The ability to search through files and folders can greatly improve your productivity using Unix. First we'll cover searching through text files. I recently downloaded a list of the names of the states in the US, which you can find here. Let's take a look at this file:

```
 1  cd ~/Documents
 2  ls
 3
 4  ## canada.txt
 5  ## states.txt
 6
 7  wc states.txt
 8
 9  ## 50      60      472 states.txt
```

It makes sense that there are 50 lines, but it's interesting that there are 60 total words. Let's a take a peak at the beginning of the file:

```
 1  head states.txt
 2
 3  ## Alabama
 4  ## Alaska
 5  ## Arizona
 6  ## Arkansas
 7  ## California
 8  ## Colorado
 9  ## Connecticut
10  ## Delaware
11  ## Florida
12  ## Georgia
13
```

This file looks basically how you would expect it to look! You may recall from Chapter 3 that the kind of shell that we're using is the bash shell. Bash treats different kinds of data differently, and we'll dive deeper into data types in Chapter 5. For now all you need to know is that text data are called strings. A string could be a word, a sentence, a book, or a file or folder name. One of the most effective ways to search through strings is to use regular expressions. Regular expressions are strings that define patterns in other strings. You can use regular expressions to search for a sub-string contained within a larger string, or to replace one part of a string with another string.

One of the most popular tools for searching through text files is grep. The simplest use of grep requires two arguments: a regular expression and a text file to search. Let's see a simple example of grep in action and then I'll explain how it works:

```
1   grep "x" states.txt
2
3   ## New Mexico
4   ## Texas
5
```

In the command above, the first argument to grep is the regular expression "x". The "x" regular expression represents one instance of the letter "x". Every line of the states.txt file that contains at least one instance of the letter "x" is printed to the console. As you can see New Mexico and Texas are the only two state names that contain the letter "x". Let's try searching for the letter "q" in all of the state names using grep:

```
1   grep "q" states.txt
2
```

Nothing is printed to the console because the letter "q" isn't in any of the state names. We can search for more than individual characters though. For example the following command will search for the state names that contain the word "New":

```
1   grep "New" states.txt
2
3   ## New Hampshire
4   ## New Jersey
5   ## New Mexico
6   ## New York
7
```

In the previous case the regular expression we used was simply "New", which represents an occurrence of the string "New". Regular expressions are not limited to just being individual characters or words, they can also represent parts of words. For example I could search all of the state names that contain the string "nia" with the following command:

```
1   grep "nia" states.txt
2
3   ## California
4   ## Pennsylvania
5   ## Virginia
6   ## West Virginia
7
```

All of the state names above happen to end with the string "nia".

Mark as completed