Overview and Motivation:
Social networks have changed the way people communicate with each other. Today the world is lot more connected than it was before. People are also more open towards voicing their opinions regarding issues of their interests. These issues can be related to sports, careers, politics, art etc.

Social network websites like Twitter and Facebook provide great resources to learn about the preferences and interests of people in different demographics. Based on those interests and preferences, it's easy to know about the sentiments and the location of the people. The data collected from these social networking websites is mostly in the form of raw text, but based on the context it's easy to know the sentiments and opinions of people residing in different demographics. Our goal for the project is to gather, group and analyze such text data and provide meaningful insights in the form of visualizations.

For our project we have decided to work with twitter data which was collected for over a week. Tweets are complex and have many different attributes. It would be very interesting to visualize such data and discover any patterns in terms of how people tweet, where they tweet and when they tweet.

Related Work:
We worked with a lot of numerical data during the class homework assignments, but never worked with text data. We felt that using the text data for our project would also give us the opportunity to think outside the box and come up with visualizations that draw meaningful insights on the underlying dataset. Also, since everyone uses social media, we felt that the text data gives us to have a larger target audience. Our main goal is to develop a tool which can be used by people of all age groups.
Visualizations are often used to understand the data on social media. For example, https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/. This website can give a sense of people's opinions on current events. We also would like to create visualizations which will help data make sense for people of all age groups and connect the data with the events and provide a bigger picture for everyone.

Data:
The twitter data comes from a project of the cse530s class. A script has been implemented to programmatically stream tweets from twitter streaming APIs. Approximately 1 million tweets have been downloaded for over a week. We are streaming more tweets from twitter and hope to discover more interesting results. To process the data, we implemented scripts to extract hashtags from the tweets. To find meaningful features, we also leveraged database and SQL queries to filter and aggregate the tweets.

Links:
Twitter Streaming API: https://dev.twitter.com/streaming/overview
NHGIS: https://www.nhgis.org/

Exploratory Data Analysis:
Initially we decided to use a Map with bubbles where the size of the bubble represents the number of tweets and the color of the states represent the sentiment for that state. We also decided to provide a timeline feature where the user can filter the map based on the timeline. Later we bound the the timeline feature with a search bar and a word cloud to add more usability features for the map. This gives the user more freedom to narrow down the visualization based on his/her visualization. We feel the map and the word cloud would be the must have visualizations for our project as they help to capture the sentiment of the people. We'll also implement theme river and/or parallel coordinates if we get the time towards the end to implement it.

Design Evolution:
We have kept the same design layout as the one mentioned in the proposal. We plan to implement a dashboard which has 3 pages in total. The first page which provides the summary of the project, the second the map visualization and the third implementing the world clouds. Initially we were planning to implement the zoom feature to display the counties, but later we decided to drop the idea mainly because the word clouds which will be generated are very less likely to be dense and informative. If we manage to include a bigger dataset of over a month, then we might decide to implement it.

Implementation:
So far we have implement the Map feature and a word cloud for the map with a timeline slider for the week and a search bar. The user can interact with the map, timeline, search bar and word cloud all at once or one at a time. The helps the user to narrow down the visualization based on his/her preferences . The color of the state will represent the sentiment for that state and the size of the bubble on the map will represent the total number of tweets for that state. The sentiment and the bubble size changes with the timeline, search bar and word cloud. Word clouds are also generated whenever a user hovers over a particular state on types in his/her search term in the search box. As far as tools are concerned we are using, bootstrap, javascript, d3, python and sql. We have built a JSON object which helps to filter the data based on the user interaction.

Evaluation:
From the map visualization developed so far, we have seen a change in the sentiment for each state. For ex: on St. Patrick day's (03/17) people seem to be in a more happy mood. Whereas for other days, the sentiment for the states was pretty neutral. Also, we found out that California and Texas were the two states which tweeted a lot during that week. We can improve our visualization by creating a bar plot to show the number of positive, neutral and negative tweet. We can also provide a gradient legend for the varying sentiment and a legend as a reference to

the bubble size for the number of tweets. For our next presentation, we plan to have a fully functional summary page and a map visualization.