1. You will be provided with a file of driver speeds, and if they were going recklessly or not.
The two columns are the speed vehicles were observed travelling at, and if the driver was trying to be reckless. The recklessness was based on an officer painstakingly interviewing the drivers. Some of these were pulled over for aggressive driving, such as following too closely or neglecting to signal lane changes.

a.) (1⁄2) Considering that we are trying to maximize public safety, how would you break a tie if two different speed thresholds have the same lowest misclassification rate?

If there are two speeds that they have the same lowest misclassification rate then I will pick the “lowest speed” among those two speeds as default value to classify the speedy drivers. I am primarily applying the domain knowledge of lower the speed higher the safety. Also based on the data, when the speed increases I observed the number of drivers who have intention to speeding also increases.

b.) (1⁄2) Consider that you are trying to maximize how much trust the public has in the police officers, how would you break a tie if two different speed thresholds have the same lowest misclassification rate?

In this case, I would again check the speeds that they have lowest misclassification rate and I will pick “highest speed” for the classification. This definitely reduce the misclassification rate and increase the trust.

c.) (1⁄2) What design decision will you use for your threshold? Will the value below the threshold be for speeds <, or <= the threshold?

I am using **<** instead of <= to identify the non-speedy drivers. By using < I was able to find a higher threshold value that will eliminate the false alarms (for an example when I used the <= I am getting the threshold value as 58 but when I used the < I am getting the threshold value as 59). When we deal with people we need to reduce the false alarm to increase the trust.

```
    leftClass <- newDataFrame[newDataFrame$Speeds < clusterCenter,]
    rightClass <- newDataFrame[newDataFrame$Speeds >= clusterCenter,]
```

Again my consideration is reduce the false alarm (due to human involvement) and increase the trust.

d.) We define the number of misclassifications as the total of number false alarms + number of missed speeders.

Using the techniques covered in class, write a program to find a threshold for a police officer to set their laser speed detector at so that it beeps in such a way that it minimizes the misclassifications.

In case of ties, maximize the public's trust that a police officer is not pulling people over for the fun of it. (Minimize false alarms over misses.)

Here, I want you to round the speeds to the nearest mph. Sort them, and then try the speeds from slowest to the fastest. Compute the misclassification rate for each threshold.

Def Misclassification Rate = (FP + FN)/(TP+TN+FP+FN)

Output from the function "calculateMissClassificationRate" in the attached program

No	Speeds	MissClassRate
1	47	0.5000
2	48	0.4950
3	49	0.4900
4	50	0.4800
5	51	0.4675
6	52	0.4550
7	53	0.4175
8	54	0.3675
9	55	0.3075
10	56	0.2300
11	57	0.1500
12	58	0.0725
13	**59**	**0.0150**
14	60	0.0825
15	61	0.1675
16	62	0.2375
17	63	0.3075
18	64	0.3550
19	65	0.3925
20	66	0.4175
21	67	0.4450
22	68	0.4650
23	69	0.4750
24	70	0.4900
25	71	0.4950
26	72	0.4975

e.) (1⁄2) What threshold value did you compute? (To the nearest mph)

Confusion matrix for 59 mph

	Actual		
Observed		True	False
	Trule	(TT) 200	(FP) 6
	False	(FN) 0	(TN) 194

I would use 59 mph

f.) (1⁄4) For the given training data, how many reckless drivers does this let through?

0 Drivers (FN)

g.) (1⁄4) For the given training data, how many non-reckless drivers would be pulled over?

6 Drivers (FP)

h.) (1⁄2) How does this value compare to the value you found using Otsu's method?

It is very interesting both clustering and classification gave me the same results.

i.) (1) Plot the misclassification rate as a function of the threshold used. Mark the points with the lowest misclassification rate.

