

Core tools I

Lecture 02

Quiz

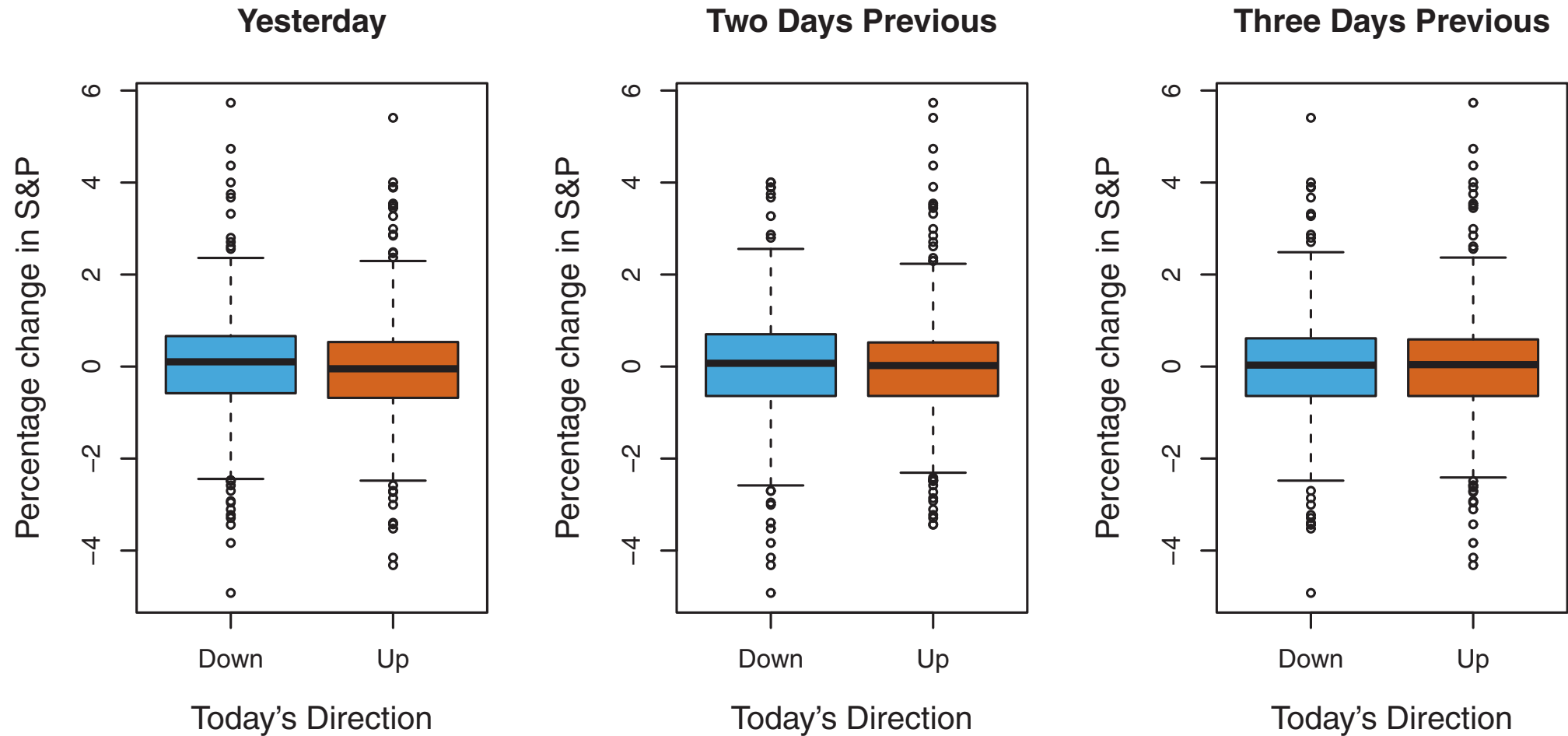


FIGURE 1.2. Left: *Boxplots of the previous day's percentage change in the S&P index for the days for which the market increased or decreased, obtained from the Smarket data.* Center and Right: *Same as left panel, but the percentage changes for 2 and 3 days previous are shown.*

