Unix in-class activity

Aman Kumar 17025

12th September 2020

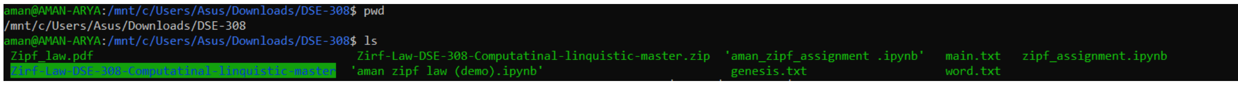
Download the file “genesis.txt” from piazza (from the Holy Bible corpus). Use the commands you learned today to perform the following operations:

* Output a list of words in the file with freq counts
* Find the 50 most common words in the corpus * Merge upper and lower case by downcasing everything * How common are different sequences of vowels (aeiou)?

1 Change directory

Use command for change directory (cd) **cd /mnt/c/Users/Asus/Downloads/DSE-308/**

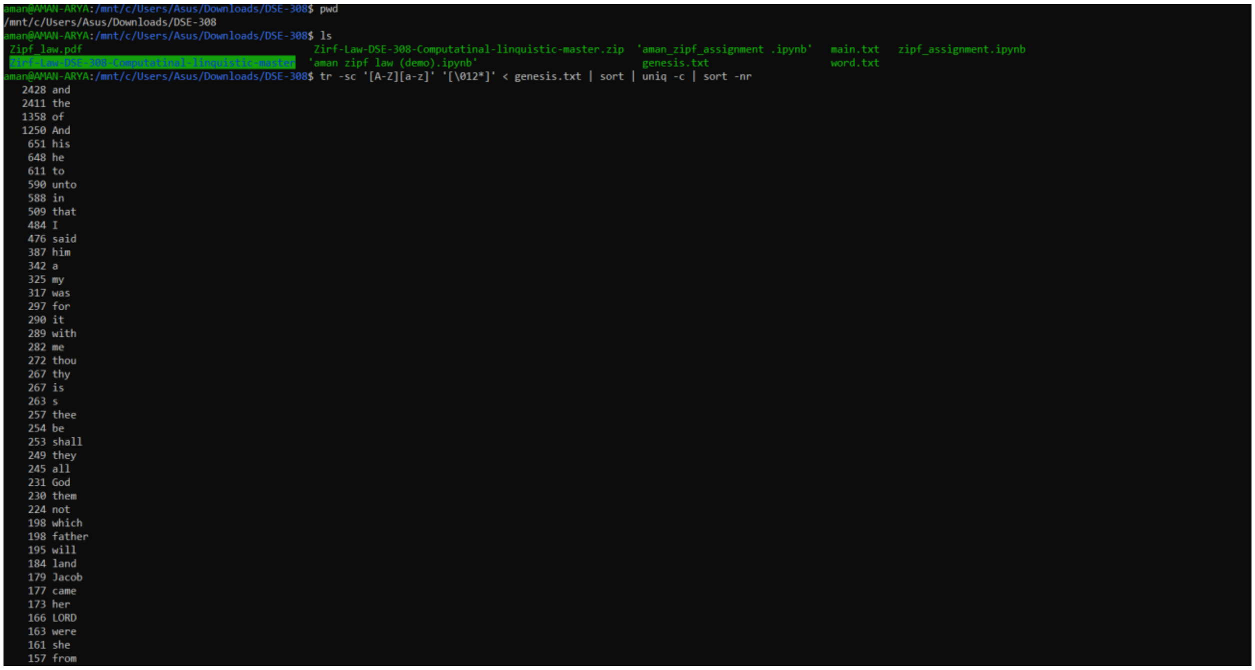
.



```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ pwd
/mnt/c/Users/Asus/Downloads/DSE-308
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ ls
 Zipf_law.pdf                                Zirf-Law-DSE-308-Computatinal-linquistic-master.zip  'aman_zipf_assignment .ipynb'   main.txt   zipf_assignment.ipynb
 Zirf-Law-DSE-308-Computatinal-linquistic-master  'aman zipf law (demo).ipynb'                     genesis.txt                   word.txt
```

2 Finding frequency corresponding to each word

```
tr -sc ’[A-Z][a-z]’ ’[\012*]’ < genesis.txt | sort | uniq -c | sort –nr
```