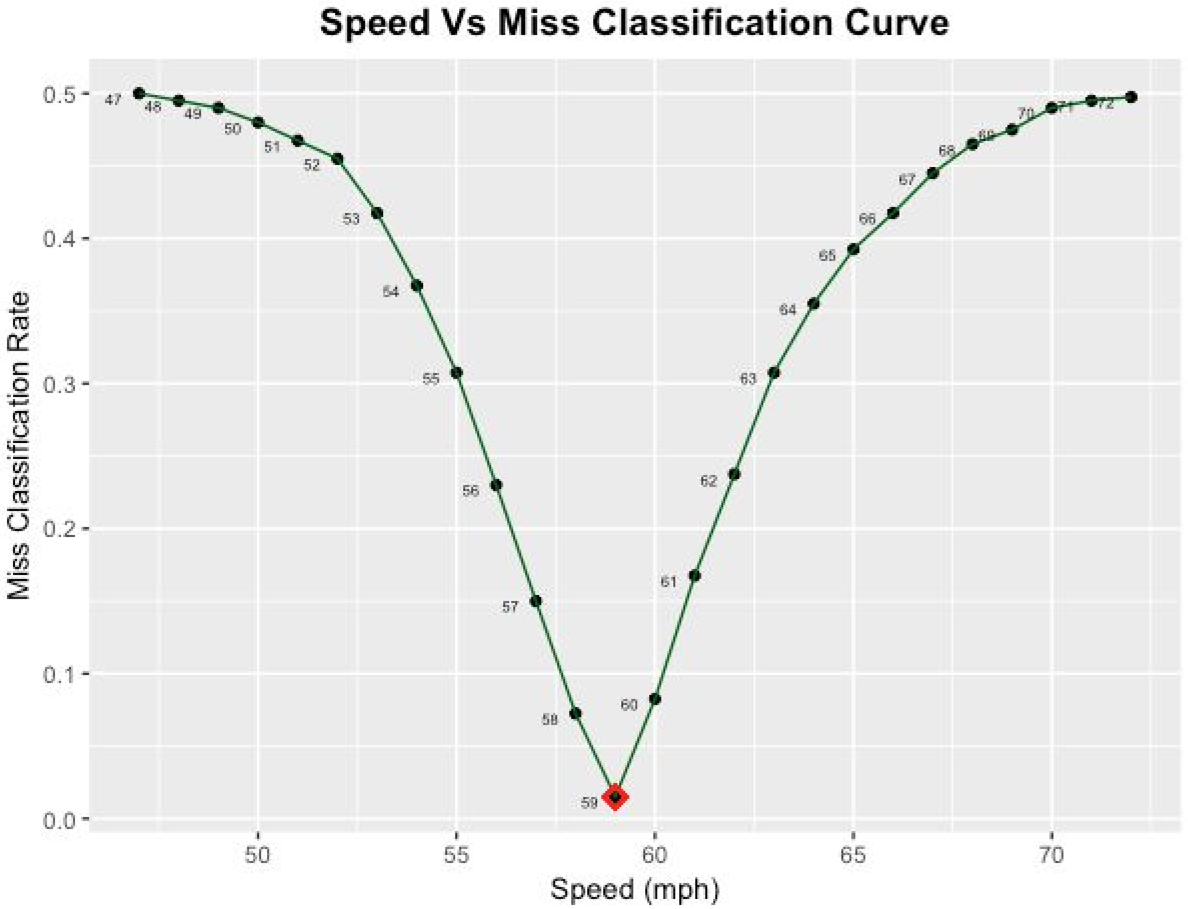

Figure 1: The lowest misclassification rate is happening at speed **59 mph**

j.) (2) Conclusion:
Write up what you learned here.

This assignment taught me a lot of valuable lessons such as
a.) The classification strategy may vary depends on the application(here we deal with people therefore we need to reduce the misclassification rate)
b.) This also taught me different indices we use for confusion matrix (TP, TN, FP, FN)
c.) Instead of remembering these formula, this taught me a way to think these numbers and understand them
d.) Dimension reduction: Reduce multidimensional data to single dimension and do the classification.
e.) Finally I was also able to play with Receiver-Operator Curve

Was this a complete waste of time?

No, it is very useful to understand the one dimensional classification system. Specially this exercise taught me how to manipulate the confusion matrix. Honestly this also gave the practical understanding about how to change the threshold based on different applications.

How might you use a one-dimensional classifier with multi-dimensional data? Was there anything particularly challenging?

Honestly this gave me a knowledge that we can always do a dimension reduction and use the 1-D classifier for multi-dimensional data. It is really nice to see that misclassification rate is low at 59 mph also ROC curve plot the 59 mph as the ideal point.

