from a string:

```
>>> words = 'The Lion, The Witch and The Wardrobe'.split()
>>> words
```

```

['The', 'Lion,', 'The', 'Witch', 'and', 'The', 'Wardrobe']
>>> from collections import defaultdict
>>> all_words = defaultdict(list)
>>> num_words = 1
>>> for i in range(0, len(words) - num_words):
...     l = words[i:i + num_words + 1]
...     all_words[tuple(l[:-1])].append(l[-1])

```

In the resulting `all_words` structure, we see that ‘The’ has the expected three options:

```

>>> from pprint import pprint as pp
>>> pp(all_words)
defaultdict(<class 'list'>,
            {('Lion,',): ['The'],
             ('The',): ['Lion,', 'Witch', 'Wardrobe'],
             ('Witch',): ['and'],
             ('and',): ['The']})

```

In creating the list of options, I chose not to unique the values. Some texts may have the same word following a given sequence many times which will result in that word being randomly selected more often, but this is a consequence of the training data influencing the outcome. If you used a `set` instead of a `list`, you would lose that influence. The data structure matters!

On line 80, I need to find a place to start. I use `random.choice` to select from the list of `training` words that start with a capital letter. I can `filter` the keys of the `training` dict which you should recall are tuples. In the `lambda`, I reference `t[0][0]` (t for “tuple”) to index the zeroth element in the tuple and then the zeroth character of that word. This will return a `str` object which I can use to call the `isupper` method to tell me if the character is an uppercase letter. Remember that `filter` will only allow to pass those elements for which the `lambda` evaluates to `True`.

The `while True` begins the actual generation of text. I get the previous `num_words` by multiplying that value by `-1` and indexing from the end of the `words` list. I need to turn that list into a `tuple` to use for the key to `training`. (A note here that you cannot use a `list` as a key to a `dict` because it’s not immutable whereas strings and tuples are.) I need to `break` out of the loop if I happened down deadend; otherwise, I can use `random.choice` again to select a new word from the list of options and `append` that to the `words` I’ve selected already.

To figure out if we’ve gone far enough, I need to count up how many characters I’ve got in `words`, so I `map` the `words` into the `len` function and `sum` them up:

```

>>> words
['The', 'Lion,', 'The', 'Witch', 'and', 'The', 'Wardrobe']
>>> sum(map(len, words))

```

But there will be spaces in between each word, so I account for them by adding on the `len(words)`. If I have matched or exceeded the `char_max`, then I need to find a stopping point by looking to see if the `new_word` I've selected ends with an ending punctuation like `.`, `!`, or `?`. If so, we `break` out of the loop.

At this point, the `words` list needs to be turned into text. It would be ugly to just `print` out one long string, so I use the `textwrap.wrap` to break up the long string into lines that are no longer than the given `text_width`. That function returns a list of lines that need to be joined on newlines to print.

Chapter 38: Hamming Chain

Write a Python program called `chain.py` that takes a `-s|--start` word and searches a `-w|--wordlist` argument (default `/usr/local/share/dict`) for words no more than `-d|--max_distance` Hamming distance for some number of `-i|--iteration` (default 20). Be sure to accept a `-S|--seed` for `random.seed`.

If the given word is not found in the word list, exit with an error and message. While searching for the next word in the chain, be sure not to repeat any words previously found or you might just go in circles! If you fail to find any new words before the end of the iterations, exit with an error and message as such.

```
$ ./chain.py -h
usage: chain.py [-h] [-s START] [-w FILE] [-d int] [-i int] [-S int] [-D]
```

Hamming chain

optional arguments:

```
-h, --help            show this help message and exit
-s START, --start START
                        Starting word (default: )
-w FILE, --wordlist FILE
                        File input (default: /usr/share/dict/words)
-d int, --max_distance int
                        Maximum Hamming distance (default: 1)
-i int, --iterations int
                        Random seed (default: 20)
-S int, --seed int    Random seed (default: None)
-D, --debug           Debug (default: False)
```

```
$ ./chain.py -s foobar
```

Unknown word "foobar"

```
$ ./chain.py -s bike -S 1 -i 5
```

```
1: bike
2: bikh
3: Sikh
4: sith
5: sithe
```

```
$ ./chain.py -s bike -S 1 -i 5 -d 2
```

```
1: bike
2: bit
3: net
4: yot
5: ye
```

```
$ ./chain.py -S 1 -s bicycle
```

Failed to find more words!

```
1: bicycle
2: bicycler
$ ./chain.py -S 1 -s bicycle -d 2 -i 5
1: bicycle
2: bicyclic
3: bicyclism
4: dicyclist
5: bicyclist
```

Use the `uscities.txt` file to plan a trip!

```
$ ./chain.py -S 1 -w ../inputs/uscities.txt -s Clinton -d 3
1: Clinton
2: Flint
3: Fritz
4: Unity
5: Union
6: Mason
7: Oasis
8: Nash
9: Zag
10: Guy
11: Gaza
12: Jay
13: Ely
14: Egan
15: Aden
16: Alta
17: Ada
18: Nyac
19: Pyatt
20: Plato
$ ./chain.py -S 1 -w ../inputs/uscities.txt -s 'Calumet City' -d 4
Failed to find more words!
1: Calumet City
2: Calumet Park
3: Palomar Park
4: Hanover Park
5: Langley Park
6: Stanley Park
7: Kearney Park
```

Solution

```
1  #!/usr/bin/env python3
2  """Hamming chain"""
3
4  import argparse
5  import logging
6  import random
7  import re
8  from dire import die, warn
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14
15      parser = argparse.ArgumentParser(
16          description='Hamming chain',
17          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
18
19      parser.add_argument('-s', '--start', type=str, help='Starting word', default='')
20
21      parser.add_argument('-w',
22                          '--wordlist',
23                          metavar='FILE',
24                          type=argparse.FileType('r'),
25                          help='File input',
26                          default='/usr/share/dict/words')
27
28      parser.add_argument('-d',
29                          '--max_distance',
30                          metavar='int',
31                          type=int,
32                          help='Maximum Hamming distance',
33                          default=1)
34
35      parser.add_argument('-i',
36                          '--iterations',
37                          metavar='int',
38                          type=int,
39                          help='Random seed',
40                          default=20)
41
42      parser.add_argument('-S',
43                          '--seed',
```

```

44             metavar='int',
45             type=int,
46             help='Random seed',
47             default=None)
48
49     parser.add_argument('-D', '--debug', help='Debug', action='store_true')
50
51     return parser.parse_args()
52
53
54 # -----
55 def dist(s1, s2):
56     """Given two strings, return the Hamming distance (int)"""
57
58     return abs(len(s1) - len(s2)) + sum(
59         map(lambda p: 0 if p[0] == p[1] else 1, zip(s1.lower(), s2.lower())))
60
61
62 # -----
63 def main():
64     """Make a jazz noise here"""
65
66     args = get_args()
67     start = args.start
68     fh = args.wordlist
69     distance = args.max_distance
70
71     random.seed(args.seed)
72
73     logging.basicConfig(
74         filename='.log',
75         filemode='w',
76         level=logging.DEBUG if args.debug else logging.CRITICAL)
77
78     logging.debug('file = %s', fh.name)
79
80     words = fh.read().splitlines()
81
82     if not start:
83         start = random.choice(words)
84
85     if not start in words:
86         die('Unknown word "{}".format(start))
87
88     def find_close(word):
89         l = len(word)

