Learning goals: Linux OS commands grep, tar, gzip

NOTE: Save screenshots of each exercise, and upload your work to your github account as Lab5.pdf by end of Friday Aug 1.

For this lab, download 'BrainCancer.csv' to a folder called 'Lab5'. If the file download to the "Downloads" folder by default, move the file to the folder 'Lab5'.

The data file contains the following columns for the survival times of patients diagnosed with brain cancer.

• Sex Male or Female patients
• diagnosis Meningoma, LG glioma, HG glioma and Other
• loc Location of the cancer- infratentorial or supratentorial
• ki Karnofsky index is a scale to assess the functional impairment of a patient with cancer. 100 is normal and 0 is dead.
• gtv gross tumor volume in cc
• stereo stereotactic methods 'SRS' and 'SRT'
• status whether the patient is alive at the end of the study 0=Yes, 1=No
• time survival time in months

EXERCISES Start a fresh terminal for this Lab.

(1) Look for the pattern 'Meningoma' in the data file. Redirect the output to a new file called 'LinesWithMeningoma.out'. Show the command used and the contents of this new file in your screenshot.

(Snap No.1)

Ans:- Command :-grep Meningioma BrainCancer.csv > LinesWithMeningoma.out

(2) How many males and females were present in this study? How will you create a grep filter to find this out?

(Snap No.2 Male:43 , Female :45)

(3) How many males have supratentorial brain cancer and how many females have infratetorial cancer? Which option in grep makes you ignore the case whilesearching? Make use of this in this exercise.

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep "Female" BrainCancer.csv | grep Infratentorial
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
3,Female,Meningioma,Infratentorial,70,7.95,SRS,0,26.46
12,Female,Meningioma,Infratentorial,90,6.54,SRS,0,34.26
13,Female,Meningioma,Infratentorial,90,0.63,SRS,0,32.98
17,Female,Other,Infratentorial,60,24,SRT,1,6.82
23,Female,Other,Infratentorial,80,2.11,SRS,0,10.49
24,Female,Meningioma,Infratentorial,100,2.13,SRS,1,51.02
33,Female,Other,Infratentorial,70,13.45,SRT,1,10.82
52,Female,Meningioma,Infratentorial,90,9.24,SRT,0,26.85
80,Female,Meningioma,Infratentorial,70,2.94,SRS,0,1.54
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -c "Female" BrainCancer.csv | grep Infratentorial
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -e Male -e Supratentorial -c BrainCancer.csv
78
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -e Male -e Supratentorial -v -c BrainCancer.csv
11
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap No.2)

ANS- Command for getting no of Female patient having Infratentrorial brain cancer is **grep -e Male -e Supratentorial -v -c BrainCancer.csv** **= 10 patient**
what this command dose its ignore all Male & Supratentorial -v
& -c count the number(One minus from count beacuse command will also count the head line)

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -e Female -e Infratentorial -v -c BrainCancer.csv
35
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap N0.3)

ANS- Command for getting no of Male patient having Supratentrorial brain cancer is **grep -e Female -e Infratentorial -v -c BrainCancer.csv** **= 34 patient**
what this command dose its ignore all Female & Infratentorial -v
& -c count the number (One minus from count beacuse command will also count the head line)

(4) Find out which patient has the lowest ki value. (Hint: Use sort for this, it's easier)

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ sort -t ',' --key 5 -n -r  BrainCancer.csv
75,Male,Other,Supratentorial,100,24.91,SRT,0,19.74
83,Male,LG glioma,Infratentorial,90,30.41,SRT,0,1.18
2,Male,HG glioma,Supratentorial,90,19.35,SRT,1,8.98
```

(Snap No.4)

Ans:- the 75 th no male having highest ki value of 100
Command :- **sort -t ',' --key 5 -n -r BrainCancer.csv**

(5) Print out the output by searching for the patterns 'SRS' as well as 'infratentorial'. How many such entries are there?

