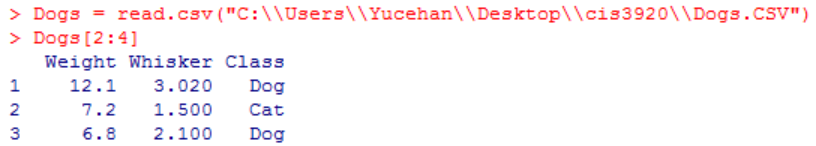
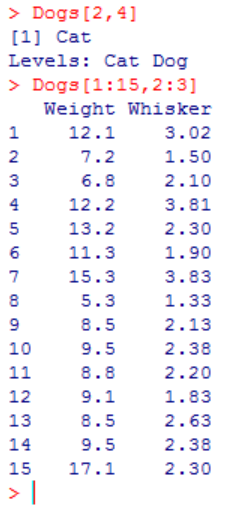
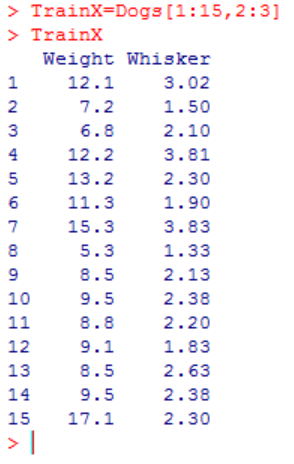
Yucehan Kucukmotor

CIS3920, Data Mining

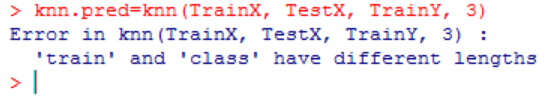
Prof. Lawrence Tatum

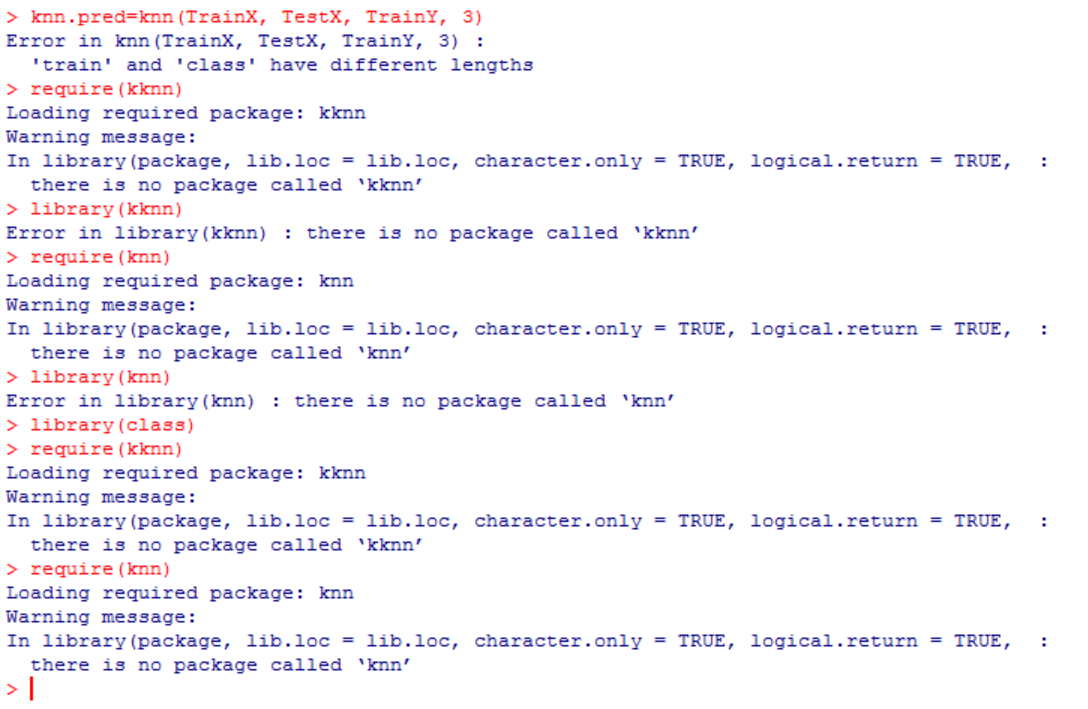
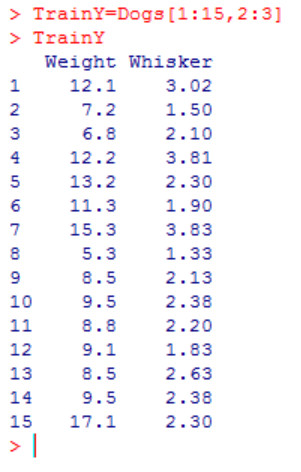
Due 09/09

