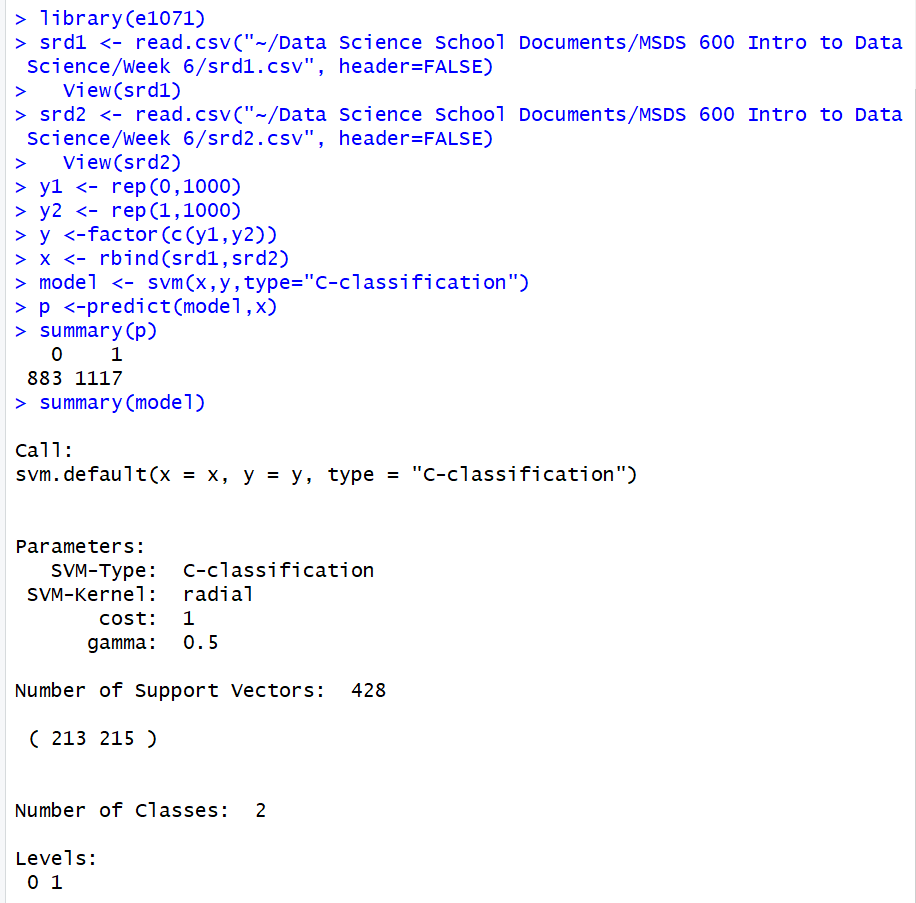
# Exercise SVM 2

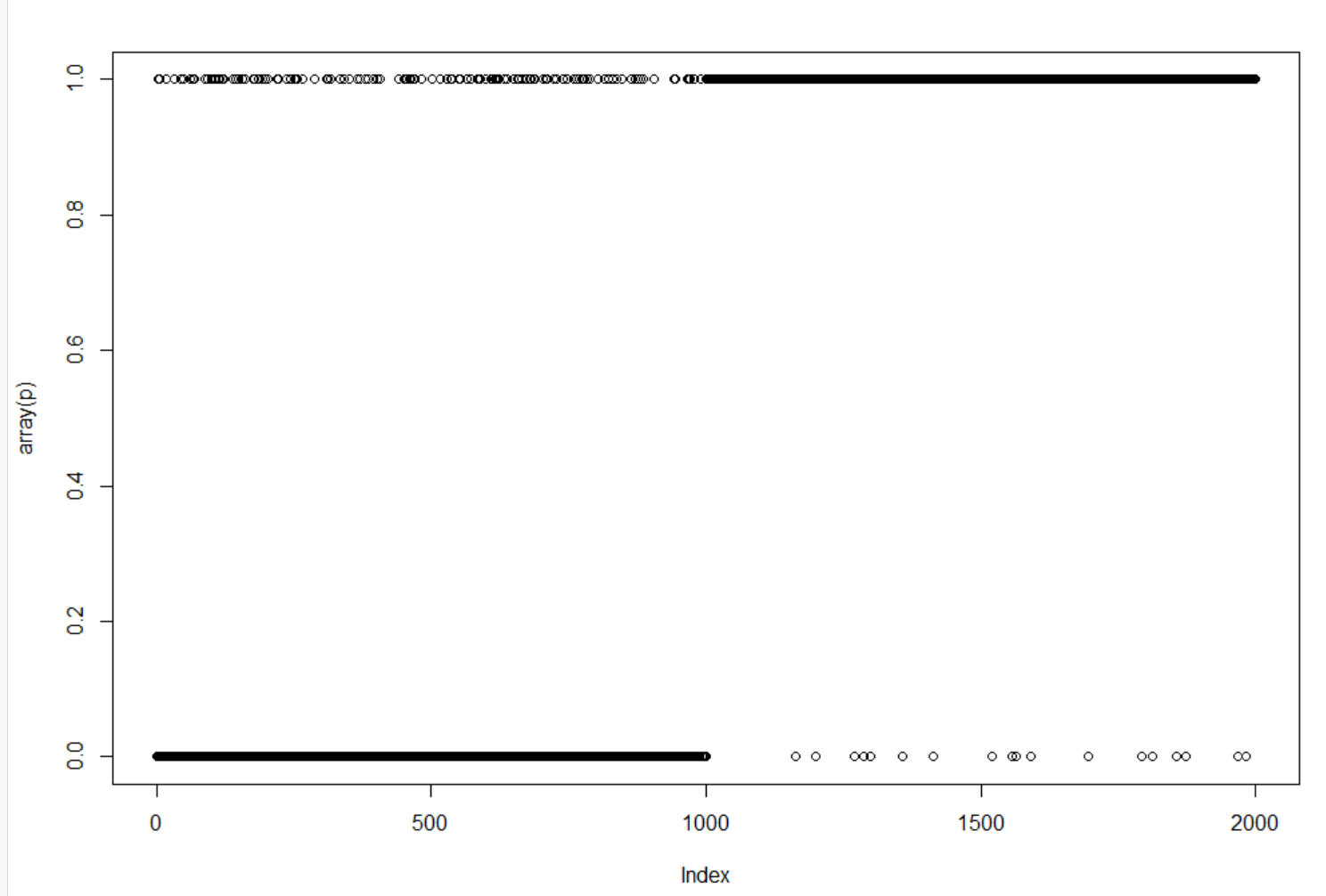
## Objective:

Go through the steps of the second half of the Sector Vector Machine Project, then research and discuss