

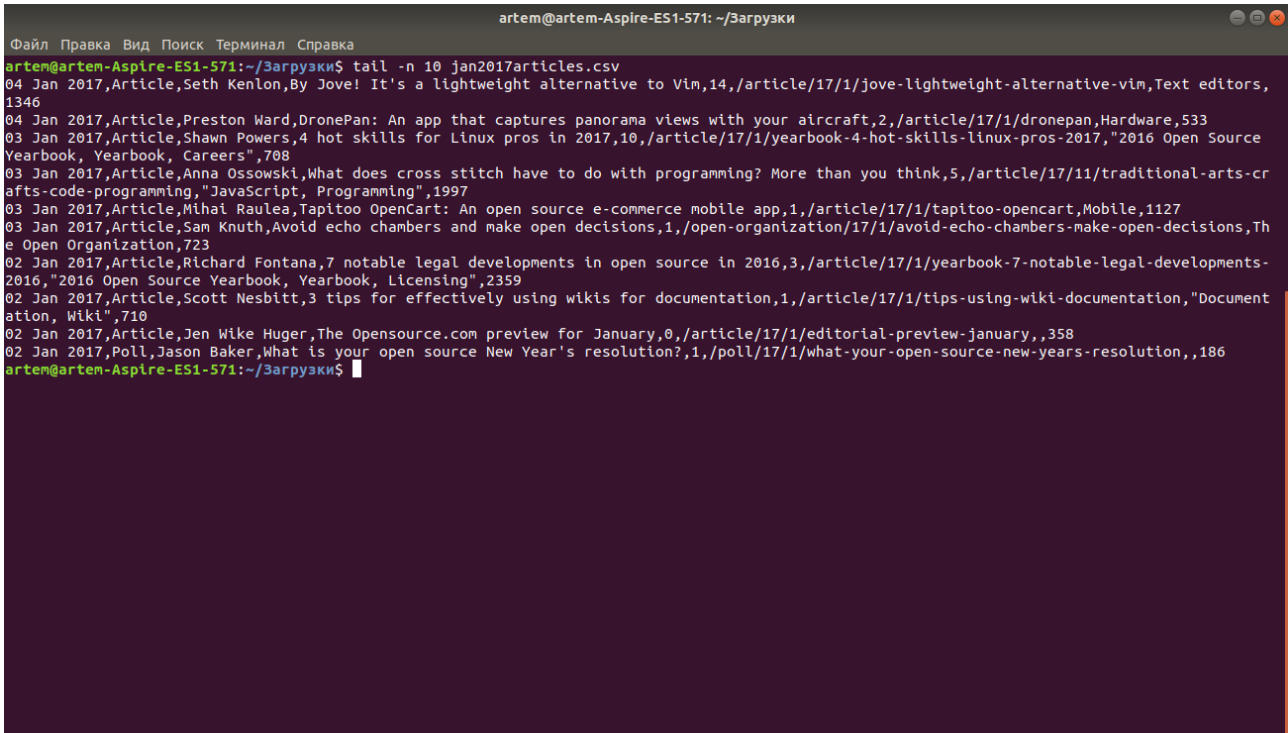
Parallel and distributed systems. Assignment #1.

Artem Reshetnikov, Lawrence Adu-Gyamfi

1. Take a look at the last 10 lines of the file. Which command are you going to use? Modify the command to show just the last line of the file.

