

# Chapter 1 Exercises

*Erin M. Buchanan*

*1/1/2019*

## Get Started

- Create a Jupyter notebook with the following items. You can upload a compiled version of the notebook and the ipython or a script file.
  - Remember, use Markdown cells to answer text questions. Paste the questions into the cells so it's clear what you are answering.
- Import the nltk and the pre-loaded texts as shown in the lecture.

## Basic Statistics

- How many total words are there in the document contained in the variable text3?
- How many distinct words are there?

## Dispersion Plots

- Produce a dispersion plot of the four main protagonists in Sense and Sensibility: Elinor, Marianne, Edward, and Willoughby.
- What can you observe about the different roles played by the males and females in this novel?
- Can you identify the couples?

## Collocations

- Find the collocations in the text contained in the variable text5.

## Searching Text

- Find all the four-letter words in the Chat Corpus (text5).
  - With the help of a frequency distribution (FreqDist), show these words in decreasing order of frequency.

## Understanding Python

### Functions

- Consider the following Python expression: len(set(text4)).
  - State the purpose of this expression.
  - Describe the two steps involved in performing this computation.
- Define a function called vocab\_size(text) that has a single parameter for the text, and which returns the vocabulary size of the text.

## Working with Strings

- Define two variables containing lists of words, e.g., `phrase1`, `phrase2`.
  - Join them together using the plus operator to form whole sentences.
  - What is the relationship between `len(phrase1 + phrase2)` and `len(phrase1) + len(phrase2)`?

## Indexing

- The first sentence of `text3` is provided to you in the variable `sent3`. The index of the in `sent3` is 1, because `sent3[1]` gives us 'the'.
  - What are the indexes of the two other occurrences of this word in `sent3`?
- Use `text9.index()` to find the index of the word `sunset`.
  - You'll need to insert this word as an argument between the parentheses.
  - By a process of trial and error, find the slice for the complete sentence that contains this word.