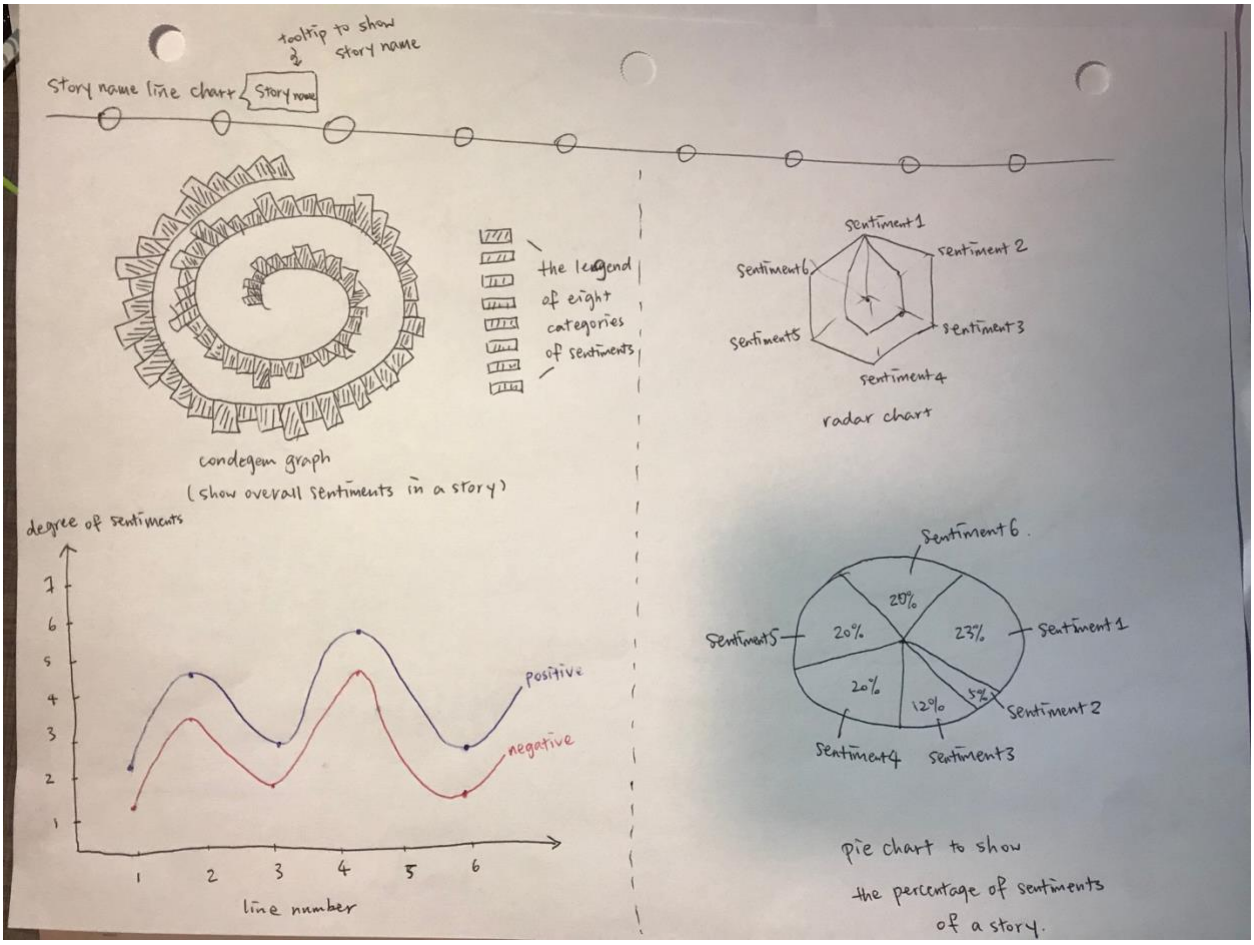
Design:

a) What do you want to achieve with the visualization?
What I want to achieve in my visualization is to analyze the text data by using python and R library, then I can get the sentiment data about each story. The goal is to let users have better understanding of the sentiment trend of each story in the English fairy tales.

b) What tasks do you want to support?
At first, I will design a story name line chart that users can click on it and see the detail analysis of the story. And then when users want to see the detail of the data/information by hovering mouse on the graph.

c) What designs will help you achieve these designs? Name at least two.
1. For all five charts, red represents negative sentiments and blue represents positive sentiments.
2. For the story name line chart, it is designed for users to click on it and view different stories separately.
3. A condegram graph which is designed for showing the overall sentiments from the first line of the story till the end.
4. The radar chart can help users to see which sentiment has the highest amount among the story.
5. The line chart can also help users to know the trend of positive sentiment and negative sentiment and compare the difference from the very beginning till the end of story.
6. The pie chart can help users to know the percentage of each sentiment in the entire story.

Sketches:



Justification for the final design:

a) I use the python and R library (getSentiment library from R and the reference is https://www.tidytextmining.com/sentiment.html), the data I used in this assignment is the twelve stories of the English fairy tales, and the first column of the data file is the line number of the story, the second column is the type of sentiment, third column is the appear times of the sentiment in this line and the forth column is weight of the each sentiment. The weight of sentiment is designed based on the degree of positive and negatives sentiments. I put all these data files in the folder called "sentiments_new". The parseData folder is to show how I use library to parse data.

b) The story name line chart is to allow users to click it and choose the story and see the detailed results of four charts which analyze data in four different parts and it will update the result when you click the other circles.

c) For the first part, I designed and implemented the condegram chart to show the overall sentiments through the whole story, from the beginning to the end, and you can see the "start" sign and the "end" sign on the graph. Each color represents different sentiment and I showed the

color legend on the right. From the top (means extremely negative) to the bottom (means extremely positive) with the name of each sentiment. The bar height represents the number of times the sentiment appears in each line. Through the condegram chart, you can clearly see the changes of sentiments through the whole story. Users can hover mouse on each circle to see the name of the story.

d) For the top right chart which is the radar chart, I designed it to show the total amount of each sentiment in the story. It is the second part of sentiments analysis that users can obviously see the total number of times of each sentiment appears in the story, and I also colored the polygon based on the sum of positive sentiments (surprise, anticipation, trust and joy) and the sum of negative sentiments (fear, disgust, anger and sadness). If the sum of the positive sentiments is larger than the sum of the negative sentiments, then it will be colored by blue, otherwise red. Users can hover mouse over the circle points to see the detailed information/data of each sentiment.

e) For the bottom left chart, I designed and drew a line chart to show the trend of positive sentiments and negative sentiments in the whole story. The x axis is the line number, which represents each line of the story, the y axis is the degree of sentiments. For the each data point, I use the formula weights (aka. e_depth from the data file) times the number of appear times, and then calculate the sum of the positive sentiment for each line and the same as for negative sentiment. In this way, it can better show the trend of the sentiment in this story. If a really bad word shows up many times in this story, it will possibly lead this story to a negative sentiment. The weight here is just as the cost function. Users can interact with the chart by hovering the mouse over the line to see the trend of the line clearly and users can also hover the mouse over the circle points to see the details of the data.

f) For the bottom right chart, I designed a pie chart to show the percentage of each sentiment in the story. How I calculate it is based on the sum of each sentiment in the story divided by the total amount of the sentiments in the story. Users can clearly see which sentiment (positive or negative) is the dominance of the story. Users can interact with the chart by hovering the mouse over pie chart or the name of the sentiment, it will show the detail data in the center of the pie.

Based on these four charts, users can easily know the trend sentiment of the story and understand the reason why the circle is colored by red or blue in the story name line chart. And also, users can have a better understanding of the story sentiments before or after they read the story.