Answer: `tail -n 10 jan2017articles.csv`

`tail -n 1 jan2017articles.csv`

A screenshot of a terminal window with a dark background. The title bar at the top reads "artem@artem-Aspire-ES1-571: ~/Зарпрудки". The terminal shows the command `tail -n 10 jan2017articles.csv` being executed. The output consists of 10 lines of CSV data, each starting with a date in "DD Jan 2017" format. The lines are truncated on the right side of the terminal window. The last line of the output is `02 Jan 2017,Poll,Jason Baker,What is your open source New Year's resolution?,1,/poll/17/1/what-your-open-source-new-years-resolution,,186`.

```
Файл Правка Вид Поиск Терминал Справка
artem@artem-Aspire-ES1-571:~/Зарпрудки$ tail -n 10 jan2017articles.csv
04 Jan 2017,Article,Seth Kenlon,By Jove! It's a lightweight alternative to Vim,14,/article/17/1/jove-lightweight-alternative-vim,Text editors,1346
04 Jan 2017,Article,Preston Ward,DronePan: An app that captures panorama views with your aircraft,2,/article/17/1/dronepan,Hardware,533
03 Jan 2017,Article,Shawn Powers,4 hot skills for Linux pros in 2017,10,/article/17/1/yearbook-4-hot-skills-linux-pros-2017,"2016 Open Source Yearbook, Yearbook, Careers",708
03 Jan 2017,Article,Anna Ossowski,What does cross stitch have to do with programming? More than you think,5,/article/17/11/traditional-arts-crafts-code-programming,"JavaScript, Programming",1997
03 Jan 2017,Article,Mihai Raulea,Tapitoo OpenCart: An open source e-commerce mobile app,1,/article/17/1/tapitoo-opencart,Mobile,1127
03 Jan 2017,Article,Sam Knuth,Avoid echo chambers and make open decisions,1,/open-organization/17/1/avoid-echo-chambers-make-open-decisions,The Open Organization,723
02 Jan 2017,Article,Richard Fontana,7 notable legal developments in open source in 2016,3,/article/17/1/yearbook-7-notable-legal-developments-2016,"2016 Open Source Yearbook, Yearbook, Licensing",2359
02 Jan 2017,Article,Scott Nesbitt,3 tips for effectively using wikis for documentation,1,/article/17/1/tips-using-wiki-documentation,"Documentation, Wiki",710
02 Jan 2017,Article,Jen Wike Huger,The Opensource.com preview for January,0,/article/17/1/editorial-preview-january,,358
02 Jan 2017,Poll,Jason Baker,What is your open source New Year's resolution?,1,/poll/17/1/what-your-open-source-new-years-resolution,,186
artem@artem-Aspire-ES1-571:~/Зарпрудки$
```

2. Extract all lines that belong to January 6th from the file and store them in a new file named "reyes.csv". Check that the first line of the new file has the expected values.

Answer: `grep "06 Jan" jan2017articles.csv > reyes.csv`

`- head -1 reyes.csv`

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ grep "06 Jan" jan2017articles.csv
06 Jan 2017,Article,Jen Wike Huger,"Top 5: Hot programming trends, How Linux got to be Linux, and more",0,/article/17/1/top-5-january-6,Top 5,
241
06 Jan 2017,Article,Rikki Endsley,Publisher's picks: Top 2016 open source books ,0,/article/16/12/yearbook-publishers-picks,"2016 Open Source
Yearbook, Books, Yearbook",2034
06 Jan 2017,Article,Marcus D. Hanwell,"Using Clang-format to ensure clean, consistent code",5,/article/17/1/coding-style,,627
06 Jan 2017,Article,Rikki Endsley,Top 10 and editor's picks: December review,0,/article/17/1/December-2016-review-top-10,Top 10,260
артем@артем-Aspire-ES1-571:~/Заргузки$
```

3. Use the original csv to find which entries have 0 at the comment count only for those entries from january 25th

Answer: `grep "25 Jan" jan2017articles.csv | awk -F "," '$5 == "0"'`

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ grep "25 Jan" jan2017articles.csv | awk -F "," '$5 == "0"'
25 Jan 2017,Article,Ben Cotton,24 Pull Requests challenge encourages fruitful contributions,0,/article/17/1/24-Pull-Requests-challenge,Program
ming,429
25 Jan 2017,Article,Rikki Endsley,Announcing the 2016 Open Source Yearbook: Download now,0,/article/17/1/announcing-2016-open-source-yearbook,
2016 Open Source Yearbook,58
артем@артем-Aspire-ES1-571:~/Заргузки$
```

4. Now count the number of entries of Q3 and compare with the total number of entries

Answer: `grep "25 Jan" jan2017articles.csv | awk -F "," '$5 == "0"' | wc -l`
`wc -l jan2017articles.csv`

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ grep "25 Jan" jan2017articles.csv | awk -F "," ' $5 == "0" ' | wc -l
2
артем@артем-Aspire-ES1-571:~/Заргузки$
```

5. Now use example.bed file. In this file, we are interested in the exon sizes of each entry. They are located in field number 11. Now you have to get the exon sizes of the first 10 entries of the file.

Answer: `head -n 10 example.bed | cut -f 11`

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ head -n 10 example.bed | cut -f 11
541,322,429,
385,143,144,186,125,573,
258,19,143,144,186,125,573,
370,107,97,101,57,77,163,98,80,263,
315,113,97,101,57,77,163,98,101,
370,113,97,101,57,77,163,98,80,257,
370,104,97,101,57,77,163,98,80,263,
370,113,97,101,57,77,163,98,80,263,
293,93,81,72,132,87,72,86,133,189,275,
203,96,81,72,132,87,72,86,133,189,275,
артем@артем-Aspire-ES1-571:~/Заргузки$
```

6.How would you remove the last comma?

Answer: `head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'`

```
артем@артем-Aspire-ES1-571: ~/Заргуэки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргуэки$ head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'
```

541,322,429
385,143,144,186,125,573
258,19,143,144,186,125,573
370,107,97,101,57,77,163,98,80,263
315,113,97,101,57,77,163,98,101
370,113,97,101,57,77,163,98,80,257
370,104,97,101,57,77,163,98,80,263
370,113,97,101,57,77,163,98,80,263
293,93,81,72,132,87,72,86,133,189,275
203,96,81,72,132,87,72,86,133,189,275

```
артем@артем-Aspire-ES1-571:~/Заргуэки$
```

7. How would get the smallest size from each of the records? The result should provide a number for each line of the input

Answer 1 : `head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'` `gawk ' {split($0, a, ","); asort(a); printf("%s ", a[1]); printf("\n"); }'`

Answer 2: `head -10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'` `awk -F “,” ‘{min=$1; for(i=2; i<=NF; i++) if($i<min) min=$i; print min}’`

```
артем@артем-Aspire-ES1-571: ~/Заргуэки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргуэки$ head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//'
```

