

Project Overview: I chose to use Project Gutenberg as my data source. My goal was to generate a chapter of text in the style of Jane Austen using Markov analysis. I hoped to learn how to implement Markov analysis on text, gather data from the web, and use string methods to format my text data to make a more believable looking chapter.

Implementation: I first had to gather my data from Project Gutenberg. I did this by using the `pattern.web` package to download the text. I then pickled the data in order to ensure that I was only working with text and saved it to the disk to avoid going over Gutenberg's download limit. Once I had a file I could play with an unlimited amount of times, I unpickled the file and wrote the text between the Gutenberg Project introduction and conclusion to another text file (without pickling to simplify my functions later). I at first only downloaded the novels "Emma" and "Pride and Prejudice", but after realizing that I needed more data, I opted to download "The Complete Works of Jane Austen" volume.

To generate my random chapter, I first used a Markov analysis algorithm to map 'prefixes' to suffixes of "The Complete Works of Jane Austen". I then randomly created a title out of words that came from Jane Austen novel titles and words that sounded like they could have come from her titles. Finally, I used the function `new_text` to format my chapter with a title, chapter number, and the body text. I generated the body of the chapter by starting with a random prefix key from the suffix mapping dictionary and then pulling a random suffix out of the list corresponding to that key. By encapsulating this and a reassignment of the prefix in a loop, I could generate the chapter. I chose to have my loop skip over words that I chose to exclude such as "CHAPTER" in order to make my final text more believable. I also chose to have my program keep searching until it found a capitalized word to begin the chapter.

Results: At first, I only analyzed two novels of Jane Austen, "Emma" and "Pride and Prejudice". However, I found that as I increased the order, my generated chapter became rapidly less random. At an order of 4, my chapter would exactly quote large sections of one of the original books. When I started using "The Complete Works", the order slightly increased to 5 before directly quoting a book. I found that an order of 4 made sense mostly, but did not directly quote one book for more than a sentence or two. Below is a sample of a chapter generated by my program with an order of 3:

SENSE AND MARRIAGE
by Jane Austen

Chapter 15

I sincerely wish I could do something. But to be candid in my judgment of every body. Willoughby may undoubtedly have very sufficient reasons for his conduct, and I will visit them all." Mr. Bennet was among the earliest thoughts of both. It was also a very early period of my return to England. I had thought it right to put it off, I could have wished, but dared not depend upon General Tilney." Catherine sighed. "Well," continued her philosophic mother, "I am glad you are come, for there is a secret subterraneous communication between your apartment and the chapel of St. Anthony, scarcely two miles off? Could you shrink from so simple an adventure? No, no, you will proceed into this small vaulted room, however, your eyes will be attracted towards a large, old-fashioned cabinet of ebony and gold, which, though narrowly examining the furniture before, you had passed unnoticed. Impelled by an irresistible presentiment, you will eagerly advance to it, unlock its folding doors, and without

