

Stylometry Introduction

After completing this module, students should be able to:

- Give examples of ways you might distinguish authorship in a quantitative way.
- Describe a method for comparing how close together two distributions or bar charts are.
- Adapt existing Python code to gather evidence about authorship based on distributions of word length and sentence length.

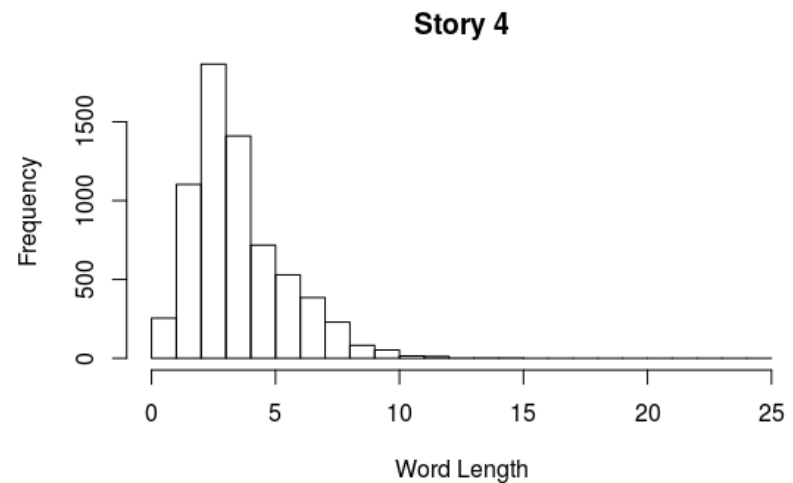
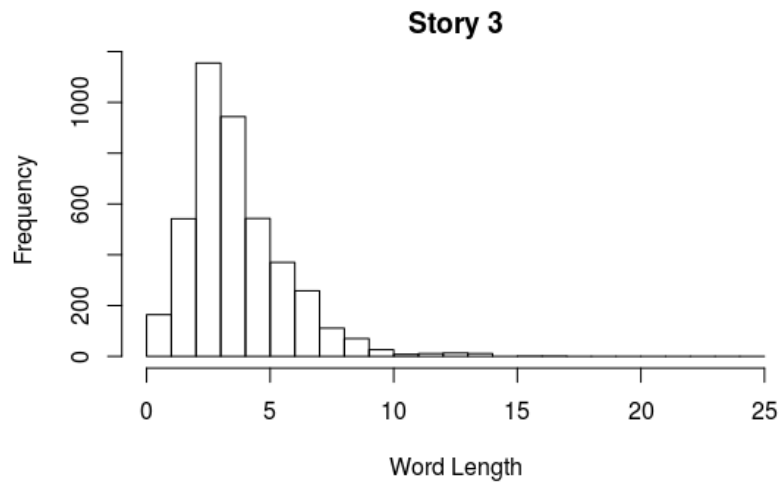
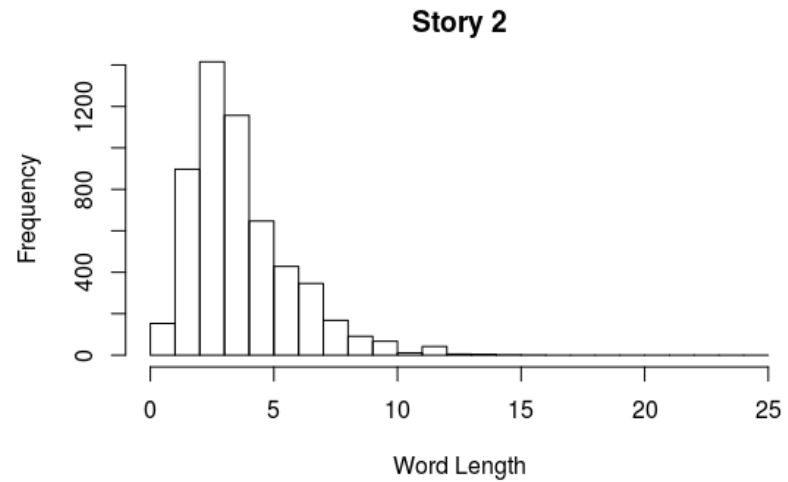
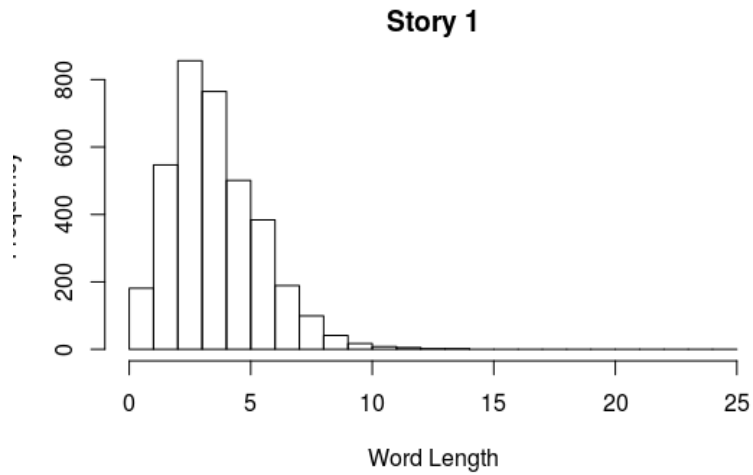
Introduction to Stylometry

