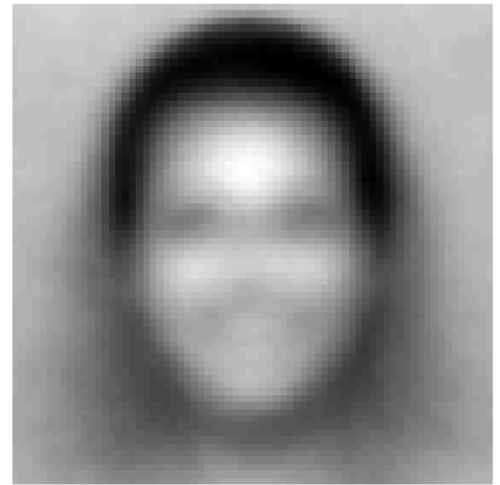


Step 1: Applying PCA to the Survivor Face Dataset

To reduce the survivor dataset dimensionality we utilized the scikit-learn decomposition library to implement their prebuilt PCA algorithm. Using this function we could fit the PCA to the survivor dataset that we loaded into our code. The PCA works on reducing the image while keeping a variance threshold of 90%. The actual number of principal components we used was 89, which allowed a variance of 90.0676%. To test our fitted PCA we used the mean function to get the average face of the survivor dataset. The average survivor's face was then projected into the face space and visualized through a matplotlib reconstruction. In the end, we were able to reduce the survivor dataset dimensionality using the PCA.

Mean Survivor Face in PCA Space



Step 2: Comparing Professors to Survivor Faces

This step seeks to find the professor who looks the least like a survivor by calculating the Euclidean distance between the original images of each of the professors and reconstructed images made from the PCA in the last step, and choosing the face with the largest distance. In order to do that, we started by making the reconstructions by projecting them into the face space. After this was done, we used matplotlib to display a montage of the reconstructed professor's faces to act as a visual measure of how each reconstructed face looked. We then calculated the Euclidean distance between the original photos and the reconstructed photos of each professor. The face that had the largest computed distance (at a distance of 10.6537) from the original photo was the face of Professor Burke indicating that he looked least like a face based on the PCA. Based on the general image of his reconstructed face, we'd agree that it doesn't quite look like a face. While the rest of the professors' reconstructions hold at least some resemblance to their original image, we struggle to see Dr. Burke's reconstruction showing that same resemblance.

Roberson | Ngo | Cazalas | Burke | Eicholtz

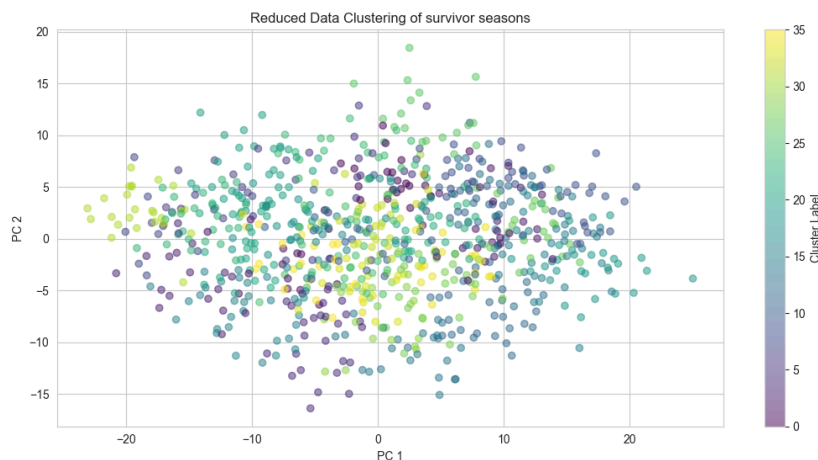
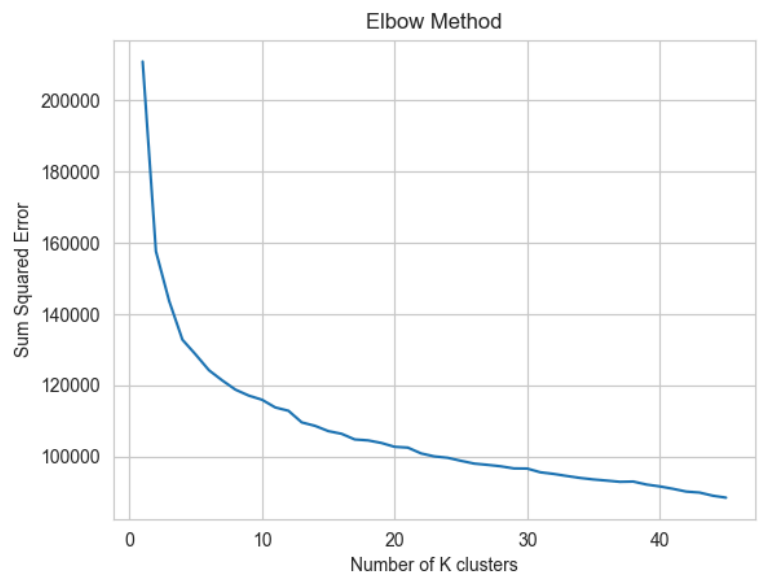


