Assignment #5: Evaluation and Visualization
Spring 2016
CS3943
Prof. Rumi Chunara
Total: 30 points

All questions must be completed in R. Implement and comment your code so that anyone reading the file can reproduce the code easily (e.g. set the file path once at the beginning of the script where it can be easily changed). Save the code as an R markdown file, and upload it to NYU classes.

1. Visualization of Health Data (19 points)

I have shared data from my Apple Health Kit on NYU Classes/Resources. This data includes the number of steps at different time intervals, for a little over a year time period. The iPhone includes some simple visualizations (example screen shot pictured here), however there are many opportunities to improve these and convey more information from this data. The objective of this question is for you to create a useful visualization of this data back to me, the consumer.

a) What metrics do you think are informative for me to understand from the data? How do you determine which metrics are important? (2 points)
b) For each of the Daily and Hourly data sets, what form of graph and graphical objects are appropriate? Why? (5 points)
c) Generate two visualizations, keeping in mind principles we discussed in class. (12 points)

2. Evaluation (11 points)
In this problem we will think about evaluation methods – there is no coding in this question, just written answers. You may want to implement some of these evaluation methods in your course project. Here are some helpful resources about ROC and AUC:
`https://www.youtube.com/watch?v=OAl6eAyP-yo`
`http://www.navan.name/roc/`

a) If you have a classification model that outputs predicted probabilities, how could you convert those probabilities to class predictions? (2 points).
b) Why are predicted probabilities (rather than just class predictions) required to

generate an ROC curve? (1 point).

c) Could you use an ROC curve for a regression problem? Why or why not? (2 points).

d) What's another term for True Positive Rate? (1 point).

e) If I wanted to increase specificity, how would I change the classification threshold? (1 point).

f) Is it possible to adjust your classification threshold such that both sensitivity and specificity increase simultaneously? Why or why not? (2 points).

g) What are the primary benefits of ROC curves over classification accuracy?

h) What should you do if you have a low AUC value like 0.15? (1 point).

i) What's a real-world scenario in which you would prefer high specificity (rather than high sensitivity) for your classifier? (1 point).