NLP - Assignment 3

In this assignment you will analyze the emotional arc of several books.

Preparations

1) Select five books that you find interesting and that do not drastically differ in length. The smallest book should be more than 50% of the size of the largest book.

2) Following the steps of the the previous assignments, read in the texts, extract the main text from them, and create a tibble with five rows that looks like this.

```
## # A tibble: 5 x 2
##   book             text
##   <chr>            <chr>
## 1 Alice in Wonder~ "\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\nALICE'S AD~
## 2 Dorian Gray      "\r\n\r\n\r\n\r\n\r\nProduced by Judith Boss.  HTML ver~
## 3 Huckleberry Finn "\r\n\r\nProduced by David Widger\r\n\r\n\r\n\r\n\r\n\r~
## 4 Peter Pan        "\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\n\r\nPETER PAN\~
## 5 Treasure Island  "\r\n\r\n\r\n\r\n\r\nProduced by Judy Boss, John Hamm a~
```

2) Next use `unnest_tokens` to tokenize the text.

3) Now use `tidyverse`'s `group_by()` and `mutate()` functions to add a variable `pos` that codes the position of a word inside the respective books.

```
# add pos variable
token_tbl <- token_tbl %>%
  group_by(XX) %>%
  mutate(pos = 1:n(),
         rel_pos = pos / max(pos)) %>%
  ungroup()
```

Sentiment analysis

1) Extract the *afinn* sentiment dictionary using the `get_sentiments` function and store it in an object called `afinn`.

2) Use `inner_join` to combine your `token_tbl` with `afinn`.

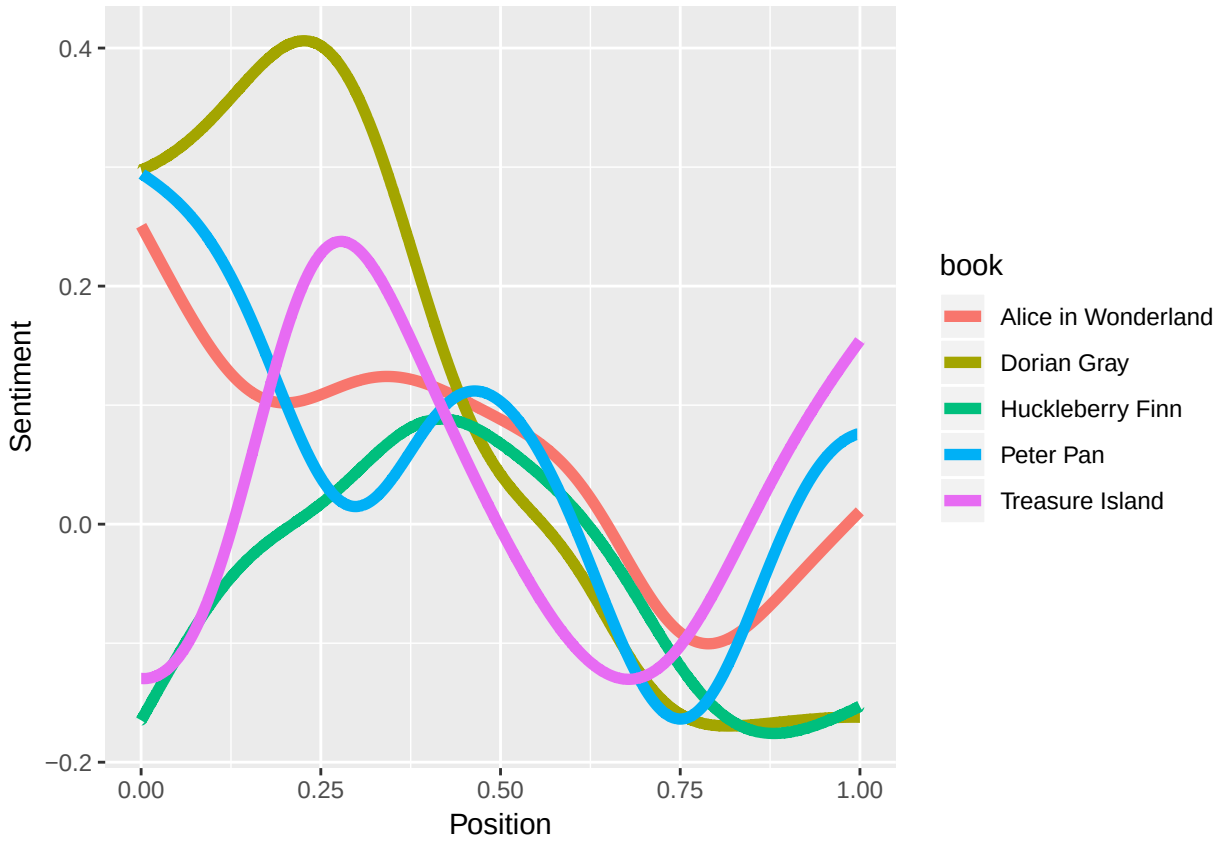
Smoothing

1. Use the `group_by` - `summarize` idiom along with the smooth function below to calculate more interpretable, smoothed sentiment scores for each of the books.

```
# smoothing function
smooth = function(pos, score){
  sm = sapply(pos, function(x) {
    weights = dnorm(pos, x, max(pos) / 10)
    sum(score * (weights / sum(weights)))
    })
  }
```

2. Use the code below to create a plot like this:

```
ggplot(token_tbl,
       aes(rel_pos, smooth_score,color=book)) +
  geom_line(lwd=2) +
  labs(x = "Position", y = 'Sentiment')
```



Project proposal

Come up with 1 or 2 project proposals, each about half a page long. Address which question you would like to address and which data you want or would like to use for it.