

Introduction to Data Science: CptS 483-06 – Fall 2015

Assignment 1: Create a Data Science Profile of Yourself

Date: 2015-08-27.

As mentioned in class on Wed. Aug. 26, the purpose of this assignment is to create a visual data science profile of yourself. Specifically, you will create two profiles. The first will show the way you see yourself now. The second will show how you would like to see yourself by the end of this course.

The profile is simple. On the horizontal-axis you will have seven “domains (areas)” that could generally be considered important to Data Science:

1) Computer Science. 2) Math. 3) Statistics. 4) Machine Learning. 5) Domain expertise. 6) Communication and presentation skills. 7) Data visualization.

On the vertical-axis you will have a relative scale (think percentage) of your skill level in each of the domains. The domain in which you have the strongest skill will be close to 100, and the domain in which you think you have very little skill would be close to zero.

As an example, see the figure on page 11 of the slides from Wednesday’s lecture (posted on OSBLE) that shows the data science profile of the author of the main textbook we are using for the course. (Recall that the author has a PhD in Stat and has held several senior and executive-level Data Science positions in industry.)

Your task is to create your own profile – two to be exact, one showing current and the other projected. You are still a student and you may not feel you have a lot of skill in any of these domains. Allow yourself a liberal interpretation of skill level and keep in mind that this is on a relative scale. Also, keep in mind that it is perfectly okay to have zero skill level in some of these domains.

You can use any tool (Excel, Matlab, R, etc) you wish to make the plots.

Here are a few associated presentation considerations and discussion points you are asked to address as part of this assignment.

1. The domains in the horizontal axis could be ordered in a number of different ways. What ordering in your opinion would be most effective and why? Create your profile in the order you chose and justified. In addition, create your profile in which the ordering is as listed earlier (CS, Math, Stat, ML, Domain Expertise, Communication, Data Viz). A common way to order the domains could for example make comparison with other profiles easier.
2. Is there a skill (bucket) you think should be added to this data science profile? A skill you think should be removed? Specify and justify briefly.
3. Can you think of a different scale or visualization method for this purpose? If yes, describe the approach in a few sentences, or better yet, show how that would look like in your case.
4. Do you see any drawbacks in self-reporting? Discuss if you see any.