**Second Capstone Project Proposal: Bob Ross' Happy Trees**

1. **What is the problem you want to solve?**

Understand patterns in viewership of Bob Ross painting shows.

2. **Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?**

YouTube or the manager of the Bob Ross YouTube channel could decide which episodes to promote to YouTube visitors based on content.

3. **What data are you going to use for this? How will you acquire this data?**

I will use data containing episode information, and paintings categorized ('tagged') by content, e.g. mountains, streams, etc. These data are found here: https://github.com/fivethirtyeight/data/tree/master/bob-ross. I will combine these data with data pulled from the Bob Ross YouTube channel, such as number of views: https://www.youtube.com/channel/UCxcnsr1R5Ge_fbTu5ajt8DQ.

4. **In brief, outline your approach to solving this problem (knowing that this might change later).**

- Gather data from YouTube Bob Ross channel (API)
- Combine with category data set
- Initial data summary:
    - Are some characteristics more often painted together?
    - Did Mr. Ross have a *most favorite* topic?
    - Run images of paintings through a pretrained image net type classifier to create a deep learning based feature representation
    - Cluster paintings based on provided categorizations as well as the categories from the image classifier
- Use a machine learning model to predict popularity on YouTube based on painting subject matter.
- Use an unsupervised machine learning model to develop customer segmentation of YouTube viewers

5. **What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

The final products will include a python notebook, and a blog post describing the outcomes of the analysis.