

# Data Manipulation With Unix/Linux BASH

Note: *file* means `path/to/file` in the examples below

I want to ...	Example
<b>Wrangle Data Files</b> <ul style="list-style-type: none"> <li>Convert a FASTA file to tidy, row 1-2 become cols 1-2</li> <li>Convert a FASTQ file to tidy, rows 1-4 become cols 1-4</li> <li>Find &amp; replace or remove text using regex pattern</li> <li>Modify column delimiters (aka field separators)</li> <li>Keep Columns 1, 2, 3 and 6 in tsv file, remove the rest</li> <li>Keep Columns 1, 2, 3 and 6 in csv file, remove the rest</li> <li>Return lines that match regex pattern</li> <li>Remove column names (first line) from any tidy file</li> <li>Bind columns together into same tidy file</li> <li>Bind rows together from two sources into one tidy file</li> <li>Sort rows by column 2 alphabetically</li> <li>Sort rows by col3 from lrg to sml numbers</li> <li>Keep only consecutive unique values, optionally also count the number of duplicates</li> <li>Count the number of rows in a file, or the number of characters</li> </ul>	<pre> paste - - file paste - - - - file sed "s/FindPattern/ReplacePattern/g" tr -s "PresentDelimiter" "NewDelimiter" cut -f1-3,6 file cut -d, -f1-3,6 file grep "Pattern" file tail -n+2 file paste &lt;(commands) &lt;(commands) paste file1 file2 cat &lt;(commands) &lt;(commands) cat file1 file2 sort -k2,2 file sort -nrk3,3 file uniq file uniq -c file wc -l file wc -c file </pre>
<b>Use wild cards (applies to file paths)</b> <ul style="list-style-type: none"> <li>Any and all characters repeated any number of times</li> <li>Any single character</li> <li>List all files ending with ".txt"</li> </ul>	<pre> * ? ls *.txt </pre>
<b>Create regular expression search patterns</b> <ul style="list-style-type: none"> <li>Escape character, change meaning of special character</li> <li>Match a special character like * or .</li> <li>0-infinite repeats of the previous character</li> <li>Any single character</li> <li>Beginning of line or End of line</li> <li>Match any one of the following characters: a, b, c</li> </ul>	<pre> \ \* or \. * . ^ or \$ [abc] </pre>
<b>do math</b> <ul style="list-style-type: none"> <li>add 1+1 and save to variable named X</li> <li>subtract 1-1 and save to variable name Y</li> <li>1+11-2 and print to screen</li> </ul>	<pre> X=\$((1+1)) Y=\$((10-5)) echo \$((1+11-2)) </pre>