1. **Recreate the R part of this experiment using your computer.**

> setwd('c:/users/bmack/downloads')

> getwd()

[1] "c:/users/bmack/downloads"

> srd1 <- read.csv("srd1.csv",header=FALSE)

> srd2 <- read.csv("srd2.csv",header=FALSE)

> y1 <- rep(0,1000) #repeats 0 1000 times

> y2 <- rep(1,1000) #repeats 1 1000 times

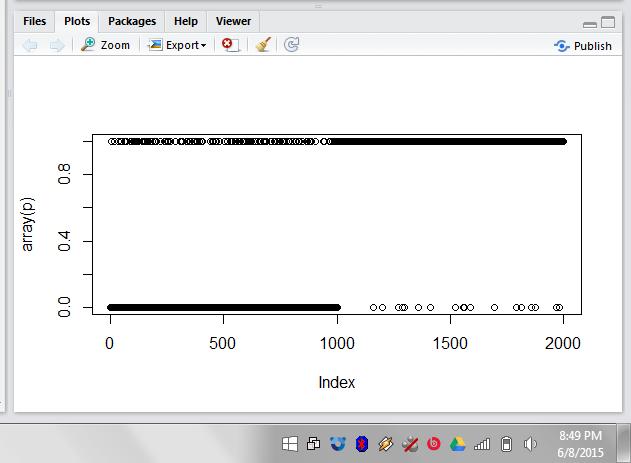
> y <- factor(c(y1,y2)) #our possible choices are 0 and 1

> x <-rbind(srd1,srd2) #combines them, so now we have 2000x2 instead of 1000x2

> model <- svm(x,y,type="C-classification")

> p <- predict(model,x)

> plot(array(p))



1. **See if you can find other ways to display the predictions.**

Plot(p) will simply make a bar chart:



I realize Paul’s alternative blows mine out of the water!

1. **Train the svm model using the 1000 length data and then predict the 10,000 length data. Compare the predictions with the data. See the ideas of cross validation.**

So first I want to load up my rd1 & rd2 data sets:

> rd1.csv <-read.csv("rd1.csv",header=FALSE)

> rd2.csv <-read.csv("rd2.csv",header=FALSE)

Then I am going to create x2, which is a bind of my two new files

> x\_10k <- rbind(rd1.csv,rd2.csv)

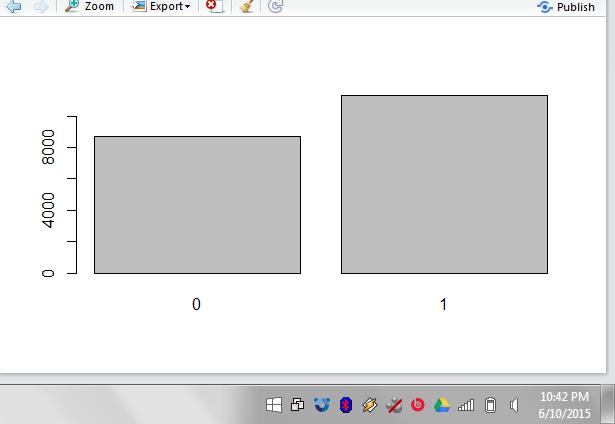
My training is already set up from last time within the y1 and y2 variables

So now I create a model using my 10k rows (concatenated, so really a 20k rows set), and use my same training data.

> model2 <- svm(x\_10k,y,type="C-classification")

> p2 <- predict(model2,x\_10k)

* Plot(p2)



The predictions of the 10k rows is very similar to the 1k rows prediction, only more data.

1. **Is there anyway to remove the ambiguity between the two classes give the existing data? What about if you could add additional measurements for each data point?**