Common tools for data science



- Free
- Plentiful, easy-to-use statistical packages
- Easy for beginners, especially with stats background
- RStudio IDE
- Ggplot2 for plotting, shiny for web apps



- Easy for beginners
- Extensive built in toolboxes
- Extensive and well-maintained documentation
- Matrix-based language – linear algebra is as simple as it gets
- Debugging and code profiling is superb
- MATLAB central community



- Free
- A **full-fledged programming language**
- Can be used from exploratory work and production environments
- numpy, scipy, matplotlib, pandas, scikit-learn, statsmodels
- Great for object oriented programming
- Ipython and Jupyter notebooks

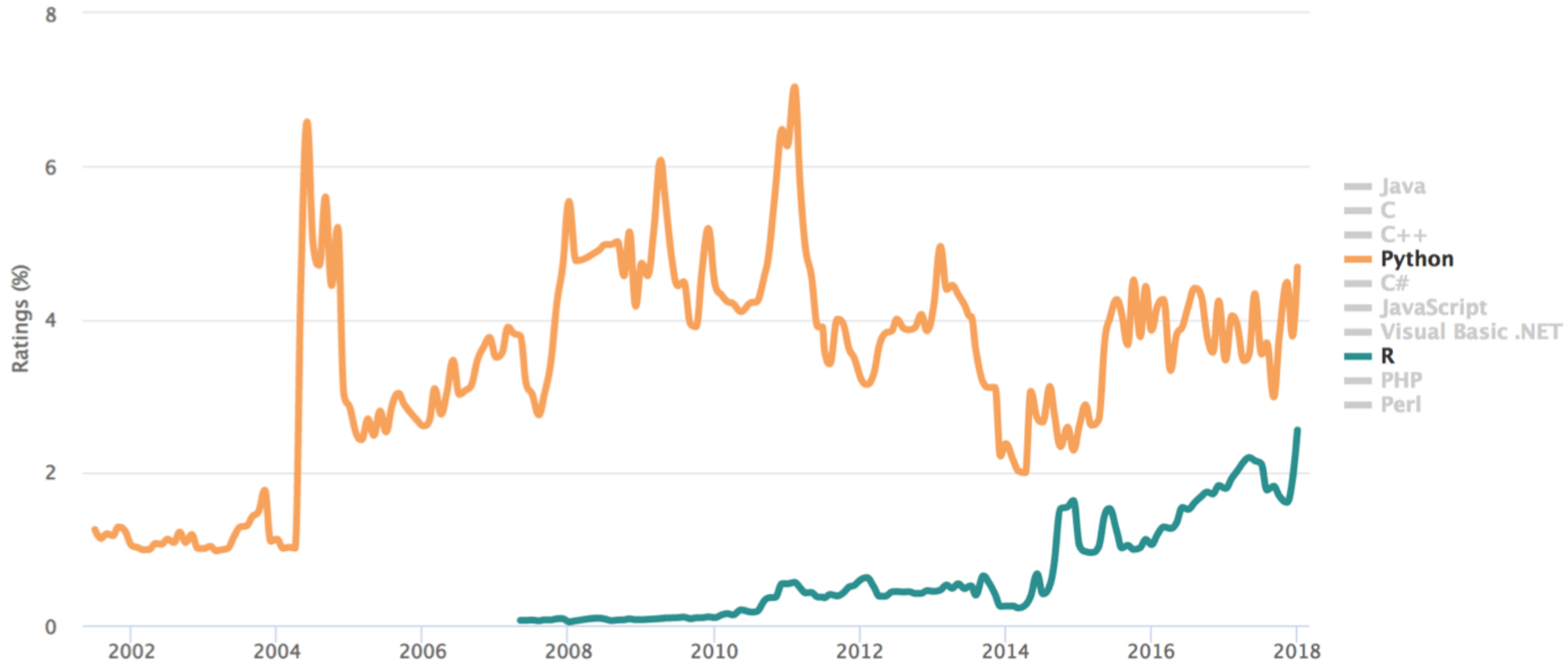
TIOBE Programming Language Index

	Jan 2018	Jan 2017	Change	Programming Language	Ratings	Change
1		1		Java	14.215%	-3.06%
2		2		C	11.037%	+1.69%
3		3		C++	5.603%	-0.70%
4		5	^	Python	4.678%	+1.21%
5		4	v	C#	3.754%	-0.29%
6		7	^	JavaScript	3.465%	+0.62%
7		6	v	Visual Basic .NET	3.261%	+0.30%
8		16	^^	R	2.549%	+0.76%
9		10	^	PHP	2.532%	-0.03%
10		8	v	Perl	2.419%	-0.33%
11		12	^	Ruby	2.406%	-0.14%
12		14	^	Swift	2.377%	+0.45%
13		11	v	Delphi/Object Pascal	2.377%	-0.18%
14		15	^	Visual Basic	2.314%	+0.40%
15		9	vv	Assembly language	2.056%	-0.65%
16		18	^	Objective-C	1.860%	+0.24%
17		23	^^	Scratch	1.740%	+0.58%
18		19	^	MATLAB	1.653%	+0.07%
19		13	vv	Go	1.569%	-0.76%
20		20		PL/SQL	1.429%	-0.11%