```

```

90         low, high = l - distance, l + distance
91         test = filter(lambda w: low <= len(w) <= high, words)
92         return filter(lambda w: dist(word, w) <= distance, test)
93
94     chain = [start]
95     for _ in range(args.iterations - 1):
96         close = list(filter(lambda w: w not in chain, find_close(chain[-1])))
97         if not close:
98             warn('Failed to find more words!')
99             break
100
101     next_word = random.choice(close)
102     chain.append(next_word)
103
104     for i, link in enumerate(chain, start=1):
105         print('{:3}: {}'.format(i, link))
106
107     # -----
108     if __name__ == '__main__':
109         main()

```

Chapter 39: Morse Encoder/Decoder

Write a Python program called `morse.py` that will encrypt/decrypt text to/from Morse code. The program should expect a single positional argument which is either the name of a file to read for the input or the character `-` to indicate reading from STDIN. The program should also take a `-c|--coding` option to indicate use of the `itu` or standard `morse` tables, `-o|--outfile` for writing the output (default STDOUT), and a `-d|--decode` flag to indicate that the action is to decode the input (the default is to encode it).

```
$ ./morse.py
usage: morse.py [-h] [-c str] [-o str] [-d] [-D] FILE
morse.py: error: the following arguments are required: FILE
$ ./morse.py -h
usage: morse.py [-h] [-c str] [-o str] [-d] [-D] FILE
```

Encode and decode text/Morse

positional arguments:

FILE Input file or "-" for stdin

optional arguments:

```
-h, --help            show this help message and exit
-c str, --coding str  Coding version (default: itu)
-o str, --outfile str Output file (default: None)
-d, --decode          Decode message from Morse to text (default: False)
-D, --debug           Debug (default: False)
```

```
$ ./morse.py ../inputs/fox.txt
```

```
[cholla@~/work/python/playful_python/morse]$ ./morse.py ./inputs/fox.txt | ./morse.py -d -
THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
```

Solution

```
1  #!/usr/bin/env python3
2  """Morse en/decoder"""
3
4  import argparse
5  import logging
6  import random
7  import re
8  import string
9  import sys
10
11
12  # -----
13  def get_args():
14      """Get command-line arguments"""
15
16      parser = argparse.ArgumentParser(
17          description='Encode and decode text/Morse',
18          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
19
20      parser.add_argument('input',
21                          metavar='FILE',
22                          help='Input file or "-" for stdin')
23
24      parser.add_argument('-c',
25                          '--coding',
26                          help='Coding version',
27                          metavar='str',
28                          type=str,
29                          choices=['itu', 'morse'],
30                          default='itu')
31
32      parser.add_argument('-o',
33                          '--outfile',
34                          help='Output file',
35                          metavar='str',
36                          type=str,
37                          default=None)
38
39      parser.add_argument('-d',
40                          '--decode',
41                          help='Decode message from Morse to text',
42                          action='store_true')
43
```

```

44     parser.add_argument('-D', '--debug', help='Debug', action='store_true')
45
46     return parser.parse_args()
47
48
49 # -----
50 def encode_word(word, table):
51     """Encode word using given table"""
52
53     coded = []
54     for char in word.upper():
55         logging.debug(char)
56         if char != ' ' and char in table:
57             coded.append(table[char])
58
59     encoded = ' '.join(coded)
60     logging.debug('encoding "{}" to "{}"'.format(word, encoded))
61
62     return encoded
63
64
65 # -----
66 def decode_word(encoded, table):
67     """Decode word using given table"""
68
69     decoded = []
70     for code in encoded.split(' '):
71         if code in table:
72             decoded.append(table[code])
73
74     word = ' '.join(decoded)
75     logging.debug('dedoding "{}" to "{}"'.format(encoded, word))
76
77     return word
78
79
80 # -----
81 def test_encode_word():
82     """Test Encoding"""
83
84     assert encode_word('sos', ENCODE_ITU) == '... --- ...'
85     assert encode_word('sos', ENCODE_MORSE) == '... .. ...'
86
87
88 # -----
89 def test_decode_word():

```



```

90     """Test Decoding"""
91
92     assert decode_word('... --- ...', DECODE_ITU) == 'SOS'
93     assert decode_word('... .,. ...', DECODE_MORSE) == 'SOS'
94
95
96     # -----
97     def test_roundtrip():
98         """Test En/decoding"""
99
100         random_str = lambda: ''.join(random.sample(string.ascii_lowercase, k=10))
101         for _ in range(10):
102             word = random_str()
103             for encode_tbl, decode_tbl in [(ENCODE_ITU, DECODE_ITU),
104                                           (ENCODE_MORSE, DECODE_MORSE)]:
105
106                 assert word.upper() == decode_word(encode_word(word, encode_tbl),
107                                                    decode_tbl)
108
109
110     # -----
111     def main():
112         """Make a jazz noise here"""
113         args = get_args()
114         action = 'decode' if args.decode else 'encode'
115         output = open(args.outfile, 'wt') if args.outfile else sys.stdout
116         source = sys.stdin if args.input == '-' else open(args.input)
117
118         coding_table = ''
119         if args.coding == 'itu':
120             coding_table = ENCODE_ITU if action == 'encode' else DECODE_ITU
121         else:
122             coding_table = ENCODE_MORSE if action == 'encode' else DECODE_MORSE
123
124         logging.basicConfig(
125             filename='.log',
126             filemode='w',
127             level=logging.DEBUG if args.debug else logging.CRITICAL)
128
129         word_split = r'\s+' if action == 'encode' else r'\s{2}'
130
131         for line in source:
132             for word in re.split(word_split, line):
133                 if action == 'encode':
134                     print(encode_word(word, coding_table), end=' ')
135                 else:

```

```

136         print(decode_word(word, coding_table), end=' ')
137     print()
138
139
140 # -----
141 def invert_dict(d):
142     """Invert a dictionary's key/value"""
143
144     #return dict(map(lambda t: list(reversed(t)), d.items()))
145     return dict([(v, k) for k, v in d.items()])
146
147
148 # -----
149 # GLOBALS
150
151 ENCODE_ITU = {
152     'A': '.-.', 'B': '-...', 'C': '-.-.-', 'D': '-.-.', 'E': '. ', 'F': '-.-.-',
153     'G': '-.-.-', 'H': '....', 'I': '...', 'J': '-.-.-', 'K': '-.-.-', 'L': '-.-.-',
154     'M': '-.-.-', 'N': '-.-.', 'O': '---', 'P': '-.-.-', 'Q': '---.-', 'R': '-.-.-',
155     'S': '...-', 'T': '-.', 'U': '...-', 'V': '...-.-', 'W': '-.-.-', 'X': '-.-.-',
156     'Y': '-.-.-', 'Z': '-.-.-', '0': '-----', '1': '-.-.-.-', '2': '-.-.-.-', '3':
157     '-.-.-.-', '4': '-.-.-.-', '5': '.....', '6': '-.-.-.-', '7': '-.-.-.-', '8':
158     '-.-.-.-', '9': '-----', ',': '-.-.-.-', ':': '-.-.-.-', '?': '-.-.-.-', '!':
159     '-.-.-.-', '&': '-.-.-.-', ';': '-.-.-.-', ':': '-.-.-.-', '"': '-.-.-.-', '/':
160     '-.-.-.-', '-': '-.-.-.-', '(': '-.-.-.-', ')': '-.-.-.-',
161 }
162
163 ENCODE_MORSE = {
164     'A': '.-.', 'B': '-...', 'C': '....', 'D': '-.-.', 'E': '. ', 'F': '-.-.-', 'G':
165     '-.-.-', 'H': '....', 'I': '...', 'J': '-.-.-', 'K': '-.-.-', 'L': '+', 'M':
166     '-.-.-', 'N': '-.-.', 'O': '...', 'P': '.....', 'Q': '-.-.-', 'R': '.,..', 'S':
167     '...-', 'T': '-.', 'U': '...-', 'V': '...-.-', 'W': '-.-.-', 'X': '-.-.-', 'Y':
168     '.,..', 'Z': '....', '0': '+++++', '1': '-.-.-', '2': '-.-.-', '3':
169     '-.-.-', '4': '-.-.-', '5': '---', '6': '.....', '7': '-.-.-', '8':
170     '-.-.-', '9': '-.-.-', ',': '-.-.-.-', ':': '-.-.-.-', '?': '-.-.-.-', '!':
171     '-.-.-.-', '&': '.,..', ';': '.,..', ':': '-.-.-.-', '"': '-.-.-.-', '/':
172     '...-', '-': '....,.-.-', '(': '.....,.-.-', ')': '.....,.-.-',
173 }
174
175 DECODE_ITU = invert_dict(ENCODE_ITU)
176 DECODE_MORSE = invert_dict(ENCODE_MORSE)
177
178 # -----
179 if __name__ == '__main__':
180     main()

```

Chapter 40: ROT13 (Rotate 13)

Write a Python program called `rot13.py` that will encrypt/decrypt input text by shifting the text by a given `-s|--shift` argument or will move each character halfway through the alphabet, e.g., “a” becomes “n,” “b” becomes “o,” etc. The text to rotate should be provided as a single positional argument to your program and can either be a text file, text on the command line, or `-` to indicate STDIN so that you can round-trip data through your program to ensure you are encrypting and decrypting properly.

The way I approached the solution is to think of adding time. If it’s 8 in the morning and I want to know the time in 6 hours on a 12-hour (not military/24-hour) clock, I need to think in terms of 12 when the clock rolls over from AM to PM. To do that, I need to know the remainder of dividing by 12, which is given by the modulus `%` operator:

```
>>> now = 8
>>> (now + 6) % 12
2
```

And 6 hours from 8AM is, indeed, 2PM.

Similarly if I want to know how many hours (in decimal) are a particular number of minutes, I need to mod by 60:

```
>>> minutes = 90
>>> int(minutes / 60) + (minutes % 60) / 60
1.5
>>> minutes = 204
>>> int(minutes / 60) + (minutes % 60) / 60
3.4
```

If you `import string`, you can see all the lower/uppercase letters

```
>>> import string
>>> string.ascii_lowercase
'abcdefghijklmnopqrstuvwxyz'
>>> string.ascii_uppercase
'ABCDEFGHIJKLMNOPQRSTUVWXYZ'
```

So I think about “rot13” like adding 13 (or some other shift interval) to the position of the letter in the list and modding by the length of the list to wrap it around. If the shift is 13 and we are at “a” and want to know what the letter 13 way is, we can use `pos` to find “a” and add 13 to that:

```
>>> lcase = list(string.ascii_lowercase)
>>> lcase.index('a')
0
>>> lcase[lcase.index('a') + 13]
'n'
```

But if we want to know the value for something after the 13th letter in our list, we are in trouble!

```
>>> lcase[lcase.index('x') + 13]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
IndexError: list index out of range
```

% to the rescue!

```
>>> lcase[(lcase.index('x') + 13) % len(lcase)]
'k'
```

It's not necessary in this algorithm to shift by any particular number. 13 is special because it's halfway through the alphabet, but we could shift by just 2 or 5 characters. If we want to round-trip our text, it's necessary to shift in the opposite direction on the second half of the trip, so be sure to use the negative value there!

```
$ ./rot13.py
usage: rot13.py [-h] [-s int] str
rot13.py: error: the following arguments are required: str
$ ./rot13.py -h
usage: rot13.py [-h] [-s int] str
```

Argparse Python script

positional arguments:

str	Input text, file, or "-" for STDIN
-----	------------------------------------

optional arguments:

-h, --help	show this help message and exit
-s int, --shift int	Shift arg (default: 0)

```
$ ./rot13.py AbCd
```

```
NoPq
```

```
$ ./rot13.py AbCd -s 2
```

```
CdEf
```

```
$ ./rot13.py fox.txt
```

```
Gur dhvpx oebja sbk whzcf bire gur ynml qbt.
```

```
$ ./rot13.py fox.txt | ./rot13.py -
```

```
The quick brown fox jumps over the lazy dog.
```

```
$ ./rot13.py -s 3 fox.txt | ./rot13.py -s -3 -
```

```
The quick brown fox jumps over the lazy dog.
```

Solution

```
1  #!/usr/bin/env python3
2
3  import argparse
4  import os
5  import re
6  import string
7  import sys
8
9
10 # -----
11 def get_args():
12     """get command-line arguments"""
13     parser = argparse.ArgumentParser(
14         description='ROT13 encryption',
15         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
16
17     parser.add_argument('text',
18                         metavar='str',
19                         help='Input text, file, or "-" for STDIN')
20
21     parser.add_argument('-s',
22                         '--shift',
23                         help='Shift arg',
24                         metavar='int',
25                         type=int,
26                         default=0)
27
28     return parser.parse_args()
29
30
31 # -----
32 def main():
33     """Make a jazz noise here"""
34     args = get_args()
35     text = args.text
36
37     if text == '-':
38         text = sys.stdin.read()
39     elif os.path.isfile(text):
40         text = open(text).read()
41
42     lcase = list(string.ascii_lowercase)
43     ucase = list(string.ascii_uppercase)
```

```

44     num_lcase = len(lcase)
45     num_ucase = len(ucase)
46     lcase_shift = args.shift or int(num_lcase / 2)
47     ucase_shift = args.shift or int(num_ucase / 2)
48
49     def rot13(char):
50         if char in lcase:
51             pos = lcase.index(char)
52             rot = (pos + lcase_shift) % num_lcase
53             return lcase[rot]
54         elif char in ucase:
55             pos = ucase.index(char)
56             rot = (pos + ucase_shift) % num_ucase
57             return ucase[rot]
58         else:
59             return char
60
61     print(''.join(map(rot13, text)).rstrip())
62
63
64     # -----
65     if __name__ == '__main__':
66         main()

