

Homework #1 – due Thu, Sep 6

<http://bigdata-madesimple.com/what-is-a-data-scientist-14-definitions-of-a-data-scientist/>

<https://jakevdp.github.io/blog/2014/08/22/hacking-academia/>

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

<https://datascience.stackexchange.com/questions/2403/data-science-without-knowledge-of-a-specific-topic-is-it-worth-pursuing-as-a-ca>

<https://www.oreilly.com/ideas/data-engineers-vs-data-scientists>

Read the articles in the links above, and those on our webpage www.agapie.net → COSC 3360, under the section JOBS, TRENDS, DEGREES AND CAREERS IN DATA SCIENCE.

Answer the following questions based on the articles:

1. Explain DS as the “fourth paradigm” of science.
2. What were the first two professional journals in the field of DS?
3. When was the “inflection point” for DS software, and what did it mean, exactly?
 - Include a relevant graph here!
4. In today’s (2017) DS practice, which is more heavily used, proprietary software or open-source software?
5. What company invented the “People you may know” ad?
6. How is a data scientist different from a statistician?
7. How is a data scientist different from a data manager?
8. How are data scientists similar to Star Trek’s Mr. Spock?
9. What does the word “quant” mean? What were the “quants” of Wall Street in the 1980s and 1990s?
10. What was the “Netflix Prize”, who won it, when and how?

- A paragraph-length answer is required for each question.
- Submit in electronic format (.docx or .pdf).
- Question 3 requires a graph, but it’s OK to include graphics in your replies to other questions if you want.

?