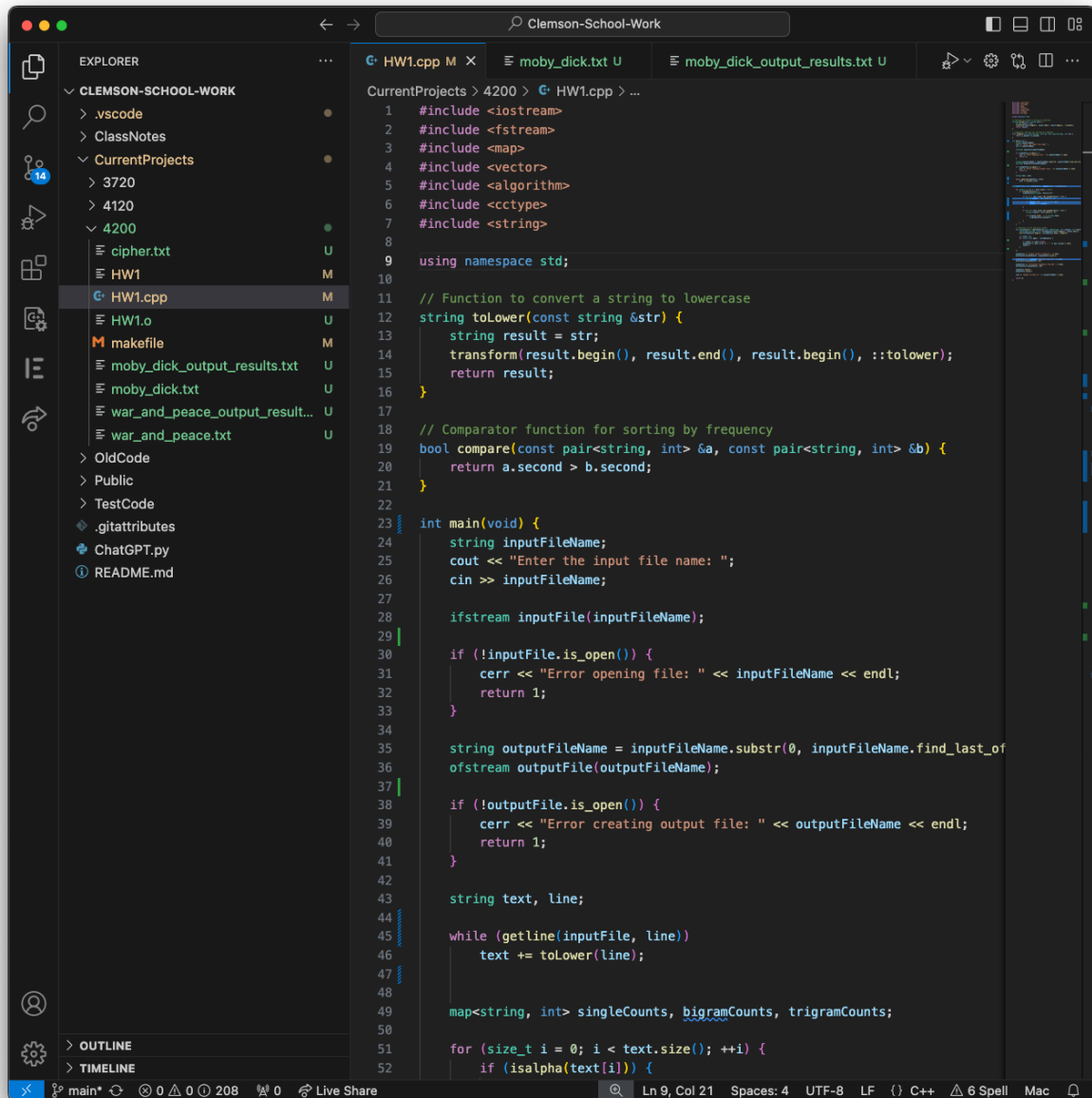


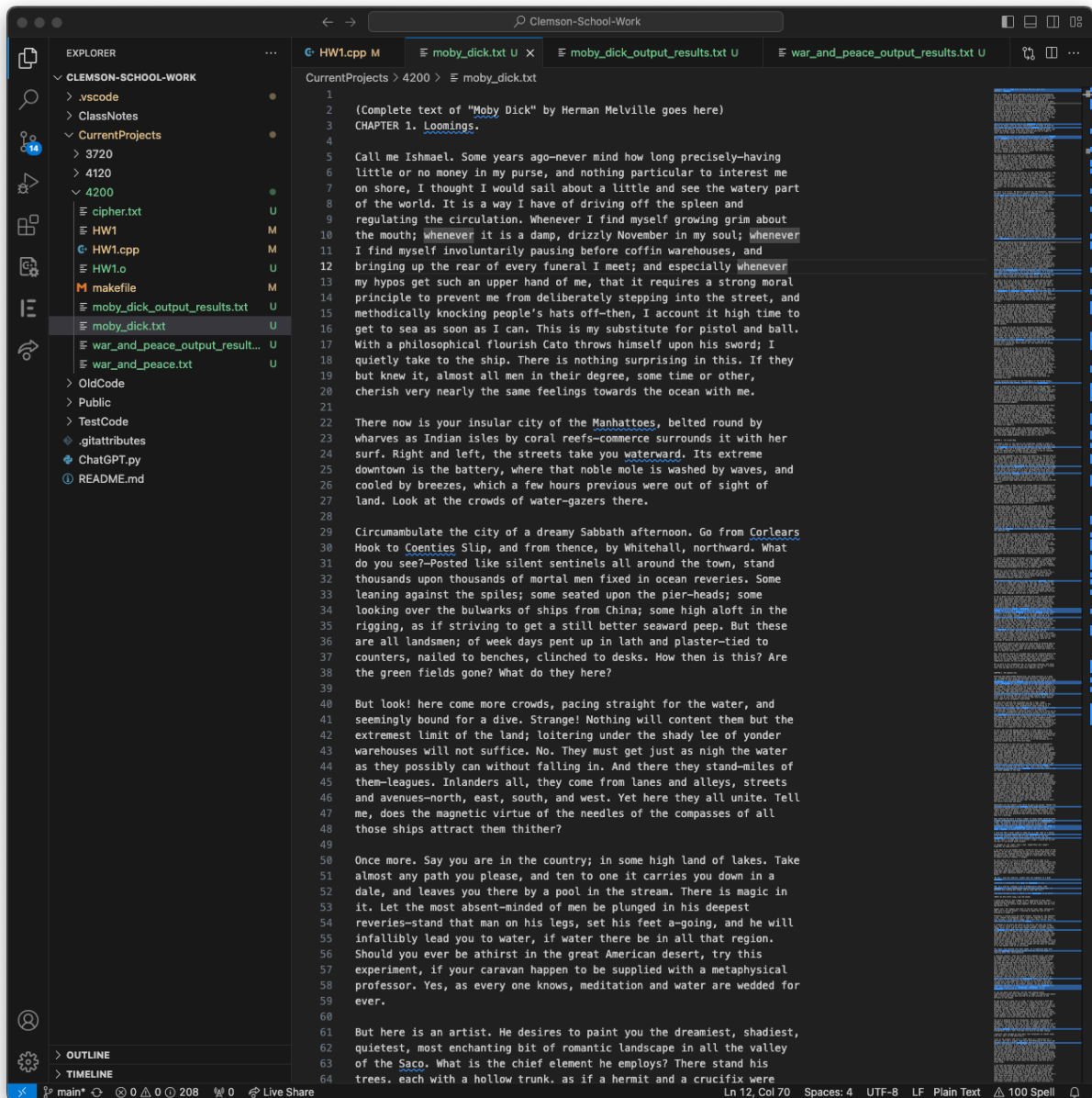
