

```

Assignment 1 % tr -sc '[:alpha:]' '\n' <ny_article.txt | gsort | uniq -c | gsort -r
]
5802 that
5773 s
4864 for
4266 is
4035 on
3809 said
3455 was
3322 with
3291 The
3254 he
3031 it
2843 at
2728 as
26561 the
2595 his
2584 I
2460 be
2432 have
2186 are
2183 by
2178 has
2034 from
1994 not
1978 an
1853 who
1588 they
1580 had
1497 t
1487 this
1448 their
1398 about
1372 but
1364 or
1346 will
1300 more
12859 a
1263 were
12766 to
12514 of
1197 one
1191 But
1188 would
11816 and
1176 Obama
1144 been
1137 we
1116 you
1083 up
1070 It
10397 in
1012 out
981 He
977 her
965 which
955 year
954 than
941 can
938 she
928 when
915 like
913 all
859 its
859 McCain
858 time

```

```

Assignment 1 % tr -sc 'A-Za-z' '\n' < ny_article.txt > tokens.txt
Assignment 1 % tail -n +2 tokens.txt > file1.txt
Assignment 1 % sed 's/ /' tokens.txt > file2.txt
Assignment 1 % paste file1.txt file2.txt > bigrams.txt
Assignment 1 % gsort bigrams.txt | uniq -c | gsort -r
]
]
]
]
]
3186 the of
2525 the in
1184 the to
1090 the on
884 the for
811 the and
763 a in
733 be to
679 the at
565 the with
535 the that
532 said he
514 a of
498 the from
446 the by
444 a as
429 a with
426 a is
425 a for
415 s it
362 s It
362 a and
355 t don
354 been has
350 than more
344 of one
342 first the
341 a to
336 to going
330 York New
321 be will
316 been have
312 a was
308 a on
293 m I
293 his of
287 he that
281 the is
279 the about
278 the
273 was it
273 be would
268 was he
269 of out
242 to have
237 a be
234 a at
221 lot a
230 t didn
230 same the
230 get to
226 he said
225 the into
223 the said
218 most the
218 a s
217 not did
217 is it
215 the In

```

```
Assignment 1 % tr -sc 'aeiou' '\n' < ny_article.txt | tr '[:upper:]' '[:lower:]' | gsort | uniq -c | gsort -r
```

```
7915 ee
6345 ie
5269 ia
4949 oo
34834 u
3254 ei
239921 e
2365 ue
2136 au
1978 oi
1763 ua
168485 a
16532 ou
1608 oa
1561 eo
1520 ui
14047 ea
132685 o
128438 i
10331 ai
10369 io
726 oe
585 iou
301 eu
218 iu
204 ae
131 eau
79 uo
68 so
60 oui
60 eou
59 eei
56 uee
44 uie
41 aa
25 eau
23 eii
22 uou
19 iou
10 uu
10 ii
7 uea
7 uei
7 eei
7 eiaa
6 ooe
5 ooi
5 oia
4 ioia
3 eoi
3 eie
3 sou
2 uoi
2 ueu
2 oue
2 ouu
2 ioe
2 isi
2 eia
1 ueui
1 oua
1 ooooo
1 ooo
```