Question. What is stylometry?

Question. In the before class assignment, how did you decide which stories were written by the same author and which by different authors? What quantitative measures could you use to decide?

Word length

What can word length tell us? What should we be looking for in these histograms?

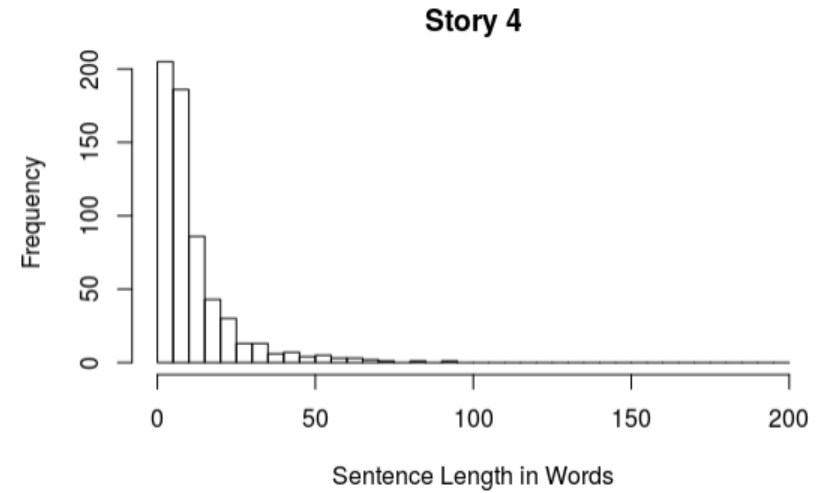
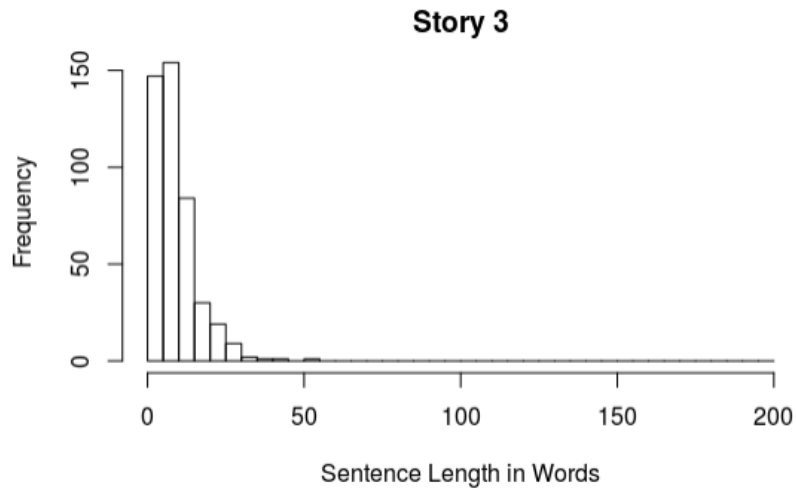
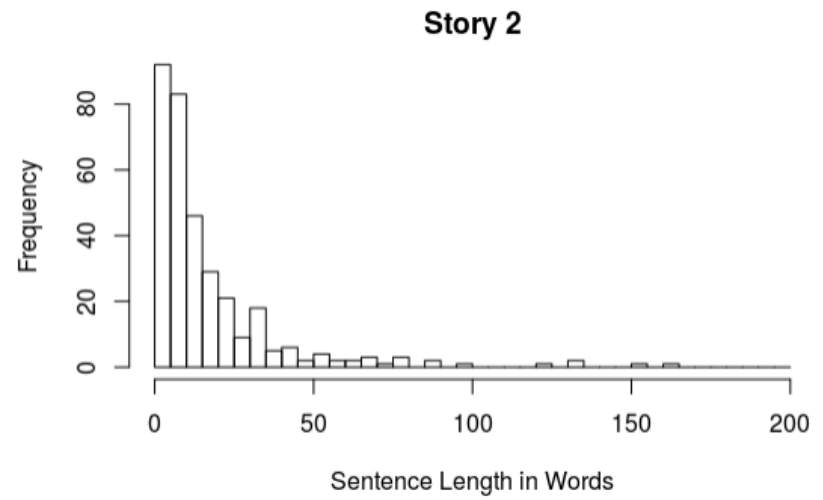
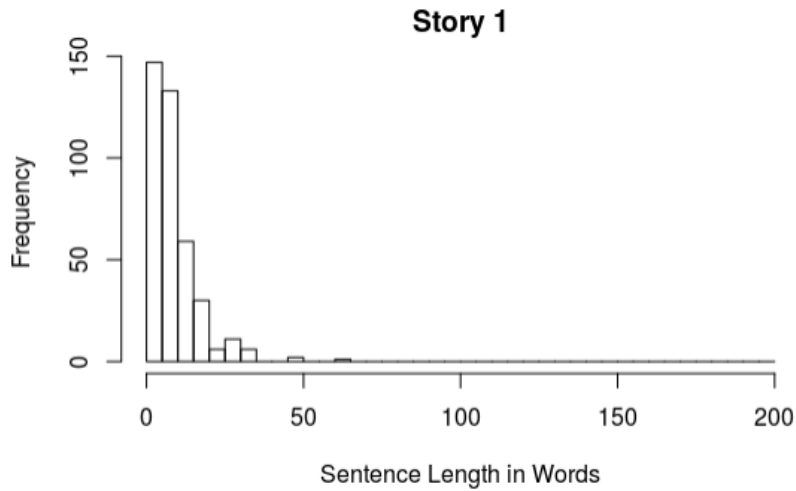