cross validation and ambiguous data and their roles in analytics and how they can apply to this exercise.

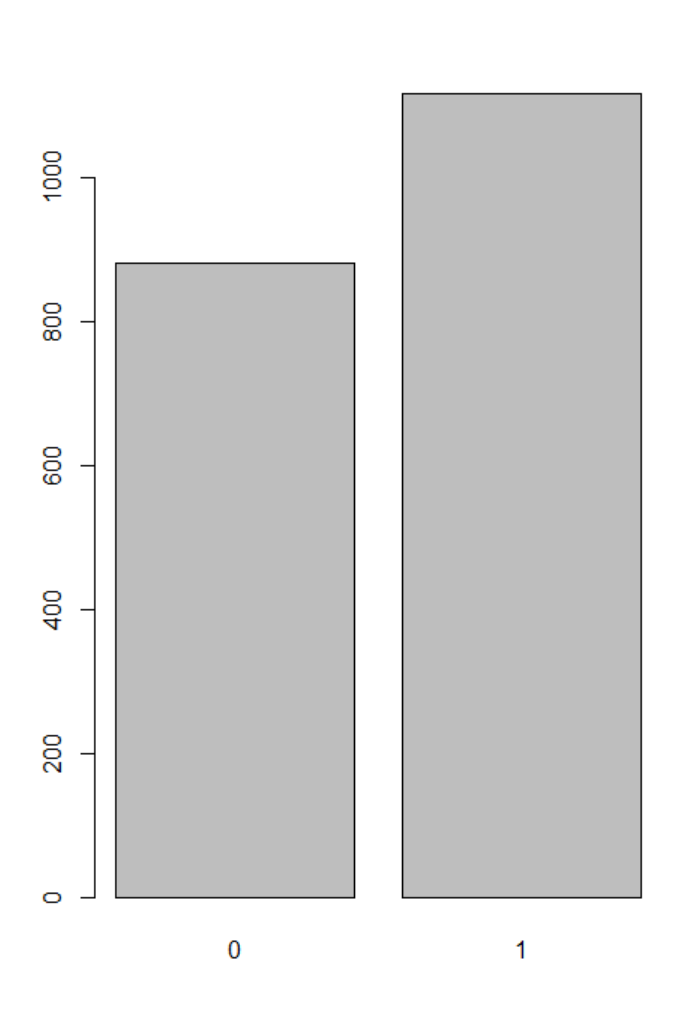
## Part 1: Upload the csv’s into R and plot the data points.



