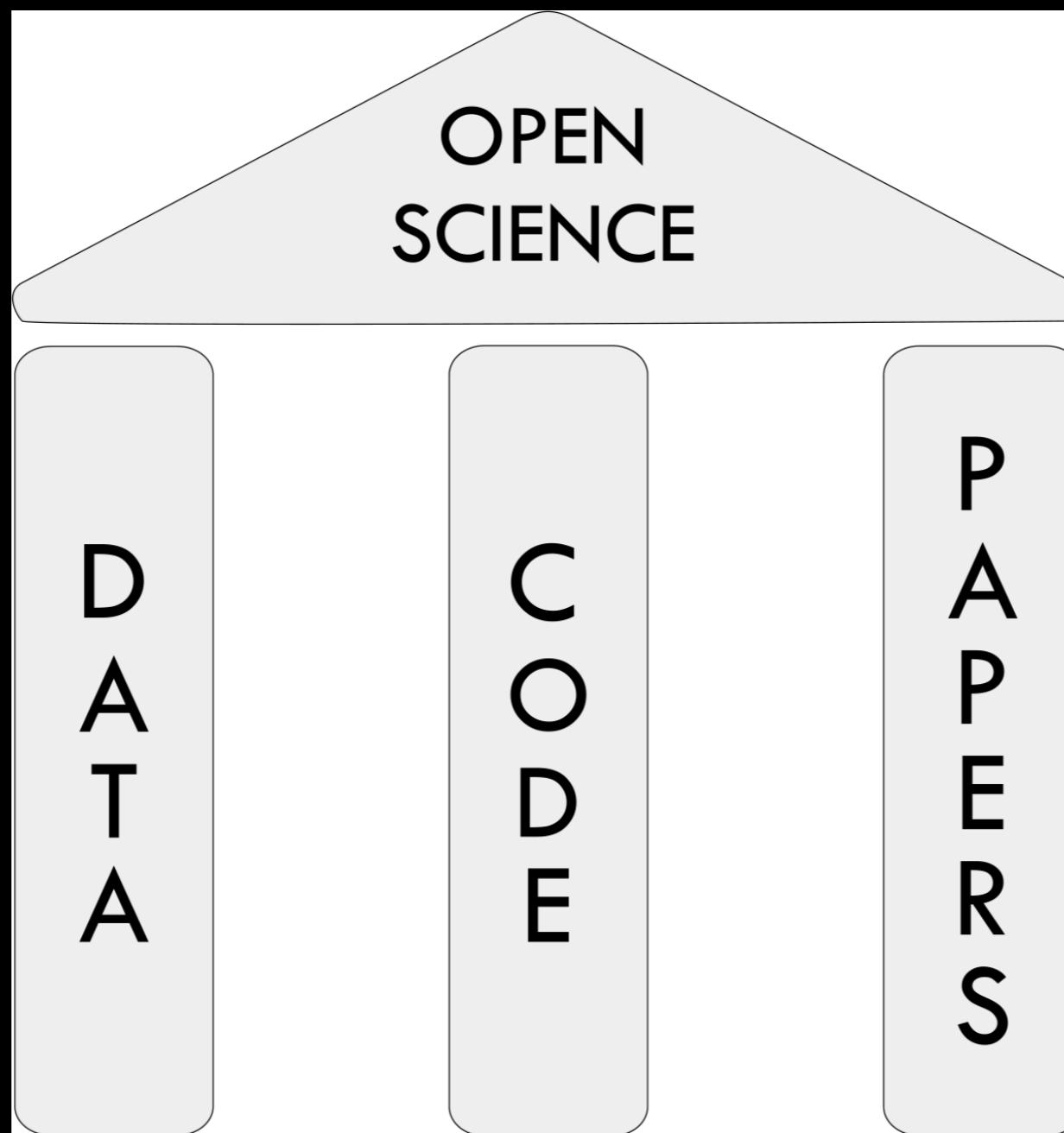


Scientific Computing with Python

Luke Chang
MIND 2018 Hanover, NH

Pillars of Open Science



Old Software Stack

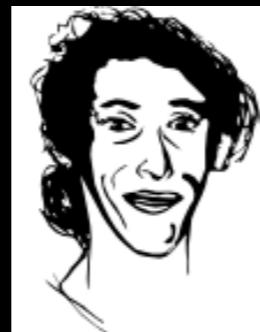
- Neuroimaging: Matlab
- Computational Modeling: Matlab
- Statistics: R
- Plotting: R
- Scripting: Bash (awk, sed)

Problems

- **Switching costs** get worse when you have less time (professor)
- Matlab is extremely **expensive** and network authentication can be a nightmare (e.g., rural NH, plane)
- Can't easily **share code** and packages with others in Matlab
- Can't easily deploy matlab on **cloud computing** infrastructure
- Can't automate GitHub **webhooks**
- Matlab doesn't play well with **containers**
- Matlab skills are not particularly useful in **industry**
- What do you want your **students** to learn?

