# Part 1: Define the objective and scope of the project. Gather and organize the data for the project.

## Conduct exploratory data analysis such as visualizing the data through graphs, tables, summary statistics, and other means to understand the data.

We have a dataset of 4544 labeled images of 447 individual whales. The images will have to be quantified for further analysis and creation of the model. That is a work in progress.

## Identify any issues associated with data gap, data size, data type, data manipulation, data storage and data retrieval for analysis. Structured or unstructured data?

Data extraction is a challenge: raw images contain large amount of irrelevant data (image background). We will have to utilize image processing techniques to extract relevant areas of the image.

Volume of the data may present a challenge: image processing will have to be performed on over 8Gb of images

## Describe the high level analytic problem needs to be resolved: supervised learning, unsupervised learning.

Ours is a supervised learning problem: once the images are converted into numeric data, we will identify features for the model(s) for identification of individual whales.