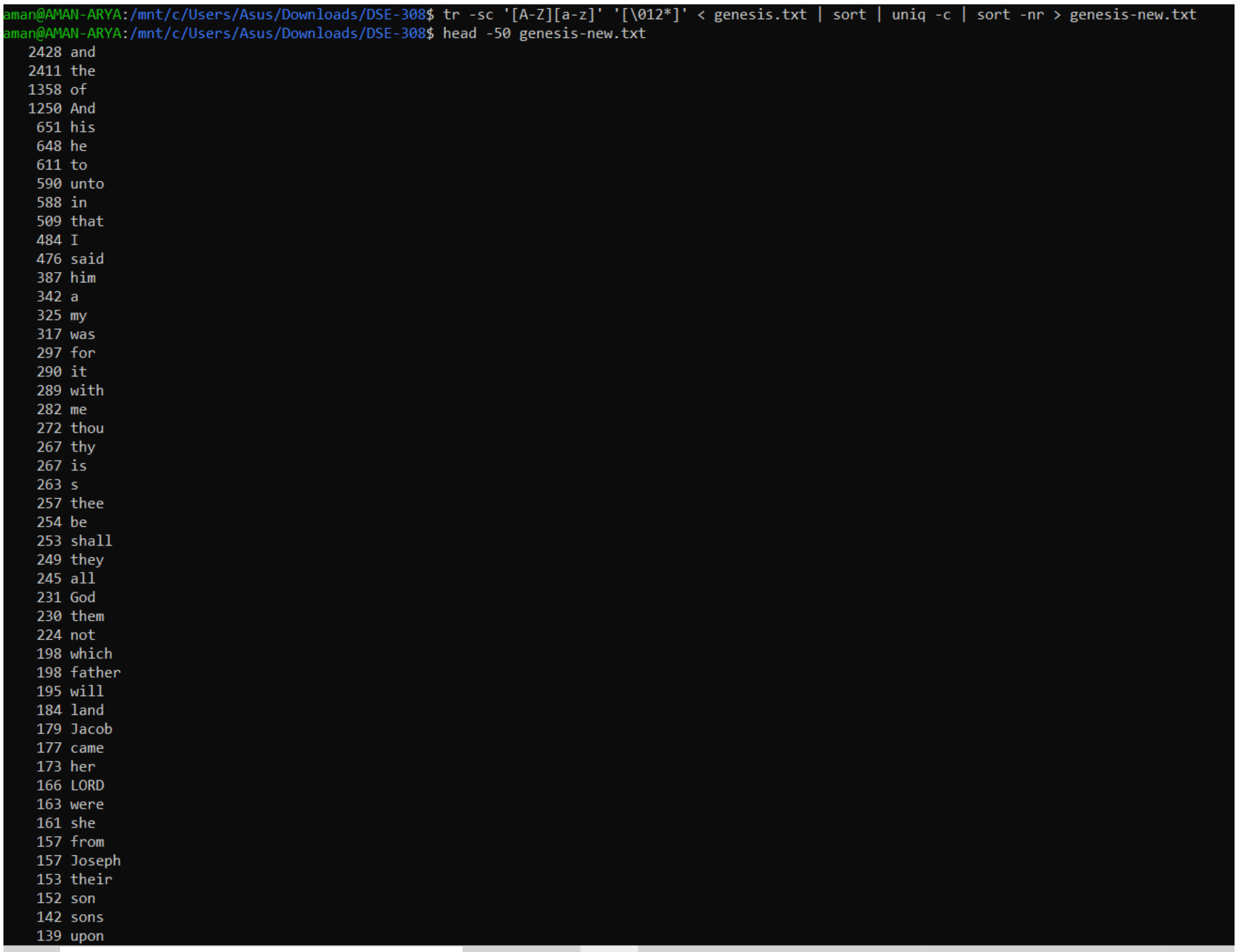
```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ pwd
/mnt/c/Users/Asus/Downloads/DSE-308
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ ls
 Zipf_law.pdf                                Zirf-Law-DSE-308-Computatinal-linquistic-master.zip  'aman_zipf_assignment .ipynb'   main.txt   zipf_assignment.ipynb
 Zirf-Law-DSE-308-Computatinal-linquistic-master  'aman zipf law (demo).ipynb'                      genesis.txt                  word.txt
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ tr -sc '[A-Z][a-z]' '[\012*]' < genesis.txt | sort | uniq -c | sort -nr
   2428 and
   2411 the
   1358 of
   1250 And
    651 his
    648 he
    611 to
    590 unto
    588 in
    509 that
    484 I
    476 said
    387 him
    342 a
    325 my
    317 was
    297 for
    290 it
    289 with
    282 me
    272 thou
    267 thy
    267 is
    263 s
    257 thee
    254 be
    253 shall
    249 they
    245 all
    231 God
    230 them
    224 not
    198 which
    198 father
    195 will
    184 land
    179 Jacob
    177 came
    173 her
    166 LORD
    163 were
    161 she
    157 from
```

