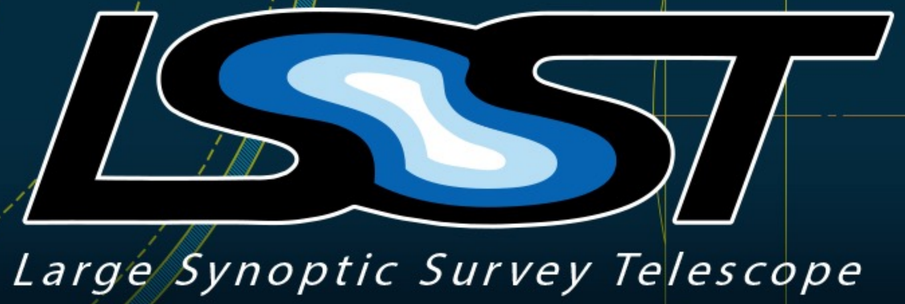
Science Notebooks

J. Matt Peterson, Senior Astronomy Software Developer

Education & Public Outreach Acceptance Review



September 26, 2017

Goals

One System. Many Aspects.

- Engaging and accessible.
- Authentic Data and Tools.
- Reproducible Transparent Science.
- Flexible systems.

Demo

Just Works.

- Same unified experience for all students and teachers.
- Appropriate resources available for all users.
- The environment is correctly configured.
- Code can be hidden. Or code can provide transparency to how it’s really working.

Technology

- Jupyter Project
- Jupyter Lab
- Jupyter Notebooks
- Jupyter Kernel
- Python

jupyter

python™

Jupyter Project

The Organization.

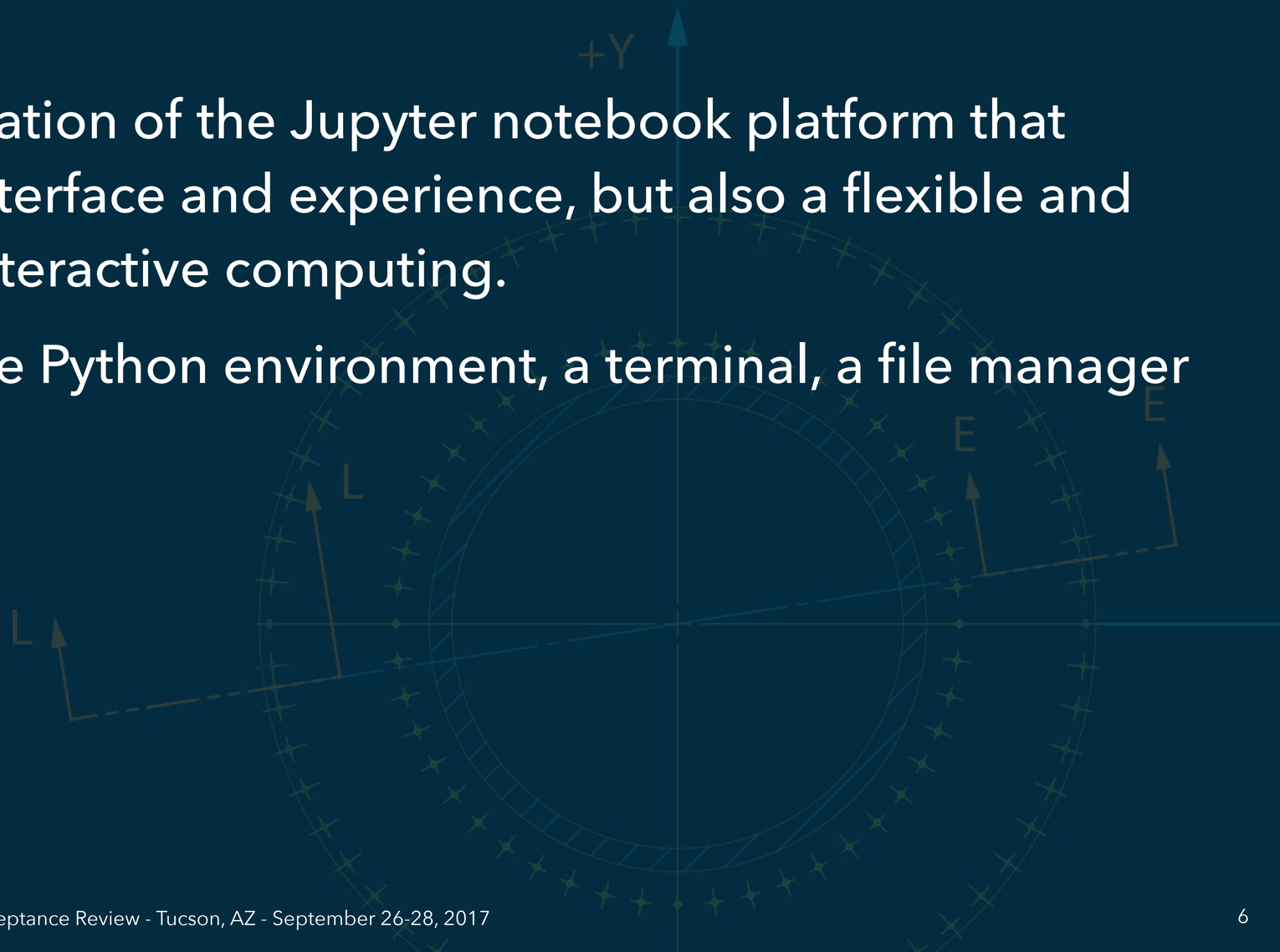
- The **Jupyter Project** is an open source project born out of the IPython project in 2014 as it evolved to support interactive data science and scientific computing across all programming languages.
- Jupyter will always be 100% open source software, free for all to use and released under the liberal terms of the modified BSD license.
- Sustainable funded. $6M funding at UC Berkeley and Cal Poly from More Foundation, Sloan Foundation and Helmsley Trust.
- 25 full-time employees. Co-developed with Bloomberg and Continuum Analytics.
- Initially driven by researchers then adopted by educators.
- Working relationship with co-PI Brian Granger.