1. C:\Users\Yucehan\AppData\Local\Microsoft\Windows\INetCache\Content.Word\1.1.pngI started off typing library(class) to be able to have access to the knn function that is readily available. Then I read Dogs.csv file into R.

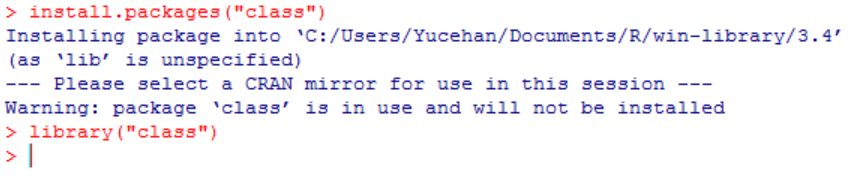
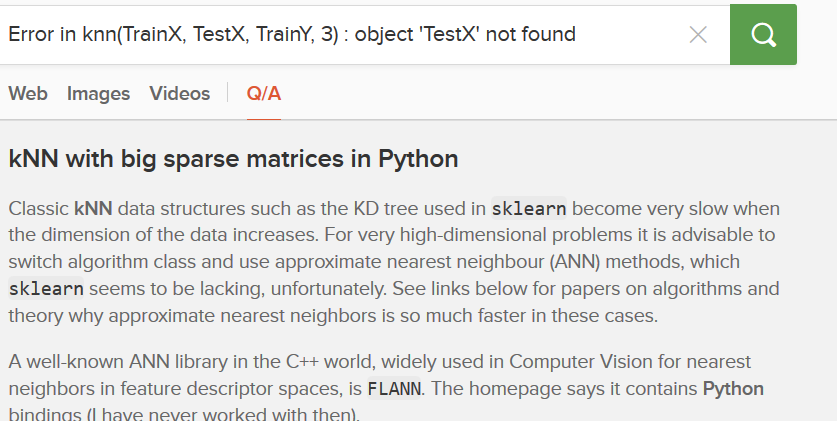
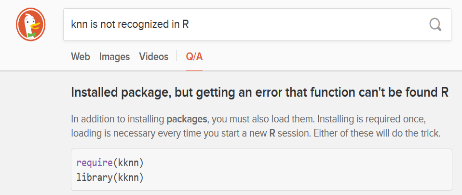
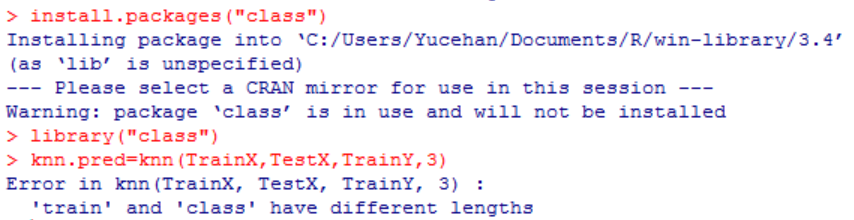
Then I typed Dogs[2:4] to make sure I had the data read successfully. I got the results and they matched the data in the excel file (Dogs.csv). After displaying all my data on R, I made sure that the data from the excel file and the data I had on R was a perfect match with each other. Afterwards, I defined train X with proper row and column values by writing down as I did on the graph to the right. Then by typing TrainX, I made sure no mistake was made along the way and correct items were included in TrainX (our first argument). Then I did the same for Train Y.

