Diff finding a needle in a haystack

Take a look at the two files haystack.txt and needle_haystack.txt. It is difficult to manually find small errors in large output files.

Finding Differences

To find any differences in the files we can simply run the `diff` command

```
$ diff haystack.txt needle_haystack.txt
122c122
< hay
---
> needle
```

This outptu shows that the files differ at line 122 of each file. The difference is that at line 122 haystack.txt has hay and haystack_needle.txt has needle. We know this because of the directions of the angle brackets.

Adding Some Color

I personally like color coding in my tools. `diff`, like some other utilities has a `--color` option. There are 3 settings `always`, `never`, and `auto`. While you might always want color you probably actually want auto. `auto` will always show colors in the terminal and won't show colors when outputting to a file.

You can make a `diff` alias to always turn on colors for you by editing your shell's rc file, which is probably ~/.bashrc

```
alias diff="diff --color=auto"
```

Side-by-Side Comparison

If we want to compare the two files side-by-side, which may be more intuitive for some, we can use the `-y` option (there is also a long form `--side-by-side`).

```
$ diff -y haystack.txt needle_haystack.txt
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
hay                                          hay
...
```

This generates a full side-by-side comparison of the two files, but maybe you only want to see the lines that are different. to do that you can also add the `--suppress-common-lines` option.

```
$ diff -y --suppress-common-lines haystack.txt needle_haystack.txt
hay                                                      | needle
```

Getting Some Context

Maybe you want to see the lines around the line that is different to know why it may be different. We could just open the files in an editor and look for the difference there, but there'a an easy way to view the lines around the differences. The -c option shows the lines from each file separately wheras the -u option shows the lines in a unified view

```
$ diff -c haystack.txt needle_haystack.txt
*** haystack.txt        2019-02-01 10:41:28.722803889 -0500
--- needle_haystack.txt 2019-02-01 10:41:19.236118886 -0500
***************
*** 119,125 ****
  hay
  hay
  weird data
! hay
  hay
  hay
  hay
--- 119,125 ----
  hay
  hay
  weird data
! needle
  hay
  hay
  hay
```

Here we can see the files from line 119-125. The lines that are different are marked with a !. You can also pass an argument to either -c or -u to specify the number of lines to show

```
$ diff -u5 haystack.txt needle_haystack.txt
--- haystack.txt        2019-02-01 10:41:28.722803889 -0500
+++ needle_haystack.txt 2019-02-01 10:41:19.236118886 -0500
@@ -117,11 +117,11 @@
 hay
 hay
 hay
 hay
 weird data
-hay
+needle
 hay
 hay
 hay
 hay
 hay
```

Here we can see from line 117 in each file +11 lines.