JupyterLab

The Platform.

- **JupyterLab** is the next generation of the Jupyter notebook platform that provides an improved user interface and experience, but also a flexible and extensible environment for interactive computing.
- Features include an interactive Python environment, a terminal, a file manager and a text editor.

JupyterLab is just a website.



Jupyter Notebooks

The Format.

- Jupyter Notebooks (.ipynb) are an open document format based on JSON. They contain a complete record of users's sessions and embed code, narrative text, equations and rich output.
- Supports Unicode.
- Programming language agnostic.
- New clients and views of notebooks continue to be created independently.

Jupyter Kernel

Language And Operating System Independent.

- Kernels run interactive code in a particular programming language and returns output to the user.
- Kernels are simple from a user perspective. They just work. No configuration.
- The notebook communicates with computational kernels using the Interactive Computing Protocol, an open network protocol based on JSON data over ZMQ and Websockets.
- Programming language and operating system independent.

Not just Python.

Python

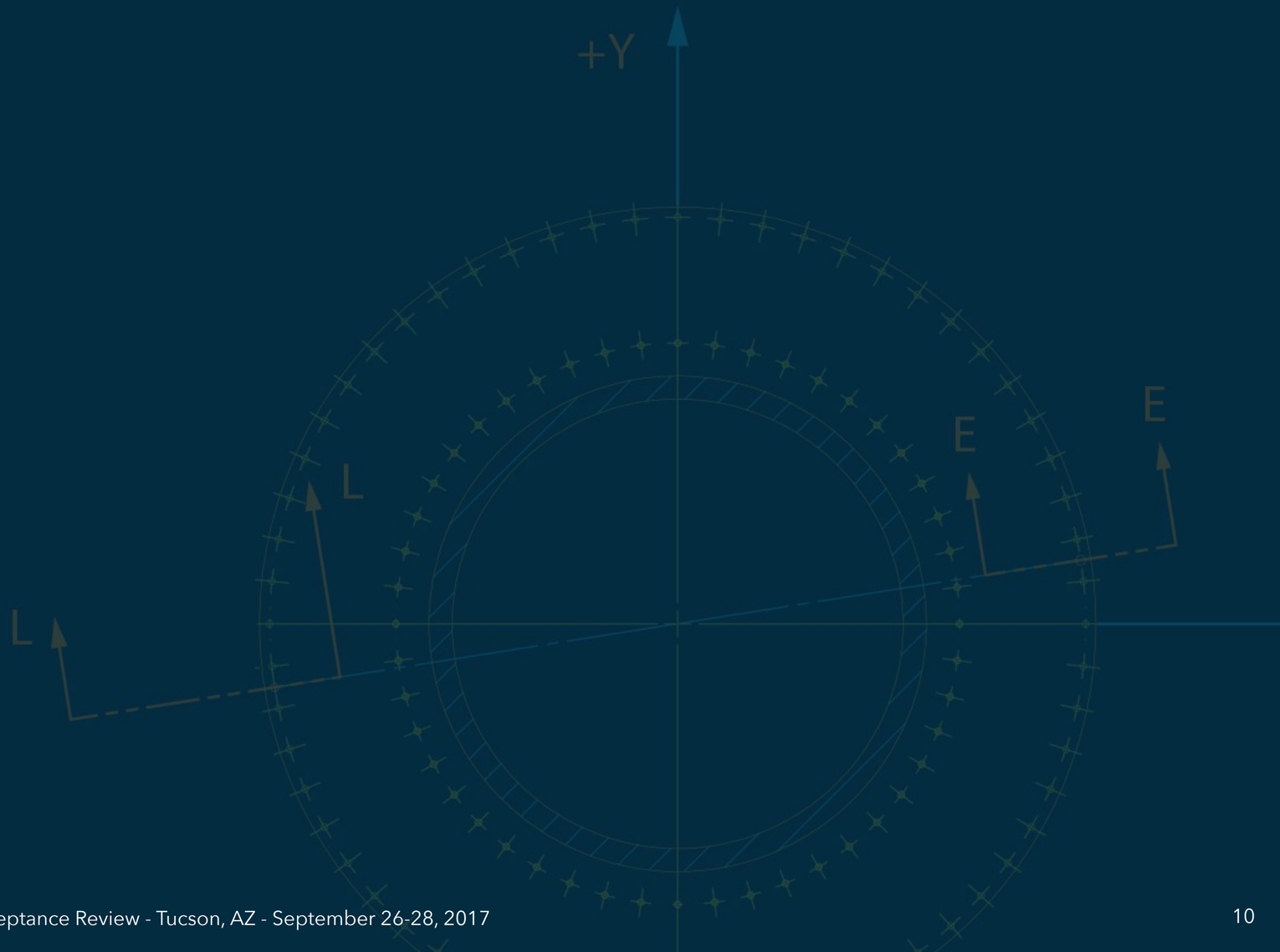
The Second Best Language.

- As the adage goes, Python may not be the best language for any specific task but it is likely to be the second best language.
- Python continues to gain popularity in astronomy, science and general programming.
- Python science libraries are driving adoption. Such as astropy, pandas, scipy, numpy, Matplotlib and ipython.

Trends

What About In 2021?