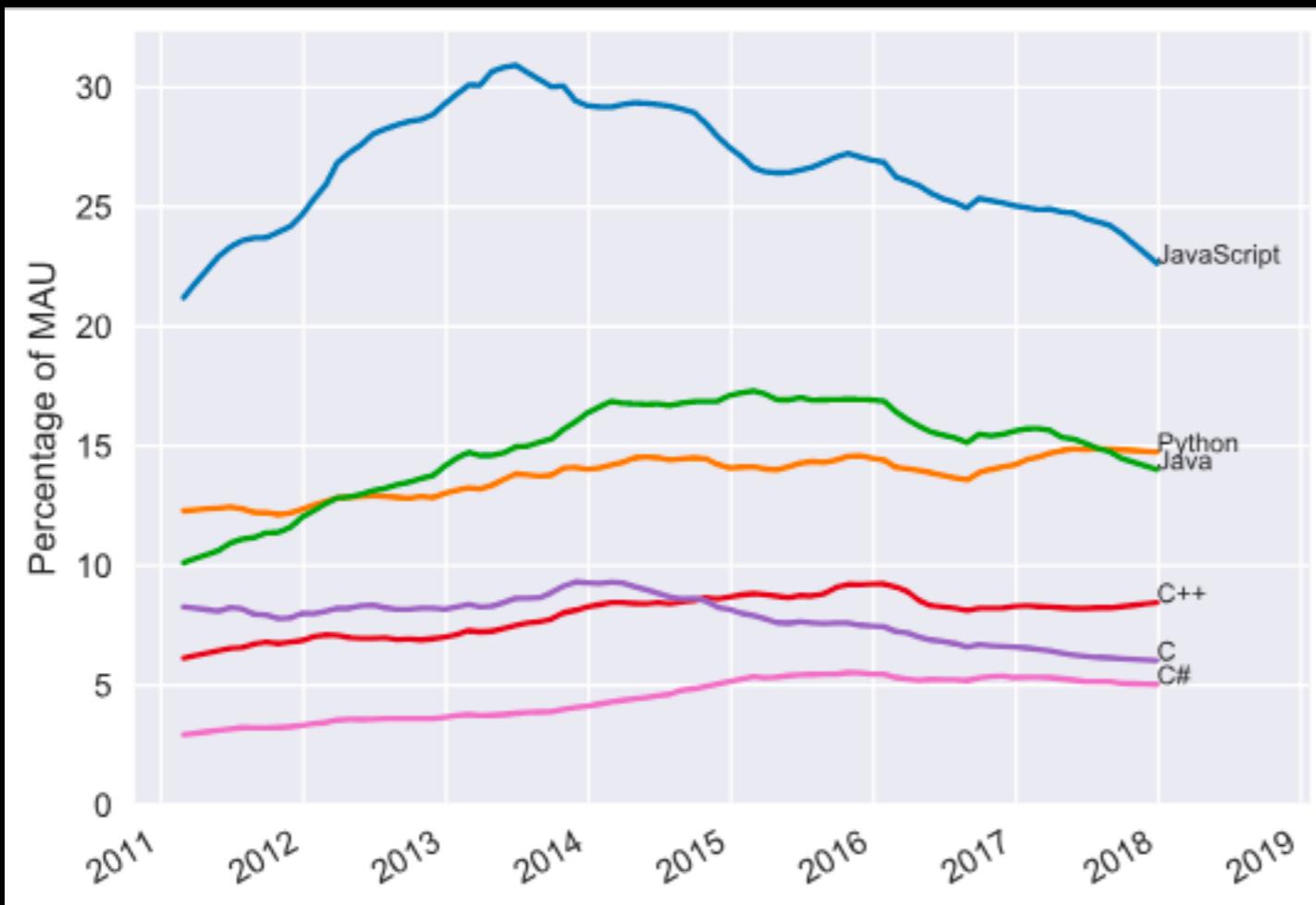
Scientific Python

- Homogenization of scientific computing (Tal Yarkoni)
 - <https://www.talyarkoni.org/blog/2013/11/18/the-homogenization-of-scientific-computing-or-why-python-is-steadily-eating-other-languages-lunch/>
- Unexpectedness effectiveness of python for science (Jake Vanderplas)
 - <https://www.youtube.com/watch?v=ZyjCqQEUA8o>
- Personalities!