3 Save word vs freq in txt file

```
tr -sc ’[A-Z][a-z]’ ’[\012*]’ < genesis.txt | sort | uniq -c | sort -nr > genesis-new.txt
```

4 Find the 50 most common words in the corpus

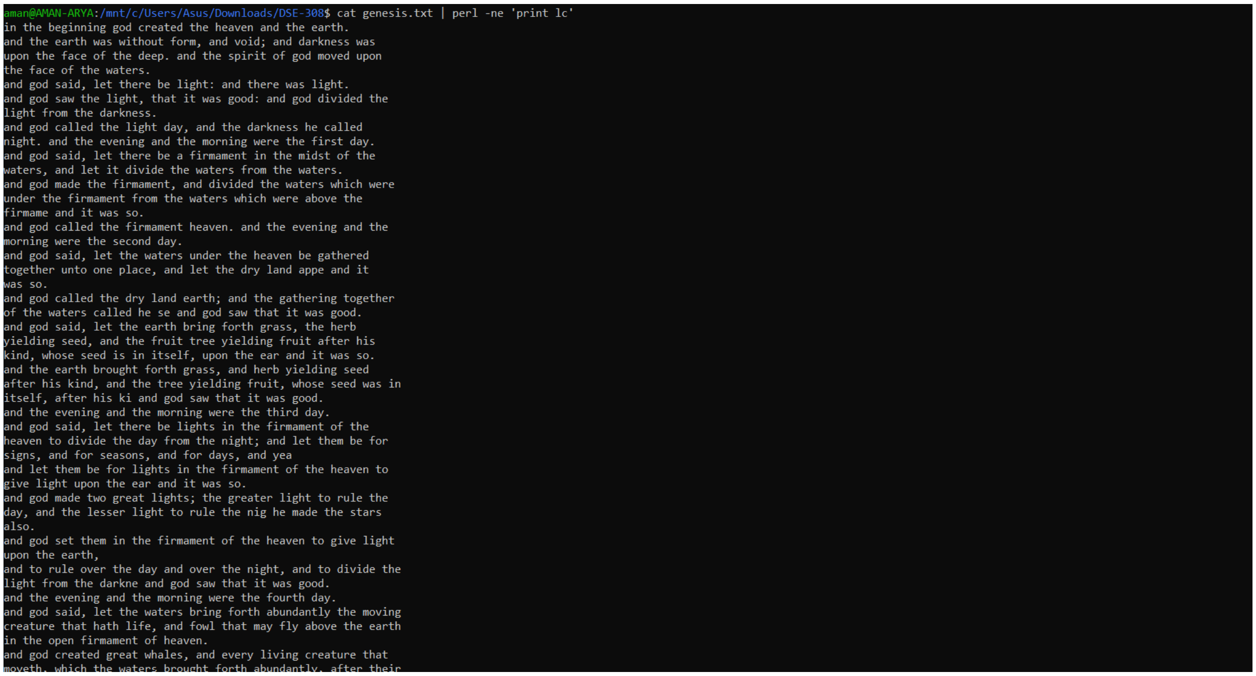
```
head -50 genesis-new.txt
```



```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ tr -sc '[A-Z][a-z]' '[\012*]' < genesis.txt | sort | uniq -c | sort -nr > genesis-new.txt
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ head -50 genesis-new.txt
   2428 and
   2411 the
   1358 of
   1250 And
    651 his
    648 he
    611 to
    590 unto
    588 in
    509 that
    484 I
    476 said
    387 him
    342 a
    325 my
    317 was
    297 for
    290 it
    289 with
    282 me
    272 thou
    267 thy
    267 is
    263 s
    257 thee
    254 be
    253 shall
    249 they
    245 all
    231 God
    230 them
    224 not
    198 which
    198 father
    195 will
    184 land
    179 Jacob
    177 came
    173 her
    166 LORD
    163 were
    161 she
    157 from
    157 Joseph
    153 their
    152 son
    142 sons
    139 upon
```