Step 3: Comparing Professors to Jeff Probst

The professor who we found most likely to be the host was calculated in step 3. We first projected the professors into the already reduced survivor face space, using that as a metric to fit our nearest neighbors algorithm. The image of Jeff Probst was also projected into the Survivor face space, which was then used by the nearest neighbors algorithm as a point to find the nearest neighbors of. Since this returned the distances of each professor to Jeff Probst, we determined that the nearest neighbor to Jeff Probst was Dr. Eicholtz, as he had the smallest distance to his image (15.7798), meaning that he would be the host of Survivor.

Step 4: K-means to Determine Which Cluster of People the Professors Look Alike to.

Step 4 tasked us to determine which survivor season each professor should be placed in based on k-means clustering. To solve this we fit the survivor data to the PCA and used the k-means function from scikit-learn. To determine which number of k (number of clusters) was the best we calculated the sum square error of k clusters from 1 to the number of survivors seasons. We then took those calculations and plotted the results to determine which value of k provided the least drastic change in the sum of square error (meaning the graph started a linear trend). The optimal number of clusters we chose was 36, as it was the most linear. After determining the optimal number of K we fit k to k-means and got the cluster labels for each image. Given the



professor's data reduced in the face space, the k-means model predicted which cluster each professor would be assigned to. Then we labeled each professor which survivor season they would most likely be in by calculating the average season label found in the clusters they were assigned to. The results were:

Professor Roberson is closest to Season 13.

Professor Ngo is closest to Season 3.
 Professor Cazalas is closest to Season 17.
 Professor Burke is closest to Season 13.
 Professor Eicholtz is closest to Season 15.

Step 5: Testing Which Professor is the Winner



For step 5, several steps were taken to determine who out of the professors would be most likely to win Survivor. To start, a list of all the winners' faces' indices within the survivor data set was created and sorted. After that these indices were used to calculate the number of winners within each cluster the professors' resided in (which were found in the previous step). Two professors were found at this point, so we tried to further narrow down the winners' indexes by making sure nobody with a common last name had accidentally snuck their way into the winners' list. This did not reduce the number of winners in any of the professors' clusters, so another step had to be taken.

Mean Winner Face in PCA Space



The two professors who had made it to the final round were Dr. Roberson and Dr. Eicholtz. To determine who out of the two of them would win, we used the winners' indices to create a smaller data set of only the winners. This winners data set was used to create a mean face based on a PCA of only the winners. From there we calculated the Euclidean distance of each of the professor's faces to the mean winner face, and found that Dr. Eicholtz's face had the smallest distance to the mean winner's face. Dr. Eicholtz had won.