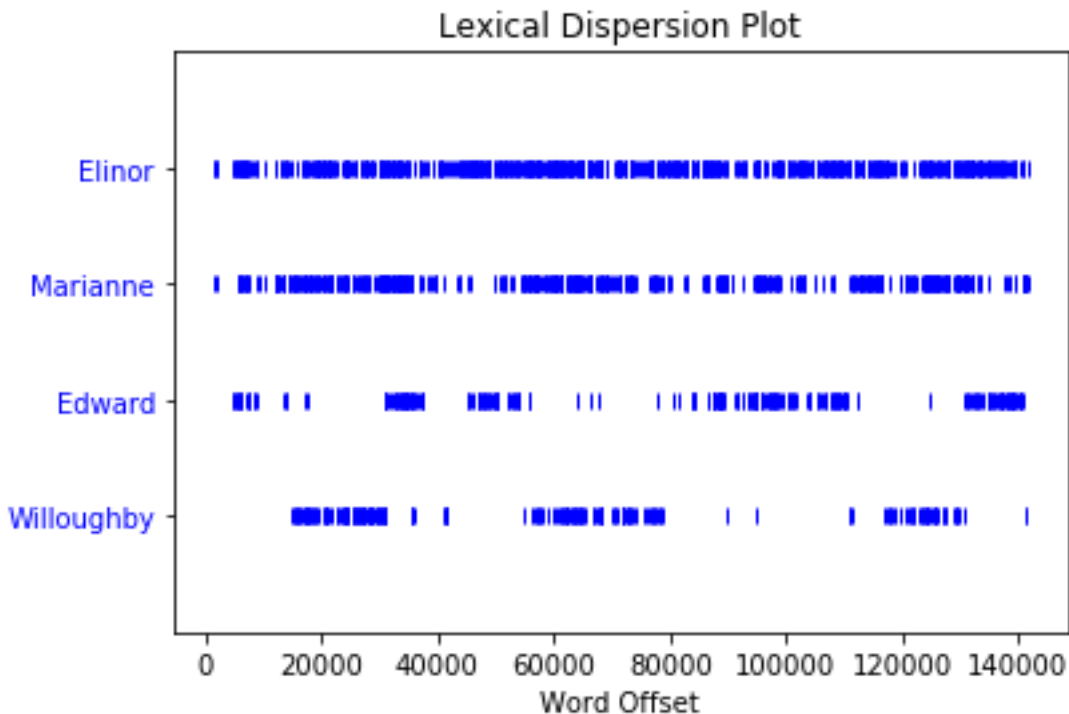Brody Vogel NLP HW #1 Output File
A few answers are truncated to keep the document short

4. There are 141576 words in text2, and 6833 unique words

6. It looks like the females are more prominent characters than the males; they appear more frequently. It looks like Willoughby and Marianne may be a couple, because Willoughby never shows up without Marianne. The same could - possibly - be said about Elinor and Edward.

Lexical Dispersion Plot

22. Here are the 50 (truncated for space) most common four-letter words, in order of number of occurrences:

[('JOIN', 1021), ('PART', 1016), ('that', 274), ('what', 183), ('here', 181), ('....', 170), ('have', 164), ('like', 156), ('with', 152), ('chat', 142), ('your', 137), ('good', 130), ('just', 125), ('lmao', 107),('know', 103), ('room', 98), ('from', 92), ('this', 86), ('well', 81), ('back', 78), ('hiya', 78) ('they', 77), ('dont', 75), ('yeah', 75), ('want', 71), ('love', 60), ('guys', 58), ('some', 58),('been', 57), ('talk', 56), ('nice', 52), ('time', 50), ('when', 48), ('haha', 44), ('make', 44), ('girl', 43), ('need', 43), ('U122', 42), ('MODE', 41), ('will', 40), ('much', 40), ('then', 40), ('over', 39), ('work', 38) ('were', 38), ('take', 37), ('U121', 36), ('U115', 36), ('song', 36), ('even', 35)]

24.
a) []
b) ['zone', 'amazes', 'Fetchez', 'Fetchez' ,'zoop', 'zoo', 'zhiv', 'frozen', 'zoosh']

c) ['empty', 'aptly', 'Thpppppt', 'Thppt', 'Thppt', 'empty', 'Thppppt', 'temptress', 'temptation', 'ptoo', 'Chapter', 'excepting', 'Thpppt']
d) (truncated for space)
['Cherries', 'Mud', 'Churches', 'Churches', 'Lead', 'Lead', 'A', 'Oooh', 'Exactly', 'So', 'If', 'And', 'A', 'A', 'A', 'A', 'Here', 'Use', 'We', 'Ohh', 'Ohh', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Ahh', 'Ahh', 'Right', 'Remove', 'A', 'A', 'A', 'It', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Burn', 'Who', 'I', 'Arthur', 'King', 'Britons', 'My', 'Good', 'Sir', 'Knight', 'Camelot', 'Round', 'Table', 'My', 'I', 'What', 'Bedevere', 'Then', 'I', 'Sir', 'Bedevere', 'Knight', 'Round', 'Table', 'The', 'Sir', 'Bedevere', 'King', 'Arthur', 'Sir', 'Lancelot', 'Brave', 'Sir', 'Gallahad', 'Pure', 'Sir', 'Robin', 'Sir', 'Lancelot']

28.
```
# returns the percentage as a float rather than long decimal; example follows
def percent(word, text):
    return(100 * (text.count(word) / len(text)))

percent('I', text6)
```