5 Merge upper and lower case by downcasing everything

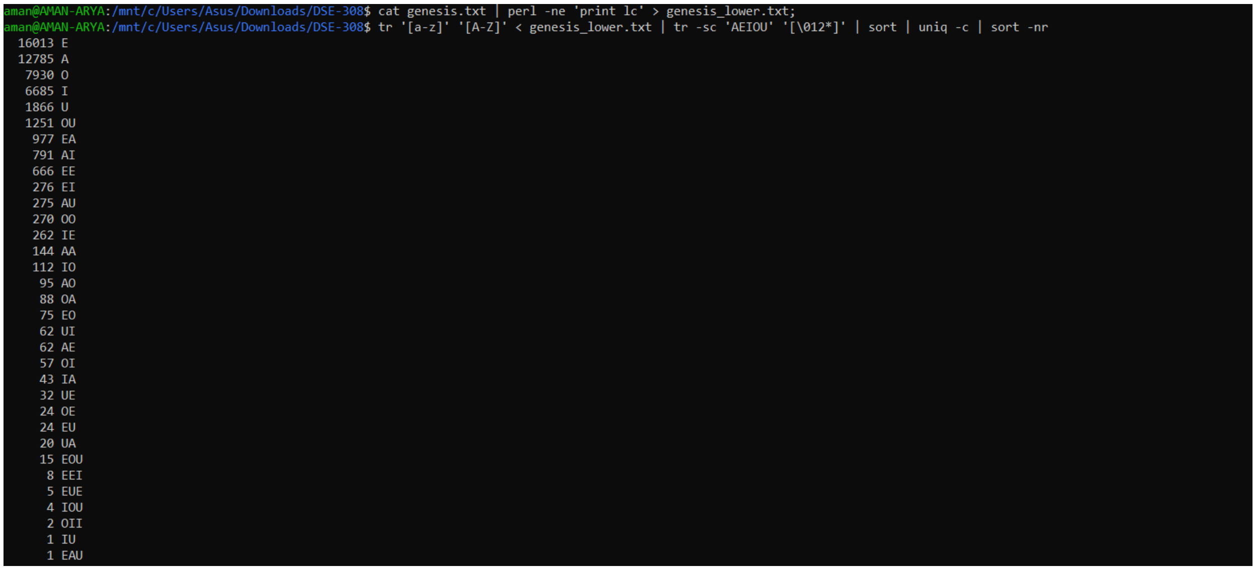
```
genesis.txt | perl -ne 'print lc' > genesis_lower.txt;
```