```

Chapter 41: Tranpose ABC Notation

Write a Python program called `transpose.py` that will read a file in ABC notation (https://en.wikipedia.org/wiki/ABC_notation) and transpose the melody line up or down by a given `-s|--shift` argument. Like the `rot13` exercise, it might be helpful to think of the space of notes (ABCDEFG) as a list which you can roll through. For instance, if you have the note `c` and want to transpose up a (minor) third (`-s 3`), you would make the new note `e`; similarly if you have the note `F` and you go up a (major) third, you get `A`. You will not need to worry about the actual number of semitones that you are being asked to shift, as the previous example showed that we might be shifting by a major/minor/augmented/diminished/pure interval. The purpose of the exercise is simply to practice with lists.

```
$ ./transpose.py
usage: transpose.py [-h] [-s int] FILE
transpose.py: error: the following arguments are required: FILE
$ ./transpose.py -h
usage: transpose.py [-h] [-s int] FILE
```

Tranpose ABC notation

positional arguments:

FILE	Input file
------	------------

optional arguments:

-h, --help	show this help message and exit
-s int, --shift int	Interval to shift (default: 2)

```
$ ./transpose.py foo
"foo" is not a file
$ ./transpose.py songs/legacy.abc -s 1
--shift "1" must be between 2 and 8
$ ./transpose.py songs/legacy.abc
<score lang="ABC">
X:1
T:The Legacy Jig
M:6/8
L:1/8
R:jig
K:A
AGA CBC | aga abc | AGA CBC | e2B BGE |
AGA CBC | aga abc | baf feC |1 eCB BGE :|2 eCB BCe |:
fgf feC | eCB BCe | fgf feC | aeC BCe |
fgf feC | e2e efg | agf feC |1 eCB BCe :|2 eCB BGE |]
</score>
```

A sample ABC song is given:

```
$ cat songs/legacy.abc
<score lang="ABC">
X:1
T:The Legacy Jig
M:6/8
L:1/8
R:jig
K:G
GFG BAB | gfg gab | GFG BAB | d2A AFD |
GFG BAB | gfg gab | age edB |1 dBA AFD :|2 dBA ABd |:
efe edB | dBA ABd | efe edB | gdB ABd |
efe edB | d2d def | gfe edB |1 dBA ABd :|2 dBA AFD |]
</score>
```

If you use `new_py.py` to create your new program with the `file` as a single positional argument, you can use this code to get the input file and check that it is, indeed, a file:

```
args = get_args()
file = args.file

if not os.path.isfile(file):
    die("{} is not a file".format(file))
```

Now that you have a file, you can use a `for` loop to read it. Each line will still have a newline attached to the end, so you can use `rstrip()` to remove it:

```
for line in open(file):
    line = line.rstrip()
```

If a line starts with `<` and ends with `>` (cf. `str.startswith` and `str.endswith`), you can just print the line as-is. If the line starts with `K:`, then you have the key signature and should transpose it, e.g., if you have `K:A` and you are shifting a fifth, you should print `K:E`. If you have a line that starts with any other single uppercase letter and a colon, just print the line as-is. Finally, if you have a line that doesn't match any of the above conditions, you have a line of melody that needs to be transposed.

If you are unfamiliar with musical transposition, you may be a bit confused by the notion of an interval. A “second” equals a `--shift` of one note; that is, the distance from `A` to `B` is one note, but we call that a “second.” Therefore, assume that the `--shift` argument is the name of the interval, e.g., `4` (a “fourth”) is actually a move of three notes. That means the argument provided by the user should be in the range 2 to 8, inclusive, so complain if it is not.

Note that the transposition of a tune up a fourth is the same as down a fifth:

```
$ ./transpose.py songs/legacy.abc -s 4
```



```

<score lang="ABC">
X:1
T:The Legacy Jig
M:6/8
L:1/8
R:jig
K:C
CBC EDE | cbc cde | CBC EDE | g2D DBG |
CBC EDE | cbc cde | dca agE |1 gED DBG :|2 gED DEg |:
aba agE | gED DEg | aba agE | cgE DEg |
aba agE | g2g gab | cba agE |1 gED DEg :|2 gED DBG |]
</score>
$ ./transpose.py songs/legacy.abc -s -5
<score lang="ABC">
X:1
T:The Legacy Jig
M:6/8
L:1/8
R:jig
K:C
CBC EDE | cbc cde | CBC EDE | g2D DBG |
CBC EDE | cbc cde | dca agE |1 gED DBG :|2 gED DEg |:
aba agE | gED DEg | aba agE | cgE DEg |
aba agE | g2g gab | cba agE |1 gED DEg :|2 gED DBG |]
</score>

```

Solution

```
1  #!/usr/bin/env python3
2  """Tranpose ABC notation"""
3
4  import argparse
5  import os
6  import re
7  import sys
8  from dire import die
9
10
11  # -----
12  def get_args():
13      """get command-line arguments"""
14      parser = argparse.ArgumentParser(
15          description='Tranpose ABC notation',
16          formatter_class=argparse.ArgumentDefaultsHelpFormatter)
17
18      parser.add_argument('file', metavar='FILE', help='Input file')
19
20      parser.add_argument('-s',
21                          '--shift',
22                          help='Interval to shift',
23                          metavar='int',
24                          type=int,
25                          default=2)
26
27      return parser.parse_args()
28
29  # -----
30  def main():
31      """Make a jazz noise here"""
32      args = get_args()
33      file = args.file
34      shift = args.shift
35      ucase = 'ABCDEFG'
36      lcase = 'abcdefg'
37      num_notes = 7
38
39
40      if not 1 < abs(shift) <= 8:
41          die('--shift "{}" must be between 2 and 8'.format(shift))
42
43      if not os.path.isfile(file):
```

```

44         die("{} is not a file".format(file))
45
46     # account for interval where a 2nd (-s 2) is a move of one note
47     shift = shift - 1 if shift > 0 else shift + 1
48
49     def transpose(note):
50         if note in lcase:
51             pos = lcase.index(note)
52             tran = (pos + shift) % num_notes
53             return lcase[tran]
54         elif note in ucase:
55             pos = ucase.index(note)
56             tran = (pos + shift) % num_notes
57             return ucase[tran]
58         else:
59             return note
60
61     for line in open(file):
62         line = line.rstrip()
63
64         if line.startswith('K:'):
65             key = line[2]
66             print('K:' + transpose(key))
67         elif (line.startswith('<') and line.endswith('>')) or re.match(
68             '[A-Z]:\s?', line):
69             print(line)
70         else:
71             for char in line.rstrip():
72                 print(transpose(char), end='')
73
74             print()
75
76
77     # -----
78     if __name__ == '__main__':
79         main()

```

Chapter 42: Word Search

Write a Python program called `search.py` that takes a file name as the single positional argument and finds the words hidden in the puzzle grid.

