What do you want to achieve with your visualization?

The main goal of my visualization is to help people better understand what makes a text "difficult" to read. This is done by educating users about the underlying factors that influence the overall reading difficulty score of a given passage as calculated by the Flesch Reading Ease Test. Just quantifying the reading difficulty of a passage only provides an overview of how hard a passage is to read. Using my visualization, people will be able to see how a passage's Flesch Reading Ease score is influenced based off a combination of total word count, total sentence count, total syllable count, and total complex word count, where a complex word is defined as having three or more syllables.

What tasks do you want to support?

My visualization will achieve its goal by accomplishing the following tasks:

1. show users the Flesch Reading Ease scores for all Fairy Tales given
2. illustrate the following different factors that influence the overall Flesch Reading Ease score:
   a. Total Number of Words
   b. Total Number of Sentences
   c. Total Number of Syllables
   d. Total Number of Complex Words
3. allow users to compare two stories at a time in order to observe how the different readability scores are the result of different values of factors.

Factors (a), (b), and (c) were determined to influence the overall Flesch Reading Ease score because they are all variables used to directly calculate the score. Even though Factor (d) is not a direct variable used in the equation for calculating Flesch Reading Ease, it still directly affects the overall score since a higher number of complex words will contribute to a higher number of total syllables, which increases the reading difficulty of a passage.

What designs will help you achieve these designs (Name at least two)?

Because my visualization aims to visualize four variables at the same time, my first idea is to use a parallel coordinates chart that has four y-axes to represent the values of the four factors previously described for each story. The color of each line on the parallel coordinates chart will represent the flesch reading ease score. The parallel coordinates chart will accommodate multiple lines for multiple fairy tales that the user can use to compare.

My second idea is to create a radar chart that creates a polygon whose vertices represent the values of each of the four factors. The color of the polygon will correspond to its flesch reading ease score. Users will be able to select up to two stories that my visualization will generate polygons for so that they can compare the stories using the radar chart.