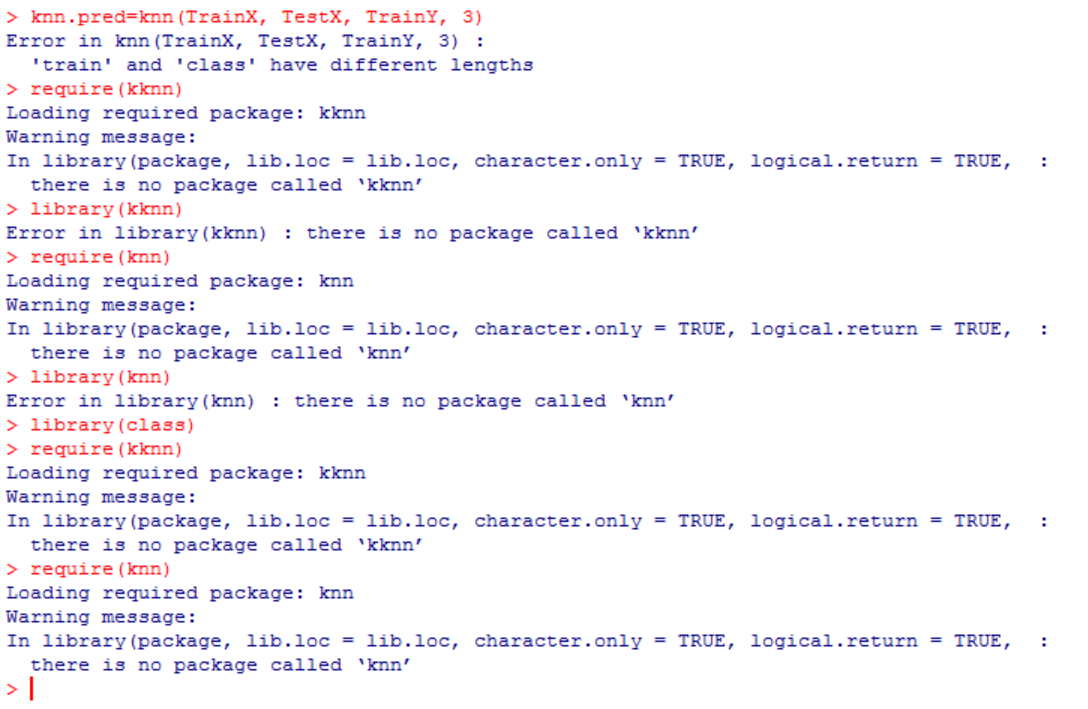
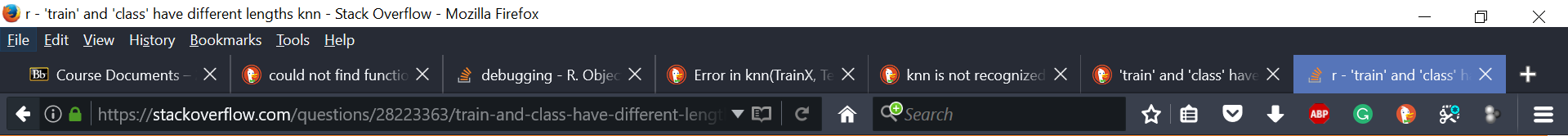
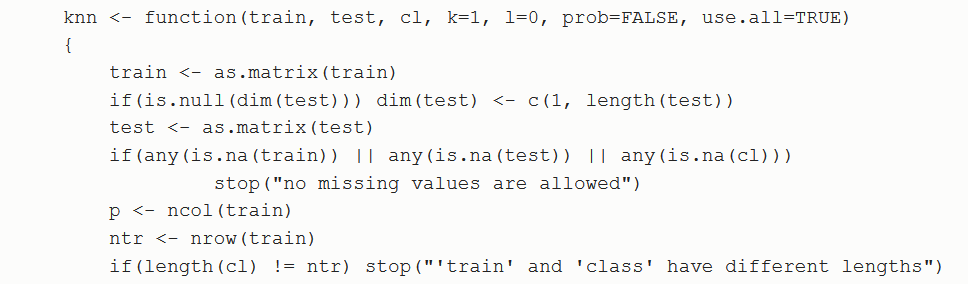
This is where everything got out of hand.



