

# Homework #3

Juliana Dau

2025-02-26

## #Example 1

### #Input:

First String	Second	1.22	3.4
Second	More Text	1.555555	2.2220
Third	x	3	124

#Find: \s{2,}

#Replace: ,

#Reason: to find consecutive single spaces that are 2 or more characters and replace with , to match the .csv file. Needed this way so the single space between "First string is not detected"

#Result:

First String,Second,1.22,3.4  
Second,More Text,1.555555,2.2220  
Third,x,3,124

## #Example 2

### #Input:

Ballif, Bryan, University of Vermont  
Ellison, Aaron, Harvard Forest  
Record, Sydne, Bryn Mawr

#Find: (\w\*),\s(\w\*),\s(.\*)

#Replace: \2 \1 \(\3\)

#Reason: (\w\*) capture the first word, and followed by \s(\w\*) finds the first word comma and second word. \s(.\*) will indicate all the rest (university name). On replace the order matches the desired organization bringing the second capture to be first and the first capture to second position while the third is added parentheses.

#Result:

Bryan Ballif (University of Vermont)  
Aaron Ellison (Harvard Forest)  
Sydne Record (Bryn Mawr)

## #Example 3

### #Input:

0001 Georgia Horseshoe.mp3 0002 Billy In The Lowground.mp3 0003 Winder Slide.mp3 0004 Walking Cane.mp3

#Find: .mp3\s

#Replace: .mp3\n

#Reason: Indicated to find mp3 and space (.mp3\s) since that is the end of each line. In turn indicated to the place a line break at this point so each new line start after .mp3

#Result:

0001 Georgia Horseshoe.mp3

0002 Billy In The Lowground.mp3  
0003 Winder Slide.mp3  
0004 Walking Cane.mp3

#Example 4

#Input:

0001 Georgia Horseshoe.mp3  
0002 Billy In The Lowground.mp3  
0003 Winder Slide.mp3  
0004 Walking Cane.mp3

#Find: (\d{4}) (\w\*) (\w+)

#Replace: (\2) (\3) (\_\1)(\4)

#Reason: with (\d{4}) I'm marking the 4 digit sequence as the first item of capture. Followed by the file name

#Result:

Georgia Horseshoe \_0001.mp3  
Billy In \_The Lowground \_0002.mp3  
Winder Slide \_0003.mp3  
Walking Cane \_0004.mp3

#Example 5

#Input:

Camponotus,pennsylvanicus,10.2,44  
Camponotus,herculeanus,10.5,3  
Myrmica,punctiventris,12.2,4  
Lasius,neoniger,3.3,55

#Find: (\w)(\w+),(\w+),(\d{1,}).(\d),(\d{1,})

#Replace: \1\_\3,6

#Reason: Using the first part (\w)(\w+), captures genus with (\w) marking the first letter that will be replaced

#Result:

C\_pennsylvanicus,44  
C\_herculeanus,3  
M\_punctiventris,4  
L\_neoniger,55

#Example 6

#Input:

Camponotus,pennsylvanicus,10.2,44  
Camponotus,herculeanus,10.5,3  
Myrmica,punctiventris,12.2,4  
Lasius,neoniger,3.3,55

#Find: (\w)(\w+),(\w{4})(\w+),(\d{1,}).(\d),(\d{1,})

#Replace: \1\_\3,\7

#Reason: Using the first part (\w)(\w+), captures genus with (\w) marking the first letter that will be replaced

#Result:

C\_penn,44  
C\_herc,3  
M\_punc,4  
L\_neon,55

#Example 7

#Input:

Camponotus,pennsylvanicus,10.2,44

```

Camponotus,herculeanus,10.5,3
Myrmica,punctiventris,12.2,4
Lasius,neoniger,3.3,55
#Find: (\w{3})(\w+),(\w{3})(\w+),(\d{1,}).(\d),(\d{1,})
#Replace: \1\3, \7, \5.\6
#Reason: To capture only the first 3 letters of the genus used (\w{3}) which will be position 1 and (w+)
#Result:
Campen, 44, 10.2
Camher, 3, 10.5
Myrpun, 4, 12.2
Lasneo, 55, 3.3