- Python
- Jupyter Notebooks



Trends

Methods

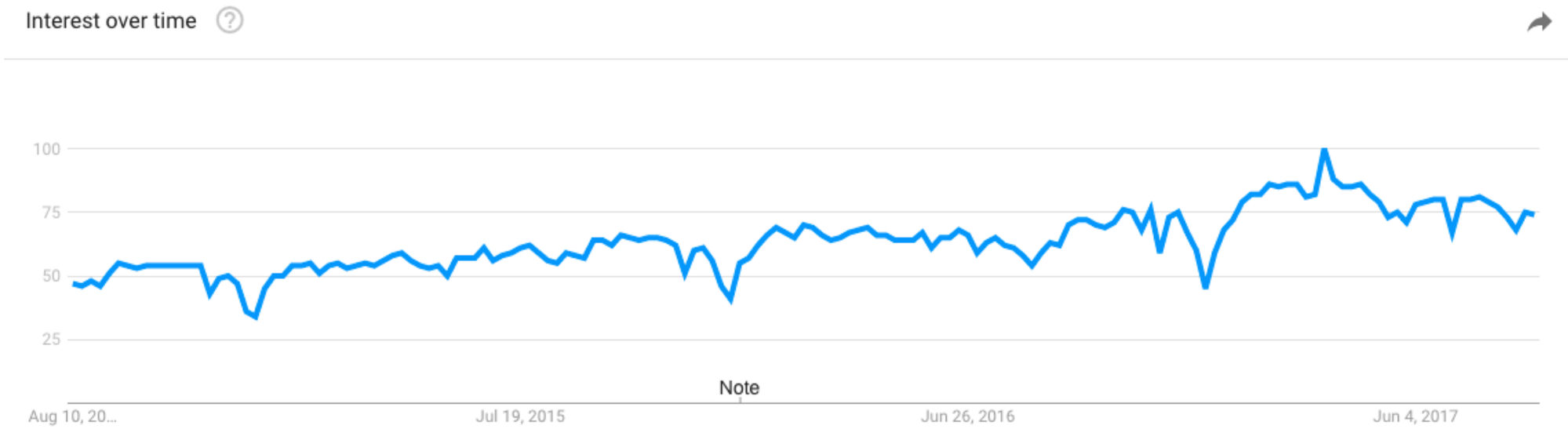
- Google Trends in the United States.
- StackOverflow question counts over time.
- StackOverflow is an online question and answer site for software developers. As the top site, it is considered representative of the software industry overall.

Trends

Jupyter Google Trends In The U.S.