```
BrainCancer.csv.31
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -e Supratentorial -e SRT -v -c BrainCancer.csv
12
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap No. 5)

Ans :- Command : grep -e Supratentorial -e SRT -v -c BrainCancer.csv
'SRS' as well as 'infratentorial' total 11 entries it excludes all Supratentorial &
SRT -c for Count see Snap no 5 & snap no 6

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep -e Supratentorial -e SRT -v  BrainCancer.csv
,sex,diagnosis,loc,ki,gtv,stereo,status,time
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
3,Female,Meningioma,Infratentorial,70,7.95,SRS,0,26.46
12,Female,Meningioma,Infratentorial,90,6.54,SRS,0,34.26
13,Female,Meningioma,Infratentorial,90,0.63,SRS,0,32.98
23,Female,Other,Infratentorial,80,2.11,SRS,0,10.49
24,Female,Meningioma,Infratentorial,100,2.13,SRS,1,51.02
34,Male,Other,Infratentorial,80,6.81,SRS,0,57.11
56,Male,Other,Infratentorial,90,0.48,SRS,0,54.43
57,Male,HG glioma,Infratentorial,80,0.22,SRS,0,33.67
68,Male,Other,Infratentorial,70,0.01,SRS,0,23.67
80,Female,Meningioma,Infratentorial,70,2.94,SRS,0,1.54
```

(Snap No. 6)

(6) List the patients who have survival times more than 50 months.

```
,sex,diagnosis,loc,ki,gtv,stereo,status,time
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ sort -t ',' -k9 -n -r BrainCancer.csv -o sortedfor50.csv
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls
 BrainCancer.csv   lab5.pdf   LinesWithMeningoma.out   sortedfor50.csv  'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ cat sorted,csv
cat: sorted,csv: No such file or directory
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ cat sortedfor50.csv
18,Male,HG glioma,Supratentorial,90,10.8,SRT,0,82.56
65,Male,LG glioma,Supratentorial,80,9.58,SRT,0,78.75
77,Female,Meningioma,Supratentorial,80,2.39,SRS,0,73.74
64,Male,Meningioma,Supratentorial,80,4.77,SRT,0,67.38
30,Female,Meningioma,Supratentorial,90,4.72,SRS,0,65.02
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
76,Female,Meningioma,Supratentorial,80,31.74,SRT,0,57.25
```

(Snap No. 7)

Ans :- a) Command :- sort -t ',' -k9 -n -r BrainCancer.csv -o sortedfor50.csv
This Command will sort the fifth column in dicendiing order & create a file
named sortedfor50.csv see snap No. 7b)After that command cat sortedfor50.csv
to check the content of file see snap No.8

(Snap No. 8)

C) After seeing data i got the above age 50 value is contained only in first 14 lines so i used the head command :- head -n14 sortedfor50.csv see snap No.9

(Snap No. 9)

(7) List the patients who have HG glioma and have survival times more than 50 months.How many are there?

(Snap No. 10)

Ans :- a) Command:- sort -t ',' -k9 -n -r BrainCancer.csv -o sortedfor50.csv
This Command will sort the fifth column in dicendiing order & create a file named sortedfor50.csv see snap No. 10
b)After that command cat sortedfor50.csvto check the content of file see snap No.11

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ head -n14 sortedfor50.csv > sortedtill50.csv
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls
 BrainCancer.csv   lab5.pdf   LinesWithMeningoma.out   sortedfor50.csv   sortedtill50.csv  'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ cat sortedtill50.csv
18,Male,HG glioma,Supratentorial,90,10.8,SRT,0,82.56
65,Male,LG glioma,Supratentorial,80,9.58,SRT,0,78.75
77,Female,Meningioma,Supratentorial,80,2.39,SRS,0,73.74
64,Male,Meningioma,Supratentorial,80,4.77,SRT,0,67.38
30,Female,Meningioma,Supratentorial,90,4.72,SRS,0,65.02
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
76,Female,Meningioma,Supratentorial,80,31.74,SRT,0,57.25
34,Male,Other,Infratentorial,80,6.81,SRS,0,57.11
7,Male,Meningioma,Supratentorial,80,3.19,SRT,0,55.8
56,Male,Other,Infratentorial,90,0.48,SRS,0,54.43
6,Female,Meningioma,Supratentorial,80,4.82,SRS,0,52.75
66,Female,Meningioma,Supratentorial,100,4,SRT,0,52.23
24,Female,Meningioma,Infratentorial,100,2.13,SRS,1,51.02
14,Male,NA,Supratentorial,90,6.38,SRT,0,50.85
```

