Course title: Digital Humanities Prof. J. Savoy

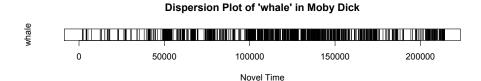
Serie 3. Literature

After reading the novel *Moby Dick*, a reader asks for some overall view of this (long) novel with some statistics and graphical representation. (This novel is available in data folder of the second chapter).

The questions:

- 1. How many distinct word-tokens do you count in this novel?
- 2. What is the vocabulary size of this novel?
- 3. How many occurrences of "Swiss" do you count in this novel?

As additional question, the reader wants to know where and with which intensity Melville (the author of *Moby Dick*) is talking about 'whale' (or 'whales', 'Whale', ...). One of his friends produces this picture:



- 4. Can you produce something similar with the word 'Ahab' (the captain name) and the string 'whale'?
- 5. Can you produce something similar with the word 'upon?
- 6. Do you see similarities and differences between these three views?

PS: The above picture was produced with R... don't try to obtain exactly the same (it is impossible with the time you have to do it).