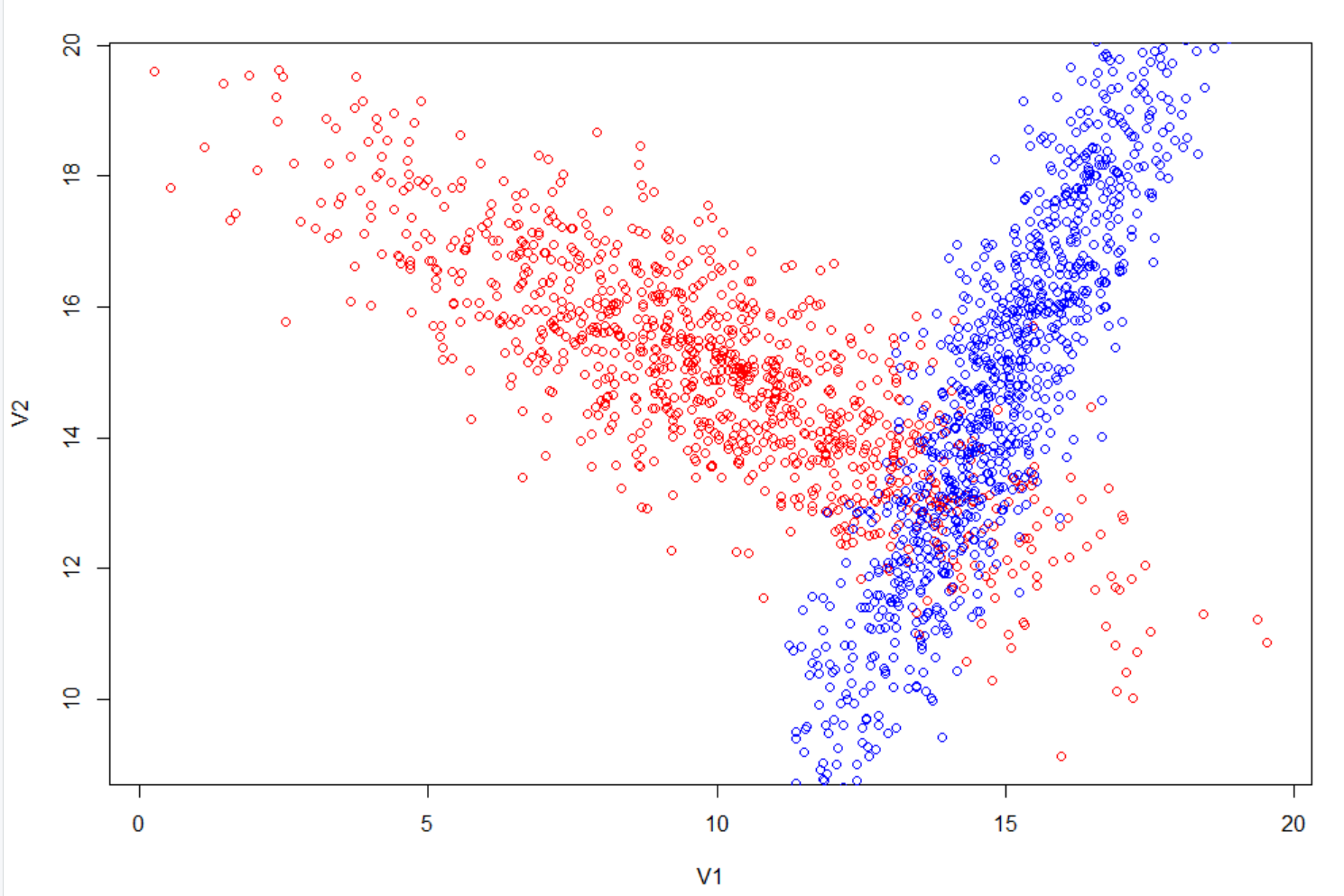




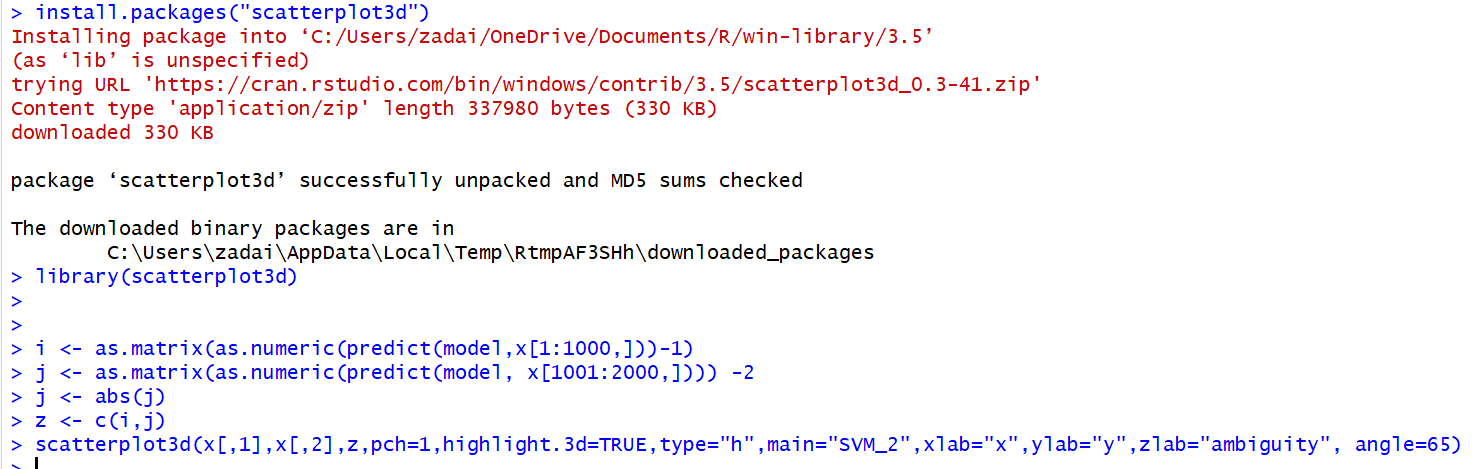