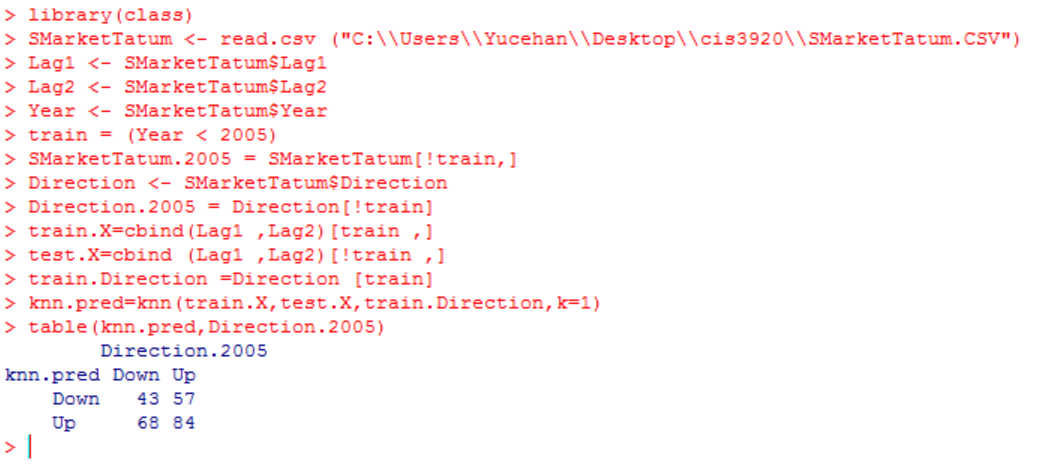
When I wrote

knn.pred=knn(TrainX, TestX, TrainY, 3) to have my function ready, I encounter an error which can be seen above. I spent hours here and there try to understand the concept better in my head (which I think I’ve accomplished), trying to solve this issue by googling errors, reading lecture notes as well as the ISL text, and reading online material from e-resources. I am, very likely, missing a fundamental knowledge here which prevents me from moving on. I googled others way to have knn function imported into my R. I used require(kknn), library(kknn), require(knn), library(knn) by hoping to fetch the function into my R in a different way. Although I don’t have the screenshots for it (command history got pretty messy), I encountered numerous errors and mistakes. I’ve tried to learn along the way. “Object not found error” to “-knn function not found error”. I, somehow, moved ahead of those problems.

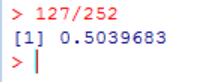


I typed install.packages(“class”) thinking maybe I was missing the library, somehow. I tried knn.pred=knn(TrainX,TestX,TrainY,3) again, and again, every time I tried something new to fix the problem. Last problem I came across to and could not get past of was “'train' and 'class' have different lengths” problem. This is most likely caused by a conflict between columns and rows. But I could not find a way how to precisely match datas with each other to know for sure. I took a look at the source code for

the knn function. It is as shown above. I was not sure how to read this and make use from this though. I worked on this for some hours. However, I could not move ahead. I checked the code from your work as a final solution to move along, but answer was not there, obviously.

out of 100 tries, there were 43 correct D&D, and 57 incorrect forecasts D&U. And there were 84 U&U (correct forecast) and 68 U&D (incorrect corecast)

127(corrects)/252(total)=



Rate of Correct Forecasts out of all attempts = 0.5039683.