Fundamentals of Data Science

Introduction: Core Concepts and Technologies

Southampton Data Science Academy

The Data Science toolkit

Southampton Data Science Academy

Technologies and tools

Programming languages

R, Python, Java, JavaScript

Big data processing

Hadoop, HDFS, MapReduce, Spark, Storm, Hive, Pig

Data management

SQL, NoSQL, Mongo DB, Cassandra

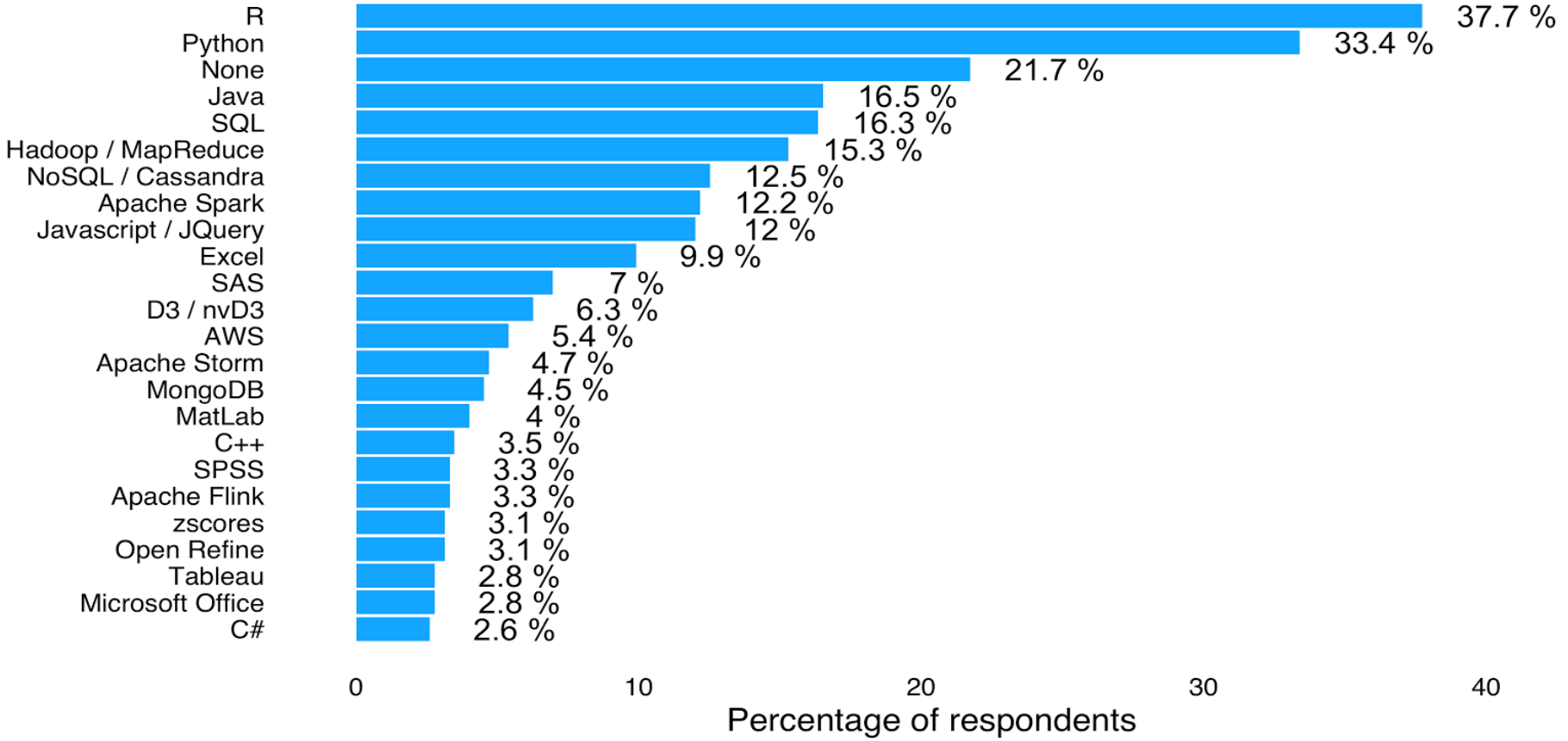
Visualisation

D3.js, Gephi, Tableau, Shiny, Excel, Gapminder

Data analysis

R, Excel, Weka, RapidMiner, OpenRefine, SAS, SPSS, Watson Analytics, Open Calais, Matlab

Southampton Data Science Academy



What tools should be taught or covered on data science courses?

