Exercise 3 - Literature and Moby Dick

First name: Brian

Last name: Schweigler

Matriculation number: 16-102-071

Q1: How many word-tokens do you count in this novel?

General imports and set-up and solving the question:

```
In [1]:
%load_ext autoreload
%autoreload 2
%matplotlib inline
import matplotlib.pyplot as plt
import re
import numpy as np

with open('data/MobyDick.txt') as f:
    lines = f.read()

words = re.split(r'\W+',lines.lower())
```

Note that a part at the beginning and end do not belong to the book, so we adapted the text file and removed those. The adapted file is included in the upload.

```
In [2]:
print("Word-tokens: ",len(words))
```

```
Word-tokens:  214890
```

Q2: What is the vocabulary size of this novel?

```
In [3]:
print("Distinct words: ",len(set(words)))
```

```
Distinct words:  16873
```

Q3: How many occurrences of "Swiss" do you count in this novel?

```
In [4]:
swiss = [w for w in words if w == "Swiss"]
# swiss = re.findall('swiss',words)
print("the word Swiss occurs:" ,len(swiss))
```

```
the word Swiss occurs: 0
```

Q4: Can you count the number of plays per author?

This was done in exercise 2 (the last series), thus I assume this is a mistake as we only have one author here.
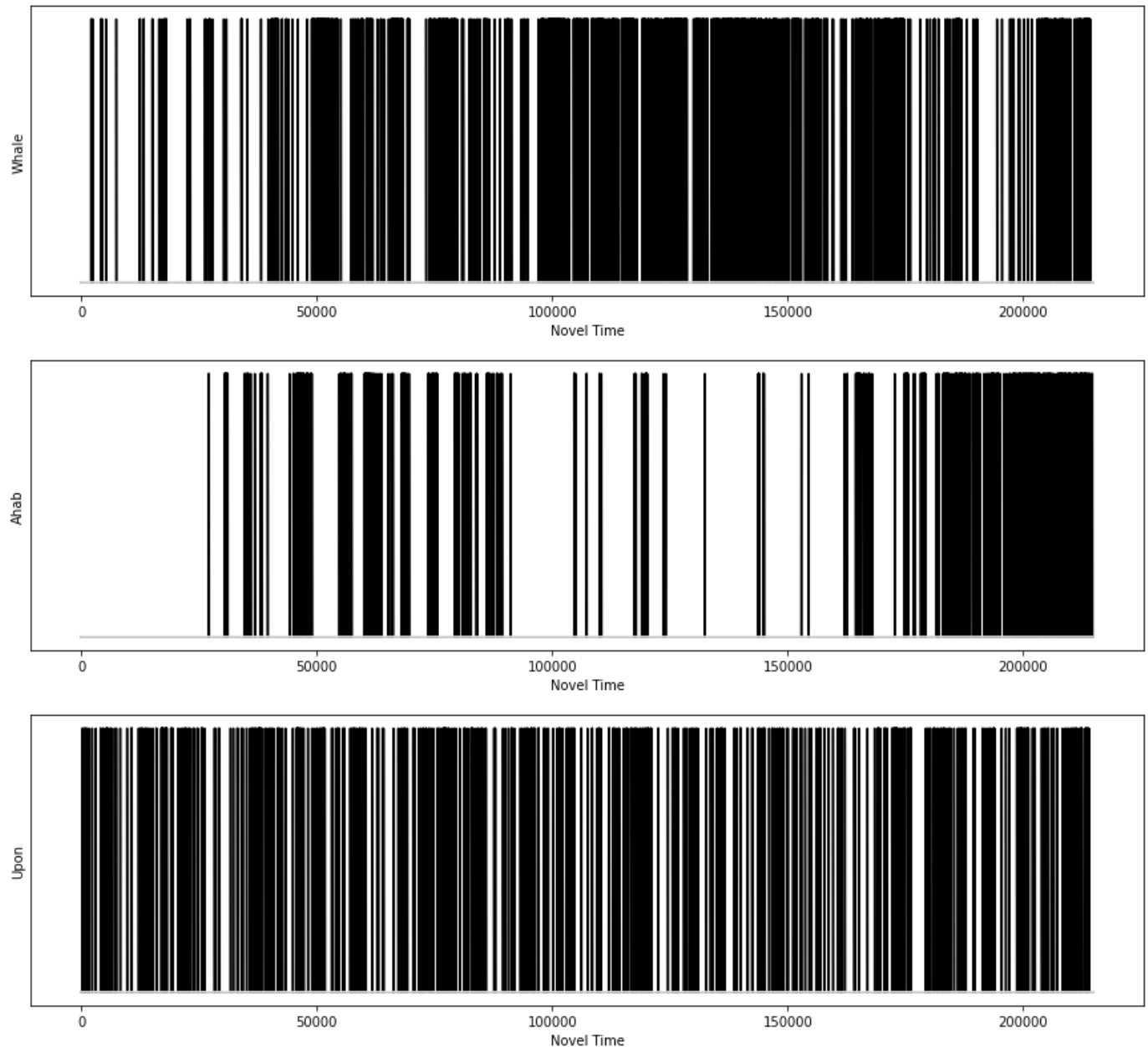
Q5: Can you generate a similar view for the word "whale"? (+Q6 "Ahab", and Q7 "upon")

Define a plot and then we can also do question 6 and 7.

```
In [5]:
def dispersionPlot(string_to_match: str):
    mask = [ 0.0000001 if w == string_to_match else 0 for w in words]
    fig = plt.figure(figsize=(15, 4))
    plt.plot( mask, color = 'black' )
    plt.ylabel(string_to_match.capitalize())
    plt.xlabel("Novel Time")
    plt.yticks([])
    plt.plot( np.zeros(len(mask)), color = 'white')
    plt.show()

    #from the text above the exercise i conclude that we also should match Whale
```

```
for word in ['whale','ahab','upon']:
    dispersionPlot(word)
```


Q8: Do you notice any similarities/differences between these three views?

Upon is quite evenly distributed.

Whale appears a lot more frequently in the middle of the book, probably towards the climax (and whale encounters).

Ahab appears almost only towards the end; maybe a change from first-person to third-person in narrating style?