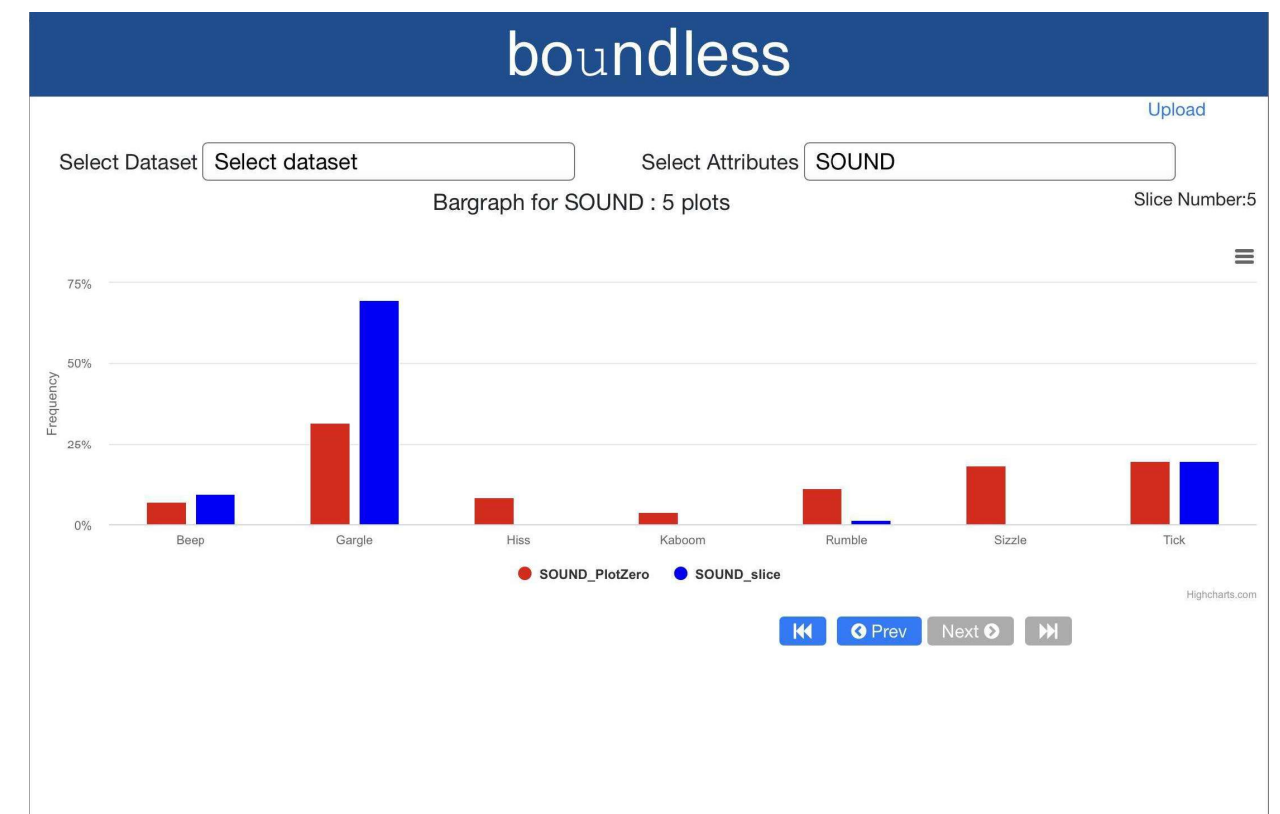# HW12 - Mysterious Box found on the beach

Atharva Patil

Approach

- The goal was to predict which type of sound there was based on the variables inputs 1-4 and switch.
- I tried to get a feel for the data but there weren't any easy patterns to spot. I used boxplots of sound and inputs 1,2,3, and 4 but there was nothing significant.
- On boundless analytics I noticed that interesting pattern for Gargle with the switch as HIGH.
- In the end I decided to use rpart as a model to be my predictor because it is effective for predicting a categorical variable like SOUND

Models

- Started off with just rpart using all variables of the data
  - Error - .3382092
  - Cross Validation accuracy_subset   accuracy_all
    - 1   0.6570637   0.6570637
    - 2   0.6573407   0.6573407

- I decided to then play around with the controls of rpart including minbucket, minsplit, and cp.
- New Model included all variables of the data, minbucket = 30, minsplit = 10, and cp = 0.0001
  - I chose these parameters through trial and error by checking what would happen if I were to increase or decrease anything, and ultimately chose the ones that gave me the lowest error.
  - Error - 0.3069038
  - Cross Validation accuracy_subset   accuracy_all
    - 1   0.6795014   0.6570637
    - 2   0.6891967   0.6703601