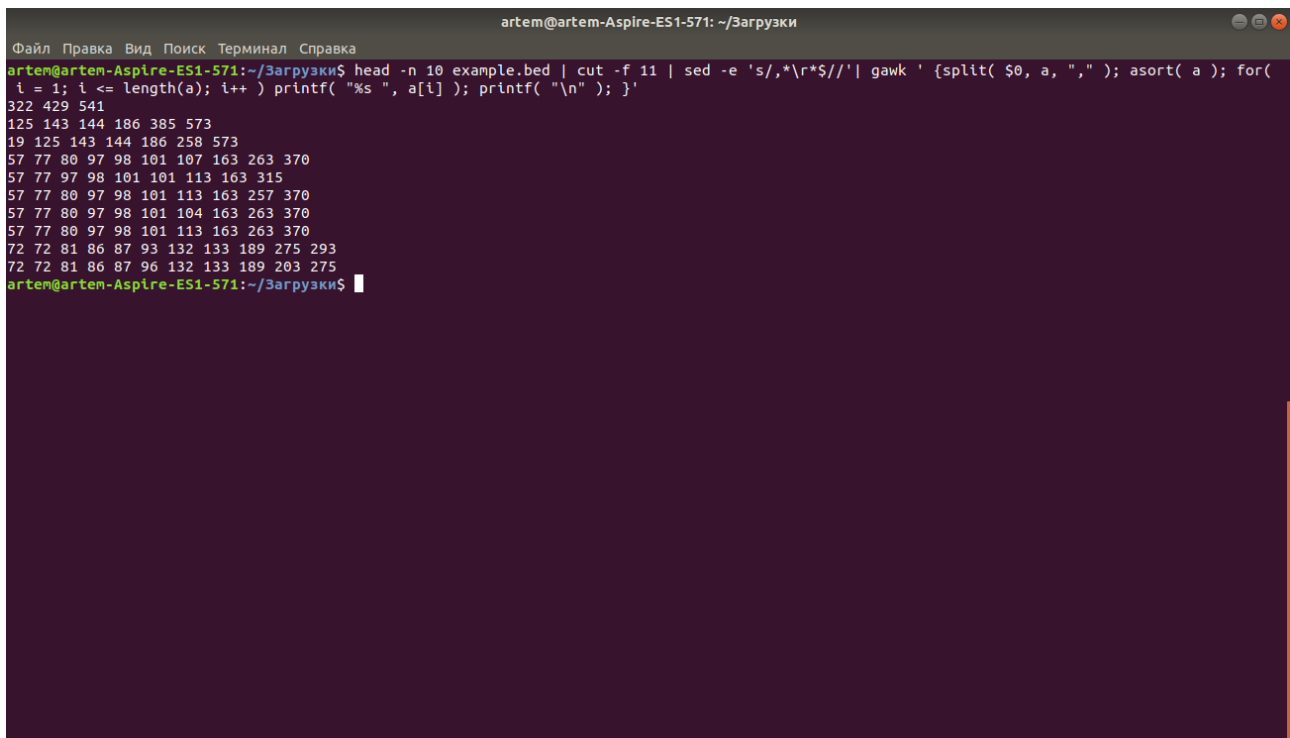
322
125
19
57
57
57
57
57
72
72

```
артем@артем-Aspire-ES1-571:~/Заргуэки$
```

8. How would you now sort the records so that the first number shown is the smallest exon size? Again, the answer must provide a sorted list of numbers for each line of the input

Answer 1: `head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//|' | gawk ' {split($0, a, ","); asort(a); for(i = 1; i <= length(a); i++) printf("%s ", a[i]); printf("\n"); } '`

Answer 2: `head -10 example.bed | cut -f11 | sed 's/,$/g' | awk -F ',' -v OFS=' ' '{j=0; while (j<NF) {for (i=1; i<NF;i++) if ($(i+1) < $(i)) {temp=$(i+1); $(i+1)=$i; $i=temp;} j++} print $0}'`



```
артем@артем-Aspire-ES1-571: ~/Заргуаки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргуаки$ head -n 10 example.bed | cut -f 11 | sed -e 's/,*\r*$//|' | gawk ' {split( $0, a, "," ); asort( a ); for(
i = 1; i <= length(a); i++ ) printf( "%s ", a[i] ); printf( "\n" ); } '
322 429 541
125 143 144 186 385 573
19 125 143 144 186 258 573
57 77 80 97 98 101 107 163 263 370
57 77 97 98 101 101 113 163 315
57 77 80 97 98 101 113 163 257 370
57 77 80 97 98 101 104 163 263 370
57 77 80 97 98 101 113 163 263 370
72 72 81 86 87 93 132 133 189 275 293
72 72 81 86 87 96 132 133 189 203 275
артем@артем-Aspire-ES1-571:~/Заргуаки$
```

9. Now Get the 10 largest exons of chr1 stored in example.bed.

Answer 1: `grep "chr1" example.bed | cut -f11 | sort -n -r | cut -d "," -f1 | head -n 10`

Answer 2: `grep -i 'chr1' example.bed | cut -f11 | awk -F ',' '{max=$1; for (i=1; i<NF; i++) if ($i > max) max = $i; print max}' | sort -rn | head -10`

```
артем@артем-Aspire-ES1-571: ~/Загрузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Загрузки$ grep "chr1" example.bed | cut -f11 | sort -n -r | cut -d "," -f1 | head -n 10
5616
5239
3897
3757
3647
3586
3575
3539
3444
3439
артем@артем-Aspire-ES1-571:~/Загрузки$
```

