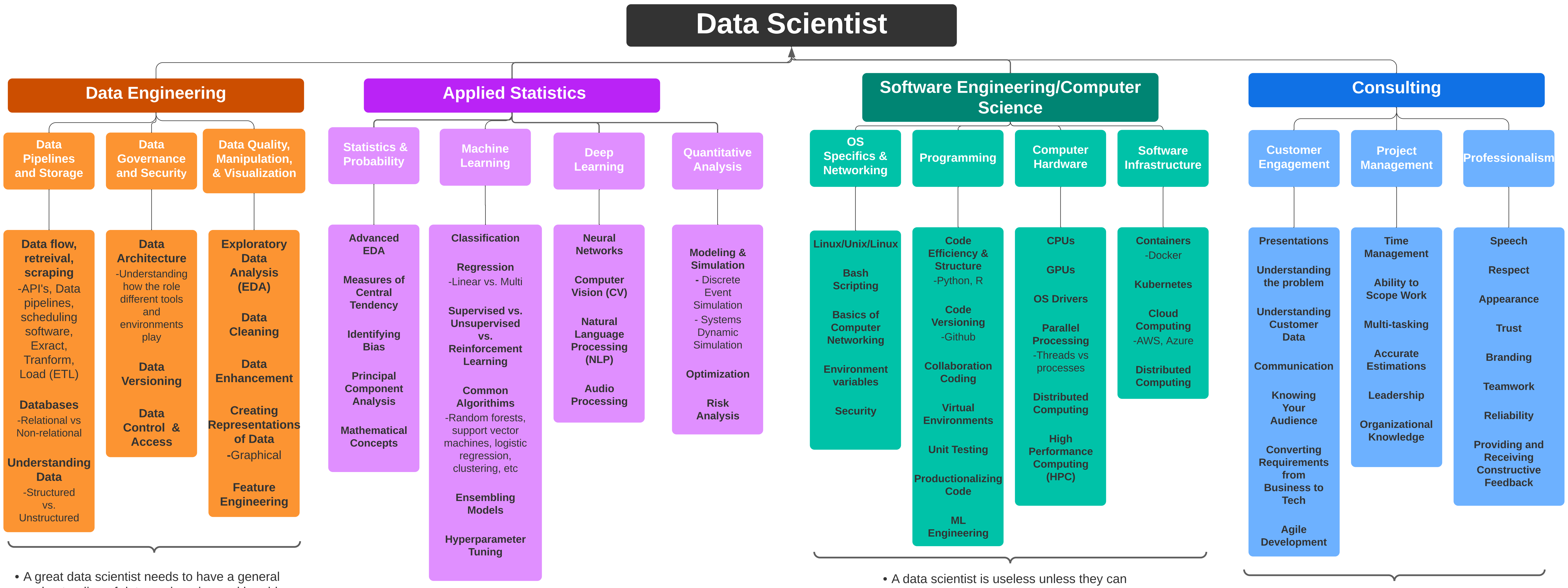
Data Science Diagram

Steven Cocke | February 18, 2022

What does it mean to be a data scientist? What are the different skillsets one must have to be successful in a data science centric role?

The definition of a data scientist has notoriously always been broken down into two areas: someone who has knowledge in programming and statistics. While I believe that both of those areas are necessities, this can be expanded upon to include the below areas: data engineering, applied statistics, software engineering/computer science, and consulting. By no means is the below diagram all encompassing, but should serve as a quick overview of the different buckets of knowledge individuals should concentrate on when pursuing a career in data science. In some areas that seemed ambiguous, I tried to provide a few examples beneath the topic, but again this diagram is not all emcompassing. I also combined some sections for space's sake that I would probably break out separately if I wanted to make this more of an eye-sore.



- A great data scientist needs to have a general understanding of data engineering and be able to apply data engineering at least at an individual level
- Being able to communicate and, therefore, work more closely with data engineers is extremely important for project success
- Being able to understand how data is stored and having the ability to "self-serve" is crucial (a data engineer might not always be at your exposal) - have the ability to understand how to pull and manipulate data to suit the need of your analysis

- This is the meat of a data scientist's usefullness in an overall team. A data scientist should have the ability to summarize datasets via basic statistics and probability, as well as apply/understand the application of common machine learning algorithms
- Data scientists should know when to use which algorithms and the pros and cons of each, depending on the scenario

- A data scientist is useless unless they can apply their algorithms and experiments via a computer - this requires the ability to code.
- A data scientist does not need to be a full blown software engineer, but needs to understand basic software engineering processes and skills (althought the more the better)
- Having an understanding of computers operate, the different pieces of hardware, and how to communicate with them will ultimately make life easier, and is necessary for applying algorithms at large

- A data scientist who cannot communicate effectively to a customer, within a team, across teams, or to leadership will have an extremely hard time completing projects successfully
- Consulting is one of the most important skills to have to not only advance your career, but to grow as a professional and be someone other people want to work with/for
- This is the one piece of a data scientist that is considered an art and not a skill, and usually individuals who come from a strong techincal background struggle with this the most - practice makes perfect
- Motivate yourself to take constructive feedback and improve yourself in this area - you represent your organization and yourself