# Assignment_4 rubric
# Text Analysis

**Note: This rubric is in addition to the rubric included in the course syllabus.**

---

**Submission Pts: 15 - submission downloads and runs, 7 - doesn't download or run, 0 - no submission**

(1) Assignment is submitted using an assignment-specific GitHub link. Remember, this is an individual assignment. You may consult with your classmates, but reference classmates who share techniques, methods, or code that you use. You will not lose points for these references and may gain points if your commentary includes modifications and improvements that you have been able to make. Of course, you must reference all other reference material, too.

---

**Task TWO: bag of words analysis**

**(2,3,5) 15 Outstanding, 10 Adequate, 5 Needs more work**

**(4) 10 Good**

(2) Submit a sentiment analysis of the book you have chosen. The submission should include a report in PDF and the Rmd file you wrote to produce the report.

Use the approach to sentiment analysis detailed in Text Mining with R. The analysis should include visualizations, including a visualization of the sentiment flow in the book. You should compare the results when different lexicons are used. Lexicons differ in the kind of output they produce -- signed real numbers, binary outcomes, multi-dimensional indicators, and so you. What do these outputs mean? Discuss how you understand and use them.

(3) Commentary in the Rmd file should explain how you organized the data used in the analysis. Be sure that your commentary describes important data structures (e.g., data frames, lists, and so on ) and is sufficient for reproducing your analysis with different books.

(4) **Extra credit:** Text Mining with R discusses three different lexicons for sentiment analysis, but there are many more. Find additional lexicons and include them in your analysis. Explain how you selected the additional lexicons you used and reflect on the results.

Produce a visualization, using the data frame mentioned above in (2), has been created that shows the sentiment progression from start to finish of the book.

(5) Provide a verbal description of how your visualization matches (or does not match) the plotline of your book.

---

**Task THREE: sentence-level analysis**

**(6, 7, 8, 9) 10 Outstanding, 8 Adequate, 5 Needs more work**

(6) Use the tnum ingester to load your book into the test2 number space. Then explore your data using the tnum package. You should verify that you can access the text

(7) Produce an analysis of your book. Start with using sentimentr. Keep it simple.

(8) Now compare the bag of words analysis you did in Task TWO with what you can do from TN. You may need to re-do the bag of words analysis so you can get sentiment scores for sentences or paragraphs.

(9) Prepare another report that is parallel to the analysis you did in Task TWO. Use as much visualization as you can, being sure to fully label and discuss your visualizations. Your commentary should include questions, notes, and implementation details. If you see a need for features the tnum package does not yet have, include them in your report.

---

**EXTRA CREDIT: character analysis**

**(8, 9) Good**

(8) The Truenumbers tag feature can be used flexibly. For text analysis, tags can be used to mark where characters appear, speak, or act. You can tag passages where characters appear, interact, or speak with other characters.

Visualizing when, where, with whom characters appear in a book can guide your understanding of how the book tells its story, delivers its message, builds or releases tension. When combined with sentiment analysis, character analysis helps you understand the sentiments in the book and identify the characters with the sentiment.

(9) A visualization has been created using character occurrence and sentiment values* has been created to provide information about the storyline of the book.

*You will need to decide the best way to tokenize your book and compute sentiment values, to meet the needs of your character level analysis.