10. Now modify Q9 script to receive as a parameter the number of exons to search for

Answer: *Text of script*

```
#!/bin/bash
```

```
n=$1
```

```
grep "chr1" example.bed | cut -f11 | sort -n -r | cut -d "," -f1 | head -n $n
```

Text of commands:

```
chmod +x q10.bach
```

```
./q10.bash 15
```

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ ./q10.bash 15
5616
5239
3897
3757
3647
3586
3575
3539
3444
3439
3305
3298
3276
3251
3251
артем@артем-Aspire-ES1-571:~/Заргузки$
```

11. Get the first 10 records of jan2017articles.csv with largest number of comments from the original csv file

Answer: `sort -t "," -k5 -n -r jan2017articles.csv | head -n 10`

```
артем@артем-Aspire-ES1-571: ~/Заргузки
Файл Правка Вид Поиск Терминал Справка
артем@артем-Aspire-ES1-571:~/Заргузки$ sort -t "," -k5 -n -r jan2017articles.csv | head -n 10
17 Jan 2017,Poll,Opensource.com,What is your favorite Linux distribution?,174,/poll/17/1/what-your-favorite-linux-distribution,"Linux, Poll",278
26 Jan 2017,Article,David Both,Using rsync to back up your Linux system,19,/article/17/1/rsync-backup-linux,"Linux, Command line",1838
25 Jan 2017,Article,Don Watkins,Solid state drives in Linux: Enabling TRIM for SSDs,19,/article/17/1/solid-state-drives-linux-enabling-trim-ssds,Linux,573
