Applied Data Science Skills for Research

	A non-computational non-data driven researcher. Practices and methods used often prevent scaling research questions to meet ambitions.	Desktop-based computational researcher able to apply local computer methods and techniques. Able to explore automation and computational modeling.	HPC user or beginning Cloud user. Has a mental model for computing at-a-distance. Is able to integrate multiple remote systems into workflows.	A skilled computational practicioner who is able to connect discrete systems for business and research purposes.
Compute	Spreadsheet skills, mostly unstructured use of spreadsheets and other data sets. Beginning to think about data structures and models.	Able to implement scripts and simple programs which aid the researcher in autommating repetitive tasks.	Able to think about discrete servers and systems and how they could contribute to a research or business workflow.	Can leverage multiple cloud systems and or geographic locations to research or economic benefit. Understands DevOps and orchestration.
Storage	Consider data and storage as something local and contained within a device caried around. Limited models for data-at-a-distance or backup	Able to consider throughput and redundancy as important considerations to achieve research objectives.	Able to make decisions about storage based on performance and latency and design systems to consider I/O workflows.	Able to build/use tiered storage systems which trade off economics, timeliness and accessibility/persistence of data.
Software	Mostly point and click interface usage. Limited models for reasoning about scripting and programming.	Beginer use of version control and tracking changes in scripts and code. Collaboration beginning to draw upon these tools.	Able to collaborate on version control systems and begin using continuous integration and automation workflows.	Able to build integrated systems which are continuously integrated and invite external contribution and collaboration. Applies mature software practices.
Workflows	Checklists with little to no integration into how items interelate or are scripted. Manual sets of point-click tasks.	Able to connect multiple scripts or programs together in simple workflows. May not have abstracted scripts to general use.	Able to consider asynchronous workflow systems and the use of queing systems to manage distributed infrastrucutre.	Able to build systems of systems, interconnected and interoperable to build and extend capability.
Methods	Statistical models over small and individual data sets. Limited ability to apply similar models over and repeatedly to 100s or 1,000s of data sets.	Growing understanding of Machine Learning (ML) and ANNs. Able to apply stat methods across splits of larger data sets more capably.	Advanced applications of ML and considerations of custom hardware, GPUs, FPGAs etc. Able to balance, compute & model complexity to achieve insight.	Able to connect and interconnect flows of machine learning and statistical insight into complex reasoning networks.

Jonah M. Duckles