(Snap No. 11)

c) After that i soretd the file for above 50 month surival & created another file named sortedtill50.csv command :- head -n14 sortedfor50.csv > sortedtill50.csv see Snap No.12

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls
 BrainCancer.csv   lab5.pdf   LinesWithMeningoma.out   sortedfor50.csv   sortedtill50.csv  'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ cat sortedtill50.csv
18,Male,HG glioma,Supratentorial,90,10.8,SRT,0,82.56
65,Male,LG glioma,Supratentorial,80,9.58,SRT,0,78.75
77,Female,Meningioma,Supratentorial,80,2.39,SRS,0,73.74
64,Male,Meningioma,Supratentorial,80,4.77,SRT,0,67.38
30,Female,Meningioma,Supratentorial,90,4.72,SRS,0,65.02
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
76,Female,Meningioma,Supratentorial,80,31.74,SRT,0,57.25
34,Male,Other,Infratentorial,80,6.81,SRS,0,57.11
7,Male,Meningioma,Supratentorial,80,3.19,SRT,0,55.8
56,Male,Other,Infratentorial,90,0.48,SRS,0,54.43
6,Female,Meningioma,Supratentorial,80,4.82,SRS,0,52.75
66,Female,Meningioma,Supratentorial,100,4,SRT,0,52.23
24,Female,Meningioma,Infratentorial,100,2.13,SRS,1,51.02
14,Male,NA,Supratentorial,90,6.38,SRT,0,50.85
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ^C
```

(Snap No. 12)

d) After that Command :- grep HG sortedtill50.csv there only one entery HG glioma and have survival times more than 50 months

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep HG sortedtill50.csv
18,Male,HG glioma,Supratentorial,90,10.8,SRT,0,82.56
```

(Snap No. 13)

(8) List the patients who have LG glioma with location supratentorial and survial times less than 30 months. How many are there?
Note :- i am condering for this prevously created file from question 7 command :- sort -t ',' -k9 -n -r BrainCancer.csv -o sortedfor50.csv   after i creat another file using tail command till 30 moth as per last column
tail -n 50 sortedfor50.csv > revsorted.csv see Snap No.14

(Snap No. 14)

(Snap No. 15)

After That i Sorted the data as per requirement file contain all data below 30 moth data
we want LG glioma Supertentorial **so by command** grep -e Other -e HG -e Infratentorial -e Meningioma -v revsorted.csv **by this command** **all data line which contain Other,HG,Infratentorial,Meningioma will get ignored beacuse of -v in command so we will get only** LG glioma **Supertentorial thus answer for these question is only 4 patient has Lg glioma Supertentorital**

(9) By default does grep ignore or not ignore case? Demonstrate this by using the appro-priate option and by running the grep command without this option.