09 Jan 2017,Article,Jeremy Garcia,Troubleshooting tips for the 5 most common Linux issues,18,/article/17/1/yearbook-linux-troubleshooting-tips,"2016 Open Source Yearbook, Yearbook, The Queue column, Linux",794
18 Jan 2017,Article,David Both,10 reasons to use Cinnamon as your Linux desktop environment,14,/article/17/1/cinnamon-desktop-environment,Linux,1296
04 Jan 2017,Article,Seth Kenlon,By Jove! It's a lightweight alternative to Vim,14,/article/17/1/jove-lightweight-alternative-vim,Text editors,1346
04 Jan 2017,Article,Daniel J Walsh,50 ways to avoid getting hacked in 2017,14,/article/17/1/yearbook-50-ways-avoid-getting-hacked,"Yearbook, 2016 Open Source Yearbook, Security and encryption, Containers, Docker, Linux",2143
30 Jan 2017,Article,David Egts,How to get up and running with sweet Orange Pi,12,/article/17/1/how-to-orange-pi,"Hardware, How-tos and tutorials",933
24 Jan 2017,Article,Stefano Maffulli,Brotli: A new compression algorithm for faster Internet,12,/article/17/1/brotli-compression-algorithm,Internet,590
25 Jan 2017,Article,Jen Wike Huger,Happy birthday to Opensource.com: 7 years of open source,11,/article/17/1/happy-birthday-7,Opensource.com community,244
артем@артем-Aspire-ES1-571:~/Заргузки$
```

12. Modify your previous script to receive a number as a parameter N and then show the top N entries with more comments

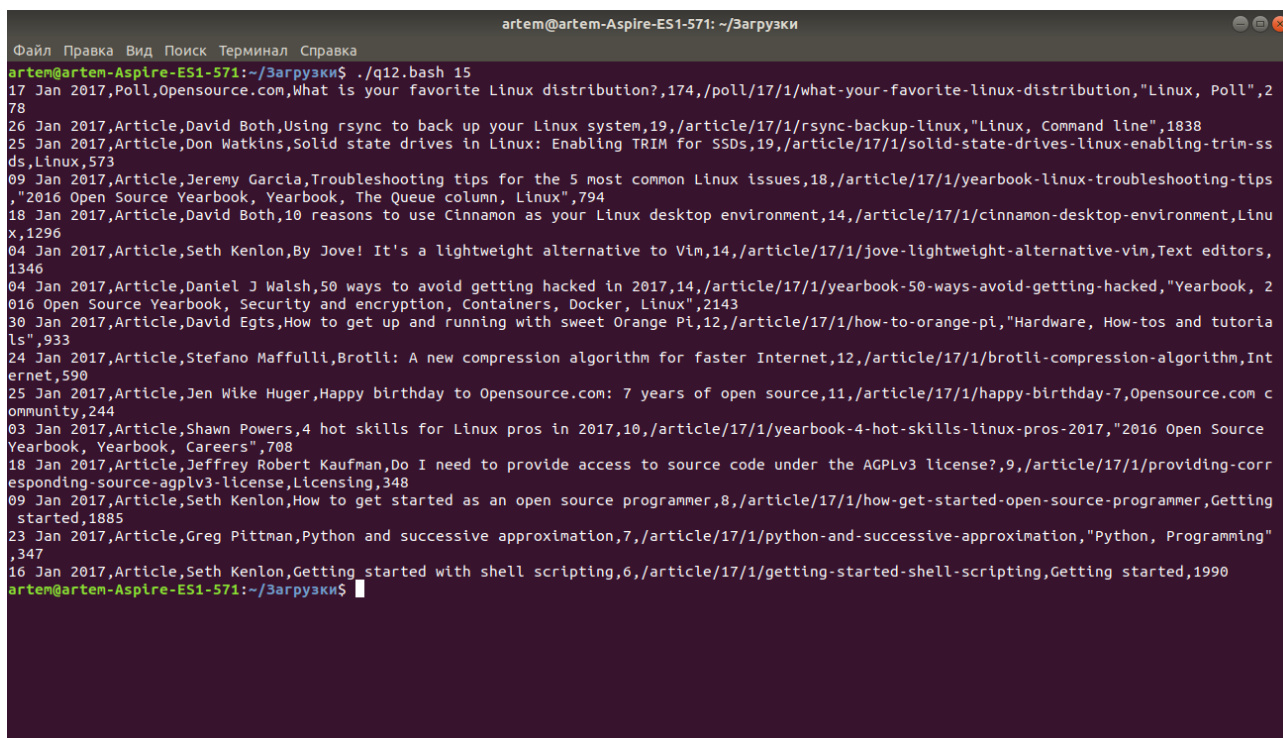
Answer:

Text of script:

```
#!/bin/bash
n=$1
sort -t "," -k5 -n -r jan2017articles.csv | head -n $n
```

Text of commands

```
chmod +x q10.bach
./q12.bash 15
```



```
artem@artem-Aspire-ES1-571:~/Зарпрузки$ ./q12.bash 15
17 Jan 2017,Poll,Opensource.com,What is your favorite Linux distribution?,174,/poll/17/1/what-your-favorite-linux-distribution,"Linux, Poll",278
26 Jan 2017,Article,David Both,Using rsync to back up your Linux system,19,/article/17/1/rsync-backup-linux,"Linux, Command line",1838
25 Jan 2017,Article,Don Watkins,Solid state drives in Linux: Enabling TRIM for SSDs,19,/article/17/1/solid-state-drives-linux-enabling-trim-ssds,Linux,573
09 Jan 2017,Article,Jeremy Garcia,Troubleshooting tips for the 5 most common Linux issues,18,/article/17/1/yearbook-linux-troubleshooting-tips,"2016 Open Source Yearbook, Yearbook, The Queue column, Linux",794
18 Jan 2017,Article,David Both,10 reasons to use Cinnamon as your Linux desktop environment,14,/article/17/1/cinnamon-desktop-environment,Linux,1296
04 Jan 2017,Article,Seth Kenlon,By Jove! It's a lightweight alternative to Vim,14,/article/17/1/jove-lightweight-alternative-vim,Text editors,1346
04 Jan 2017,Article,Daniel J Walsh,50 ways to avoid getting hacked in 2017,14,/article/17/1/yearbook-50-ways-avoid-getting-hacked,"Yearbook, 2016 Open Source Yearbook, Security and encryption, Containers, Docker, Linux",2143
30 Jan 2017,Article,David Egts,How to get up and running with sweet Orange Pi,12,/article/17/1/how-to-orange-pi,"Hardware, How-tos and tutorials",933
24 Jan 2017,Article,Stefano Maffulli,Brotli: A new compression algorithm for faster Internet,12,/article/17/1/brotli-compression-algorithm,Internet,590
25 Jan 2017,Article,Jen Wike Huger,Happy birthday to Opensource.com: 7 years of open source,11,/article/17/1/happy-birthday-7,Opensource.com community,244
03 Jan 2017,Article,Shawn Powers,4 hot skills for Linux pros in 2017,10,/article/17/1/yearbook-4-hot-skills-linux-pros-2017,"2016 Open Source Yearbook, Yearbook, Careers",708
18 Jan 2017,Article,Jeffrey Robert Kaufman,Do I need to provide access to source code under the AGPLv3 license?,9,/article/17/1/providing-corresponding-source-agplv3-license,Licensing,348
09 Jan 2017,Article,Seth Kenlon,How to get started as an open source programmer,8,/article/17/1/how-get-started-open-source-programmer,Getting started,1885
23 Jan 2017,Article,Greg Pittman,Python and successive approximation,7,/article/17/1/python-and-successive-approximation,"Python, Programming",347
16 Jan 2017,Article,Seth Kenlon,Getting started with shell scripting,6,/article/17/1/getting-started-shell-scripting,Getting started,1990
artem@artem-Aspire-ES1-571:~/Зарпрузки$
```

13. Now we are going to create a new articles.csv where we get a different output data layout using awk tool INPUT: Post date,Content type,Author,Title,Comm count,Path,Tags,Word count OUTPUT: Title;Comment count;Word count;Post date

Answer: `awk -F "," '{print $4"\t"$5"\t"$8"\t"$1}' jan2017articles.csv > articles.csv`


