Introduction/Overview

* Zooniverse engages volunteers to classify large datasets; as the number of projects seeking volunteer classifiers grows, projects need to become more efficient to keep pace with data productions.
  + Some sentences on growth of projects and beginnings of slowing on classification rates
* **Mobile platforms and machine learning** provide two avenues to speeding the pace of data processing and enabling citizen science projects to keep pace with the deluge of data.
  + Mobile phones provide a way to engage more people at more opportunities throughout the day.
  + Machine learning allows for automatic image identification, thus reducing the number of images that require human input.
* However, not all projects are immediately suited to these two approaches. For example, many ecology projects ask volunteers to identify which of > 40 species is visible in an image.
  + While good interface design can make these complex tasks accessible for untrained volunteers, the complexity of the interface does not translate well to the small screen of a mobile phone.
  + Similarly, identifying which of dozens of species is present in a photograph is a largely unsolved machine learning problem.
* We have identified an alternative approach to image identification that could allow projects to better leverage growing mobile platforms and existing machine learning techniques. The “collaborative filtering” approach converts a complex question into a series of simple questions which are presented separately to users.
  + For example, a species identification task would be broken down into a series of separate workflows consisting of a single yes/no question. The first question might be “Is there an animal in the photo?” Images that are retired as containing animals would be passed on to a second question, such as “Is the animal a whale?” Images retired as containing a whale could be passed to increasingly refined questions such as “Is the whale a humpback whale?” or “Is the whale a blue whale?” or “Is the whale doing a dance?” Each question successively “filters out” images of interest, whittling down the dataset to the images the research team most needs to extract detailed information for.
  + Simplifying a complex task into single yes/no questions not only provides more portability to mobile devices, but also enables the integration of machine learning for automated image identification at every step of the filtering process. The questions that a machine classifier is asked to answer at any stage means that the machine learning routines could be much simpler.
  + Thus, although images would ultimately require more classifications to produce the same amount of final information, the ease of classification and ability to integrate existing machine learning routines could provide benefits to efficiency that far outweigh the increase in classification counts.
* However, changing the structure and nature of a project so dramatically could have unanticipated and profound impacts on volunteer communities and classification accuracy.
  + For example, Supernova Hunters posts new data every day at 2pm and announces when the data are available. [Volunteers complete a new dataset withn X minutes, meaning that only a small percentage of would-be contributors ever have the opportunity to participate, and driving the demographics to overwhelmingly older white male volunteers.]??
  + Any broader citations that might cause us to expect a change in community?
* Here we explore the potential consequences of implementing a collaborative filtering approach in terms of user contributions, engagement, and accuracy by simultaneously running both types of workflow simultaneously on an existing Zooniverse citizen science project, Snapshots at Sea.
  + More on SAS…

Methods:

We launched Snapshots at Sea in X-X-201X to identify humpback whales

* We acquired 13,264 images taken by cruise passengers (?) [Need info from Ted]
* We randomly assigned these into two subsets of 6,632 images each, which were assigned to a “Survey” workflow and a “Yes/No” workflow.