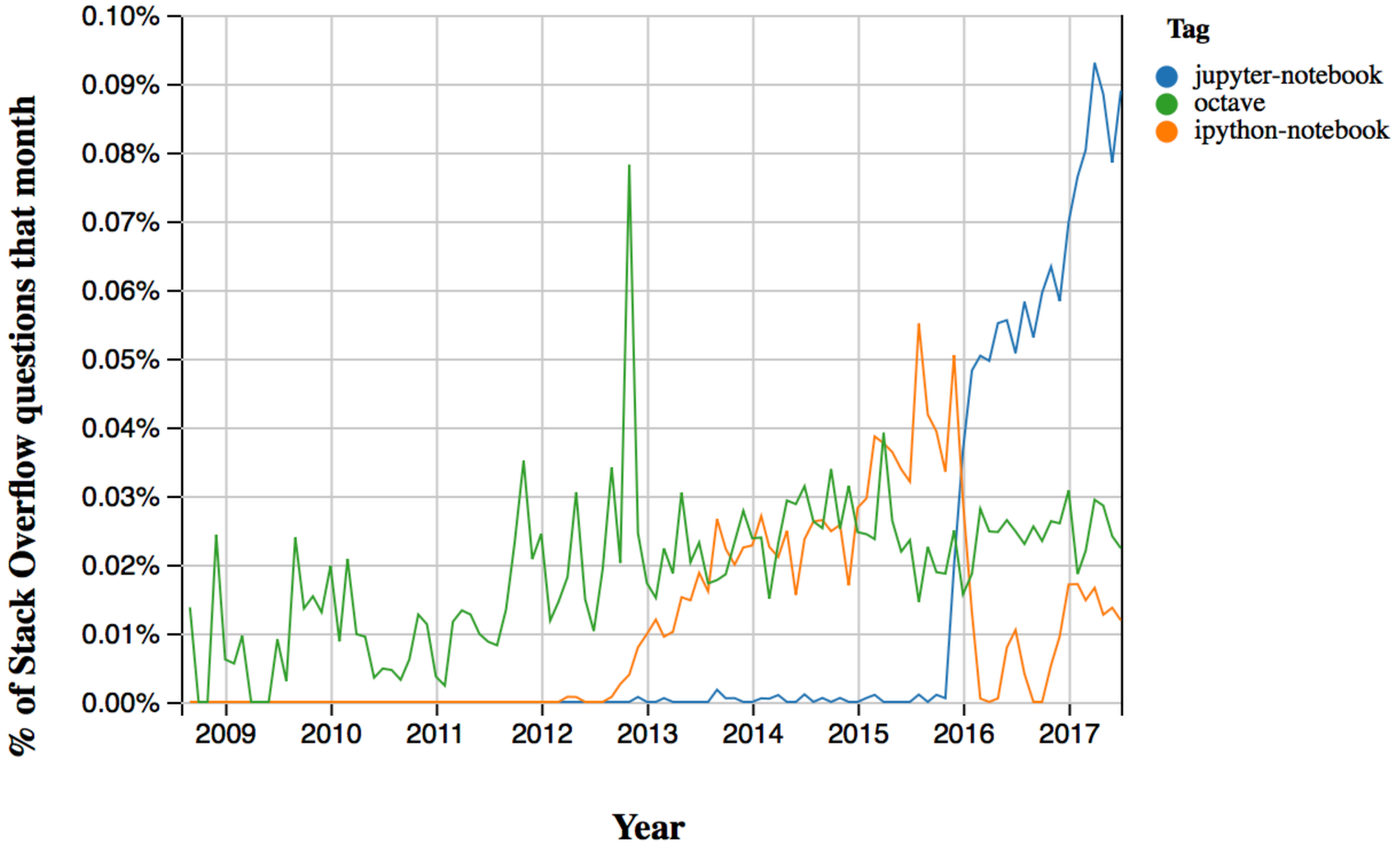
Trends

Python Google Trends In The U.S.