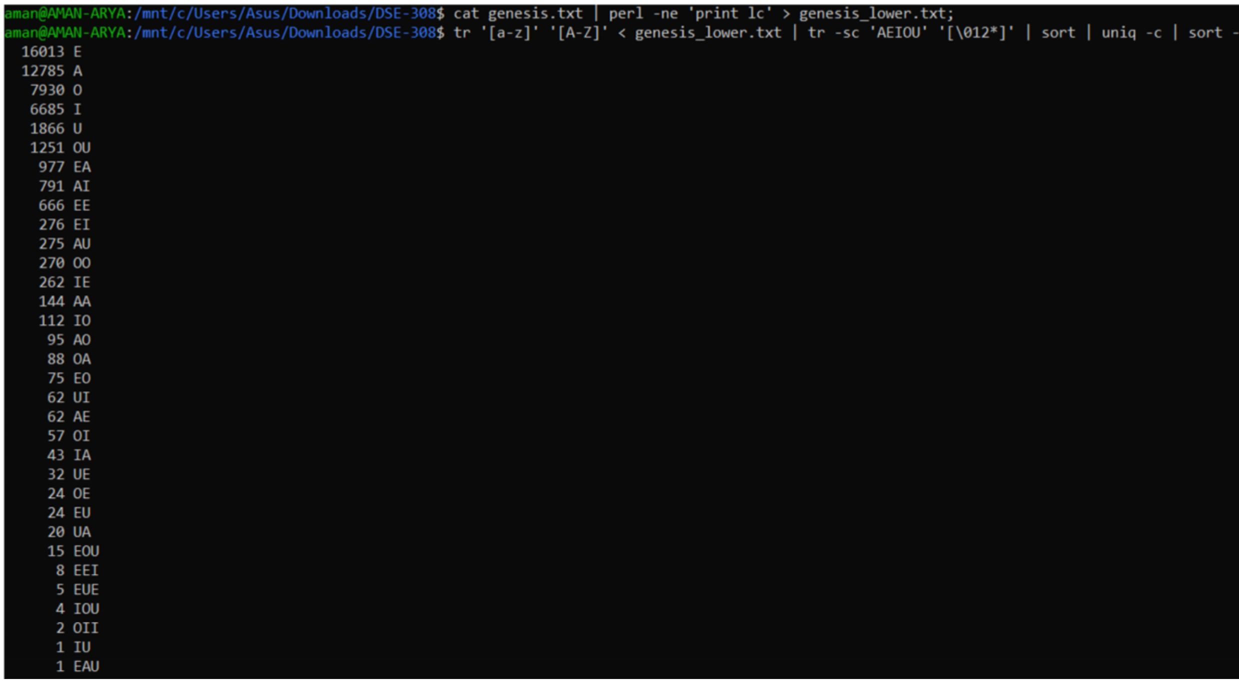
```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ cat genesis.txt | perl -ne 'print lc'
in the beginning god created the heaven and the earth.
and the earth was without form, and void; and darkness was
upon the face of the deep. and the spirit of god moved upon
the face of the waters.
and god said, let there be light: and there was light.
and god saw the light, that it was good: and god divided the
light from the darkness.
and god called the light day, and the darkness he called
night. and the evening and the morning were the first day.
and god said, let there be a firmament in the midst of the
waters, and let it divide the waters from the waters.
and god made the firmament, and divided the waters which were
under the firmament from the waters which were above the
firmame and it was so.
and god called the firmament heaven. and the evening and the
morning were the second day.
and god said, let the waters under the heaven be gathered
together unto one place, and let the dry land appe and it
was so.
and god called the dry land earth; and the gathering together
of the waters called he se and god saw that it was good.
and god said, let the earth bring forth grass, the herb
yielding seed, and the fruit tree yielding fruit after his
kind, whose seed is in itself, upon the ear and it was so.
and the earth brought forth grass, and herb yielding seed
after his kind, and the tree yielding fruit, whose seed was in
itself, after his ki and god saw that it was good.
and the evening and the morning were the third day.
and god said, let there be lights in the firmament of the
heaven to divide the day from the night; and let them be for
signs, and for seasons, and for days, and yea
and let them be for lights in the firmament of the heaven to
give light upon the ear and it was so.
and god made two great lights; the greater light to rule the
day, and the lesser light to rule the nig he made the stars
also.
and god set them in the firmament of the heaven to give light
upon the earth,
and to rule over the day and over the night, and to divide the
light from the darkne and god saw that it was good.
and the evening and the morning were the fourth day.
and god said, let the waters bring forth abundantly the moving
creature that hath life, and fowl that may fly above the earth
in the open firmament of heaven.
and god created great whales, and every living creature that
moveth, which the waters brought forth abundantly, after their
```

6 How common are different sequences of vowels (aeiou)?

```
tr ’[a-z]’ ’[A-Z]’ < genesis_lower.txt | tr -sc ’AEIOU’ ’[\012*]’ | sort | uniq -c | sort -nr
```



```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ cat genesis.txt | perl -ne 'print lc' > genesis_lower.txt;
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ tr '[a-z]' '[A-Z]' < genesis_lower.txt | tr -sc 'AEIOU' '[\012*]' | sort | uniq -c | sort -nr
  16013 E
  12785 A
   7930 O
   6685 I
   1866 U
   1251 OU
    977 EA
    791 AI
    666 EE
    276 EI
    275 AU
    270 OO
    262 IE
    144 AA
    112 IO
     95 AO
     88 OA
     75 EO
     62 UI
     62 AE
     57 OI
     43 IA
     32 UE
     24 OE
     24 EU
     20 UA
     15 EOU
      8 EEI
      5 EUE
      4 IOU
      2 OII
      1 IU
      1 EAU
```



```
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ cat genesis.txt | perl -ne 'print lc' > genesis_lower.txt;
aman@AMAN-ARYA:/mnt/c/Users/Asus/Downloads/DSE-308$ tr '[a-z]' '[A-Z]' < genesis_lower.txt | tr -sc 'AEIOU' '[\012*]' | sort | uniq -c | sort -
  16013 E
  12785 A
   7930 O
   6685 I
   1866 U
   1251 OU
    977 EA
    791 AI
    666 EE
    276 EI
    275 AU
    270 OO
    262 IE
    144 AA
    112 IO
     95 AO
     88 OA
     75 EO
     62 UI
     62 AE
     57 OI
     43 IA
     32 UE
     24 OE
     24 EU
     20 UA
     15 EOU
      8 EEI
      5 EUE
      4 IOU
      2 OII
      1 IU
      1 EAU
```