Did anything go wrong?

Not that I know of.

Bonus:
2. (1 pts) BONUS:
Always attempt the bonuses if time allows.

Generate a receiver-operator (ROC) curve for this training data.
Plot it, and put the location of the any tie-thresholds on the ROC curve.
Label the axes correctly. Circle the points with the lowest misclassification rate.

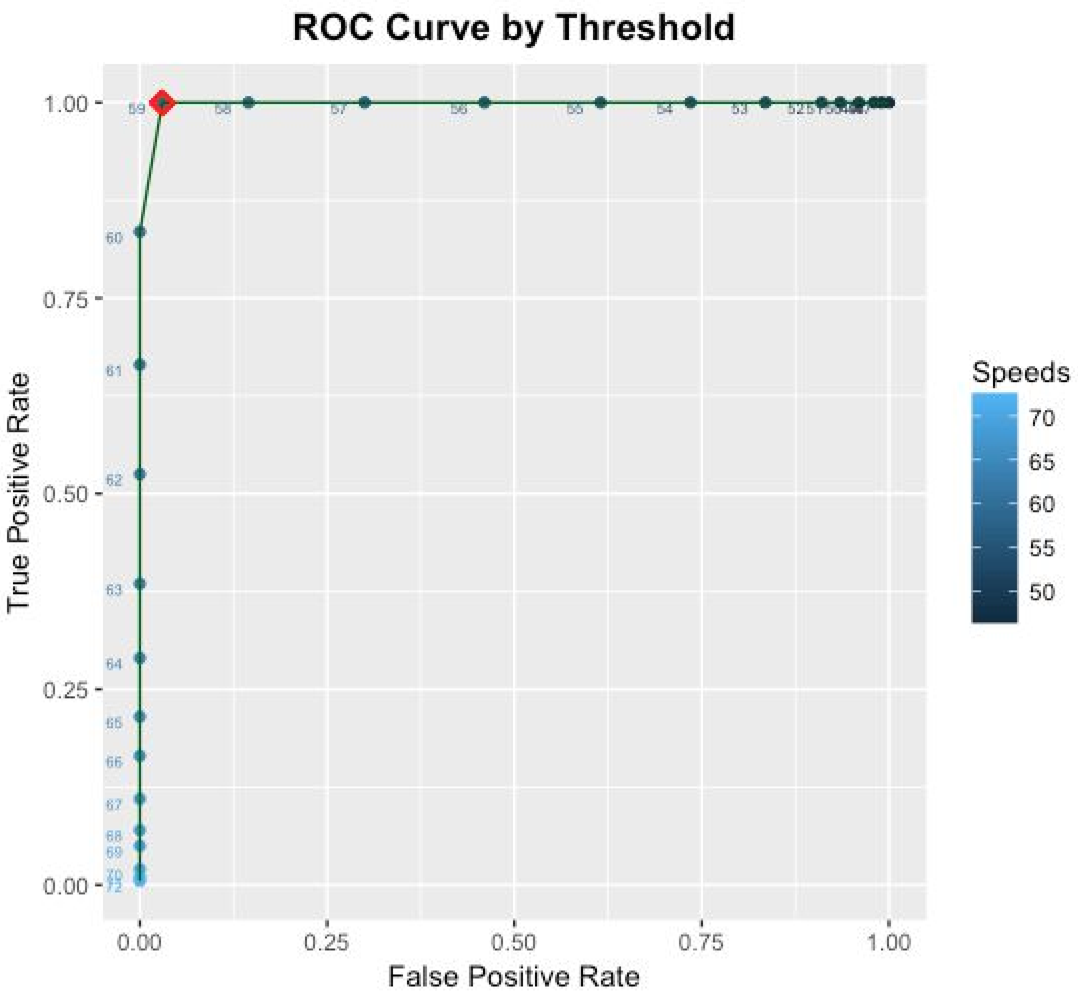

Figure 2: ROC curve with threshold of 59 mph

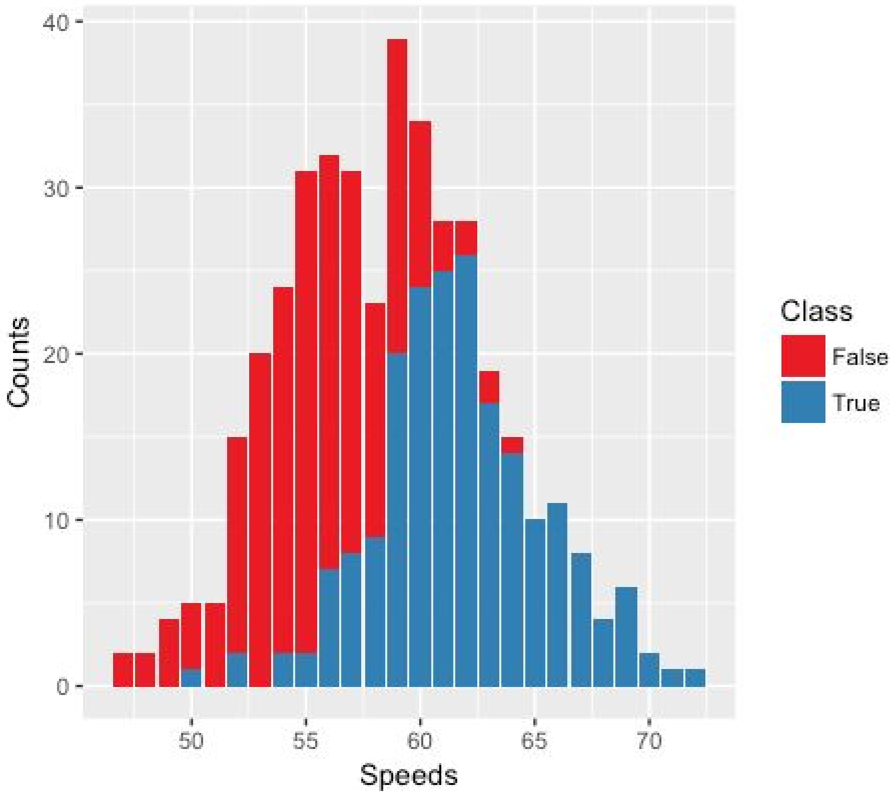

Figure 3: Actual Speeds

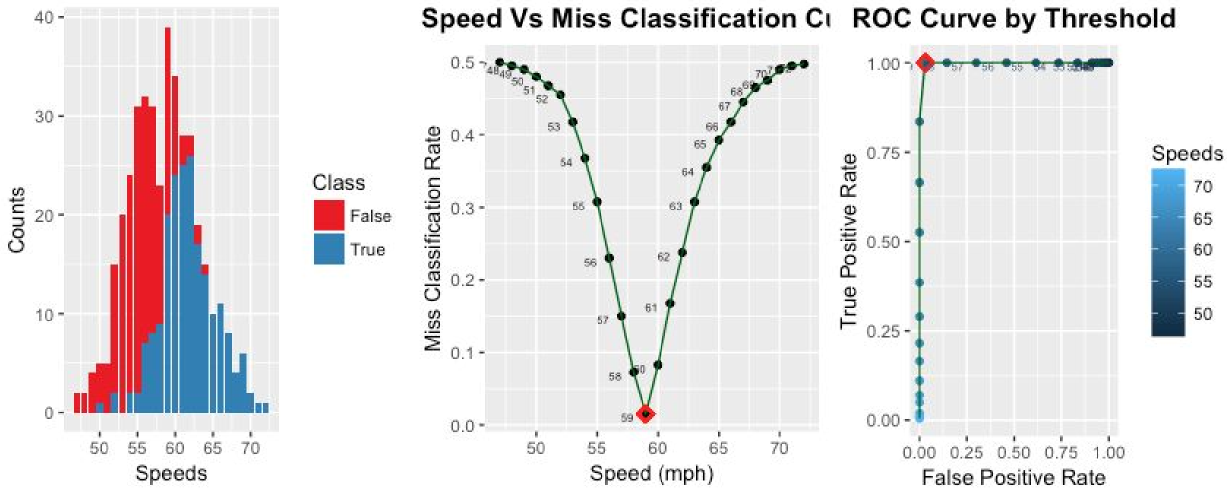

Figure 4 : All three charts together