<https://www.tiobe.com/tiobe-index/>

TIOBE Programming Community Index

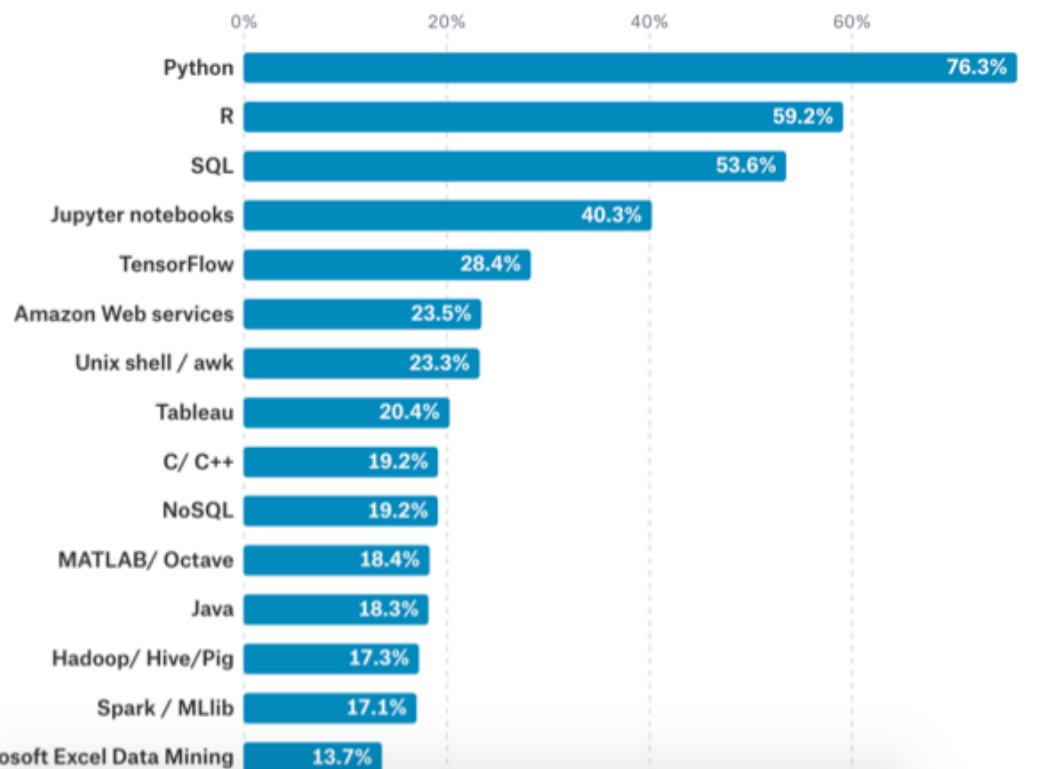
Source: www.tiobe.com



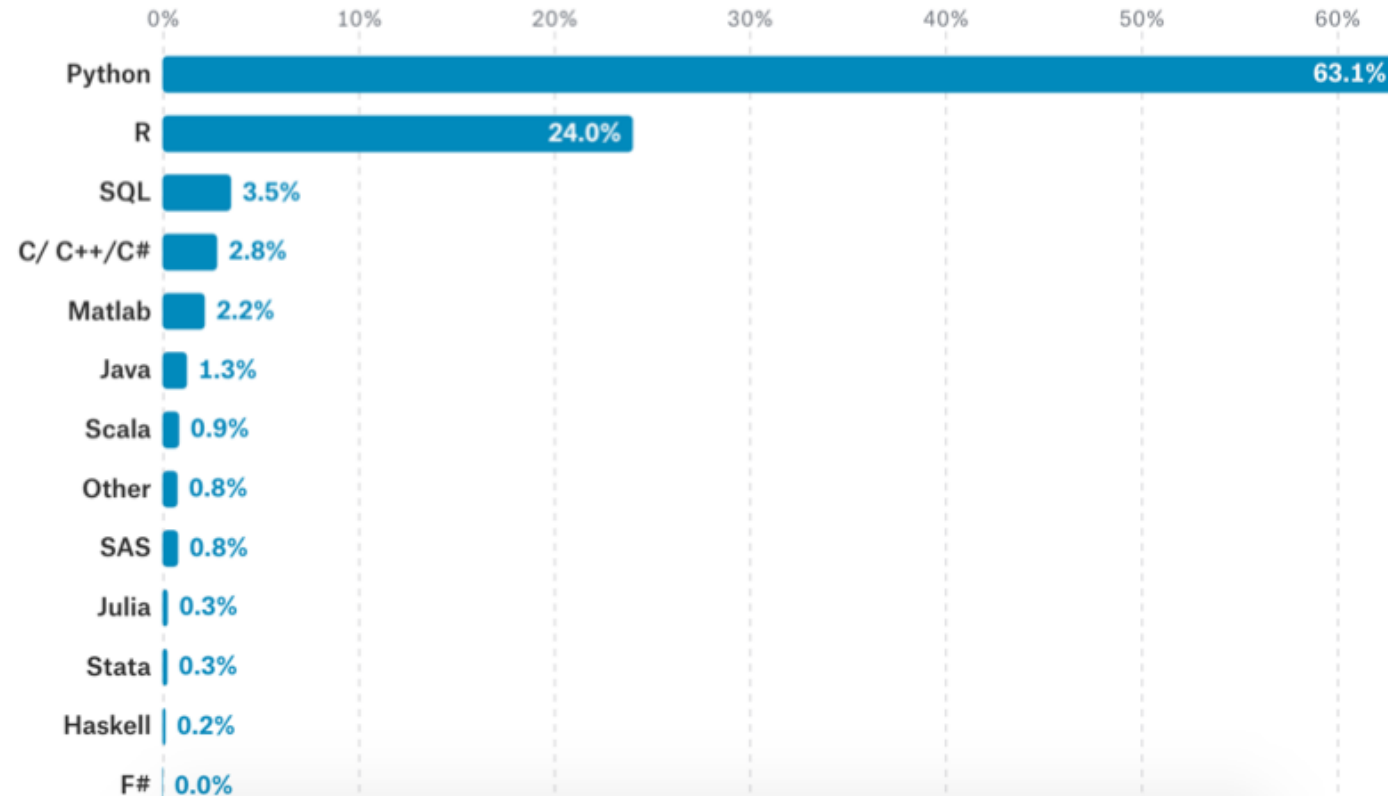
<https://www.tiobe.com/tiobe-index/>

2017 The State of Data Science & Machine Learning

What tools are used at work?



What language would you recommend new data scientists use first?

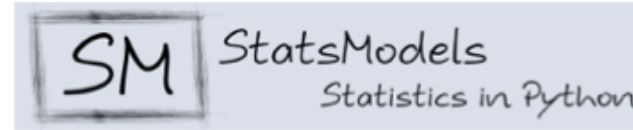


Data Science Packages

High Level Libraries



machine learning



statistical models and tests



high level data manipulation
and analysis



image processing

Fundamental Packages



numerical array and
matrix processing



Scientific computing:
optimization, linear algebra,
integration, etc.



plotting

Core Language



Environments



create and share
docs with live
code, equations,
visualizations, and
narrative text



integrated
development
environment



interactive
command
line



command
line

Topics

Python (command line, iPython, Spyder)

Jupyter notebooks

Common data science packages (numpy, matplotlib, pandas)

Tutorials