R vs Python vs MATLAB vs Octave

As machine learning engineer I would like to compare 4 machine learning programming languages(tools). Let's take this a bit deeper. Since most of us are concerned with ML and analysis being a big part of why we are using these programs.I want to list a few advantages and disadvantages of each for who want to start learning them as a data scientist.

MJ Bahmani

R

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

Advantages :

- End To End development to execution (some brokers packages allows execution, IB)
- Rapid development speed (60% fewer lines vs python, ~500% less than C)
- A large number of Open Source Packages
- Mature quantitative trading packages(quantstrat, quantmod, performanceanalyitics, xts)
- Largest Community
- Can integrate into C++/C with rcpp

Disadvantages :

- Slow vs Python especially in iterative loops and non vectorized functions
- Worse plotting than Matlab and difficult to implement interactive charts
- Limited capabilities in creating stand-alone applications

Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Advantages :

- End To End development to execution (some brokers packages allows execution, IB)
- Open source packages(Pandas, Numpy, scipy)
- Trading Packages(zipline, pybacktest, pyalgotrade)
- best for general programming and application development
- can be a "glue" language to connect R, C++, and others (python)
- Fastest general speed especially in iterative loops

Disadvantages :

- immature packages especially trading packages
- some packages are not compatible with others or contain overlap
- smaller community than R in finance
- More code required for same operations vs R or Matlab
- Silent errors that can take a very long time to track down (even with visual debuggers / IDE)

MATLAB

MATLAB (**mat**rix **lab**oratory) is a multi-paradigm numerical computing environment. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, a creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran, and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

Advantages :

- Fastest mathematical and computational platform especially vectorized operations/ linear matrix algebra
- Commercial level packages for all fields of mathematics and trading
- Very short scripts considering the high integration of all packages
- Best visualization of plots and interactive charts
- Well tested and supported due to it being a commercial product
- Easy to manage multithreaded support and garbage collection
- Best debugger

Disadvantages :

- Can not execute - must be translated into another language
- Expensive ~1000 per license and 50+ per additional individual package
- Can not integrate well with other languages
- Hard to detect biases in trading systems (it was built for math and engineering simulations) so extensive testing may be required. EG. look ahead bias
- Worst performance for iterative loops
- Can not develop stand-alone applications at all.

Octave

Octave is sort of the GNU answer to the commercial language MATLAB. That is, it is a scripting matrix language, and has a syntax that is about 95% compatible with MATLAB. It's a language designed by engineers, and thus is heavily loaded with routines commonly used by engineers. It has many of the same time series analysis routines, statistics routines, file commands, and plotting commands of the MATLAB language.

Advantages

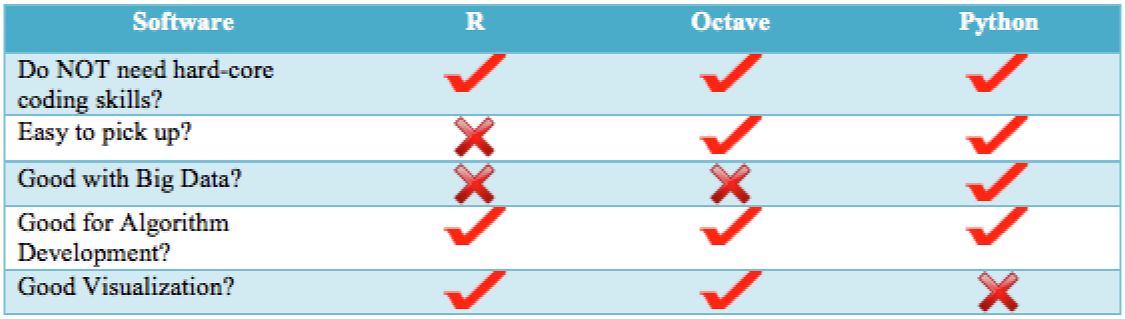
- First of all, there is no robust Octave compiler available and this is not really necessary either since the software can be installed free of charge.
- Looking at the language element the two packages are identical except for some particularities like nested functions. Octave is under constant

active development and every deviation from the Matlab syntax is treated as a bug or at least an issue to be resolved.