HW1 Report  
AJ Garner  
CPSC4200

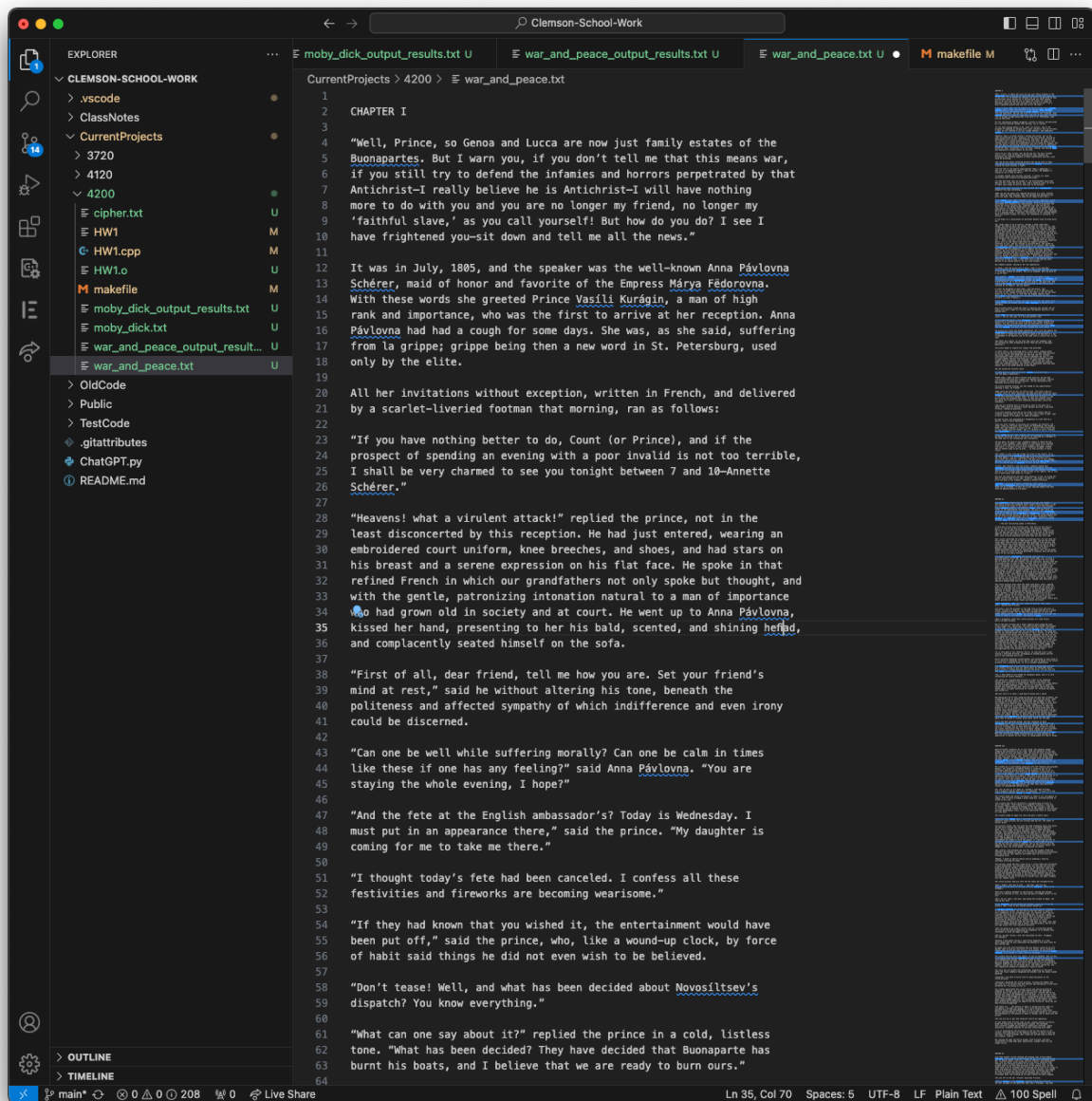
First, we make a program that will sort and count letter, bigram, and trigram frequencies



