

FIT5145

## Assignment 3

1.

unzip -q FacebookNews.zip

ls -hl (Size of the entire file 90M)

```
sai@sai-VirtualBox: ~/Desktop/FIT5145/Assignment-3
File Edit View Search Terminal Help
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3$ unzip -q FacebookNews.zip
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3$ 
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3$ ls -hl FacebookNews
total 90M
-rw-r--r-- 1 sai sai 26M Jun 24 10:11 abc-news-86680728811.csv
-rw-r--r-- 1 sai sai 3.3M Jun 24 10:20 fox-and-friends-111938618893743.csv
-rw-r--r-- 1 sai sai 34M Jun 24 10:25 the-new-york-times-5281959998.csv
-rw-r--r-- 1 sai sai 27M Jun 24 10:23 usa-today-13652355666.csv
```

2.

The commas are used to separate the columns in the file.

Abc-news

No. of Columns: 20

\$ head -1 abc-news-86680728811.csv | sed 's/[^,\$]//g'

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -1 abc-news-86680728811.csv | sed 's/[^,&]//g'
,,,,,,,,,,,,,,,,,,,,,
```

fox and friends

No. of Columns: 20

\$ head -1 fox-and-friends-111938618893743.csv | sed 's/[^,\$]//g'

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -1 fox-and-friends-111938618893743.csv | sed 's/[^,&]//g'
,,,,,,,,,,,,,,,,,,,,,
```

new york times

No. of Columns: 20

\$ head -1 the-new-york-times-5281959998.csv | sed 's/[^,\$]//g'

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -1 the-new-york-times-5281959998.csv | sed 's/[^,&]//g'
,,,,,,,,,,,,,,,,,,,,,
```

usa today

No. of Columns: 20

\$ head -1 usa-today-13652355666.csv | sed 's/[^,\$]//g'

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -1 usa-today-13652355666.csv | sed 's/[^,&]//g'
,,,,,,,,,,,,,,,,,,,,,
```

3.

Abc-news:

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -n 1 abc-news-86680728811.csv
"id","page_id","name","message","description","caption","post_type","status_type","likes_count","comments_count","shares_count","love_count","wow_count","haha_count","sad_count","thankful_count","angry_count","link","picture","posted_at"
```

fox and friends:

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -n 1 fox-and-friends-111938618893743.csv
"id","page_id","name","message","description","caption","post_type","status_type","likes_count","comments_count","shares_count","love_count","wow_count","haha_count","sad_count","thankful_count","angry_count","link","picture","posted_at"
```

new york times:

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -n 1 the-new-york-times-5281959998.csv
"id","page_id","name","message","description","caption","post_type","status_type","likes_count","comments_count","shares_count","love_count","wow_count","haha_count","sad_count","thankful_count","angry_count","link","picture","posted_at"
```

usa times:

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ head -n 1 usa-today-13652355666.csv
"id","page_id","name","message","description","caption","post_type","status_type","likes_count","comments_count","shares_count","love_count","wow_count","haha_count","sad_count","thankful_count","angry_count","link","picture","posted_at"
```

4.

Abc-news:

cat abc-news-86680728811.csv | wc -l

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cat abc-news-86680728811.csv | wc -l
43281
```

fox and friends:

cat fox-and-friends-111938618893743.csv | wc -l

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cat fox-and-friends-111938618893743.csv | wc -l
5959
```

new york times

cat the-new-york-times-5281959998.csv | wc -l

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cat the-new-york-times-5281959998.csv | wc -l
47868
```

usa today:

cat usa-today-13652355666.csv | wc -l

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cat usa-today-13652355666.csv | wc -l
38275
```

No of rows: 38275

5.

Abc-news:

cut -d "," -f 20 abc-news-86680728811.csv | head -2

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 20 abc-news-86680728811.csv | head -2  
"posted_at"  
"2012-01-01 00:30:26"
```

cut -d "," -f 20 abc-news-86680728811.csv | head -43281

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 21 abc-news-86680728811.csv | tail -1  
"2016-11-07 23:47:06"
```

fox and friends:

cut -d "," -f 20 fox-and-friends-111938618893743.csv | head -2

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 21 fox-and-friends-111938618893743.csv | head -2  
  
"2012-01-25 19:20:55"
```

cut -d "," -f 20 fox-and-friends-111938618893743.csv | tail -1

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 21 abc-news-86680728811.csv | tail -1  
"2016-11-07 23:47:06"
```

new york times:

Start Date: cat -d "," -f 26 the-new-york-times-5281959998.csv | head -2

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 26 the-new-york-times-5281959998.csv | head -2  
  
"2012-09-08 15:16:55"
```

2012-09-08 15:16:55

End Date: cat -d "," -f 24 the-new-york-times-5281959998.csv | tail -1

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 24 the-new-york-times-5281959998.csv | tail -1  
"2016-11-07 23:55:00"
```

2016-11-07 23:55:00

usa today:

Start Date: cat -d "," -f 22 usa-today-13652355666.csv | head -2

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 22 usa-today-13652355666.csv | head -2  
  
"2012-03-19 21:31:50"
```

2012-03-19 21:31:50

End Date: cat -d "," -f 20 usa-today-13652355666.csv | tail -1

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ cut -d ",  
" -f 20 usa-today-13652355666.csv | tail -1  
"2016-11-07 23:45:00"
```

2016-11-07 23:45:00

6.

Abc-news:

awk -F '"',"' '\$3 !~ /NULL/' {print \$3}' abc-news-86680728811.csv | uniq -u | wc -l

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 !~ /NULL/ {print $3}' abc-news-86680728811.csv | uniq -u | wc -l
37775
```

fox and friends:

```
awk -F '"',"' '$3 !~ /NULL/ {print $3}' fox-and-friends-111938618893743.csv | uniq -u | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 !~ /NULL/ {print $3}' fox-and-friends-111938618893743.csv | uniq -u | wc -l
1730
```

new york times:

```
awk -F '"',"' '$3 !~ /NULL/ {print $3}' the-new-york-times-5281959998.csv | uniq -u | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 !~ /NULL/ {print $3}' the-new-york-times-5281959998.csv | uniq -u | wc -l
45464
```

usa today:

```
awk -F '"',"' '$3 !~ /NULL/ {print $3}' usa-today-13652355666.csv | uniq -u | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 !~ /NULL/ {print $3}' usa-today-13652355666.csv | uniq -u | wc -l
31404
```

7.

Abc- news:

```
awk -F '","' '$3 ~ /NULL/ {print $3}' abc-news-86680728811.csv | uniq -c
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 ~ /NULL/' abc-news-86680728811.csv | wc -l
1375
```

fox and friends:

```
awk -F '',' '$3 ~ /NULL/ {print $3}' fox-and-friends-111938618893743.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 ~ /NULL/' fox-and-friends-111938618893743.csv | wc -l
120
```

new york times:

```
awk -F '',' '$3 ~ /NULL/ {print $3}' the-new-york-times-5281959998.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 ~ /NULL/' the-new-york-times-5281959998.csv | wc -l
1729
```

usa today:

```
awk -F '',' '$3 ~ /NULL/ {print $3}' the-new-york-times-5281959998.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F '"',"' '$3 ~ /NULL/' usa-today-13652355666.csv | wc -l
1159
```

8.

new york times:

```
grep -m1 'Italian food' the-new-york-times-5281959998.csv | head -1 | awk -F',' '{print $3,$21}'
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -m1 'Italian food' the-new-york-times-5281959998.csv
| head -1 | awk -F',' '{print $3,$21}'
"Remembering Marcella" "2013-11-07 22:26:29"
```

9.

Abc-news:

```
grep -o 'Hillary Clinton' abc-news-86680728811.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Hillary Clinton' abc-news-86680728811.csv | wc -l
1911
```

fox and friends:

```
grep -o 'Hillary Clinton' fox-and-friends-111938618893743.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Hillary Clinton' fox-and-friends-111938618893743
.csv | wc -l
253
```

new york times:

```
grep -o 'Hillary Clinton' the-new-york-times-5281959998.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Hillary Clinton' the-new-york-times-5281959998.c
sv | wc -l
1974
```

usa times:

```
grep -o 'Hillary Clinton' usa-today-13652355666.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Hillary Clinton' usa-today-13652355666.csv | wc -
l
851
```

For each of the files searching for the term 'Hillary Clinton' and counting the number of occurrences will fetch me the number of times Hillary Clinton was repeated and then adding them all up.

10.

abc news:

800

```
grep -o 'Donald Trump' abc-news-86680728811.csv | wc -l
```

fox and friends:

305

```
grep -o 'Donald Trump' fox-and-friends-111938618893743.csv | wc -l
```

new york times:

1445

```
grep -o 'Donald Trump' the-new-york-times-5281959998.csv | wc -l
```

usa times:

1051

```
grep -o 'Donald Trump' usa-today-13652355666.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Donald Trump' abc-news-86680728811.csv | wc -l
800
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Donald Trump' fox-and-friends-111938618893743.cs
v | wc -l
305
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Donald Trump' the-new-york-times-5281959998.csv |
wc -l
1445
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -o 'Donald Trump' usa-today-13652355666.csv | wc -l
1051
```

So overall for Donald Trump:  $800 + 305 + 1445 + 1051 = 3601$

So overall for Hillary Clinton:  $1911 + 253 + 1974 + 851 = 4989$

Hillary Clinton has more focus.

## 11.

```
awk -F",", " 'tolower($4) ~/trump/ {print $1 ":::::::::" $9}' abc-news-86680728811.csv | awk -
F':::::::::' '{if ($2 > 100) print $1 ":" $2}' | sort -t ":" -k 2n > trump1.txt
```

```
awk -F",", " 'tolower($4) ~/trump/ {print $1 ":::::::::" $9}' the-new-york-times-5281959998.csv |
awk -F':::::::::' '{if ($2 > 100) print $1 ":" $2}' | sort -t ":" -k 2n > trump2.txt
```

```
awk -F",", " 'tolower($4) ~/trump/ {print $1 ":::::::::" $9}' fox-and-friends-111938618893743.csv |
awk -F':::::::::' '{if ($2 > 100) print $1 ":" $2}' | sort -t ":" -k 2n > trump3.txt
```

```
awk -F",", " 'tolower($4) ~/trump/ {print $1 ":::::::::" $9}' usa-today-13652355666.csv | awk -
F':::::::::' '{if ($2 > 100) print $1 ":" $2}' | sort -t ":" -k 2n > trump4.txt
```

```
cat trump1.txt trump2.txt trump3.txt trump4.txt > trump.txt
```

## 12.

a.

```
awk -F",", " 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' abc-news-
86680728811.csv
```

```
awk -F",", " 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' abc-news-
86680728811.csv
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ awk -F",", " 'tolower($4) ~/donald trump/ {sum += $12} E
ND {print sum}' abc-news-86680728811.csv
19855
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ awk -F",", " 'tolower($4) ~/donald trump/ {sum += $17} E
ND {print sum}' abc-news-86680728811.csv
114927
```

```
awk -F",", " 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' abc-news-
86680728811.csv
```



```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' abc-news-86680728811.csv
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' abc-news-86680728811.csv
227040
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' abc-news-86680728811.csv
864306
```

b.

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' fox-and-friends-111938618893743.csv
```

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' fox-and-friends-111938618893743.csv
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' fox-and-friends-111938618893743.csv
1913
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' fox-and-friends-111938618893743.csv
122233
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' fox-and-friends-111938618893743.csv
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' fox-and-friends-111938618893743.csv
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' fox-and-friends-111938618893743.csv
3070
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' fox-and-friends-111938618893743.csv
82074
```

c.

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' the-new-york-times-5281959998.csv
```

```
11333
```

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' the-new-york-times-5281959998.csv
```

```
21421
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' the-new-york-times-5281959998.csv
11333
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' the-new-york-times-5281959998.csv
21421
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' the-new-york-times-5281959998.csv
```

```
174282
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' the-new-york-times-5281959998.csv
261538
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' the-new-york-times-5281959998.csv
174282
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' the-new-york-times-5281959998.csv
261538
```

d.

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' usa-today-13652355666.csv
14925
```

```
awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' usa-today-13652355666.csv
217090
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $12} END {print sum}' usa-today-13652355666.csv
14925
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/donald trump/ {sum += $17} END {print sum}' usa-today-13652355666.csv
217090
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' usa-today-13652355666.csv
17224
```

```
awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' usa-today-13652355666.csv
258625
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $12} END {print sum}' usa-today-13652355666.csv
17224
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ awk -F"," 'tolower($4) ~/hillary clinton/ {sum += $17} END {print sum}' usa-today-13652355666.csv
258625
```

### 13.

Donald Trump

```
grep -E "[Tt]rump*[Pp]utin* | [Pp]utin*[Tt]rump*" abc-news-86680728811.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/FacebookNews$ grep -E "[Tt]rump*[Pp]utin* | [Pp]utin*[Tt]rump*" abc-news-86680728811.csv | wc -l
19
```

```
grep -E "[Tt]rump*[Pp]utin* | [Pp]utin*[Tt]rump*" fox-and-friends-111938618893743.csv | wc -l
```



```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E "[Tt]rump*[Pp]utin* | *[Pp]utin*[Tt]rump*" f
ox-and-friends-111938618893743.csv | wc -l
2
```

```
grep -E "[Tt]rump*[Pp]utin* | *[Tt]rump*[Pp]utin*" the-new-york-times-5281959998.csv | wc
-l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E "[Tt]rump*[Pp]utin* | *[Pp]utin*[Tt]rump*" t
he-new-york-times-5281959998.csv | wc -l
27
```

```
grep -E "[Tt]rump*[Pp]utin* | *[Pp]utin*[Tt]rump*" usa-today-13652355666.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E "[Tt]rump*[Pp]utin* | *[Pp]utin*[Tt]rump*" u
sa-today-13652355666.csv | wc -l
14
```

Hillary Clinton

```
grep -E '[Tt]rump' abc-news-86680728811.csv | grep -v '[Cc]linton' | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E '[Tt]rump' abc-news-86680728811.csv | grep -v '[C
c]linton' | wc -l
1764
```

```
grep -E '[Tt]rump' fox-and-friends-111938618893743.csv | grep -v '[Cc]linton'
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E '[Tt]rump' fox-and-friends-111938618893743.csv |
grep -v '[Cc]linton' | wc -l
496
```

```
grep -E '[Tt]rump' the-new-york-times-5281959998.csv | grep -v '[Cc]linton'
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E '[Tt]rump' the-new-york-times-5281959998.csv | gr
ep -v '[Cc]linton' | wc -l
1891
```

```
grep -E '[Tt]rump' usa-today-13652355666.csv | grep -v '[Cc]linton'
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ grep -E '[Tt]rump' usa-today-13652355666.csv | grep -v '[
Cc]linton' | wc -l
1367
```

## Task B: Graphing the data in R

1.

Abc-news:

```
awk -F'"',"' '$4 ~/Trump/ {print $4}' abc-news-86680728811.csv | wc -l
```

```
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo
kNews$ awk -F'"',"' '$4 ~/Trump/ {print $4}' abc-news-86680728811
.csv | wc -l
2143
```

fox and friends:

```
awk -F"," ' $4 ~/Trump/ {print $4}' fox-and-friends-111938618893743.csv | wc -l  
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo  
kNews$ awk -F"," ' $4 ~/Trump/ {print $4}' fox-and-friends-1119  
38618893743.csv | wc -l  
128
```

new york times:

```
awk -F"," ' $4 ~/Trump/ {print $4}' the-new-york-times-5281959998.csv | wc -l  
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo  
kNews$ awk -F"," ' $4 ~/Trump/ {print $4}' the-new-york-times-5  
281959998.csv | wc -l  
1878
```

Usa times:

```
awk -F"," ' $4 ~/Trump/ {print $4}' usa-today-13652355666.csv | wc -l  
(base) sai@sai-VirtualBox:~/Desktop/FIT5145/Assignment-3/Faceboo  
kNews$ awk -F"," ' $4 ~/Trump/ {print $4}' usa-today-1365235566  
6.csv | wc -l  
987
```

2.

Step 1: Searched for the word “Donald Trump” in all the 4 files and extracted just the time stamp columns and wrote it to 4 new files respectively.

```
- awk -F"," ' {print $20}' abc-news-86680728811.csv > trump-abc.csv  
- awk -F"," ' {print $20}' fox-and-friends-111938618893743.csv > trump-fox.csv  
- awk -F"," ' {print $20}' the-new-york-times-5281959998.csv > trump-new.csv  
- awk -F"," ' {print $20}' usa-today-13652355666.csv > trump-usa.csv
```

Step 2: Read the data in four different variables

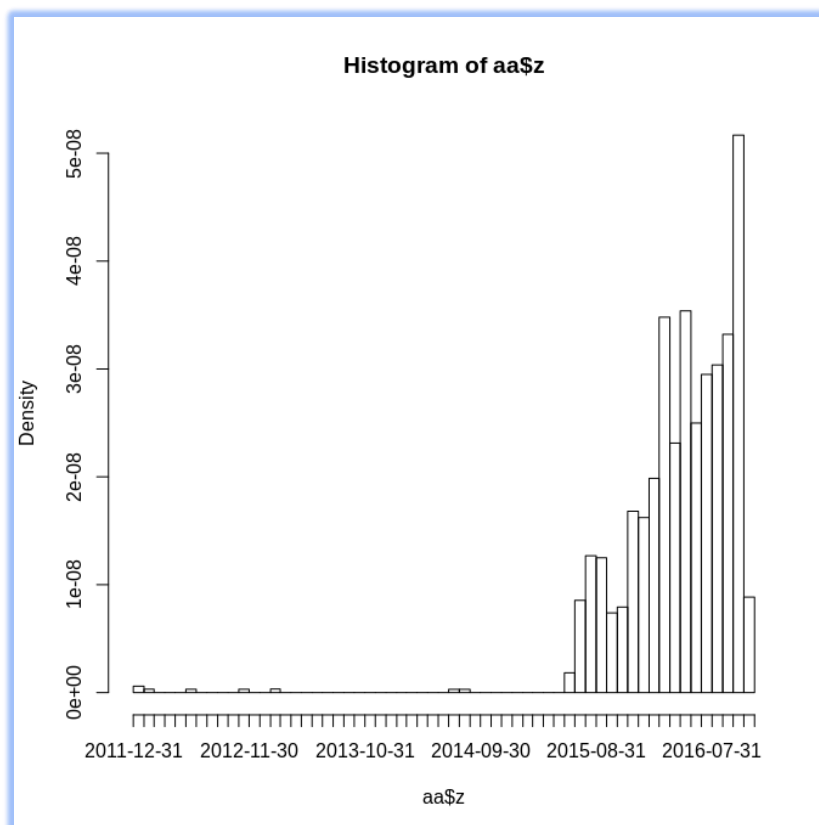
Step 3: Concatenated all the new files together using rbind()

Step 4: Then, converted datatype of the entire column to POSIXct and extracted only the date component and excluded the time stamp.

```
z <- as.POSIXct(combined$V1, format="%Y-%m-%d")
```

Step 5: Plotting the histogram

```
hist(aa$z, breaks="month")
```



The graph is plotted on a monthly basis and it can be seen that the graph shows an evident increase in trend as the years/ months pass by starting from 2015-08-31. This means from the year 2015, Donald Trump started getting mentioned on facebook more and more. This can be due to his electoral win in the year 2015.

3.

3.1

abc-news:

```
cat abc-news-86680728811.csv | grep -i "abc-news" > abc_abc.csv
cat fox-and-friends-111938618893743.csv | grep -i "abc-news" > fox_abc.csv
cat the-new-york-times-5281959998.csv | grep -i "abc-news" > new_abc.csv
cat usa-today-13652355666.csv | grep -i "abc-news" > usa_abc.csv
```

fox-news:

```
cat usa-today-13652355666.csv | grep -i "fox-news" > usa_fox.csv
cat abc-news-86680728811.csv | grep -i "fox-news" > abc_fox.csv
cat fox-and-friends-111938618893743.csv | grep -i "fox-news" > fox_fox.csv
cat the-new-york-times-5281959998.csv | grep -i "fox-news" > new_fox.csv
```

cnn:

```
cat abc-news-86680728811.csv | grep -i "cnn" > abc_cnn.csv
cat fox-and-friends-111938618893743.csv | grep -i "cnn" > fox_cnn.csv
cat the-new-york-times-5281959998.csv | grep -i "cnn" > new_cnn.csv
cat usa-today-13652355666.csv | grep -i "cnn" > usa_cnn.csv
```

### 3.2

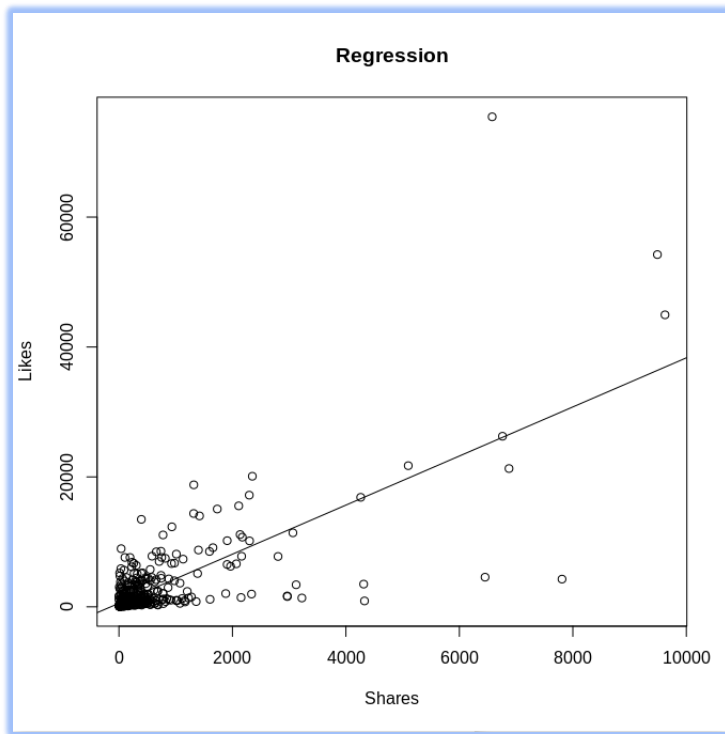
Step 1: Read in the data

Step 2: Check for relationship between likes\_count and shares\_count columns `cor.test(df$'V9', df$'V11', method=c("pearson"))`

### 3.3

Plotting

- `plot(df$V11, df$V9, main="Regression", xlab="Share No.", ylab="Likes No.")`
- `abline(lm(df$'V9' ~ df$'V11'))`



### 3.4

```
linearMod <- lm( V9 ~ V11, data=df)
summary(linearMod)
```

```
set.seed(49)
trainRowIndex <- sample(1:nrow(df), 0.8 * nrow(df))
trainingData <- df[trainRowIndex,]
testData <- df[-trainRowIndex,]
lmMod <- lm(V9 ~ V11, data = trainingData)
predic <- predict(lmMod, testData)
```

```
predict(lmMod, data.frame(shares_count = 0)) = 483
predict(lmMod, data.frame(shares_count = 100)) = 901
predict(lmMod, data.frame(shares_count = 1000)) = 4661
predict(lmMod, data.frame(shares_count = 10000)) = 42256
predict(lmMod, data.frame(shares_count = 100000)) = 418212
```