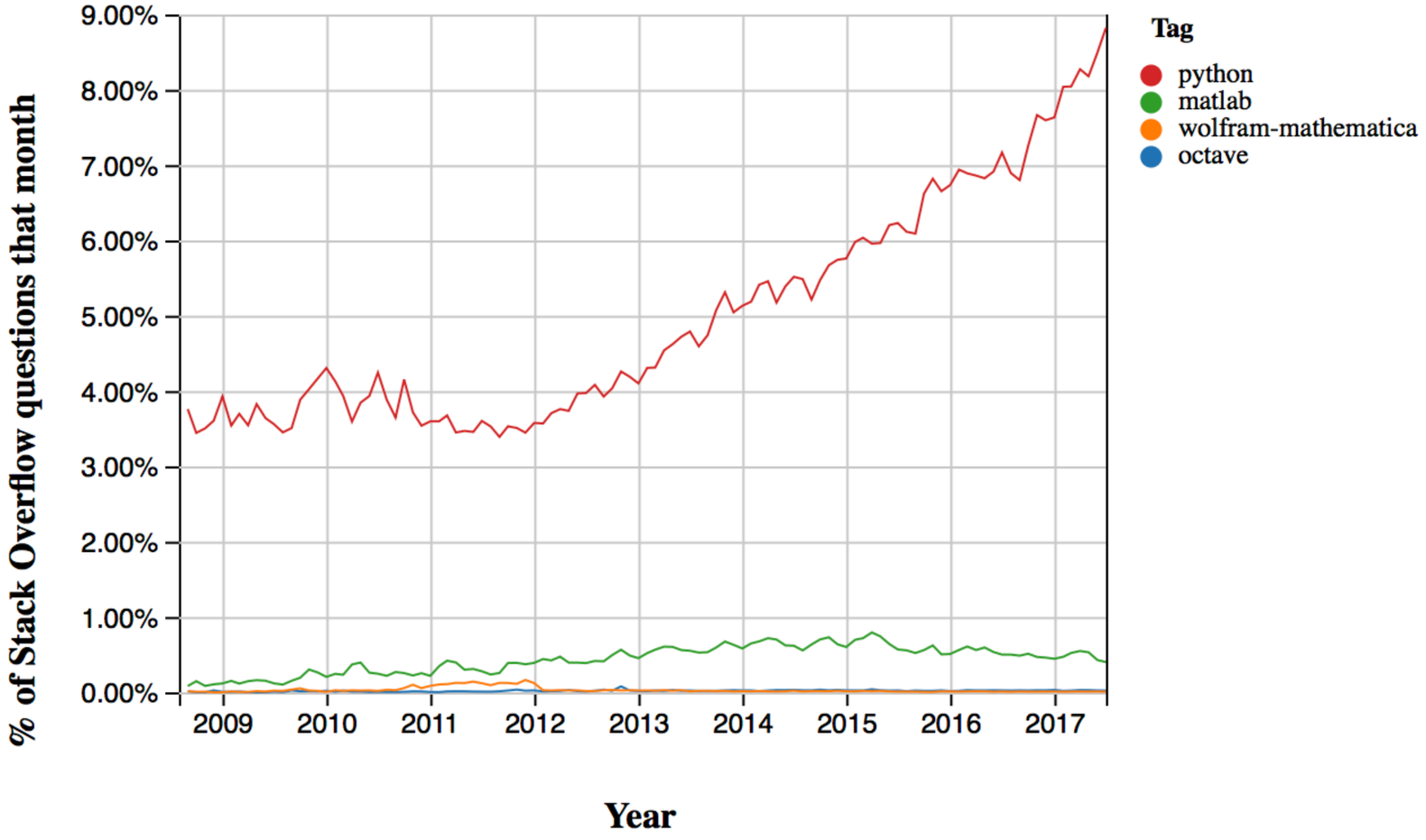
Trends

Jupyter Notebooks Comparison On Stack Overflow



