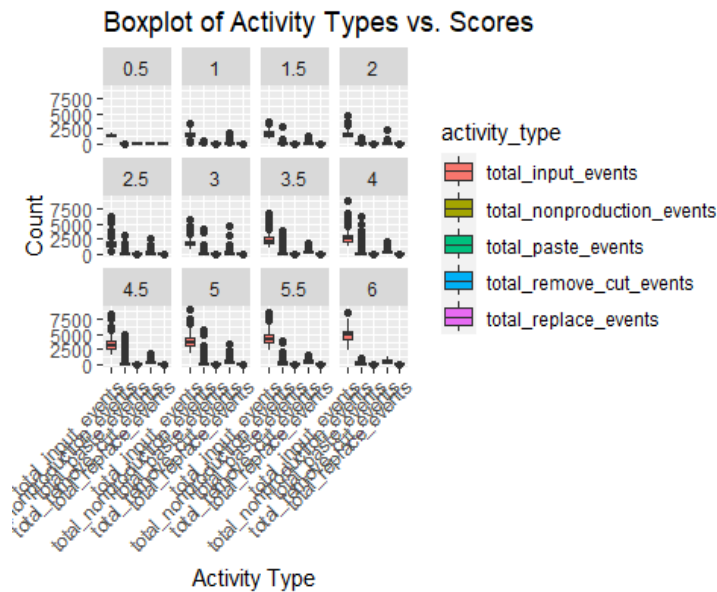
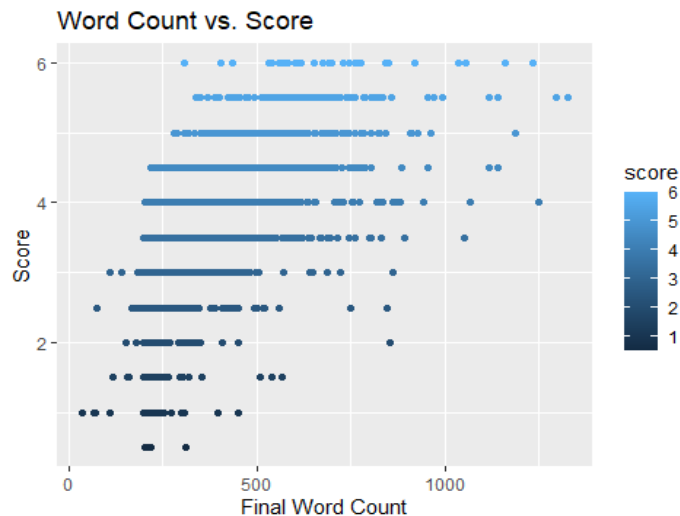


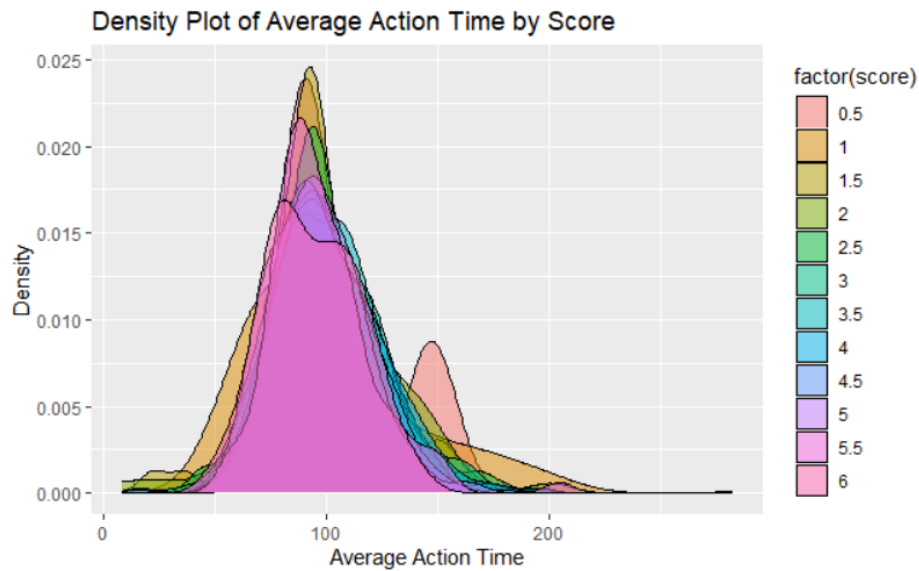
Data Processing:



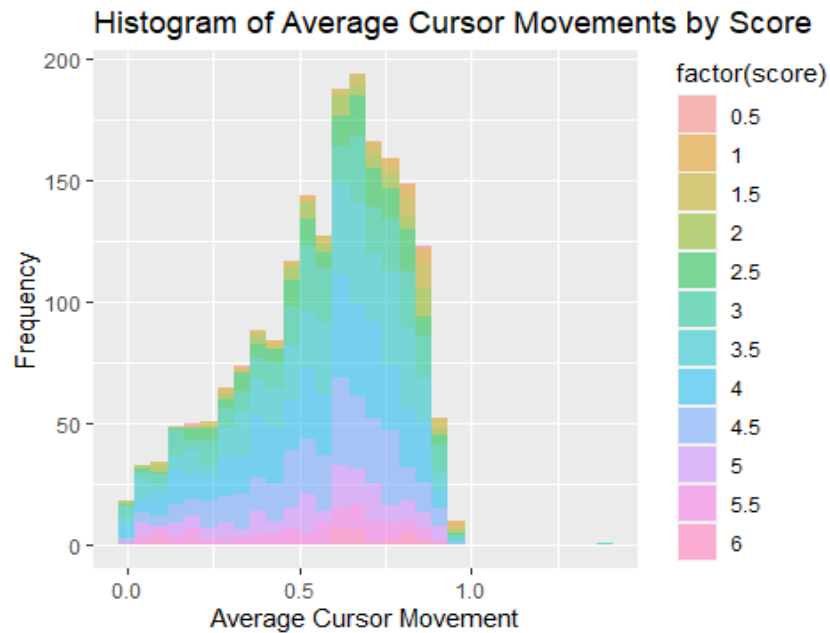
The median counts for most activities seem to increase with the scores up to a certain point, suggesting a correlation between the amount of activity and higher scores. Some activities, like “total_input_events” and “total_replace_events,” have wide variations at higher scores, indicating that these activities are more variable for essays with higher scores. “Total_remove_cut_events” and “total_nonproduction_events” appear to have a more stable count across different scores, suggesting these activities may not be as strongly correlated with the scores as others.



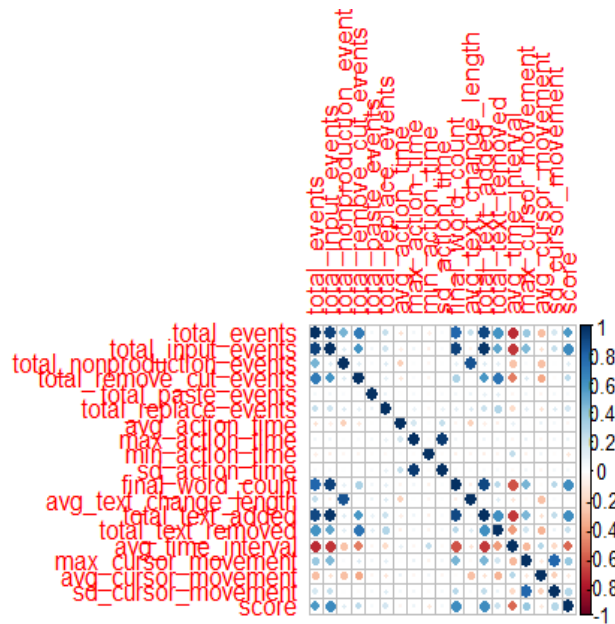
There is a potential correlation between longer essays and higher scores.



There seems to be subtle indication that the highest scoring categories tend to have slightly quicker average action times. Histogram of Cursor Movements: distribution of cursor movements for different scores.



The visual suggests that there may be an optimal range of average cursor movement that is associated with higher scores, while both very low and very high cursor movements are less frequent among higher-scoring activities.



There's a decent amount of collinearity. We will remove some variables that are kind of redundant.

After EDA, we finalized predictors. Each of these predictors can provide insights into the writing process of each essay.

total_events: This is the total number of input events for each essay. A higher number of events might indicate more effort, which could be associated with higher quality writing.

total_nonproduction_events: These events don't alter the text. A high count could imply extensive planning, which might be a trait of careful writing.

total_remove_cut_events: The number of events where text is removed. A higher number might suggest significant editing and refining of the essay, correlating with better content quality.

total_paste_events: Indicates the number of paste actions. This could reflect the use of external references.

total_replace_events: Shows how often text is replaced. Frequent replacements might indicate refinement and improvement of the essay's content and structure.

total_move_events: Counts events of moving text sections. This might be indicative of significant restructuring, suggesting an effort to improve the essay's flow and coherence.

avg_action_time, max_action_time, min_action_time, sd_action_time: These metrics give insights into the duration and variability of actions. Longer action times might indicate more thoughtful writing, while a higher standard deviation could suggest varying writing speeds, possibly reflecting different stages of the writing process.

final_word_count: The word count of the essay after the last event. Generally, longer essays might cover more content. This metric helps capture the essay's scope.

avg_text_change_length, total_text_removed: These features relate to the extent of text changes. Large amounts of text change might indicate significant editing and refinement.

max_cursor_movement, avg_cursor_movement, sd_cursor_movement: These metrics measure how much the cursor position changes, reflecting the extent of navigation and editing. Significant movement might indicate active organization and editing of the essay.