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

Word search

positional arguments:
FILE The puzzle

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

If given a non-existent file, it should complain and exit with a non-zero status:

```
$ ./search.py lkdfak
usage: search.py [-h] FILE
search.py: error: argument FILE: can't open 'lkdfak': [Errno 2] No such file or directory:
```

The format of the puzzle file will be a grid of letters followed by an empty line followed by a list of words to find delimited by newlines, e.g.:

```
$ cat puzzle06.txt
ABC
DEF
GHI
```

DH

If the input grid is uneven, the program should error out:

```
$ cat bad_grid.txt
ABC
DEFG
HIJ
```

XYZ

```
$ ./search.py bad_grid.txt
Uneven number of columns
```

The output should be the input puzzle with only the letters showing for the words that are found replacing all the other letters with . (a period):

```
$ ./search.py puzzle06.txt
```

```

...
D..
.H.
$ cat ice_cream.txt
YMTRLCHOCOLATE
ASKCARTESOOMET
PYVANILLASNOTE
MKDETDEACFANAA
CATNLINNAOCCOE
OKPOAAGODKEAET
ECULNCAEFOPLRN
DOTAEENORYWEEE
OCBOAWYOTTEOIE
COIEAAARTSAOAR
RNTTCRALETNIAG
EEGDUFOSNIOVLT
DAORYKCORUACGT
AEETUNOCOCTPES

COTTON CANDY
MAPLE WALNUT
PECAN
BANANA
TIGER TAIL
MOOSE TRACKS
COCONUT
ROCKY ROAD
GREEN TEA
FUDGE
REESES
CHOCOLATE
VANILLA
$ ./search.py ice_cream.txt
.....CHOCOLATE
.SKCARTESOOM..
.YVANILLA.N...
M.D.T..A..A..A
.A.N.IN...C..E
..P.AAG...E..T
...LNC.E..P.RN
...AE.N.R..E.E
..B..W.O.TE..E
.....A.TSA..R
.....LET.I.G
.EGDUF.SN.O.L.
DAORYKCORU.C..

```

...TUNOCOCT...

Solution

```
1  #!/usr/bin/env python3
2  """Word Search"""
3
4  import argparse
5  from dire import die
6
7
8  # -----
9  def get_args():
10     """Get command-line arguments"""
11
12     parser = argparse.ArgumentParser(
13         description='Word search',
14         formatter_class=argparse.ArgumentDefaultsHelpFormatter)
15
16     parser.add_argument('file',
17                         metavar='FILE',
18                         type=argparse.FileType('r'),
19                         help='The puzzle')
20
21     return parser.parse_args()
22
23
24  # -----
25  def read_puzzle(fh):
26     """Read the puzzle file"""
27
28     puzzle, words = [], []
29     cell = 0
30     read = 'puzzle'
31     for line in map(str.rstrip, fh):
32         if line == '':
33             read = 'words'
34             continue
35
36         if read == 'puzzle':
37             row = []
38             for char in list(line):
39                 cell += 1
40                 row.append((char, cell))
41
42             puzzle.append(row)
43     else:
```

```

44         words.append(line.replace(' ', ''))
45
46     return puzzle, words
47
48
49 # -----
50 def all_combos(puzzle):
51     """Find all combos in puzzle"""
52
53     num_rows = len(puzzle)
54     num_cols = len(puzzle[0])
55
56     if not all([len(row) == num_cols for row in puzzle]):
57         die('Uneven number of columns')
58
59     combos = []
60
61     # Horizontal
62     for row in puzzle:
63         combos.append(row)
64
65     # Vertical
66     for col_num in range(num_cols):
67         col = [puzzle[row_num][col_num] for row_num in range(num_rows)]
68         combos.append(col)
69
70     # Diagonals Up
71     for row_i in range(0, num_rows):
72         diag = []
73         col_num = 0
74         for row_j in range(row_i, -1, -1):
75             diag.append(puzzle[row_j][col_num])
76             col_num += 1
77
78         if diag:
79             combos.append(diag)
80
81     for col_i in range(1, num_cols):
82         diag = []
83
84         col_num = col_i
85         for row_num in range(num_rows - 1, -1, -1):
86             diag.append(puzzle[row_num][col_num])
87             col_num += 1
88         if col_num == num_cols:
89             break

```



```

90
91         if diag:
92             combos.append(diag)
93
94     # Diagonals Down
95     for row_i in range(0, num_rows):
96         diag = []
97         col_num = 0
98         for row_j in range(row_i, num_rows):
99             diag.append(puzzle[row_j][col_num])
100             col_num += 1
101             if col_num == num_cols:
102                 break
103
104         if diag:
105             combos.append(diag)
106
107     for col_i in range(0, num_cols):
108         diag = []
109
110         col_num = col_i
111         for row_num in range(0, num_rows):
112             diag.append(puzzle[row_num][col_num])
113             col_num += 1
114             if col_num == num_cols:
115                 break
116
117         if diag:
118             combos.append(diag)
119
120     combos.extend([list(reversed(c)) for c in combos])
121     return combos
122
123
124 # -----
125 def fst(t):
126     """Return first element of a tuple"""
127
128     return t[0]
129
130
131 # -----
132 def snd(t):
133     """Return second element of a tuple"""
134     return t[1]
135

```

```

136
137 # -----
138 def main():
139     """Make a jazz noise here"""
140
141     args = get_args()
142     puzzle, words = read_puzzle(args.file)
143     combos = all_combos(puzzle)
144     found = set()
145     reveal = set()
146     for word in words:
147         for combo in combos:
148             test = ''.join(map(fst, combo))
149             if word in test:
150                 start = test.index(word)
151                 end = start + len(word)
152                 for cell in map(snd, combo[start:end]):
153                     reveal.add(cell)
154                 found.add(word)
155                 break
156
157     for row in puzzle:
158         cells = [c[0] if c[1] in reveal else '.' for c in row]
159         print(''.join(cells))
160
161     missing = [w for w in words if not w in found]
162     if missing:
163         print('Failed to find:')
164         for i, word in enumerate(missing, 1):
165             print('{:3}: {}'.format(i, word))
166
167
168 # -----
169 if __name__ == '__main__':
170     main()

```

Discussion

The only argument to the program is a single positional `file` which I chose to define with `type=argparse.FileType('r')` on line 17 to save me the trouble of testing for a file though you could test yourself and will pass the test as long as your error message includes `No such file or directory: '{}'` for the given file.

Reading the puzzle input

I chose to define a few additional functions while keeping most of the programs logic in the `main`. The first is `read_puzzle` that reads the file given by the user. As noted in the README, this file has the puzzle grid, an empty line, and then the list of words to search, so I define `read_puzzle` to accept the file (`fh`) as an argument and return two lists that represent the `puzzle` and `words` (line 28).

There list of `words` is really most naturally represented as a list of `str` elements, but the `puzzle` is a bit more complicated. After working through a couple of solutions, I decided I would number all the characters in the grid in order to know which ones to reveal at the end and which ones to replace with a period, so I define a `cell` variable initialized to 0 to keep count of the characters.