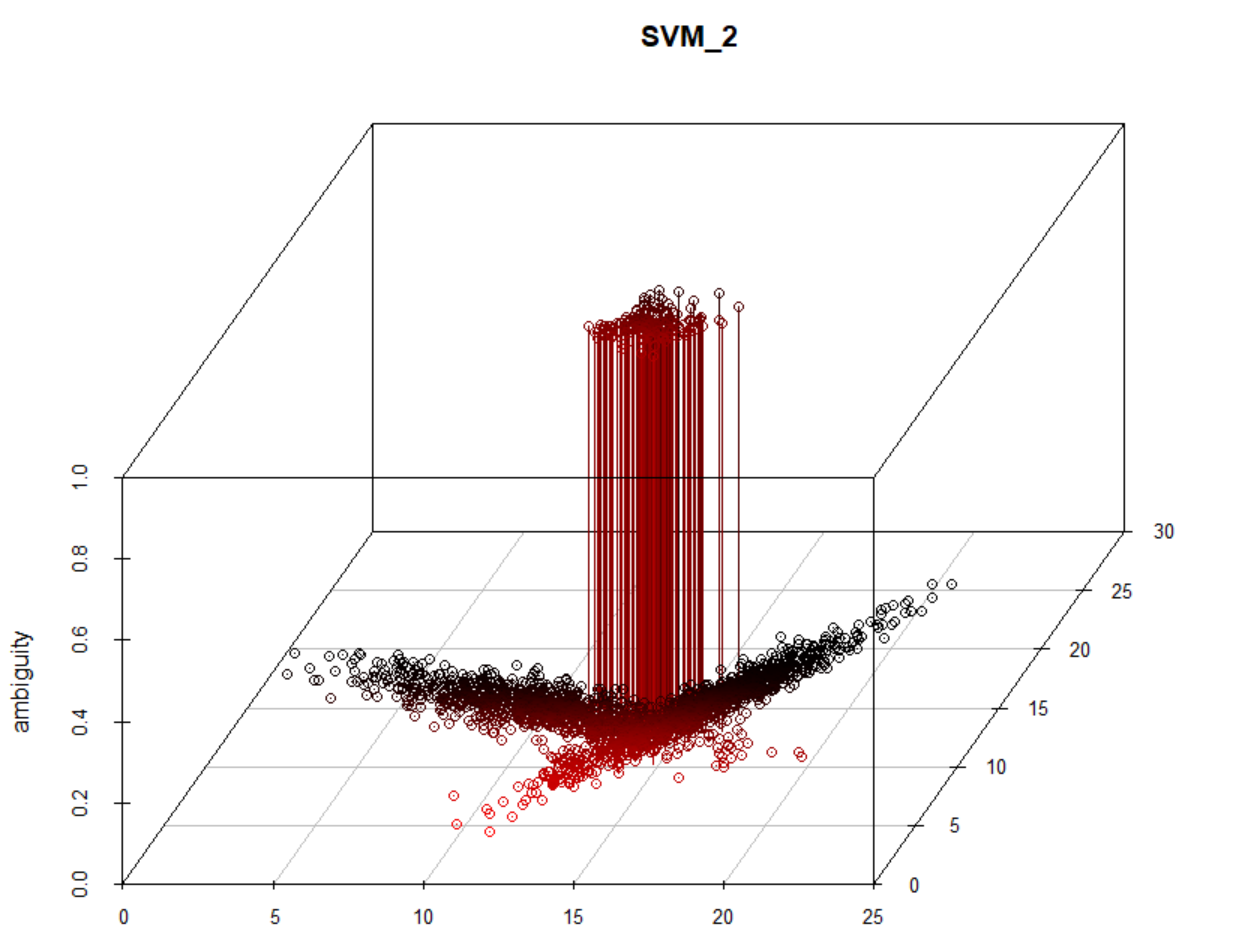




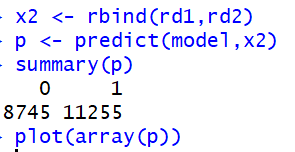


