It's important to be able to examine differences between files. First let's make two small simple text files in the Documents directory.

```
cd ~/Documents
head -n 4 states.txt > four.txt
head -n 6 states.txt > six.txt

```

If we want to look at which lines in these files are different we can use the diff command:

```
diff four.txt six.txt

## 4a5,6
## > California
## > Colorado
```

Only the differing lines are printed to the console. We could also compare differing lines in a side-by-side comparison using sdiff:

```
sdiff four.txt six.txt

## Alabama              Alabama
## Alaska               Alaska
## Arizona              Arizona
## Arkansas             Arkansas
##                    > California
##                    > Colorado
```

In a common situation you might be sent a file, or you might download a file from the internet that comes with code known as a **checksum** or a **hash**. Hashing programs generate a unique code based on the contents of a file. People distribute hashes with files so that we can be sure that the file we think we've downloaded is the genuine file. One way we can prevent malicious individuals from sending us harmful files is to check to make sure the computed hash matches the provided hash. There are a few commonly used file hashes but we'll talk about two called MD5 and SHA-1.

Since hashes are generated based on file contents, then two identical files should have the same hash. Let's test this my making a copy of states.txt.

```
cp states.txt states_copy.txt

```

To compute the MD5 hash of a file we can use the md5 command:

```
md5 states.txt

## MD5 (states.txt) = 8d7dd71ff51614e69339b03bd1cb86ac

md5 states_copy.txt

## MD5 (states_copy.txt) = 8d7dd71ff51614e69339b03bd1cb86ac
```

As we expected they're the same! We can compute the SHA-1 hash using the shasum command:

```
shasum states.txt

## 588e9de7ffa97268b2448927df41760abd3369a9  states.txt

shasum states_copy.txt

## 588e9de7ffa97268b2448927df41760abd3369a9  states_copy.txt
```

Once again, both copies produce the same hash. Let's make a change to one of the files, just to illustrate the fact that the hash changes if file contents are different:

```
head -n 5 states_copy.txt > states_copy.txt
shasum states_copy.txt

## b1c1c805f123f31795c77f78dd15c9f7ac5732d4  states_copy.txt
```

Summary

- The md5 and shasum commands use different algorithms to create codes (called hashes or checksums) that are unique to the contents of a file.
- These hashes can be used to ensure that a file is genuine.

Mark as completed