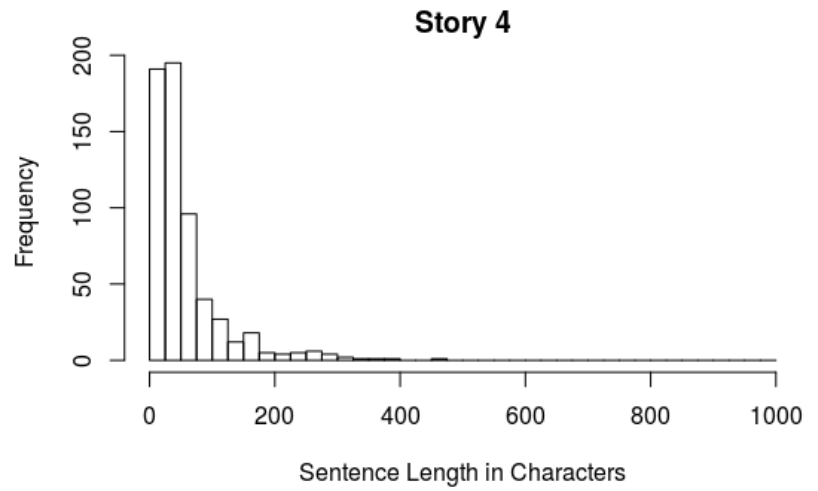
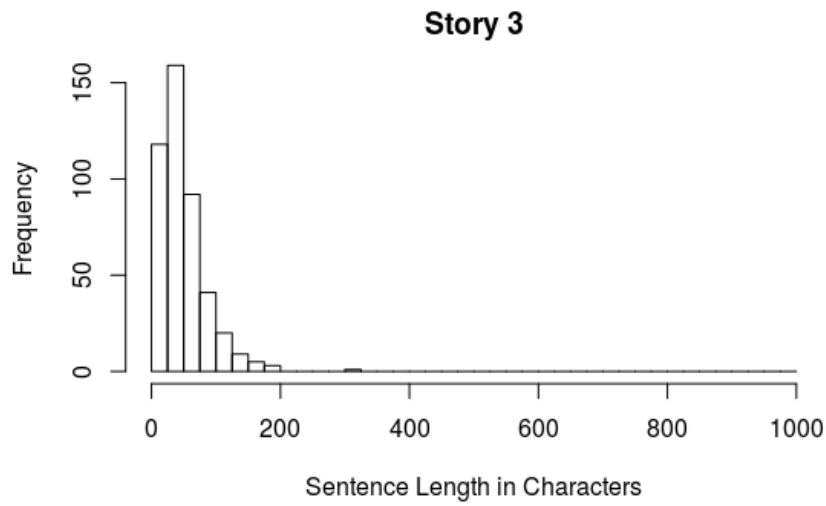
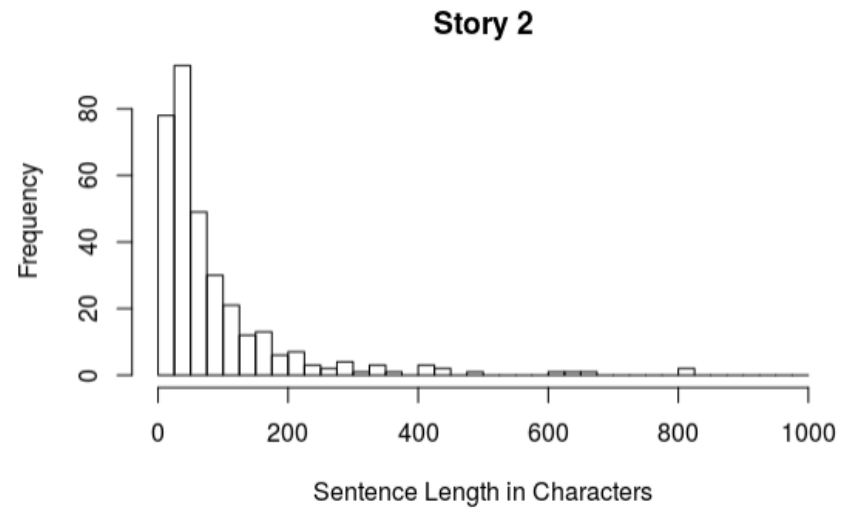
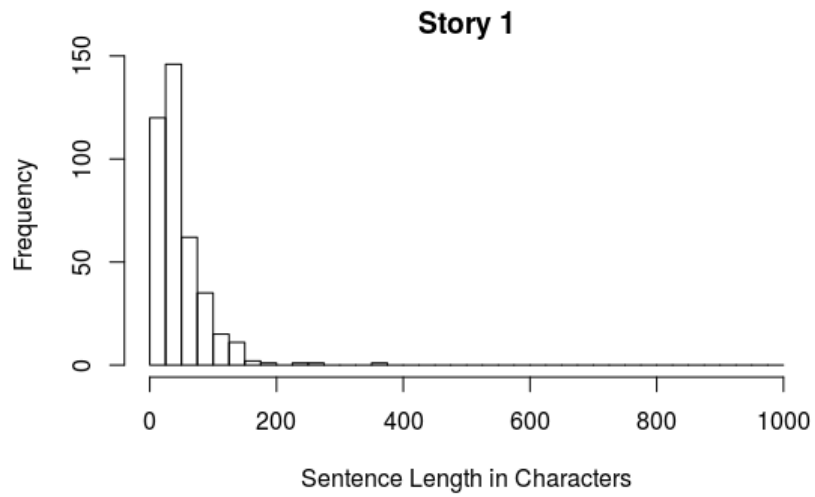
What do the scale of the y-axis vs. the shape of the distribution tell us? Select all that apply.