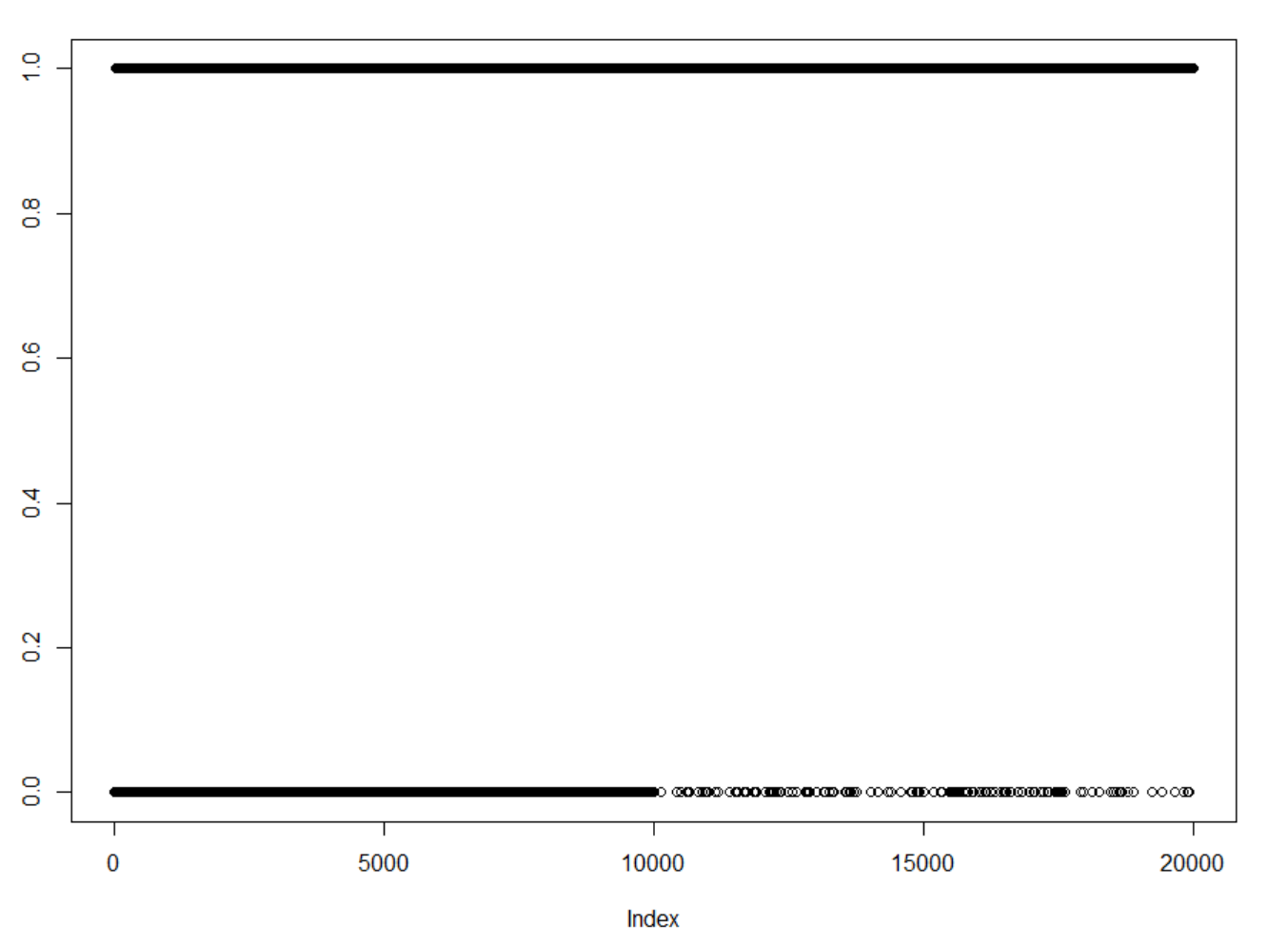
## Part 2: Use Scatter Plot3D Library to map ambiguous points