Trends

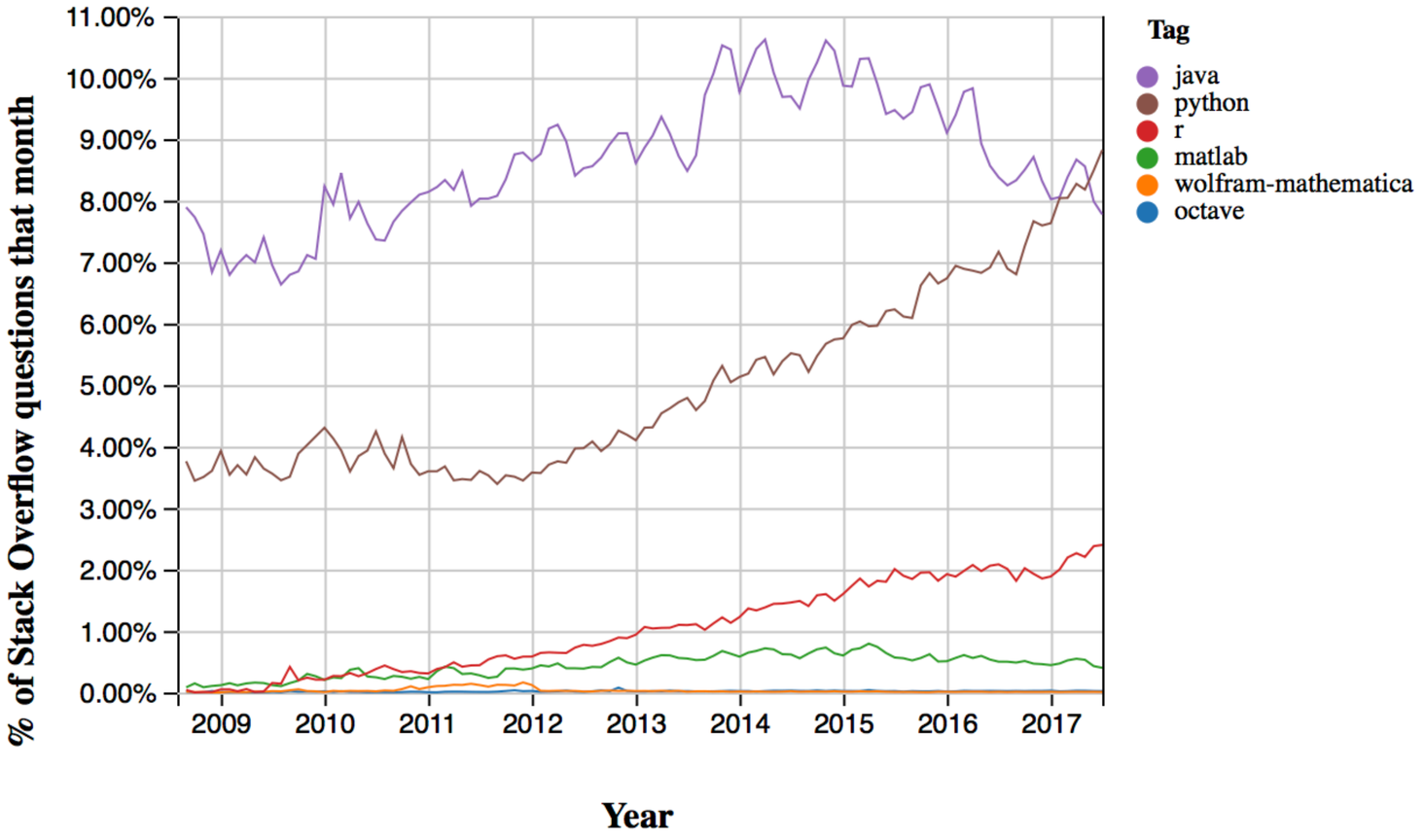
Other Scientific Languages On Stack Overflow



