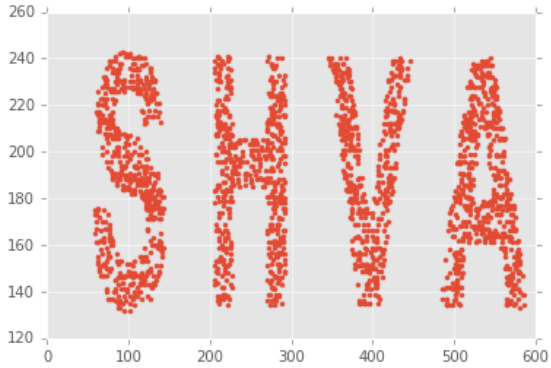
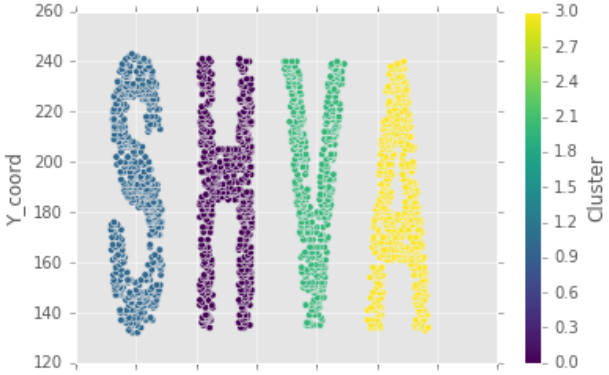
Problem 1

For getting the dataset, I first created an image with my initials written in black on white background using photoshop. Then, I imported that image as a jpeg file and read it using OpenCV library in python as a 2D matrix. For making the data points sparse, I randomly saturated pixels to 255. The x and y coordinates of the pixels with intensity equal to 0 were selected as the dataset on which the spectral clustering was applied.



(a) Initial dataset



(b) Clustering result

Problem 2

The dataset was generated using numpy library in python. The k value for bagging and random forest (m=1 and m=3) was estimated using OOB score. The value of k for bagging, random forest (m=1) and (m=3) was found to be 45, 47 and 47 respectively.