Here is my mental model of the puzzle:

Puzzle	Model		
	Col 0	Col 1	Col 2
A B C	Row 0 (A, 1)	(B, 2)	(C, 3)
D E F	Row 1 (D, 4)	(E, 5)	(F, 6)
G H I	Row 2 (G, 7)	(H, 8)	(I, 9)

Lastly, I need to know if I'm reading the first part of the file with the puzzle or the latter part with the words, so I define a `read` variable initialized to `'puzzle'` on line 30.

I start reading with `for line in` the file, but I want to chop off the trailing whitespace so I `map(str.rstrip, fh)`. Remember not to include parens `()` on `str.rstrip` as we want to *reference* the function not *call* it. The first operation in the loop is to check for an empty string `('')`, because we remove the newlines). If we find that, then we note the switch to reading the `'words'` and use `continue` to skip to the next iteration of the loop.

If I'm reading the puzzle part of the file. then I want to read each character (line 38), increment the `cell` counter, then create a new tuple with the character and it's cell number, appending this to the `row`, a list to hold all the new tuples. The `row` then gets appended to the `puzzle` list that will eventually be a list of rows, each of which is a list of tuples representing `(char, cell)`.

If we get to line 44, we must be reading the latter part of the file, so the `line` is actually a word that I will **append** to the `words` list. Before doing that, however, I will **replace** any space (' ') with the empty string ('') so as to remove spaces (cf. the `ice_cream.txt` input). Finally I **return** `puzzle`, `words` which is actually returning a tuple created by the comma , and which I immediately unpack on line 124.

Finding all the strings

I always try to make a function fit into about 50 lines of code. While my `read_puzzle` fits into 22 lines, the other function, `all_combos` is considerable longer. I couldn't find a way to shorten it, so I at least try to keep the idea fully contained to one function that, once it works, I no longer need to consider. The idea of this function is to find all the strings possible by reading each row, column, and diagonal both forward and backward. To do this, I first figure out how many rows and columns are present by checking the length (`len`) of the `puzzle` itself (the number of rows) and the length of the first row (the number of character in the first row). I double-check on line 56 that **all** of the the rows have the same `len` as the first one, using the `die` function from the `dire` module to print a message to `STDERR` and then `sys.exit(1)` to indicate a failure.

The `all_combos` will return a **list** of the characters and their cells, so I define `combos` on line 59 as an empty list (`[]`). Reading the rows is easiest on lines 61-62 as we just copy each `row` into `combo`. Reading the columns is done by moving from column 0 to the last column using the `range(num_cols)` (remembering the last number is not included which is important because if there are 10 columns then we need to move from column 0 to column 9). I can then extract each column position from each row in the puzzle by indexing `puzzle[row_num][col_num]` and appending those to the `combos`.

The diagonals are the trickiest. I chose to go up (lower-left to upper-right) first. I start in the top-left corner, row 0 and column 0. For each row, I'm going to move diagonally upwards (toward the top of the grid) which is actually counting *down* from the row I'm on, so I actually need to move `row_i` *up* and then `row_j` *down*. (I use `i` for "integer" and then `j` because "j" comes after "i". This is a typical naming convention. If I needed a third counter, I'd move to `k`.) I count `row_j` *down* by using `range(row_i, -1, -1)` (where the first `-1` is so I can count all the way to 0 and the second indicates the step should go down by one), I need to move the `col_num` over by 1. If I successfully read a diagonal, I append that to the `combos`.

The next block starts at the bottommost row of the and moves across the columns and is very similar to how I read the columns. Then moving into reading the diagonals in a downward (upper-left to bottom-right) fashion, I modified the other two blocks to handle the specifics. Finally at the end of the function (line 120), I want to **extend** the `combos` list by adding a **reversed**

version of each combo. It's necessary to coerce `list(reversed(c))` otherwise we'd end up with references to *reversed objects*.

Solving the puzzle

Once we've read the puzzle and found all the possible strings both forwards and backwards, we can then look for each of the words in each of the strings. In my `main`, I want to use sets to note all the words that are **found** as well as the cell numbers to **reveal**. Because I'll be reading lists of tuples where the character is in the first position and the cell number in the second, I define two functions `fst` and `snd` (stolen from Haskell) that I can use in `map` expressions. I iterate `for word in words` (line 146) and `for combo in combos` to check all combinations. Recall that the `combo` is a list of tuples:

```
>>> combo = [('X', 1), ('F', 2), ('O', 3), ('O', 4)]
```

so I can build a string from the characters in the `fst` position of the tuples by mapping them to `fst`:

```
>>> list(map(fst, combo))
['X', 'F', 'O', 'O']
```

and joining them on an empty string:

```
>>> test = ''.join(map(fst, combo))
>>> test
'XF00'
```

Then I check if the `word` is in the `test` string:

```
>>> word='F00'
>>> word in test
True
```

If it is, then I can find where it starts with the `str.index` function:

```
>>> start = test.index(word)
>>> start
1
```

I know then end is:

```
>>> end = start + len(word)
>>> end
4
```

I can use that information to iterate over the elements in the `combo` to extract the cell numbers which are in the `snd` position of the tuple because ultimately what I need to print is the original puzzle grid with the cells showing the hidden words and all the others masked. I can extract a list slice using `combo[start:end]`,

`map` those elements through `snd` to get the `cell` and `add` those to the `reveal` set. I can also note that I `found` the `word`.

At line 157, I start the work of printing the revealed puzzle, iterating over the original rows in the puzzle and over each cell in the row. If the cell number is in the `reveal` set, I chose the character (in the first position of the tuple); otherwise I use a period (`.`). Finally I note any missing words by looking to see if any of the original words were not in the `found` set.

Appendix 1: argparse

The `argparse` module will interpret all the command-line arguments to your program. I suggest you use `argparse` for every command-line program you write so that you always have a standard way to get arguments and present help.

Types of arguments

Command-line arguments come in a variety of flavors:

- Positional: 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.
- 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. This allows for arguments to be provided in any order or not provided in which case the program can use a reasonable default value.
- Flag: A “Boolean” value like “yes”/“no” or `True/False` usually indicated by something that looks like a named option but without a value, e.g., `-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 no `--debug` flag means there should not be debugging.

Datatypes of values

The `argparse` module can save you enormous amounts of time by forcing the user to provide arguments of a particular type. If you run `new.py`, all of the above types of arguments are present along with suggestions for how to get string or integer values:

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

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

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

```

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

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

parser.add_argument('-f',
                    '--flag',
                    help='A boolean flag',
                    action='store_true')

return parser.parse_args()

```

You should change the **description** to a short sentence describing your program. The **formatter_class** argument tells **argparse** to show the default values in the the standard help documentation.

The **positional** argument's definition indicates we expect exactly one positional argument. The **-a** argument's **type** must be a **str** while the **-i** option must be something that Python can convert to the **int** type (you can also use **float**). Both of these arguments have **default** values which means the user is not required to provide them. You could instead define them with **required=True** to force the user to provide values themselves.

The **-f** flag notes that the **action** is to **store_true** which means the value's default with be **True** if the argument is present and **False** otherwise.

