# Python

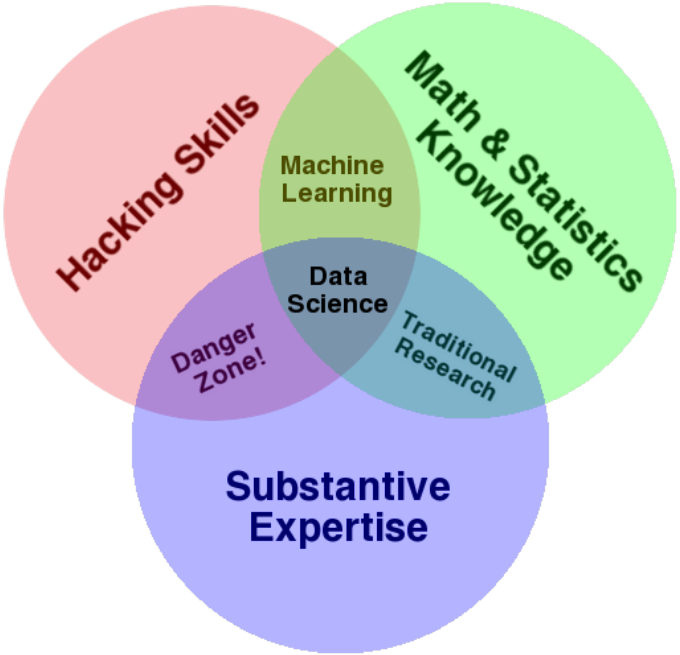
- **Why Python?**
  1. *It's easy to learn*
     - *Now the language of choice for 8 of 10 top US computer science programs (Philip Guo, CACM)*
  2. *Full featured*
     - *Not just a statistics language, but has full capabilities for data acquisition, cleaning, databases, high performance computing, and more*
  3. *Strong Data Science Libraries*
     - *The SciPy Ecosystem*

# Course Outline

1. Prerequisite Python Knowledge
2. The *pandas* Toolkit
3. Advanced Querying and Manipulation with *pandas*
4. Basic Statistical Analysis with *numpy* and *scipy*, and project

# Data Science

- Drew Conway perspective on data science:
  - *Hacking Skills*
  - *Math and Statistics Knowledge*
  - *Substantive Expertise*
- Other data science perspectives:
  - *Skepticism, experimentation, simulation, and replication*



# Data Science

- **David Donoho, "50 Years of Data Science"**
  1. *Data Exploration and Preparation*
  2. *Data Representation and Transformation*
  3. *Computing with Data*
  4. *Data Modeling*
  5. *Data Visualization and Presentation*
  6. *Science about Data Science*

# The map() function

**map**(*function*, *iterable*, ...)

Return an iterator that applies *function* to every item of *iterable*, yielding the results. If additional *iterable* arguments are passed, *function* must take that many arguments and is applied to the items from all iterables in parallel. With multiple iterables, the iterator stops when the shortest iterable is exhausted. For cases where the function inputs are already arranged into argument tuples, see `itertools.starmap()`.