Ans By deafult grep not ignore the case, grep is case sensitive
For ex data set has LG glioma is we given the lg in samller case than LG we We Wont Get any answer see Snap No.16

(Snap No. 16)

But when we give propre case it give proper result see Snap No. 17

(Snap No. 17)

For ignoring the case of letter -i before is used then it will not be the problem

(10) List the patients who are dead at the end of the study. How many are there? Find another way to listing this set of patients (look up grep man pages to look for a specific option to select non-matching lines)

(Snap No. 19)

ANS;- -v ignore the pattern thus 1 dictates in eight column live & 0 as death
grep -e 1 -v BrainCancer.csv This Command will give only patient dead at last of this expirement. See Snap No.19

(Snap No. 20)

Command grep -e 1 -v -c BrainCancer.csv No such entries required 16 patieent But it will count the head also thus minu one 15 patient dead at the end of experiment.

(11) Let's list the patients who have their cancer in infratentorial location. How many are there? How do I restrict the output to first ten patients?

```
BrainCancer.csv:89:88,Male,Other,Infratentorial,80,0.11,SRT,0,13.9
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep Infratentorial BrainCancer.csv > infsort.csv
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ head -n 10 infsort.csv
1,Female,Meningioma,Infratentorial,90,6.11,SRS,0,57.64
3,Female,Meningioma,Infratentorial,70,7.95,SRS,0,26.46
12,Female,Meningioma,Infratentorial,90,6.54,SRS,0,34.26
13,Female,Meningioma,Infratentorial,90,0.63,SRS,0,32.98
17,Female,Other,Infratentorial,60,24,SRT,1,6.82
23,Female,Other,Infratentorial,80,2.11,SRS,0,10.49
24,Female,Meningioma,Infratentorial,100,2.13,SRS,1,51.02
33,Female,Other,Infratentorial,70,13.45,SRT,1,10.82
34,Male,Other,Infratentorial,80,6.81,SRS,0,57.11
39,Male,Other,Infratentorial,80,12.51,SRT,1,29.7
```

(12) Create an archive file by combining the outputs of questions 1-5.

Note:- For question 1 only created file for 2-5 I Did the Question Without by creating any file thuse i am using othe file
so from 1 questioni have 1 fi;le called LinesWithMeningoma.out so i am creating some other file

```
-rw-rw-r-- 1 ibab ibab    760 Aug  4 22:04  sortedtill50.csv
-rw-rw-r-- 1 ibab ibab 786755 Aug  4 21:56 'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep Meningioma revsorted.csv > Filel1.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep Meningioma sortedfor50.csv > Filel2.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ grep Meningioma sortedtill50.csv > Filel3.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap No. 22)

Creation Files for Question 12 File1.out,File2.out File3.out for this question

```
-rw-rw-r-- 1 ibab ibab    760 Aug  4 22:04  sortedtill50.csv
-rw-rw-r-- 1 ibab ibab 786755 Aug  4 21:56 'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ tar -cvf question2.tar LinesWithMeningoma.out Filel1.out Filel2.out Filel3.out
LinesWithMeningoma.out
Filel1.out
Filel2.out
Filel3.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls-l
ls-l: command not found
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls -l
total 896
-rw-rw-r-- 1 ibab ibab   4843 Aug  1 15:08  BrainCancer.csv
-rw-rw-r-- 1 ibab ibab    837 Aug  4 23:53  Filel1.out
-rw-rw-r-- 1 ibab ibab   2355 Aug  4 23:54  Filel2.out
-rw-rw-r-- 1 ibab ibab    505 Aug  4 23:54  Filel3.out
-rw-rw-r-- 1 ibab ibab   1014 Aug  4 23:32  infsort.csv
-rw-rw-r-- 1 ibab ibab  65679 Aug  1 15:07  lab5.pdf
-rw-rw-r-- 1 ibab ibab   2355 Aug  4 23:51  LinesWithMeningoma.out
-rw-rw-r-- 1 ibab ibab  10240 Aug  4 23:58  question2.tar
-rw-rw-r-- 1 ibab ibab   2690 Aug  4 22:37  revsorted.csv
-rw-rw-r-- 1 ibab ibab   4843 Aug  4 22:31  sortedfor50.csv
-rw-rw-r-- 1 ibab ibab    760 Aug  4 22:04  sortedtill50.csv
-rw-rw-r-- 1 ibab ibab 786755 Aug  4 21:56 'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap No.23)

