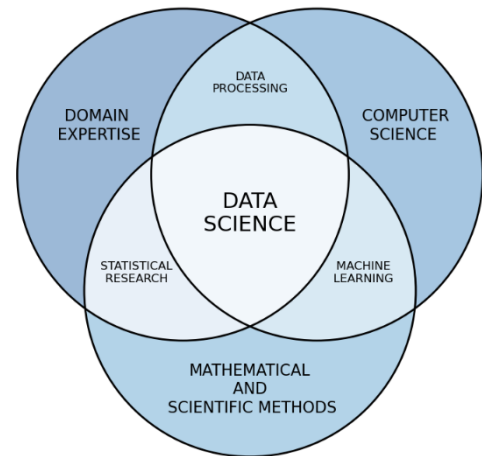# Data Science Competencies

Data science is often described as a profession at the intersection of computer science, statistics, and domain expertise. Domain expertise can be general like design thinking or business processes, but in the social sector it is often substantive areas like housing policy, environmental policy, criminal justice, or program evaluation.

A data science professional typically has a baseline level of competency in two of these areas and deep expertise in one. It's impossible to master everything, so knowing how to import knowledge into projects and collaborate with other experts is an important component of the job. Good data scientists are constantly learning and always examining trade-offs between make versus buy decisions when managing projects (execute yourself vs delegate/contract). One way to think about measuring the minimal viable competency required for a skill is determining whether you could write an RFP or a contract for a particular task, which requires a basic conceptual understanding of the work entailed and the ability to describe the requirements and desired deliverable precisely. Do you have enough expertise to manage the process or review the work of others?

The following list represents various domains of knowledge or skills that are common among data science professionals. It is meant as a jumping off point for those new to the profession to measure one's current level of expertise and strategize knowledge-acquisition accordingly.

Programming Languages / Syntax

- Data Programming (at least one of R, Python, Julia, Stata, SAS)
- Markdown (text formatting syntax) / R Markdown / Quarto
- TeX, Mathjax (mathematical notation)
- Basic HTML, CSS, Javascript (websites and document style)

Data Concepts

- Simple and Complex Data Types (integers and strings vs NAs and dates)
- Data Structures (vectors, matrices, data frames, lists, arrays)
- File Formats (CSV, RDB, JSON, XML, Shapefiles)
- Data Ingestion (Load vs Read, Format Conversion, APIs)
- Data Harvesting (web scraping, parsing)
- Data Cleaning, Wrangling, Munging, Joining, Augmentation
- Variable Transformations and Feature Engineering
- Database Concepts (tables, joins, keys, cardinality, queries)
- Assessment of Data Quality
- Validation and Documentation
- Open Data / Big Data
- Privacy and Responsible Usage

Computational Techniques / Software Design Skills
- Functions (inputs, arguments, return values) and Functional Programming
- Efficient Programming (syntax style guides, documentation, debugging)
- Robust Programming (unit testing, error handling)
- Profiling Performance / Benchmarking Runtime
- Parallelization and Distributed Computing
- Cloud Computing (AWS, R Studio Cloud)
- Big Data Platforms (Spark)
- Task Automation / Scheduling

Report Publishing and Documentation
- Markdown, R Markdown, Quarto (YAML, CSS)
- Dashboards/Apps (R Shiny, Tableau, D3.js)
- Report Automation / Batch Reporting
- Template Design (HTML, Javascript, liquidtags)

Data Science Methodologies
- Predictive Modeling, Data Mining, and Machine Learning
- Text Analysis, Natural Language Processing, and Fuzzy Matching
- Spatial Analytics and Mapping
- Imputation and Analysis with Missing Data
- Synthetic Data Generation or Anonymizing Data
- Data Visualization / Storytelling
- AI / Deep Learning / Task Automation

Collaboration / Reproducibility
- Version Control (Git, Github)
- Literate Programming (R Markdown, Quarto)
- Managing Packages / Environments (renv, docker)
- Documentation (markdown, README, roxygen, data dictionaries)

Team Management / Project Management
- SCRUM or equivalent flavor of Agile Management for Software
- Project Scoping & Planning (planning poker)
- Project Specification
  - Software Requirements Documentation (SRD)
  - Stakeholder Engagement / Client Management
- Kanban Boards / Time Management (project velocity & burndown)
- Code Review / Quality Assurance

Valued Professional Skills
- Written and Verbal Communication
- Problem Solving and Initiative
- Ethics and Leadership
- Innovation Management: Lean Design / Agile Development