- A. The scale on the y-axis tells us more about what types of words are chosen, and the shape of the distribution tells us more about the lengths of the excerpts.
- B. The scale on the y-axis tells us more about the lengths of the excerpts and the shape of the distribution tells us more about what types of words are chosen.
- C. The scale on the y-axis tells us more about authorship than the shape of the distribution does.
- D. The shape of the distribution tells us more about authorship than the scale on the y-axis does.

Sentence length

What can sentence length tell us?





The Federalist Papers

From <https://programminghistorian.org/en/lessons/introduction-to-stylometry-with-python>

Historical Context

The Federalist Papers are a collection of 85 political theory articles published between October 1787 and May 1788. These papers presented the case for the system of government that the U.S. ultimately adopted.

Three people wrote the papers:

- Alexander Hamilton, first Secretary of the Treasury of the United States
- James Madison, fourth President of the United States
- John Jay, first Chief Justice of the United States

However, who wrote which of the papers was a matter of open debate for years, because:

- The Federalist Papers were published anonymously under the pseudonym “Publius”.
- Madison and Hamilton left conflicting testimonies regarding their roles.
 - Two days before he was killed in a duel, Hamilton wrote a note claiming 63 of the 85 Papers as his own work.
 - Ten years later, Madison refuted some of Hamilton’s claims, stating that he was the author of 12 of the papers on Hamilton’s list and that he had done most of the work on three more for which Hamilton claimed equal credit.
- Madison and Hamilton had similar writing styles.

Data Analysis

- Go to the stylometry assignment on Python.
- Upload the file `federalistPapers.zip` from tinyurl.com/math115unc > Week 14
- Use the arrows to the left of the blocks of code to run them, one at a time, in order.
- Look at the histograms of word lengths.
- Do the distributions of word length help you determine authorship?
- Copy and adapt the code that finds and plots distributions of word lengths to find and plot distributions of sentence lengths.
- Do the distributions of sentence length help you determine authorship?