(13) Create a zipped and archived file by combining the outputs of questions 6-11.

```
rw-rw-r-- 1 ibab ibab 2200023 Aug  5 00:00 Untitled 1.odt
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ tar -czvf Question13.tar.gz BrainCancer.csv revsorted.csv sortedfor50.csv sortedtill50.csv
BrainCancer.csv
revsorted.csv
sortedfor50.csv
sortedtill50.csv
```

(Snap No.24)

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ls -l
total 2284
-rw-rw-r-- 1 ibab ibab    4843 Aug  1 15:08  BrainCancer.csv
-rw-rw-r-- 1 ibab ibab     837 Aug  4 23:53  Filel1.out
-rw-rw-r-- 1 ibab ibab    2355 Aug  4 23:54  Filel2.out
-rw-rw-r-- 1 ibab ibab     505 Aug  4 23:54  Filel3.out
-rw-rw-r-- 1 ibab ibab    1014 Aug  4 23:32  infsort.csv
-rw-rw-r-- 1 ibab ibab   65679 Aug  1 15:07  lab5.pdf
-rw-rw-r-- 1 ibab ibab    2355 Aug  4 23:51  LinesWithMeningoma.out
-rw-rw-r-- 1 ibab ibab    1682 Aug  5 22:04  Question13.tar.gz
-rw-rw-r-- 1 ibab ibab   10240 Aug  4 23:58  question2.tar
-rw-rw-r-- 1 ibab ibab    2690 Aug  4 22:37  revsorted.csv
-rw-rw-r-- 1 ibab ibab    4843 Aug  4 22:31  sortedfor50.csv
-rw-rw-r-- 1 ibab ibab     760 Aug  4 22:04  sortedtill50.csv
-rw-rw-r-- 1 ibab ibab 2206623 Aug  5 00:00 'Untitled 1.odt'
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$
```

(Snap No.25)

(14) List and extract the contents of the archive file created in (12) to a new directory and list the contents of the new directory.

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$  cd Ques14
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques14$ tar -xvf ../question2.tar
LinesWithMeningoma.out
Filel1.out
Filel2.out
Filel3.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques14$ ls -l
total 16
-rw-rw-r-- 1 ibab ibab  837 Aug  4 23:53 Filel1.out
-rw-rw-r-- 1 ibab ibab 2355 Aug  4 23:54 Filel2.out
-rw-rw-r-- 1 ibab ibab  505 Aug  4 23:54 Filel3.out
-rw-rw-r-- 1 ibab ibab 2355 Aug  4 23:51 LinesWithMeningoma.out
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques14$
```

(Snap No.26)

(15) List and extract the contents of the archived-zipped file in (13) to a new directory and list the contents of the new directory.

```
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ mkdir Ques15
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ./Ques
bash: ./Ques: No such file or directory
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ ./Ques15
bash: ./Ques15: Is a directory
ibab@IBAB-MSc-BDB-Comp03:~/Lab5$ cd ./Ques15
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques15$ tar -xzvf ../Question13.tar.gz
BrainCancer.csv
revsorted.csv
sortedfor50.csv
sortedtill50.csv
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques15$ ls -l
total 24
-rw-rw-r-- 1 ibab ibab 4843 Aug  1 15:08 BrainCancer.csv
-rw-rw-r-- 1 ibab ibab 2690 Aug  4 22:37 revsorted.csv
-rw-rw-r-- 1 ibab ibab 4843 Aug  4 22:31 sortedfor50.csv
-rw-rw-r-- 1 ibab ibab  760 Aug  4 22:04 sortedtill50.csv
ibab@IBAB-MSc-BDB-Comp03:~/Lab5/Ques15$ |
```

(Snap No.27)