Sketches
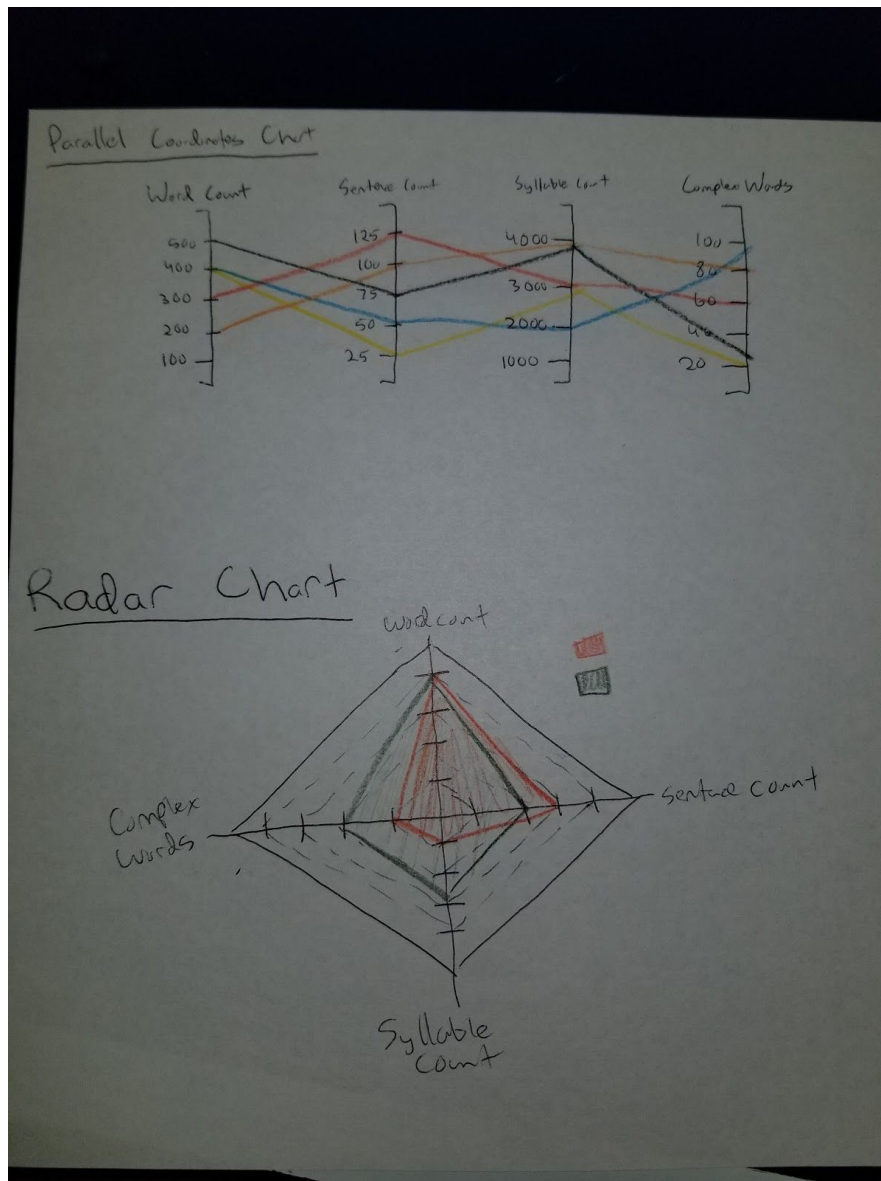
Justification for Final Design

My final design ended up being a radar chart. Initially, I set out to create a parallel coordinates chart, but found that overlaying multiple stories at once produced a lot of clutter, which made it difficult to derive insight from any single story's line. Even though I could have implemented some method of brushing and limited the number of lines that could be added to a chart, I ultimately felt that using a radar chart would do a better job

of accomplishing my main goal because its design made it easier to compare the value of one factor to another using the radar chart's polygon layout.

To reiterate, the main goal of my visualization is to help people better understand what makes a text "difficult" to read. This is done by educating users about the underlying factors that influence the overall reading difficulty score of a given passage as calculated by the Flesch Reading Ease Test. From a high level, this is accomplished by users being able to generate radar charts for fairy tale stories and to analyze how stories of specific readability scores will have certain factor values. My visualization visualizes these underlying factors by drawing a polygon on the radar chart, where the distance of each vertex from the origin represents the value of the respective factor. Users can hover over a specific polygon in order to trigger a tooltip that displays all four factor values, as well as the story title and flesch score. This accomplishes the task of *"illustrate the following different factors that influence the overall Flesch Reading Ease score"*. Another task I stated previously was to allow users to compare up to 2 stories at once in order to observe how the different readability scores are the result of different values of factors. I did this by creating a tile-mosaic of buttons that each represent a different story that a user can click on in order to generate a polygon on the radar chart. Up to 2 stories, and polygons, can be generated at the same time. Once done, the user can easily compare the different size of each story's polygon in order to observe why their reading difficulty scores might differ. At the bottom of the radar chart, two cards are generated that each contain the factor values, flesch score, and title of each of the two stories selected. That way, user's don't have to keep hovering over the polygons to look at the actual numbers. The last task was to show users the Flesch Readability score for each selected story. Though the user can already view the score using the cards or hover tooltips, my visualization also accomplishes this task by using a color scale to quickly give user's an idea of how "difficult" a story is to read compared to other stories in the corpus. Initially, my color scale function was based off of a [0,100] domain to reflect the 0-100 scale of the Flesch Test but because the flesch reading scores of all the stories were generally above 70, all the colors generated would be relatively similar. I decided that this would make it harder for users to differentiate the reading difficulty scores of one story from another, so I modified my color scale so that it would reflect the difficulty of one story compared to all the others in the corpus. Specifically, the lowest Flesch Score out of all 43 fairy tales was ~67 while the highest Flesch Score was ~99. Because of this, the new domain was set as [67,99].

Sources:

- To learn about the Flesch Readability Score, I consulted the following websites:
  - https://support.office.com/en-us/article/test-your-document-s-readability-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2#__toc342546557
  - https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests
  - Additionally, I'd like to cite **Emily Wilson** for helping to inspire my project by telling me about the possibility of using the Flesch Readability Score for my project. Thanks Emily!
- To parse the text data into the appropriate formats, I used the following python libraries:
  - Unidecode: https://pypi.org/project/Unidecode/
  - TextStat: https://pypi.org/project/textstat/
  - Textatistic: https://pypi.org/project/textatistic/
- I used the following D3 libraries:
  - D3-Tips: https://github.com/Caged/d3-tip
  - D3-Scale-Chromatic: https://github.com/d3/d3-scale-chromatic/blob/master/README.md
- In terms of actually coding my visualization, I consulted the following resources:
  - https://bl.ocks.org/pstuffa/d5934843ee3a7d2cc8406de64e6e4ea5
  - Some parts of my project include the template code from the previous Electoral Votes Assignment in this class. These code snippets are cited in my code
  - https://www.w3schools.com/howto/howto_css_cards.asp
  - https://stackoverflow.com/questions/507138/how-do-i-add-a-class-to-a-given-element