**Project Proposal**

Our group will be working with the Bob Ross dataset, <https://github.com/fivethirtyeight/data/blob/master/bob-ross/elements-by-episode.csv> collected and created by Walt Hickey, and accessed from FiveThirtyEight’s public Github repository. Bob Ross was a painter who hosted step-by-step lessons via television show in the 80’s and 90’s; tutorials with a different scene each time-- 403 in total. The data our group will be using was originally collected to see what elements were painted in Bob Ross’s 11-year television career on his PBS show, “The Joy of Painting,” and was collected through coding each episode of his series for aspects and content of his paintings. Though Ross’s TV show last aired live some 25 years ago, and Ross himself passed away in 1995, his popularity in the last few years has soared. In 2015, Twitch started streaming an archive of ‘The Joy of Painting’, which did so well that Twitch, a company mainly used by young millennials and Gen Z, began streaming old episodes of Ross’s show weekly. Additionally, Netflix added Bob Ross’s painting series in 2016. Thus, our target audience are younger, technology using individuals, estimated ages being 15-25.

Given the resurgence and somewhat unexpected success of Bob Ross within the past few years, our group wants to analyze exactly what made up the 403 paintings in the series, and what draws younger viewers to it. We expect our target audience will be interested in learning a number of things from our data, namely, which elements occurred most frequently in his paintings and in which season, which commonly occurring objects were often omitted from his work, how Ross’s subject matter changed over time and across the seasons of his show, and how the content of Ross’s paintings differed than the paintings of the various guests he brought on his show. To do so, our group will have to work with the .csv file of coded data, interpret each column, and filter columns that are relevant to answer each of our questions. We’ll need to calculate percentages of occurrence in order to conclude which elements occur most frequently, as well as which elements occur least frequently. We’ll also need these calculations to be filtered based on what the audience wants to see, in order for them to interact; we’ll need a user to be able to input a season, artist, and/or an element and see the filtered result. We’ll need to utilize ggplot2 to create meaningful and informative graphs as well as a tool to build interactive visualization, such as htmlwidgets. We’ll need statistical analyses to find the joint and conditional probabilities to make our interactive plots. Our groups expects to have challenges in calculating the joint probabilities, as well as creating interactive plots since we have yet to do so.