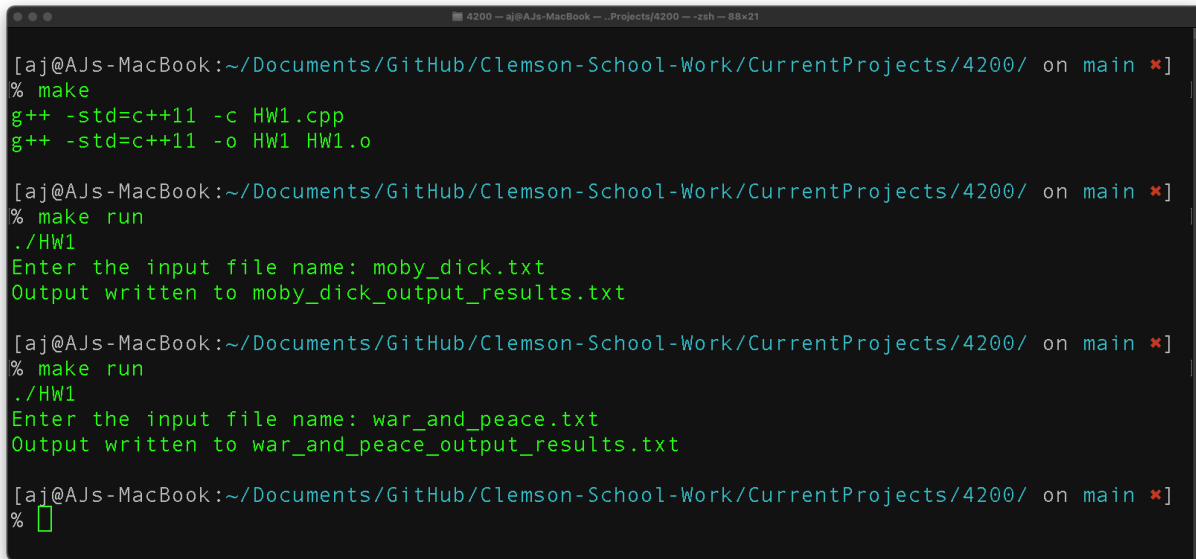
```
1 #include <iostream>
2 #include <fstream>
3 #include <map>
4 #include <vector>
5 #include <algorithm>
6 #include <cctype>
7 #include <string>
8
9 using namespace std;
10
11 // Function to convert a string to lowercase
12 string toLower(const string &str) {
13     string result = str;
14     transform(result.begin(), result.end(), result.begin(), ::tolower);
15     return result;
16 }
17
18 // Comparator function for sorting by frequency
19 bool compare(const pair<string, int> &a, const pair<string, int> &b) {
20     return a.second > b.second;
21 }
22
23 int main(void) {
24     string inputFileName;
25     cout << "Enter the input file name: ";
26     cin >> inputFileName;
27
28     ifstream inputFile(inputFileName);
29
30     if (!inputFile.is_open()) {
31         cerr << "Error opening file: " << inputFileName << endl;
32         return 1;
33     }
34
35     string outputFileName = inputFileName.substr(0, inputFileName.find_last_of('.')) + "_output_results.txt";
36     ofstream outputFile(outputFileName);
37
38     if (!outputFile.is_open()) {
39         cerr << "Error creating output file: " << outputFileName << endl;
40         return 1;
41     }
42
43     string text, line;
44
45     while (getline(inputFile, line))
46         text += toLower(line);
47
48     map<string, int> singleCounts, bigramCounts, trigramCounts;
49
50     for (size_t i = 0; i < text.size(); ++i) {
51         if (isalpha(text[i])) {
52             // Count single letters
```

Next, I used two sources for counting. They are parts from “Moby Dick” and “War and Peace” since they are in the public domain for copyrights.