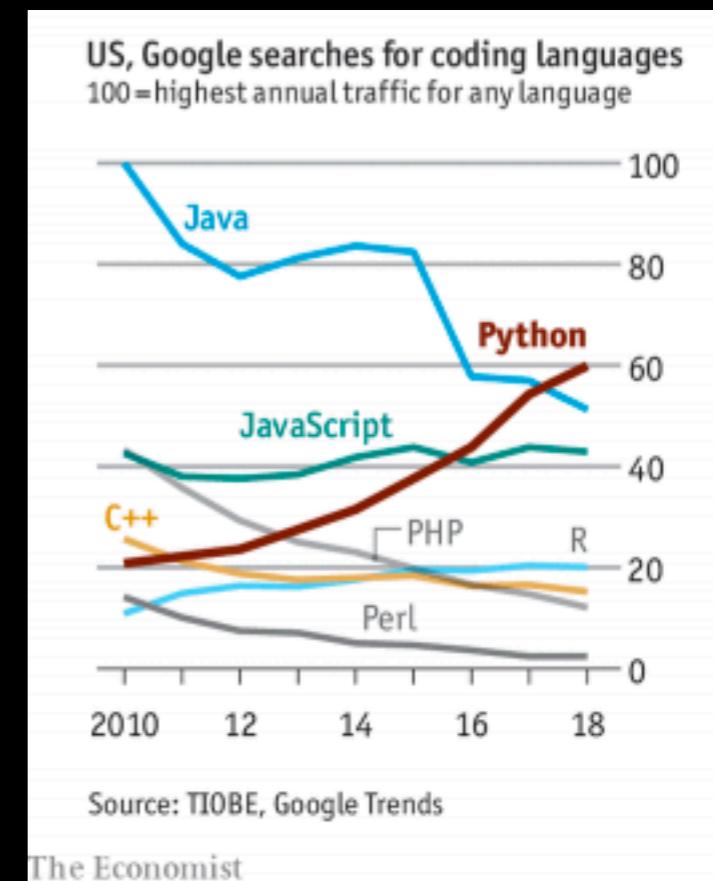


Python is Popular!