The **type** of the argument can be something much richer than simple Python types like strings or numbers. You can indicate that an argument must be a existing, readable file. Here is a simple implementation in Python of **cat -n**:

```

#!/usr/bin/env python3
"""Python version of `cat -n`"""

import argparse

# -----
def get_args():

```



```

    """Get command-line arguments"""

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

    parser.add_argument('file',
                        metavar='FILE',
                        type=argparse.FileType('r'),
                        help='Input file')

    return parser.parse_args()

# -----
def main():
    """Make a jazz noise here"""

    args = get_args()
    fh = args.file

    print('Reading "{}".format(fh.name))
    for i, line in enumerate(fh):
        print(i, line, end='')

# -----
if __name__ == '__main__':
    main()

```

The *type* of the input file argument is an *open file handle* which we can directly read line-by-line with a *for* loop! Because it's a file *handle* and not a file *name*, I chose to call the variable *fh* to help me remember what it is. You can access the file's name via *fh.name*.

```

$ ./cat_n.py ../../inputs/the-bustle.txt
Reading "../../inputs/the-bustle.txt"
0 The bustle in a house
1 The morning after death
2 Is solemnest of industries
3 Enacted upon earth,--
4
5 The sweeping up the heart,
6 And putting love away
7 We shall not want to use again
8 Until eternity.

```

Number of arguments

If you want one positional argument, you can define them like so:

```
#!/usr/bin/env python3
"""One positional argument"""

import argparse

parser = argparse.ArgumentParser(
    description='One positional argument',
    formatter_class=argparse.ArgumentDefaultsHelpFormatter)

parser.add_argument('first', metavar='str', help='First argument')
args = parser.parse_args()
print('first =', args.first)
```

If the user provides anything other exactly one argument, they get a help message:

```
$ ./one_arg.py
usage: one_arg.py [-h] str
one_arg.py: error: the following arguments are required: str
$ ./one_arg.py foo bar
usage: one_arg.py [-h] str
one_arg.py: error: unrecognized arguments: bar
$ ./one_arg.py foo
first = foo
```

If you want two different positional arguments:

```
#!/usr/bin/env python3
"""Two positional arguments"""

import argparse

parser = argparse.ArgumentParser(
    description='Two positional arguments',
    formatter_class=argparse.ArgumentDefaultsHelpFormatter)

parser.add_argument('first', metavar='str', help='First argument')

parser.add_argument('second', metavar='int', help='Second argument')

return parser.parse_args()

print('first =', args.first)
print('second =', args.second)
```

Again, the user must provide exactly this number of positional arguments:

```
$ ./two_args.py
usage: two_args.py [-h] str str
two_args.py: error: the following arguments are required: str, str
$ ./two_args.py foo
usage: two_args.py [-h] str str
two_args.py: error: the following arguments are required: str
$ ./two_args.py foo bar
first = foo
second = bar
```

You can also use the `nargs=N` option to specify some number of arguments. It only makes sense if the arguments are the same thing like two files:

```
#!/usr/bin/env python3
"""nargs=2"""

import argparse

parser = argparse.ArgumentParser(
    description='nargs=2',
    formatter_class=argparse.ArgumentDefaultsHelpFormatter)

parser.add_argument('files', metavar='FILE', nargs=2, help='Two files')

args = parser.parse_args()

file1, file2 = args.files
print('file1 =', file1)
print('file2 =', file2)
```

The help indicates we want two files:

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

And we can unpack the two file arguments and use them:

```
$ ./nargs2.py foo bar
file1 = foo
file2 = bar
```

If you want one or more of some argument, you can use `nargs='+'`:

```
$ cat nargs+.py
#!/usr/bin/env python3
"""nargs="+"""
```

```

import argparse

parser = argparse.ArgumentParser(
    description='nargs=+',
    formatter_class=argparse.ArgumentDefaultsHelpFormatter)

parser.add_argument('files', metavar='FILE', nargs='+', help='Some files')

args = parser.parse_args()
files = args.files

print('number = {}'.format(len(files)))
print('files = {}'.format(', '.join(files)))

```

Note that this will return a **list** – even a single argument will become a **list** of one value:

```

$ ./nargs+.py
usage: nargs+.py [-h] FILE [FILE ...]
nargs+.py: error: the following arguments are required: FILE
$ ./nargs+.py foo
number = 1
files = foo
$ ./nargs+.py foo bar
number = 2
files = foo, bar

```

Choices

Sometimes you want to limit the values of an argument. You can pass in a **list** of valid values to the **choices** option.

```

$ cat appendix/argparse/choices.py
#!/usr/bin/env python3
"""Choices"""

import argparse

parser = argparse.ArgumentParser(
    description='Choices',
    formatter_class=argparse.ArgumentDefaultsHelpFormatter)

parser.add_argument('color', metavar='str', help='Color', choices=['red', 'yellow', 'blue'])

args = parser.parse_args()

```

```
print('color =', args.color)
```

Any value not present in the list will be rejected and the user will be shown the valid choices:

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

Automatic help

The `argparse` module reserves the `-h` and `--help` flags for generating help documentation. You do not need to add these nor are you allowed to use these flags for other purposes. Using the above definition, this is the help that `argparse` will generate:

```
$ ./foo.py
usage: foo.py [-h] [-a str] [-i int] [-f] str
foo.py: error: the following arguments are required: str
[cholla@~/work/python/playful_python/article]$ ./foo.py -h
usage: foo.py [-h] [-a str] [-i int] [-f] 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, --flag	A boolean flag (default: False)

Notice how unhelpful a name like `positional` is?

Getting the argument values

The values for the arguments will be accessible through the “long” name you define and will have been coerced to the Python data type you indicated. If I change `main` to this:

```
# -----
def main():
```

```

"""Make a jazz noise here"""

args = get_args()
str_arg = args.arg
int_arg = args.int
flag_arg = args.flag
pos_arg = args.positional

print('str_arg = "{}" ({}).format(str_arg, type(str_arg))
print('int_arg = "{}" ({}).format(int_arg, type(int_arg))
print('flag_arg = "{}" ({}).format(flag_arg, type(flag_arg))
print('positional = "{}" ({}).format(pos_arg, type(pos_arg))

```

And then run it:

```

$ ./foo.py -a foo -i 4 -f bar
str_arg = "foo" (<class 'str'>)
int_arg = "4" (<class 'int'>)
flag_arg = "True" (<class 'bool'>)
positional = "bar" (<class 'str'>)

```

Notice how we might think that `-f` takes the argument `bar`, but it is defined as a flag and the `argparse` knows that the program take

```

$ ./foo.py foo -a bar -i 4 -f
str_arg = "bar" (<class 'str'>)
int_arg = "4" (<class 'int'>)
flag_arg = "True" (<class 'bool'>)
positional = "foo" (<class 'str'>)

```

Appendix 2: Truthiness

While it would seem Python has an actual Boolean (Yes/No, True/False) type, this idea can be seriously abused in many odd and confusing ways. First off, there are actual `True` and `False` values:

```
>>> True == True
True
>>> False == False
True
```

But they are equivalent to integers:

```
>>> True == 1
True
>>> False == 0
True
```

Which means, oddly, that you can add them:

```
>>> True + True
2
>>> True + True + False
2
```

Lots of things are `False`-ey when they are evaluated in a Boolean context. The `int 0`, the `float 0.0`, the empty string, an empty list, and the special value `None` are all considered `False`-ey:

```
>>> 'Hooray!' if 0 else 'Shucks!'
'Shucks!'
>>> 'Hooray!' if 0. else 'Shucks!'
'Shucks!'
>>> 'Hooray!' if [] else 'Shucks!'
'Shucks!'
>>> 'Hooray!' if '' else 'Shucks!'
'Shucks!'
>>> 'Hooray!' if None else 'Shucks!'
'Shucks!'
```

But note:

```
>>> 'Hooray!' if 'None' else 'Shucks!'
'Hooray!'
```

There are quotes around `'None'` so it's the literal string “None” and not the special value `None`, and, since this is not an empty string, it evaluates *in a Boolean context* to not-`False` which is basically `True`.

This behavior can introduce extremely subtle logical bugs into your programs that the Python compiler and linters cannot uncover. Consider the `dict.get`

method that will safely return the value for a given key in a dictionary, returning `None` if the key does not exist. Given this dictionary:

```
>>> d = {'foo': 0, 'bar': None}
```

If we access a key that doesn't exist, Python generates an exception that, if not caught in our code, would immediately crash the program:

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

But we can use `d.get()` to do this safely:

```
>>> d.get('baz')
```

Hmm, that seems unhelpful! What did we get back?

```
>>> type(d.get('baz'))
<class 'NoneType'>
```

Ah, we got `None`!

We could use an `or` to define a default value:

```
>>> d.get('baz') or 'NA'
'NA'
```

It turns out the `get` method accepts a second, optional argument of the default value to return:

```
>>> d.get('baz', 'NA')
'NA'
```

Great! So let's use that on the other values:

```
>>> d.get('foo', 'NA')
0
>>> d.get('bar', 'NA')
```

The call for `bar` returned nothing because we put an actual `None` as the value:

```
>>> type(d.get('bar', 'NA'))
<class 'NoneType'>
```

The key `bar` didn't fail because that key exists in the dictionary. The `dict.get` method only returns the second, default argument *if the key does not exist in the dictionary* which is entirely different from checking the *value* of the key in the dictionary. OK, so we go back to this:

```
>>> d.get('bar') or 'NA'
'NA'
```

Which seems to work, but notice this:


```
>>> d.get('foo') or 'NA'
'NA'
```

The value for `foo` is actually `0` which evaluates to `False` given the Boolean evaluation of the `or`. If this were a measurement of some value like the amount of sodium in water, then the string `NA` would indicate that no value was recorded whereas `0` indicates that sodium was measured and none detected. If some sort of important analysis rested on our interpretation of the strings in a spreadsheet, we might inadvertently introduce missing values because of the way Python coerces various non-Boolean values into Boolean values.

Perhaps a safer way to access these values would be:

```
>>> for key in ['foo', 'bar', 'baz']:
...     val = d[key] if key in d else 'NA'
...     val = 'NA' if val is None else val
...     print(key, val)
...
foo 0
bar NA
baz NA
```

Appendix 3: File Handles

A file's name is a string like `'nobody.txt'`. To read or write the contents of the file, you need a *file handle* which you can get from `open`. Think of a file name as the address of your house. It's where your house can be found, but I can't know what's in your house unless I go there and open the door. That's what `open` does – it finds the file's bits on disk and opens the door to read or write the file.

File Modes

By default, a file is opened in *read* mode which means that it can't be altered. Also, the default is to open for reading *text*. The only required argument to `open` is the file name, but a second optional argument is a combination of characters to explain how to open the file. From the documentation for `open`:

=====	
Character	Meaning

'r'	open for reading (default)
'w'	open for writing, truncating the file first
'x'	create a new file and open it for writing
'a'	open for writing, appending to the end of the file if it exists
'b'	binary mode
't'	text mode (default)
'+'	open a disk file for updating (reading and writing)
'U'	universal newline mode (deprecated)
=====	

So if you do:

```
fh = open('out.txt')
```

It's the same as doing:

```
fh = open('out.txt', 'wt')
```

Where the combination of `wt` means **write text**. We can also read and write raw bits in **binary**, e.g., if you wanted to read the bit values of the pixels in an image.

I always make a distinction in the variable names for the **file** or **filename** and the *file handle* which I usually call `fh` if there's just one or maybe `in_fh` and `out_fh` if there is one for reading and one for writing, etc.

STDIN, STDOUT, STDERR

Unix has three standard files or channels called *standard in*, *standard out*, and *standard error* which are normally written as STDIN, STDOUT, and STDERR. When you **print**, the default is that the text goes to STDOUT which you see in your terminal or REPL.

The **print** function takes some optional keyword arguments, one of which is **file** which has the default value of **sys.stdout**. If you wish to **print** to *standard error* (STDERR), you can use the **sys.stderr** file:

```
print('This is an error!', file=sys.stderr)
```

Note that you *do not* have to **open** these two special file handles. They are always available to you.

If you wish to write to a file on disc, you can **open** a file for writing and pass that:

```
print('This is an error!', file=open('error.txt', 'wt'))
```

Note that if each time you **open** a file for writing, you overwrite any existing data. If you wanted to **print** repeatedly in a program, you would either need to **open** in append mode:

```
print('This is an error!', file=open('error.txt', 'at'))
print('This is an also error!', file=open('error.txt', 'at'))
```

Or, better yet, **open** the file at the beginning of the program, **print** as often as you like, and then **close** the file:

```
fh = open('out.txt', 'wt')
print('Writing some text.', file=fh)
print('Adding more text.', file=fh)
fh.close()
```

Or use the **write** method of the file handle:

```
fh = open('out.txt', 'wt')
fh.write('Writing some text.\n')
fh.write('Adding more text.\n')
fh.close()
```

Note that **print** automatically adds a newline to the end of the text whereas **write** does not so you need to add it yourself.

You can only *read* from STDIN. Again, you do not need to **open** it as it is always available. Treat it exactly like a file handle you've opened for reading, e.g., to read lines from STDIN until you receive EOF (end of file):

```
for line in sys.stdin:
```

Appendix 4: N-grams, K-mers, and Markov Chains

Read about Markov chains:

- Claude Shannon’s 1948 MS thesis, “A Mathematical Theory of Communication” (<https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1948.tb01338.x>)
- https://en.wikipedia.org/wiki/Markov_chain
- Chapter 3 of *The Practice of Programming* by Brian Kernighan and Rob Pike where they discuss implementations in C, C++, Java, awk, and Perl
- “Computer Recreations”, A. K. Dewdney, Scientific American, 1989 (<https://archive.org/details/ComputerRecreationsMarkovChainer>)

I’d like you to consider how a Markov chain creates a graph structure. Consult the three PDFs (generated by the `mk-graphs.sh` program) that visualize the graphs created by k-mer sizes of 1, 2, 3, and 4 when given this input:

```
$ cat words.txt
maamselle
mabi
mabolo
mac
macaasim
macabre
```

Notice that sometimes the branches terminate and sometimes you can find multiple paths through the graphs. As `k` grows, there are fewer options.