We run the program twice and get two results that print to .txt files

A terminal window with a dark background and light green text. The window title bar shows '4200 - aj@AJs-MacBook - .Projects/4200 - zsh - 88x21'. The terminal content shows the user running 'make' to compile 'HW1.cpp' into 'HW1.o', then running './HW1' twice with different input files: 'moby\_dick.txt' and 'war\_and\_peace.txt'. Each run prompts for an input file name and reports the output file name. The prompt is '[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main ✖]' and the user input is shown in green. The prompt is followed by a green cursor.

```
[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main ✖]
% make
g++ -std=c++11 -c HW1.cpp
g++ -std=c++11 -o HW1 HW1.o

[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main ✖]
% make run
./HW1
Enter the input file name: moby_dick.txt
Output written to moby_dick_output_results.txt

[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main ✖]
% make run
./HW1
Enter the input file name: war_and_peace.txt
Output written to war_and_peace_output_results.txt

[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main ✖]
% 
```

Now we have two different sets of frequencies to compare. Both texts are not the whole book, but sufficiently long enough. “Moby Dick” has about 71,000 letters and “War and Peace” has about 35,000 letters.

```
HW1.cpp M    moby_dick.txt U    moby_dick_ou

CurrentProjects > 4200 > moby_dick_output_results.txt
1  Single Letters Frequency:
2  e: 7254
3  t: 5486
4  a: 4951
5  o: 4642
6  n: 4274
7  i: 4062
8  s: 3831
9  h: 3665
10 r: 3264
11 l: 2611
12 d: 2543
13 u: 1689
14 m: 1565
15 g: 1472
16 c: 1342
17 w: 1342
18 f: 1279
19 y: 1125
20 p: 1095
21 b: 1057
22 k: 530
23 v: 483
24 q: 98
25 j: 72
26 x: 58
27 z: 48
28
29 Bigrams Frequency (Top 30):
30 th: 1807
31 he: 1541
32 in: 1283
33 an: 1033
34 er: 970
35 nd: 788
36 re: 751
37 ng: 696
38 at: 654
39 ou: 639
40 ha: 639
41 ed: 616
42 on: 583
43 st: 542
44 to: 534
45 hi: 533
46 or: 523
47 es: 515
48 en: 504
49 is: 488
50 it: 487
51 ar: 476
52 of: 457
53 le: 453
54 se: 439
55 ea: 436
56 as: 433
57 me: 426
58 te: 404
59 al: 388
60
61 Trigrams Frequency (Top 30):
62 the: 1095
63 and: 596
64 ina: 570
65
66 Trigrams Frequency (Top 30):
67 the: 1095
68 and: 596
69 ing: 570
70 hat: 252
71 her: 250
72 his: 247
73 ere: 231
74 tha: 208
75 thi: 196
76 all: 165
77 ver: 161
78 for: 157
79 ght: 154
80 ter: 146
81 ent: 145
82 one: 139
83 igh: 139
84 ith: 138
85 but: 132
86 was: 132
87 wit: 130
88 you: 127
89 out: 123
90 not: 117
91 hou: 117
92 wha: 101
93 hin: 100
94 eve: 97
95 tho: 95
96 lan: 93
97
```

Above is the results for “Moby Dick”.

```

CurrentProjects > 4200 > ≡ war_and_peace_output_results.txt
1  Single Letters Frequency:
2  e: 3715
3  a: 2538
4  t: 2500
5  n: 2219
6  o: 2197
7  i: 2103
8  s: 1899
9  h: 1809
10 r: 1704
11 d: 1302
12 l: 1190
13 u: 783
14 c: 742
15 m: 725
16 p: 633
17 w: 633
18 y: 611
19 f: 582
20 g: 576
21 b: 439
22 v: 432
23 k: 223
24 x: 45
25 j: 33
26 z: 24
27 q: 23
28
29 Bigrams Frequency (Top 30):
30 he: 830
31 th: 731
32 in: 569
33 an: 548
34 er: 517
35 re: 406
36 nd: 353
37 ed: 343
38 on: 340
39 to: 316
40 ha: 304
41 at: 292
42 ou: 291
43 ng: 288
44 te: 281
45 en: 278
46 is: 269
47 hi: 269
48 as: 262
49 es: 252
50 it: 243
51 or: 226
52 ar: 207
53 st: 202
54 ce: 198
55 ti: 198
56 ne: 197
57 ve: 194
58 me: 190
59 nt: 186
60
61 Trigrams Frequency (Top 30):
62 the: 443
63 and: 273
64 ing: 238
65 her: 185
66 nce: 121
67 you: 113
68 his: 110
69 she: 100
70 ess: 100
71 hat: 97
72 rin: 95
73 ere: 88
74 pri: 83
75 inc: 83
76 thi: 80
77 ver: 80
78 tha: 80
79 ith: 79
80 ent: 76
81 res: 76
82 ter: 75
83 was: 75
84 wit: 74
85 ion: 72
86 aid: 71
87 for: 66
88 one: 65
89 ann: 64
90 not: 63
91 sai: 62
92

