Our independent variable, popcorn brand, appears to be multimodal when looking at the histogram and scatter plot. However, when looking at the kde plot the independent variable appears to be unimodal. I believe that the reason this is the case is there are only 12 trails and each mode doesn’t have enough frequence to appear as its own mode in the kde. When looking at the histogram there appears to be two modes one for each brand, because one brand appears to pop more than the other. Since there is only really one dependent and numerical variable that isn’t the trial number we can’t identify any clusters from our scatter plots. This is because it would not make sense for popcorn to pop more based off the trail number, but instead pop more based off of the brand. Looking at the scatter plot it does appear that the Seaway brand has more unpopped popcorn than Orville. It also appears that Orville brand had less unpopped as there are more trials but this is probably just a coincidence. The clusters from the scatter plot are represented in our decision tree, however; the relations have little correlation because the number of unpopped kernels are not affected by the trial number.

Our dependent variable, the number of unpopped kernels, seems to have two different levels. One for the Orville brand of popcorn and another for the Seaway brand of popcorn. Both of these levels are represented in the histogram and scatter plots that are colored by brand making the two different spreads visible.

The new decision tree is more interpretable than the old decision tree, because the previous tree had more options and it follow the general tends of the data