Angelina Lu and Sabrina van der Gracht

Milestone 1: Project Plan

- Question/Problem
    - Given a Bob Ross painting, can you make a gallery of similar Bob Ross paintings?
    - Can you cluster the paintings into different categories?
    - It's not the most important question, but Bob Ross is a sentimental figure to many so we thought this topic would be engaging.
- Data sources:
    - https://data.world/fivethirtyeight/bob-ross
        - This is a data set we found on Data World that has boolean values for whether or not a specific element (such as tree, mountain, barn) is in a specific painting.
        - It also contains the episode number and painting title
        - variables include elements in the painting, frame types, and guests on the episode
    - https://www.twoinchbrush.com/all-paintings
        - This has all of the images associated with a season number and episode
- Methods
    - We plan to use KNN
        - We'll use sklearn's k-nearest neighbors class to find similar paintings to put in a fake "gallery"
    - We also plan to use k-means clustering
        - This will be a way to see if Bob Ross paintings fall into different clusters/categories
    - Anticipated Challenges
        - Connecting the images to the paintings by title/episode number may be very challenging since we may have to do that ourselves and attach those images to the first dataset we found.
- Visualizations
    - Ideally, we'd like to be able to actually show the Bob Ross paintings together
    - For graphs, we can use bar graphs to show most popular elements.
    - We can also graph the k-means clusters of paintings we found.https://www.twoinchbrush.com/all-paintings
- Grade Targets
    - Low Target
        - Identify the names of paintings that are similar to the inputted painting
    - Medium Target
        - K-means clustering to see if we can put the paintings in groups to categorize them.

- High Target
    - Display painting images attached to the painting names in a little virtual gallery.
    - Create a slideshow of the images in Jupyter (if possible - we're still researching this)