```

Above is the results for “War and Peace”.

## Comparing them:

### Single Letters:

- Both texts have 'e' as the most frequent letter.
- The next most common letters are 't', 'a', 'o', and 'n' in both texts, though the exact order differs slightly. "War and Peace" has 'a', 't', 'n', and 'o' following 'e', while "Moby-Dick" lists 't', 'a', 'o', and 'n'.

### Bigrams:

- Common bigrams in both texts include 'th', 'he', 'in', and 'an'.
- The order of these bigrams is similar but not identical. For example, 'th' and 'he' are the top two in both, but their frequencies and the subsequent bigrams vary slightly.

### Trigrams:

- 'the', 'and', and 'ing' are the top trigrams in both texts.
- Other frequent trigrams such as 'her', 'hat', and 'his' also appear in both lists, showing a consistent pattern in common English word combinations.

### Similarities:

- Both texts display common patterns typical of the English language. The same letters, bigrams, and trigrams appear frequently, reflecting standard usage in English.

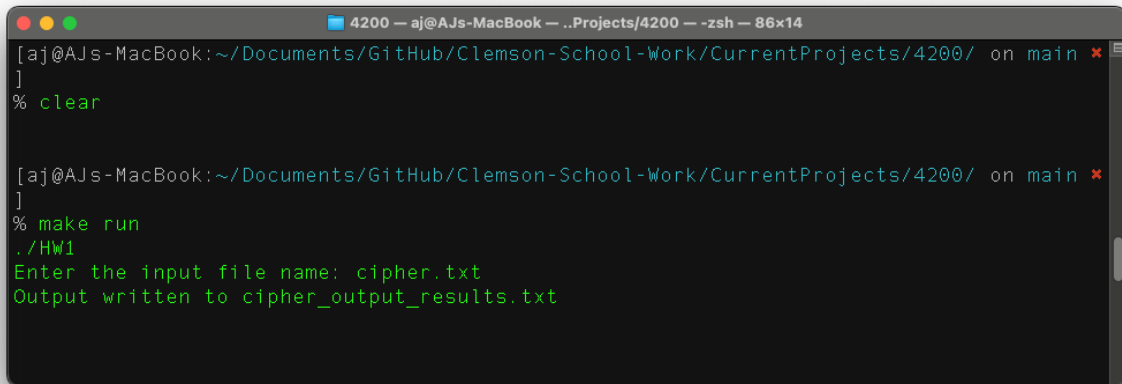
### Differences:

- Although the same elements appear frequently, the specific order and exact counts differ slightly. This reflects the unique style, vocabulary, and sentence structures of each text.

Overall, the frequencies of letters, bigrams, and trigrams in "War and Peace" and "Moby-Dick" are quite similar, showcasing common patterns in English. However, the slight differences in order and counts highlight the individual characteristics of each text sample.

## Decrypting the Cipher:

First, we will put the cipher through the program to get the frequencies.