Results from the European Data Science Academy's study of 693 data science practitioners and managers across all EU member states.

Southampton Data Science Academy

Choosing a tool

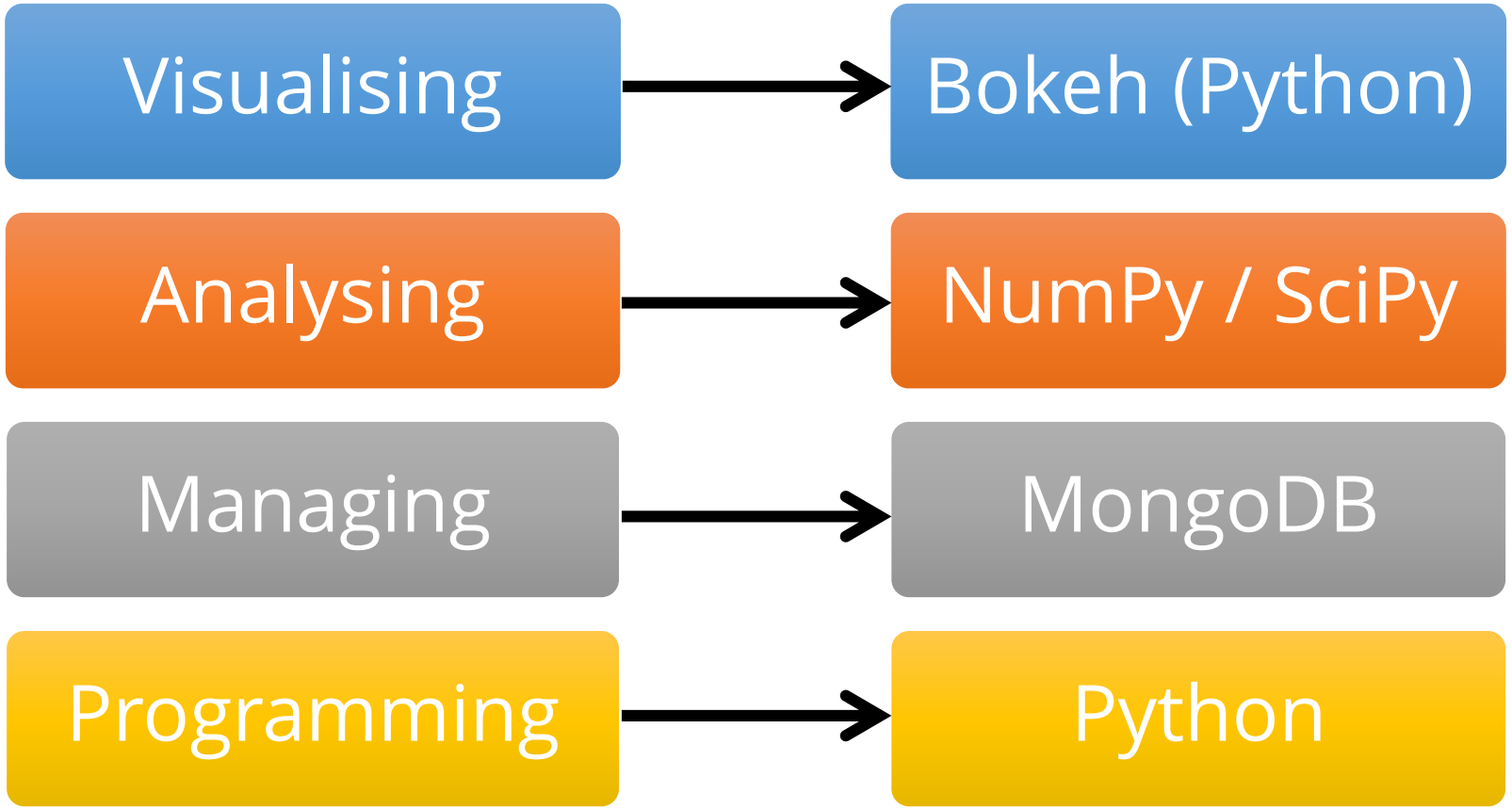
The use of each tool or technology depends on:

- type of data
- scale
- application scenario
- best practices
- etc

There is no "one size fits all" solution

Southampton Data Science Academy

Toolkit on this course



Python

- Python (as with R) is one of the most popular languages in data science jobs
- flexible language that is quick to get started with
- libraries for doing various parts of the data science process
  - Eg Bokeh for visualisation

Southampton Data Science Academy