Github Monthly Active Users



Github Monthly Active Users



<https://www.benfrederickson.com/ranking-programming-languages-by-github-users/>

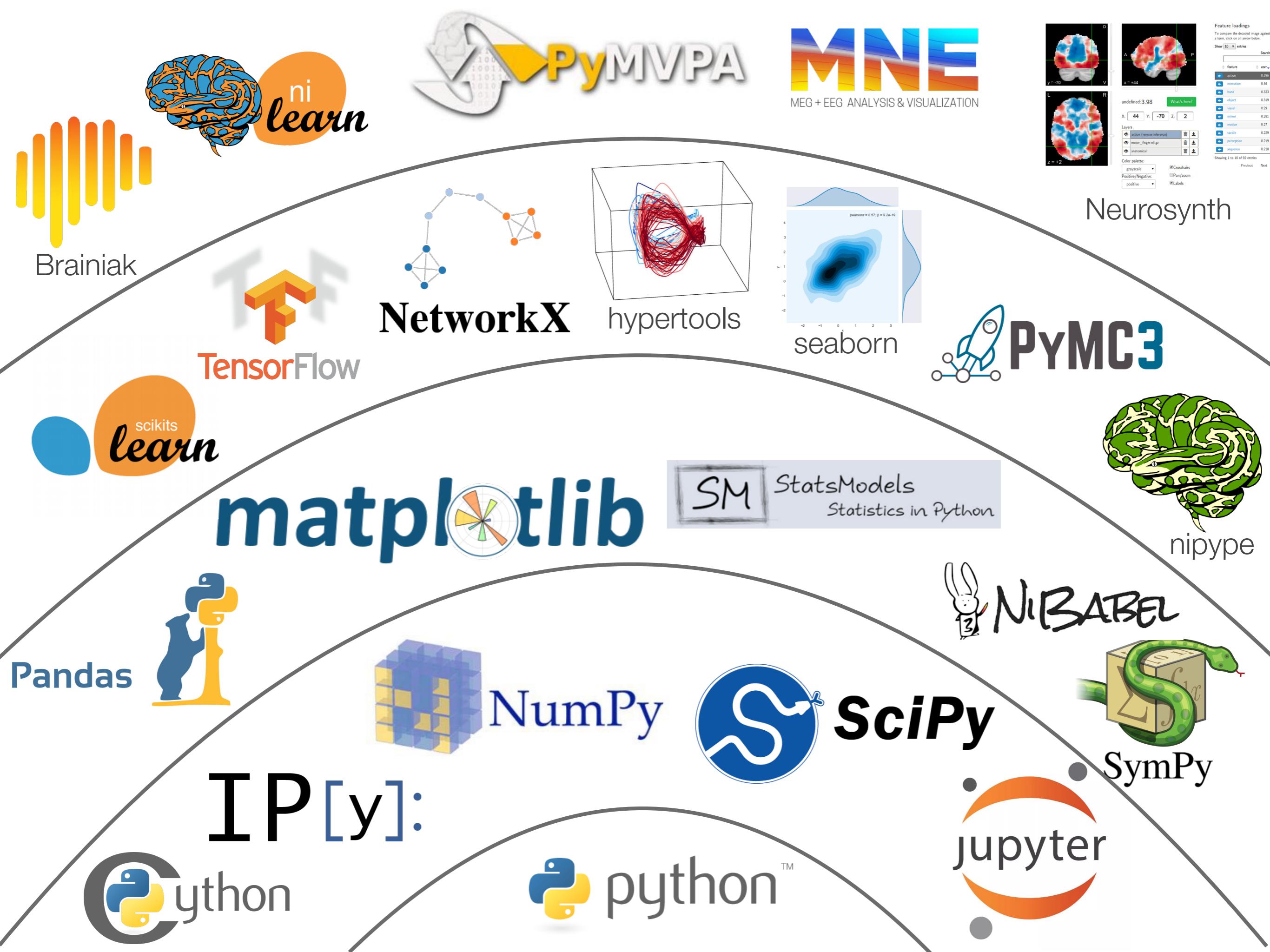
<https://www.economist.com/graphic-detail/2018/07/26/python-is-becoming-the-worlds-most-popular-coding-language>

Why is it so successful?

- Free!
- High level intuitive syntax
- “Real” programming language
- Plays well with others (C, fortran)
- Opinionated style guidelines
- Tightly integrated open source software stack
- One language to do most things (data analysis, scripting, webdev)

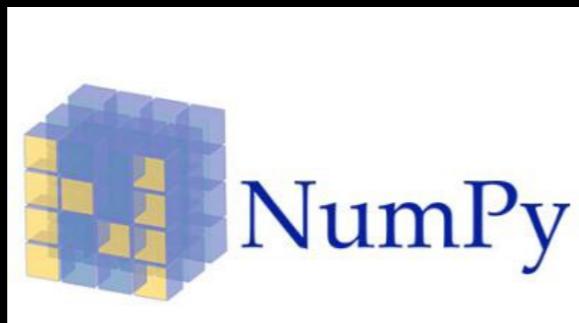
Other Bonuses

- Easy package management (PyPi)
- Easy documentation (Sphinx)
- Automated testing (pytest + continuous integration)
- Easy parallelization (multiprocessing, joblib)

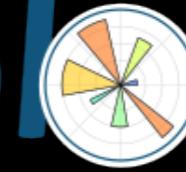


Something for everyone

Matlab



matplotlib



R



Need mixed models?

- <https://github.com/bambinos/bambi>
- <https://github.com/ejolly/pymer4>
- Don't recommend statsmodels

Interactive
Computational
Narratives





Getting Started

- Anaconda is the most straightforward scientific computing distribution
- <https://github.com/jakevdp/WhirlwindTourOfPython>
- <https://github.com/jakevdp/PythonDataScienceHandbook>
- <http://nbviewer.jupyter.org/github/dartmouth-pbs/psyc161/blob/master/classes/01a-Introduction.ipynb>
- <https://github.com/ljchang/psyc63>