Data Science Checklist

Analyzing the Analyzers

Analytics: spatial analytics, business intelligence, clustering

Spreadsheets: Excel

Algorithms: computational complexity, CS theory

Back-End Programming: JAVA/Rails/Objective C

Bayesian/Monte-Carlo Statistics: MCMC, BUGS

Big and Distributed Data: Hadoop, Map/Reduce

Business: management, business development, budgeting, communication, economics

Classical Statistics: general linear model, ANOVA

Data: regexes, R, SAS, web scraping, data mining, databases

Front-End Programming: JavaScript, HTML, CSS

Graphical Models: social networks, Bayes networks

Machine Learning: decision trees, neural nets, SVM, clustering

Math: linear algebra, real analysis, calculus

Optimization: linear, integer, convex, global

Product Development: design, project management

Science: experimental design, technical writing/publishing, social or physical science

Simulation: discrete, agent-based, continuous

Spatial Statistics: geographic covariates, GIS

Structured Data: SQL, JSON, XML

Surveys and Marketing: multinomial modeling

Systems Administration: \*nix, DBA, cloud tech

Temporal Statistics: forecasting, time-series analysis

Unstructured Data: noSQL, text mining

Visualization: Vis.js

Four types of data scientists: Data developer (programmer, engineer), Data researcher (researcher, scientist, statistician), Data businessperson (leader, businessperson, entrepreneur)

Programs/Languages: Hadoop, SVM, Scala, R, Matlab,

Interview question: Find a publicly available data set at least several hundred gigabytes in size, pose and answer an interesting question about the data, and detail all steps, assumptions, and conclusions, including code.

Data science qualifications: Defining feature of data scientists is the breadth of their skills — their ability to single-handedly do at least prototype-level versions of all the steps needed to derive new insights or build data products. Most successful data scientists are those with substantial, deep expertise in at least one aspect of data science, be it statistics, big data, or business communication.

Tools: Tools, technical expertise, mentoring, curiosity, storytelling, and cleverness are critical to data scientists effectiveness.

Working with companies: Need access to raw data and decision makers.