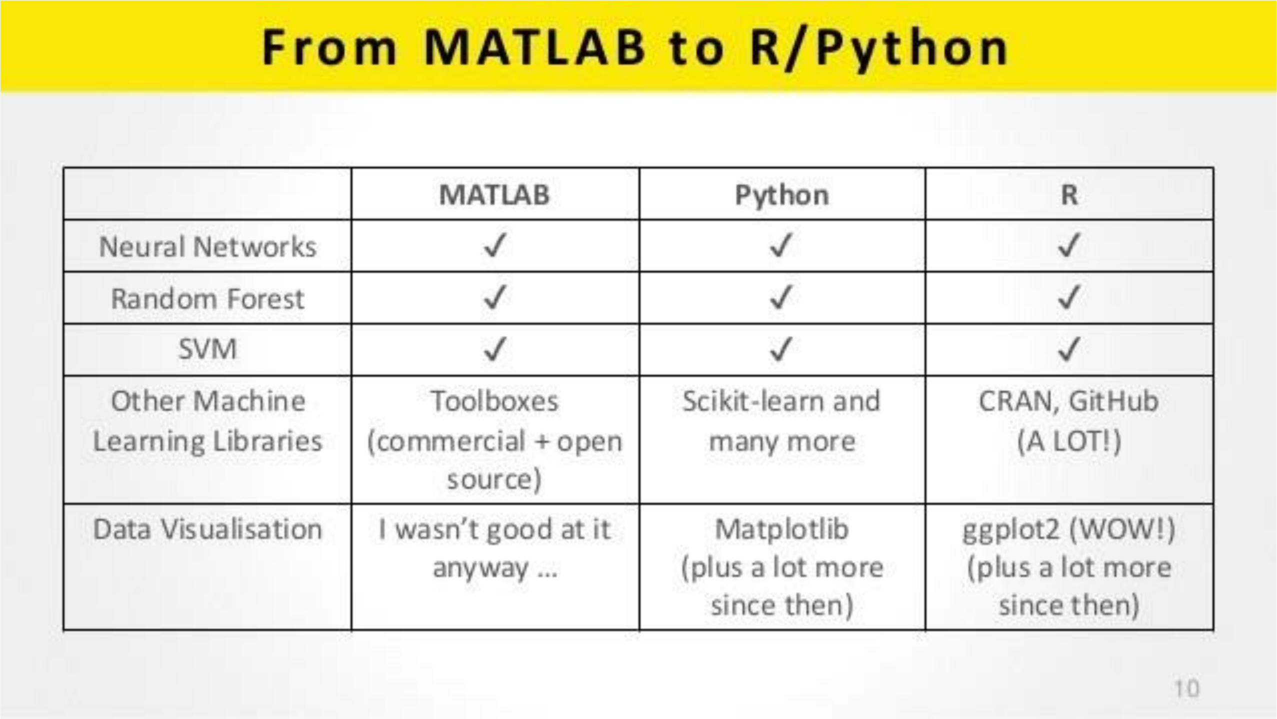
- There are also plenty of toolboxes available for octave and as long as a program does not require graphical output there is a good chance that it runs under Octave just like under Matlab without considerable modification.
- Graphics capabilities are clearly an advantage of Matlab. The latest versions include a GUI designer on top of excellent visualization features.
- Octave uses either GNU Plot or JHandles as graphics packages, where the latter is somehow closer to what Matlab provides. However, there are no Octave equivalents to a GUI designer and the visualization mechanisms are somehow limited and not Matlab compatible.
- The same holds for an integrated development environment. There is a project called QTOctave but it is still at an early stage.
- Looking at the collaborate efforts taking place around the Octave community it is likely that this software will soon provide better and possibly even compatible graphics and GUI capabilities and it is well worth a look before buying Matlab.

Disadvantages

- it just a free open source of MATLAB and don't bring us anything new
-



Software	R	Octave	Python
Do NOT need hard-core coding skills?	✓	✓	✓
Easy to pick up?	✗	✓	✓
Good with Big Data?	✗	✗	✓
Good for Algorithm Development?	✓	✓	✓
Good Visualization?	✓	✓	✗