```

артем@артем-Aspire-E51-571: ~/Зарпузки
Файл Правка Вид Поиск Терминал Справка

артем@артем-Aspire-E51-571:~/Зарпузки$ awk -F "," '{print $4"\t"$5"\t"$8"\t"$1}' jan2017articles.csv
Title Comment count Word count Post date
Book review: Ours to Hack and to Own 0 660 31 Jan 2017
5 new guides for working with OpenStack 2 How-tos and tutorials" 31 Jan 2017
Be the open source supply chain 1 1668 31 Jan 2017
Developing open leaders 1 768 31 Jan 2017
How to get up and running with sweet Orange Pi 12 How-tos and tutorials" 30 Jan 2017
4 ways to improve your security online right now 3 1242 30 Jan 2017
" WOOTConf 2017: Lockpicking Willie Nelson developers /article/17/1/women-open-technology-miniconf 30 Jan 2017
"From hobbyist to professional new analyst papers /article/17/1/openstack-news-january-30 30 Jan 2017
How communities in India support privacy and software freedom 0 453 28 Jan 2017
"New Minecraft launcher comes to Linux Tilt Brush Toolkit /article/17/1/open-gaming-news-january-28 28 Jan 2017
"Top 5: Solid state drives in Linux Brotli compression algorithm /article/17/1/top-5-january-27 27 Jan 2017
Data Privacy Day 2017: Solutions for everyday privacy 5 Security and encryption" 27 Jan 2017
How to successfully enter the FOSS emerging market 0 915 27 Jan 2017
Getting started with my new LulzBot Mini 3D printer 2 1390 27 Jan 2017
A 5-step plan to encourage your team to make changes on your project 5 1160 26 Jan 2017
Search this database for inactive patents that are now in the public domain 0 Patents" 26 Jan 2017
Using rsync to back up your Linux system 19 Command line" 26 Jan 2017
How to join a technical community 1 Community" 26 Jan 2017
Solid state drives in Linux: Enabling TRIM for SSDs 19 573 25 Jan 2017
An executive's guide to containers 4 2032 25 Jan 2017
24 Pull Requests challenge encourages fruitful contributions 0 429 25 Jan 2017
Happy birthday to Opensource.com: 7 years of open source 11 244 25 Jan 2017
Announcing the 2016 Open Source Yearbook: Download now 0 58 25 Jan 2017
Brotli: A new compression algorithm for faster Internet 12 590 24 Jan 2017
School systems desperate for standards-aligned curricula find hope 0 Education" 24 Jan 2017
What your code repository says about you 1 1085 24 Jan 2017
Unpicking the community leader 0 Six Degrees column" 23 Jan 2017
Win a LulzBot Mini 3D printer in our giveaway 1 Getting started" 23 Jan 2017
Shell scripting: An introduction to the shift method and custom functions 4 Programming" 23 Jan 2017
Python and successive approximation 7 Programming" 23 Jan 2017
2017 People's Choice Awards: Cast your vote 3 562 23 Jan 2017
"Tips for instance configuration creating a new project /article/17/1/openstack-news-january-23 23 Jan 2017
"The new Raspberry Pi Compute Module FSF's 2017 priorities /article/17/1/news-january-21 21 Jan 2017
5 ways to expand your project's contributor base 2 690 20 Jan 2017
How to write web apps in R with Shiny 2 218 20 Jan 2017
"Top 5: Shell scripting the fangsaw linux desktop environment /article/17/1/top-5-january-20 20 Jan 2017

```

14. Now create a new article2.csv format where we cut the Title text to 10 characters and we get only the last level of the Path.

Answer: `awk -F "," '{split($6,arr,"/"); print substr($4,1,10) "\t" "/"arr[4]"/"arr[5]}' jan2017articles.csv > article2.csv`

```

артем@артем-Aspire-E51-571: ~/Зарпузки
Файл Правка Вид Поиск Терминал Справка

артем@артем-Aspire-E51-571:~/Зарпузки$ awk -F "," '{split($6,arr,"/"); print substr($4,1,10) "\t" "/"arr[4]"/"arr[5]}' jan2017articles.csv
Title //
Book revie /1/review-book-ours-to-hack-and-own
5 new guid /1/openstack-tutorials
Be the ope /1/be-open-source-supply-chain
Developing /1/developing-open-leaders
How to get /1/how-to-orange-pi
4 ways to /1/4-ways-improve-your-online-security
" WOOTConf //
"From hobb //
How commun /1/how-communities-india-support-privacy-software-freedom
"New Minec //
"Top 5: So //
Data Priva /1/every-day-privacy
How to suc /1/cultivating-business-foss-market
Getting st /1/beginner-LulzBot-Mini-3D-printer
A 5-step p /1/escape-the-cave
Search thi /1/making-us-patent-system-useful-again
Using rsyn /1/rsync-backup-linux
How to joi /1/how-join-technical-community
Solid stat /1/solid-state-drives-linux-enabling-trim-ssds
An executi /1/container-strategy-for-executives
24 Pull Re /1/24-Pull-Requests-challenge
Happy birt /1/happy-birthday-7
Announcing /1/announcing-2016-open-source-yearbook
Brotli: A /1/brotli-compression-algorithm
School sys /1/open-up-resources
What your /1/repo-tells-a-story
Unpicking /1/community-leadership-matters
Win a Lulz /1/getting-started-giveaway
Shell scri /1/shell-scripting-shift-method-custom-functions
Python and /1/python-and-successive-approximation
2017 Peopl /1/2017-peoples-choice-awards
"Tips for //
"The new R //
5 ways to /1/expand-project-contributor-base
How to wri /1/writing-new-web-apps-shiny
"Top 5: Sh //

```