stopping to look about me. Yes Madam, I had the pleasure of sometimes hearing her father and Emma should remain at Randalls, while she and her sisters did not very well pleased with her conquest. Being now in her twenty-first year, Maria Bertram was beginning to be heard by no one at Mansfield. She had never heard Mrs. Tilney mentioned in the family might have prompted. After discharging his conscience on that head, he proceeded to inform them, with many rapturous expressions, of his happiness in such an instantaneous renewal of cheerfulness into the whole affair on which I want your opinion." "My opinion!" she cried, shrinking from such a compliment, highly as it gratified her. "Yes, your advice and opinion. I do not suppose the situation my friend Mrs Rooke let me thus much into the habit of spending more than they ought. His affection for her sister, or any resentment against Edward. Elinor's office was a painful consideration whenever it occurred; and eager to enter it, could not, even by the general's father, and the present situation of matters at Mansfield were known to him. This was aimed, and well aimed, at Fanny. It was pleasing to hear him say to Mr. Knightley, and turned to Mansfield. He could not see her equal. Thus much indeed he was obliged to acknowledge that it could not be unseen by the young man, and was mortified to have been shocked before. There was no getting away, no pause; and, to my certain knowledge, more than one failure of perfect symmetry in her form, he was forced to speak, and Miss Crawford were mentioned in my last letters from Mansfield. Do you find them troublesome, you must send them home again." "That is very true," said she, "at least as long as you have many a time repair to this chapel? The young Mrs. Eleanors and Mrs. Bridgets--starched up into seeming piety, but with heads full of something else." "The present always occupies you in such scenes--does it?" said he, with a sigh. "I am afraid it has torn a hole already; I shall be back by dinner." "I admire the activity of your benevolence," observed Mary, "but every impulse of feeling should be guided by second-hand conjecture is pitiful. The premises are before you. My brother is a pleasant gentlemanlike man--he is a great deal to laugh at its absurdity, or censure its impertinence, for she considered that Jane would be induced to join her--but it would not make them choose their own dinners at the inn, nothing remained to be wished that Captain Wentworth had given of him, on his return from Portsmouth, he might have been softened, his manners improved; and from his judgement, information, and knowledge of the world made her judgment very generally and deservedly looked up to him as soon as they were by themselves, "I come to talk it over; what my brother Hayter had met with it in his face. Poor man! I was really gone from Randalls, she closed with the offer of it, and that there was not a great deal of pleasure and parade. Her introductions and recommendations must all wait, and every projected party be still only talked of. So she thought at first;--but a little consideration convinced her that the forbidden gallery, in which lay the apartments of the Lodge, and to gladden the eyes of a man of principle. What will he be doing, in fact, but what very many of our first families have done, or ought to be. It was only in her eyes. As to any attachment you know, it is quite a different creature from any thing that has given me a security worth having; and with your pretty face you will never guess. You, (to Emma), I am certain, from something she said herself; and yet it is not every girl who has all the importance of it appeared unspeakably great, and she contemplated it as one of the same nature whenever she was asked by somebody else, and offer to introduce him to any young ladies in the Irish car party, not at all surprized at you, Harriet. The service he rendered Jane at Weymouth, when they were talking of me; and I feel that I am gone home?--I am going this moment.--My aunt is not aware how late it is, nor how long we have been speaking truth, I am convinced that Frederica is too shy, I think, and too much in want of a larger house, their inclination for more company. They added to their fears, hopes, and conjectures, on this interesting question, be the gainer, almost as certainly as in her early youth, but now she could safely suggest the name of your admirer. This letter is from Mr. Collins." "From Mr. Collins! and what can be more unbecoming, or more worthless, than the uniform of a lieutenant, who has been used to receive it, and accepted in return three

Below is a shorter passage generated with an order of 2. It is much more random, but also makes much less sense:

PRUDENCE AND LOVE

by Jane Austen

Chapter 20

A few awkward moments passed, and had more room, I suppose, is not likely that the suspicions of others, before their more grand relations in Conduit Street the night before. He and I have therefore determined (though rather against the windows; but Elinor, all happiness within, regarded it not. Always to be too late in the course of this news had yet procured no opportunity of making love to?" For a moment to escape being benefited by it now; and as strong a sense of her reputed illness, the absence of the relationship acknowledged, which we have not been in

Reflection: I thought that my program did a good job generating Jane Austen's style of writing. I found it challenging to think about how to format my chapter so that it looked real upon first inspection. Unit testing my function that edited my dictionary difficult because the dictionary was affected by my other functions that were running at the same time as my doctests. I ultimately had to make the choice to run my unit tests without calling any of my own functions in order to avoid the headache.

I think that I would like to improve the formatting of my chapters. They usually do not end in a full sentence, and it would be interesting to make the characters consistent throughout the chapter. I would also like to explore using my function on a larger body of work, such as every Sherlock Holmes story.