```

#### #Example 8

##### #Input:

```

id,target_name,pathogen_binary,sampling_date,site_code,field_id,bombus_spp,host_plant,bee_caste,sampling
1,BQCV,1,9/9/2020,BOST,6,imp*atiens,solidago,worker ,4,2414168.805
3,BQCV,NA,9/10/2020,MUDGE,18,impatiens,c!rown vetch,worker ,4,760793.2324
4,BQCV,NA,9/10/2020,MUDGE,11,im&patiens,c$crown vetch,worker,4,0
5,BQCV,NA,9/9/2020,BOST,11,impatiens,sol)idago,male ,4,0
6,BQCV,NA,9/10/2020,CIND,14,-i@mp#a~tiens,k*no%t w%eed,worker ,4,124236.7921
8,BQCV,NA,9/10/2020,MUDGE,4,impatiens,crown v$et(ch,worker,4,413814.5638
10,BQCV,1,9/10/2020,CIND,13,impatien$s,red clover,worker,4,1028831.605
11,BQCV,NA,7/7/2020,BOST,38,impatiens,red clover,worker,2,307696378.8
12,BQCV,NA,9/10/2020,MUDGE,5,_imp*atiens,c+crown vetch,worker,4,0
13,BQCV,1,9/9/2020,BOST,12,impatien+s,solid#ago,male,4,311873.0526
15,BQCV,0,9/9/2020,BOST,18,_impatiens,solidag)o,worker ,4,0
16,BQCV,NA,9/9/2020,BOST,23,impati+e(ns,s!olidago,male,4,1674951.391
17,BQCV,NA,9/10/2020,CIND,20,impa#tiens,red cl$over,worker,4,3214026.976
18,BQCV,1,9/10/2020,CIND,11,i~m+patiens,bi_rdsfoo_t,worker ,4,995592.63
19,BQCV,NA,9/10/2020,CIND,17,i%mpatiens,red+ @clover,worker,4,0
20,BQCV,1,9/10/2020,CIND,16,imp(atien,#red c_lover,worker,4,200804.8542
21,BQCV,1,9/9/2020,BOST,17,impatiens,solid@ago,worker ,4,228085.8104
22,BQCV,1,9/9/2020,BOST,7,impati^ens,solidago,worker,4,7261151.315
23,BQCV,NA,7/7/2020,COL,22,)impati^ens,red clov#er,worker ,2,817906.8276
24,BQCV,1,7/7/2020,BOST,45,impatiens,red_ cl)over,worker,2,858658.6884

```

1. Start fiding: [^a-zA-Z0-9/,\.\s]

Replace: leave the space empty to cut all the characters that don't match.

2. In pathogen binary since thre format required is binary it should be represented as 0 or 1. Since 1 w

Find: NA

Replace: 0

Final result:

```

id,targetname,pathogenbinary,samplingdate,sitecode,fieldid,bombusspp,hostplant,beecaste,samplingevent,p
1,BQCV,1,9/9/2020,BOST,6,impatiens,solidago,worker ,4,2414168.805
3,BQCV,0,9/10/2020,MUDGE,18,impatiens,crown vetch,worker ,4,760793.2324
4,BQCV,0,9/10/2020,MUDGE,11,impatiens,crown vetch,worker,4,0
5,BQCV,0,9/9/2020,BOST,11,impatiens,solidago,male ,4,0
6,BQCV,0,9/10/2020,CIND,14,impatiens,knot weed,worker ,4,124236.7921
8,BQCV,0,9/10/2020,MUDGE,4,impatiens,crown vetch,worker,4,413814.5638
10,BQCV,1,9/10/2020,CIND,13,impatiens,red clover,worker,4,1028831.605
11,BQCV,0,7/7/2020,BOST,38,impatiens,red clover,worker,2,307696378.8
12,BQCV,0,9/10/2020,MUDGE,5,impatiens,crown vetch,worker,4,0
13,BQCV,1,9/9/2020,BOST,12,impatiens,solidago,male,4,311873.0526
15,BQCV,0,9/9/2020,BOST,18,impatiens,solidago,worker ,4,0

```

16,BQCV,0,9/9/2020,BOST,23,impatiens,solidago,male,4,1674951.391  
17,BQCV,0,9/10/2020,CIND,20,impatiens,red clover,worker,4,3214026.976  
18,BQCV,1,9/10/2020,CIND,11,impatiens,birdsfoot,worker ,4,995592.63  
19,BQCV,0,9/10/2020,CIND,17,impatiens,red clover,worker,4,0  
20,BQCV,1,9/10/2020,CIND,16,impatiens,red clover,worker,4,200804.8542  
21,BQCV,1,9/9/2020,BOST,17,impatiens,solidago,worker ,4,228085.8104  
22,BQCV,1,9/9/2020,BOST,7,impatiens,solidago,worker,4,7261151.315  
23,BQCV,0,7/7/2020,COL,22,impatiens,red clover,worker ,2,817906.8276  
24,BQCV,1,7/7/2020,BOST,45,impatiens,red clover,worker,2,858658.6884

#Replace:

#Reason:

#Result: