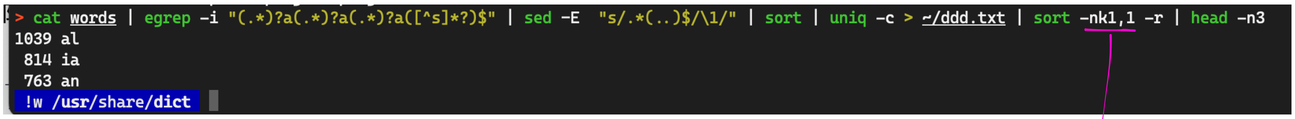
2.
统计words文件 (/usr/share/dict/words) 中包含至少三个a 且不以's 结尾的单词个数: 6830

```
cat words | egrep -i "(.*)?a(.*)?a(.*)?a([^s]*?)$" | wc -l
```

这些单词中，出现频率前三的末尾两个字母是什么?



```
> cat words | egrep -i "(.*)?a(.*)?a(.*)?a([^s]*?)$" | sed -E  "s/.*(..)$/\1/" | sort | uniq -c > ~/ddd.txt | sort -nk1,1 -r | head -n3
1039 al
 814 ia
 763 an
 !w /usr/share/dict
```

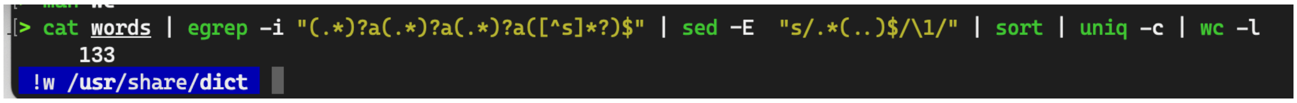
-n

对数字进行排序

-k1,1

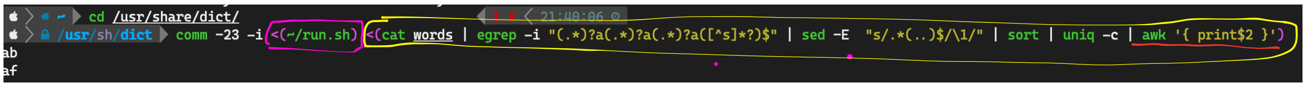
-k1,1 则表示“仅基于以空格分割的（第1列）到（第1列）为止来排序。

共存在多少种词尾两字母组合?



```
> cat words | egrep -i "(.*)?a(.*)?a(.*)?a([^s]*?)$" | sed -E  "s/.*(..)$/\1/" | sort | uniq -c | wc -l
     133
 !w /usr/share/dict
```

哪个组合从未出现过?

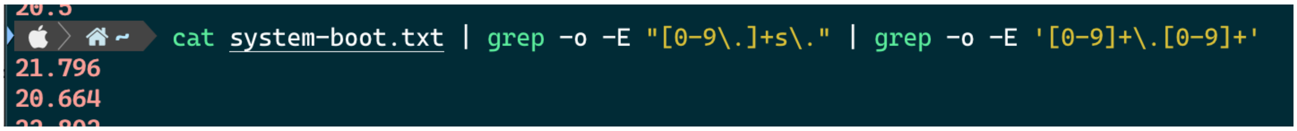


```
 ~ cd /usr/share/dict/                                            21:40:06
 /usr/sh/dict  comm -23 -i <(~/run.sh) <(cat words | egrep -i "(.*)?a(.*)?a(.*)?a([^s]*?)$" | sed -E  "s/.*(..)$/\1/" | sort | uniq -c | awk '{ print$2 }')
ab
af
```

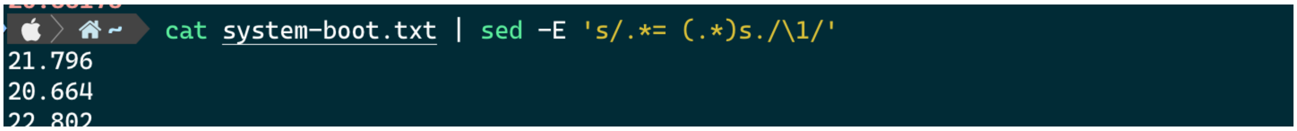
543 个

3. sed s/REGEX/SUBSTITUTION/ input.txt > input.txt 表达式中后一个 input.txt会首先被清空，而且是发生在前的。所以前面一个input.txt在还没有被 sed 处理时已经为空了。在使用正则处理文件前最好是首先备份文件。

找到文本patter并输出的该patern的办法：分别用 sed 和 grep ⚠️ !



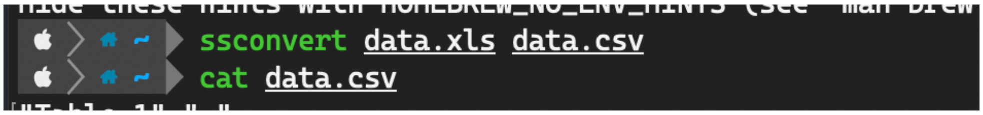
```
20.5
 ~ cat system-boot.txt | grep -o -E "[0-9\.]+s\." | grep -o -E '[0-9]+\.[0-9]+'
21.796
20.664
22.802
```



```
 ~ cat system-boot.txt | sed -E 's/.*= (.*)s./\1/'
21.796
20.664
22.802
```

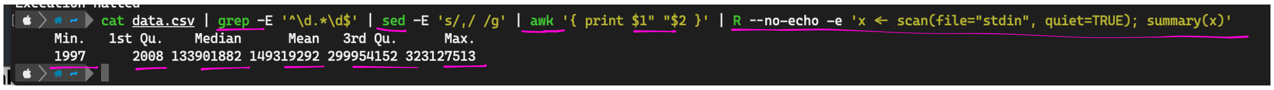
5

转换xls dao csv：



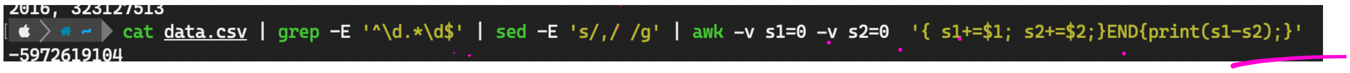
```
Hide these hints with HOMEBREW_NO_ENV_HINTS (see `man brew`
 ssconvert data.xls data.csv
 cat data.csv
["Table 1","",
```

求min and max



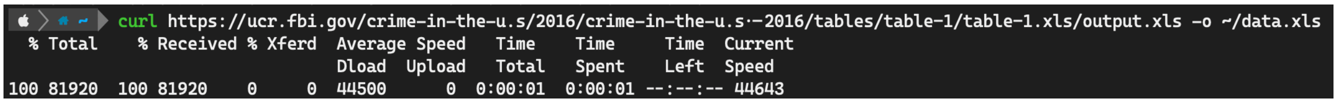
```
Execution halted
 cat data.csv | grep -E '^\d.*\d$' | sed -E 's/,/ /g' | awk '{ print $1" "$2 }' | R --no-echo -e 'x <- scan(file="stdin", quiet=TRUE); summary(x)'
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
     1997      2008 133901882 149319292 299954152 323127513
```

求两列之间差的总和



```
2016, 323127513
 cat data.csv | grep -E '^\d.*\d$' | sed -E 's/,/ /g' | awk -v s1=0 -v s2=0  '{ s1+=$1; s2+=$2;}END{print(s1-s2);}'
-5972619104
```

用curl下载文件



```
 curl https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s·-2016/tables/table-1/table-1.xls/output.xls -o ~/data.xls
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 81920  100 81920    0     0  44500      0  0:00:01  0:00:01 --:--:-- 44643
```