Figure 1: *Results from Experiment 1 (Rating the Perceptual Similarity of Scenes):* Similarity ratings for each of the 100 unique categories. Within each category, image 1 was the exemplar image, and was paired with the 4 other images in the category as well as two exemplar images from other categories. Most of the categories show a relatively evenly spaced similarity relationship (yellow to blue). Some categories are more similar to each other (all yellow/orange) or range from similar to dissimilar (yellow to dark blue). This variance in category similarity gives us the ability to capture a range of similarity ratings.
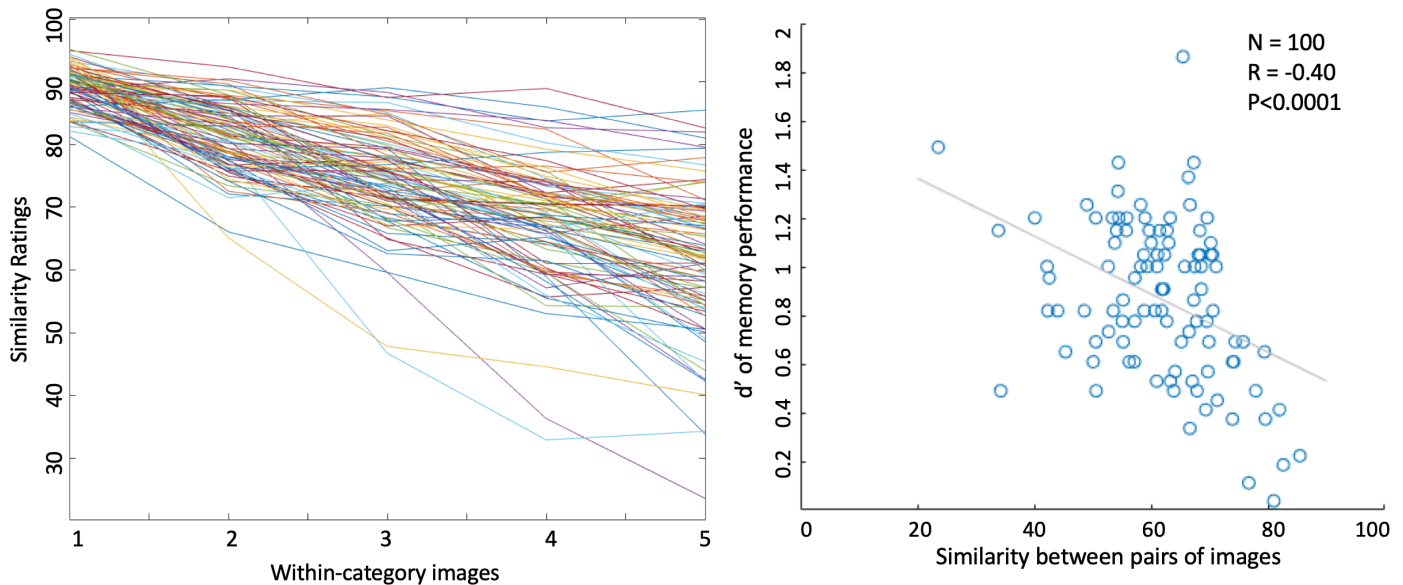

Figure 2. *Left:* Another visualization of scene similarity ratings for each within-category image. When the exemplar image (image 1) is paired with itself or the subsequent image (i.e. the next frame taken from the drone video), participants rated those as most similar. The wider distribution of lines when the 1st and the 5th image are judged reflects the variance in within-category similarity in our database. *Right: Results from Experiment 2 (Predicting Memory Confusability):* This graph depicts visual memory performance for scenes (as measured by d') as a function of the similarity ratings between pairs of images in the perceptual task. Memory for scenes is linearly accounted for by independent ratings of perceptual similarly. Given the noise ceiling is an R^2 of ~0.40, the actual R^2 of 0.16 accounts for approximately 50% of the explainable variance.