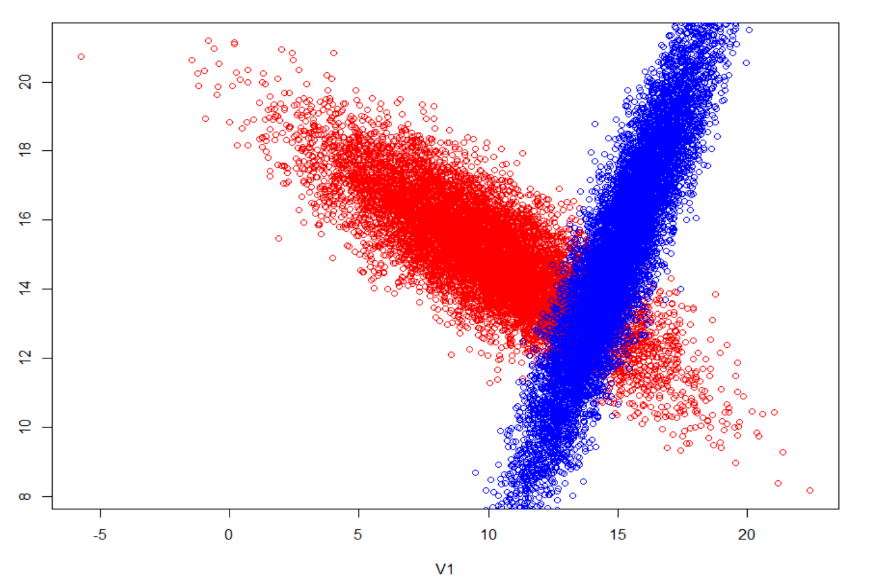


## Part 3: Train svm model using length 1000 data and then predict the 10000 length data









### Key points of Part 3:

Cross validation is important because it is a good way to confirm the results of analysis. Cross validation is a way to indirectly test a model that on its own has no test performance results, using specified bins called k-partitions, which can be measured our over your data and used to measure out how much of the data is going to be modeled and how much is going to be tested. Conventionally, an analyst would create 10 k-partitions and separate them out evenly through their data and then choose where they want to start testing and what part of the data do they want to learn or form the model from. This was done a little bit with our model above. We created our test variables with i between 1 and 1000 and j between 1001 and 2000 and then used the rest of our data to learn, so to create as accurate a prediction model as possible. One of the problems cross validation combats is something called Overfitting, which occurs when a model which was initially fit with the same data as was used to assess fit. With our data, we avoided overfitting by having our model only take in a thousand units for both our i and j variables.

### Key points of Part 4:

Removing ambiguity from our classes of existing data by setting the parameters of what makes the data setting ambiguous and then creating a classification for the ambiguous data that can then be removed from the data sets itself. For example int our SVM\_2 model, we have an x, y, and z plane which map our two classes of data, srd1 and srd2. We created two vectors from those classes, i and j and were able to map i and the absolute value of j vectors onto the x and y, and our z plane which we labeled ambiguous would measure of ambiguity based on the data point. In the graph you are able to see that our data points either measure their ambiguity at 0 or 1 but nothing else in-between, but by doing so separates our ambiguous data points from our non-ambiguous points. Separating out those ambiguous points and allowing the non-ambiguous points to be mapped to judge their overall fit to one another. By adding that third dimension to our plot, it proves that you can add additional measurements to a data point to help signal and remove ambiguous data.