A terminal window titled '4200 — aj@AJs-MacBook — ..Projects/4200 — -zsh — 86x14'. The prompt is '[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main]'. The user enters '% clear', which clears the screen. Then the user enters '% make run', which runs the program. The program prompts 'Enter the input file name: cipher.txt' and then outputs 'Output written to cipher\_output\_results.txt'.

```
[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main]
% clear

[aj@AJs-MacBook:~/Documents/GitHub/Clemson-School-Work/CurrentProjects/4200/ on main]
% make run
./HW1
Enter the input file name: cipher.txt
Output written to cipher_output_results.txt
```



Here are the results:

```
CurrentProjects > 4200 > ≡ cipher_output_results.txt
1  Single Letters Frequency:
2  x: 161
3  t: 126
4  j: 123
5  m: 114
6  p: 107
7  c: 96
8  r: 87
9  b: 83
10 n: 68
11 v: 64
12 g: 49
13 w: 38
14 y: 30
15 a: 29
16 l: 27
17 i: 26
18 u: 25
19 h: 23
20 f: 18
21 e: 15
22 q: 12
23 o: 6
24 d: 6
25 k: 3
26 s: 1
27 z: 1
28
29 Bigrams Frequency (Top 30):
30 jp: 66
31 px: 52
32 bt: 43
33 mt: 42
34 tn: 29
35 tw: 26
36 xv: 25
37 ci: 18
38 pm: 18
39 vx: 16
40 cu: 16
41 hb: 15
42 br: 15
43 jx: 14
44 xt: 13
45 xn: 12
46 pc: 12
47 pb: 12
48 tj: 12
49 gx: 12
50 mr: 12
51 tc: 10
52 mg: 10
53 gc: 9
54 rj: 9
55 lx: 9
56 xj: 9
57 rm: 9
58 cj: 9
59 rp: 9
60
61 Trigrams Frequency (Top 30):
62 jpx: 50
63 mtn: 26
64 btw: 23
65 hbt: 15
66 pbr: 8
67 tjx: 8
68 apm: 7
69 btj: 6
70 jpe: 6
71 jxv: 6
72 vxv: 6
73 xvr: 6
74 mgg: 6
75 mlx: 6
76 pxv: 6
77 jvc: 5
78 mrj: 5
79 cut: 5
80 pmg: 5
81 vxm: 5
82 pmt: 5
83 rpm: 5
84 gxj: 4
85 vqv: 4
86 acu: 4
87 ypc: 4
88 ymr: 4
89 vbj: 4
90 twx: 4
91 jxt: 4
92
```

We will now use these results to map the most common letters, bigrams, and trigrams from the cipher text to those from the frequencies.

First, here is the results from just swapping the letters.