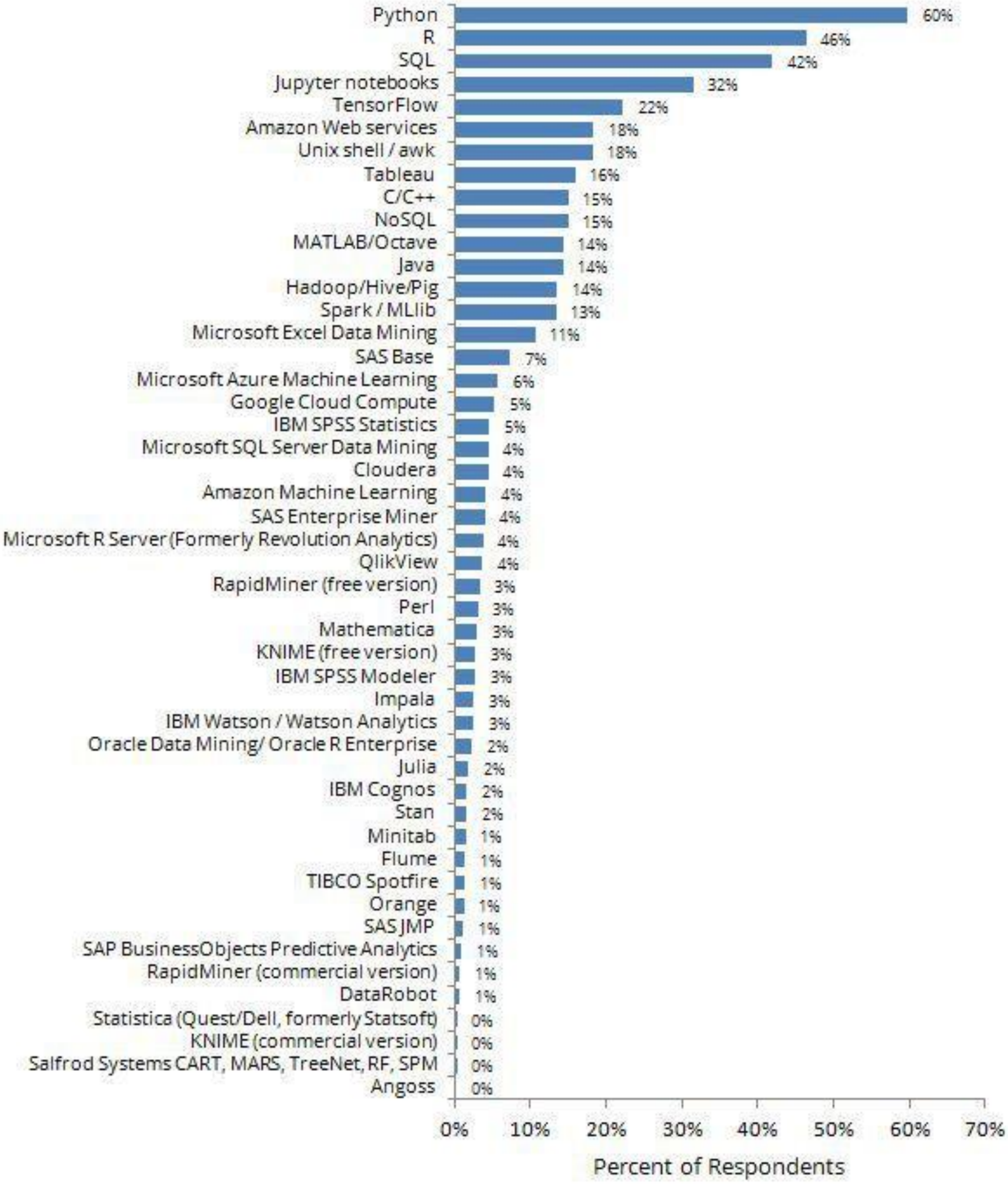


From MATLAB to R/Python

	MATLAB	Python	R
Neural Networks	✓	✓	✓
Random Forest	✓	✓	✓
SVM	✓	✓	✓
Other Machine Learning Libraries	Toolboxes (commercial + open source)	Scikit-learn and many more	CRAN, GitHub (A LOT!)
Data Visualisation	I wasn't good at it anyway ...	Matplotlib (plus a lot more since then)	ggplot2 (WOW!) (plus a lot more since then)

To sum up, there are several tools for data scientist and machine learning engineer in the below chart you can see which one is more popular than others.

Data Science / Analytics Tools, Technologies and Languages Used in Past Year



Data are from the Kaggle 2017 The State of Data Science and Machine Learning study. You can learn more about the study and download the data here: https://www.kaggle.com/surveys/2017.

Respondents were asked to indicate for work, which data science/analytics tools, technologies, and languages they used in the past year. A total of 10153 respondents answered the question.

BUSINESS BROADWAY