#### Text

ht ane soce nifl goce wilan whtuels iw o cot s notr otr mliae iqel ouohtsa ane  
BT JPX RMLX PCUV AMLX ICVJP IBTWXVR CI M LMT'R PMTN, MTN YVCJX CDXV MWMBTRJ JPX  
gotrdesahgy fkit ane kdosael iw ane modd iw ane yhtu s kodoge otr ane yhtu som  
AMTNGXRJBAH UQCT JPX QGMRJXV CI JPX YMGG CI JPX HBTW'R QMGMAX; MTN JPX HBTW RMY  
ane kola iw ane notr anoa mliae anet ane yhtu s giftaetotge mos gnotuer otr nh  
JPX QMVJ CI JPX PMTN JPMJ YVCJX. JPXT JPX HBTW'R ACUTJXTMTAX YMR APMTWXN, MTN PB  
s anifunas alifpder nhc si anoa ane xihtas iw nhs dihts mele diiser otr nhs yt  
R JPCUWPJR JVCUFGXN PBL, RC JPMJ JPX SCBTJR CI PBR GCBTR YXVX GCCRXN, MTN PBR HT  
ees sciae ite ouohtsa otianel ane yhtu glher odifr ai plhtu ht ane osalidiuels  
XXR RLCJX CTX MWMBTRJ MTCJPXV. JPX HBTW AVBXN MGCUN JC FVBW BT JPX MRJVCGCWXVR,  
ane gnodreots otr ane siiansobels otr ane yhtu skoye otr sohr ai ane mhse ce  
JPX APMGNXMTR, MTN JPX RCCJPRMEXVR. MTN JPX HBTW RQMHX, MTN RMBN JC JPX YBRX LX  
t iw popbdt mnisieqel snodd leor anhs mlhahtu otr snim ce ane htaelkleaoahit  
T CI FMFEGCT, YPCRCXDXV RPMGG VXMN JPBR YVBJBTW, MTN RPCY LX JPX BTJXVQVXJMJBCT  
aneleiw snodd pe gdianer mhan sgoldea otr noqe o gnoht iw uidr opifa nhs tegy  
JPXVXCI, RPMGG FX AGCJPXN YBJP RAMVGXJ, MTN PMDX M APMBT CI WCGN MFCUJ PBR TXAH,  
otr snodd pe ane anhlr lfdel ht ane yhturic anet goce ht odd ane yhtu s mhse c  
MTN RPMGG FX JPX JPBVN VUGXV BT JPX HBTWNCL. JPXT AMLX BT MGG JPX HBTW'R YBRX L  
et pfa aneb gifdr tia leor ane mlhahtu til coye ytimt ai ane yhtu ane htaelkle  
XT; FUJ JPXE ACUGN TCJ VXMN JPX YVBJBTW, TCV LMHX HTCYT JC JPX HBTW JPX BTJXVQVX  
aoahit aneleiw anet mos yhtu pedsnovvol uleoadb alifpder otr nhs giftaetotge m  
JMJBCT JPXVXCI. JPXT YMR HBTW FXGRPMOVM VVXMJGE JVCUFGXN, MTN PBR ACUTJXTMTAX Y  
os gnotuer ht nhc otr nhs dilrs mele osaithsner tim ane jfeet pb leosit iw an  
MR APMTWXN BT PBL, MTN PBR GCVNR YXVX MRJCTBRPXN. TCY JPX KUXXT, FE VXMRCI CI JP  
e milrs iw ane yhtu otr nhs dilrs goce htae ane potjfea nifse otr ane jfeet sk  
X YCVNR CI JPX HBTW MTN PBR GCVNR, AMLX BTJC JPX FMTKUXJ PCURX; MTN JPX KUXXT RQ  
oye otr sohr i yhtu dhqe wil  
MHX MTN RMBN, C HBTW, GBDX ICV

#### Key

g	h	i	q	b	p	d	y	w	a	j	c	o	r	v	n	k	s	x	t	f	l	u	e	m	z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

After some time and checking againts other results, we need to identify common bigram and trigrams. For example, we see that “jpx” is the most common trigram, which would line up with “the” being the most common trigram. We can now see that all the “j” would be “t” and so forth. When we apply this logic and try to solve the chipper, we find this:

In the same hour came forth fingers of a man’s hand, and wrote over against the candlestick up on the plaster of the wall of the king’s palace; and the king saw the part of the hand that wrote. Then the king’s countenance was changed, and his thoughts troubled him, so that the joints of his loins were loosed, and his knees smote one against another. The king cried aloud to bring in the astrologers, the chaldeans, and the soothsayers. And the king spake, and said to the wise men of Babylon, whosoever shall read this writing and show me the interpretation thereof, shall be clothed with scarlet, and have a chain of gold about his neck, and shall be the third ruler in the kingdom. Then came in all the king’s wise men; but they could not read the writing, nor make known to the king the interpretation thereof. Then was king Belshazzar greatly troubled, and his countenance was changed in him, and his lords were astonished. Now the queen, by reason of the words of the king and his lords, came into the banquet house; and the queen spake and said, o king live forever;

let not thy thoughts trouble thee, nor let thy countenance be changed; there is a man in thy kingdom, in whom is the spirit of the holy gods; and in the days of thy wisdom of the gods, was found in him; whom the king nebuchadnezzar thy father, the king, I say, thy father, made master of the magicians, astrologers, chaldeans, and soothsayers; forasmuch as an excellent spirit, and knowledge, and understanding, interpreting of dreams and showing of hard sentences, and dissolving of doubts, were found in the same Daniel, whom the king named Belteshazzar; now let Daniel be called, and he will show the interpretation.