Python is dominant when compared to other options.

Trends

Other Languages On Stack Overflow



Additional context.

Roadmap Jupyter

- JupyterLab 1.0 by end of year.
- JupyterLab and JupyterHub continue to see development through multiple full-time employees for the foreseeable future.

Roadmap LSST DM Science Platform

- LSST DM Science Platform baselined September 2017.
- LSST DM Science Platform part of ComCam, Science Verification and Level 3.
- Version 1.0 ready to support commissioning by end of 2019.

Roadmap EPO

- LSST EPO Alpha version by January 2018.
- LSST EPO creates at least two additional investigations a year until operations begins.
- LSST EPO continues to extend and follow the LSST DM Science Platform improvements.

Concerns and Responses

Touch Devices Have Yet To See Full Support.

- Jupyter is a huge project used by large universities. We will benefit from following best practice in education.

- The science notebook platform is fundamentally a standards compliant website. This should work to our advantage.

Common Concerns and Responses

Python May Not Be Feasible.

- It is possible to create a kernel using any general purpose programming language.

- Such a change is likely to have broad support in education. Jupyter is open source and related community kernels are typically open source. These should be useable.

- Even without any external support the EPO engineering team can implement their own Kernel.

Common Concerns and Responses

Science Notebooks May Not Be Feasible.

- Static websites can be derived from science notebooks programmatically.
- Fundamentally science notebooks are a viable environment for content creators.
- There is still value even if the end product is more website than notebook.

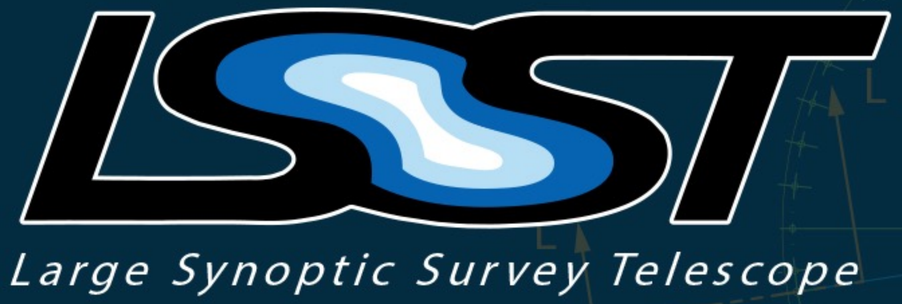
Questions?

Thanks!

Additional Slides

J. Matt Peterson, Senior Astronomy Software Developer

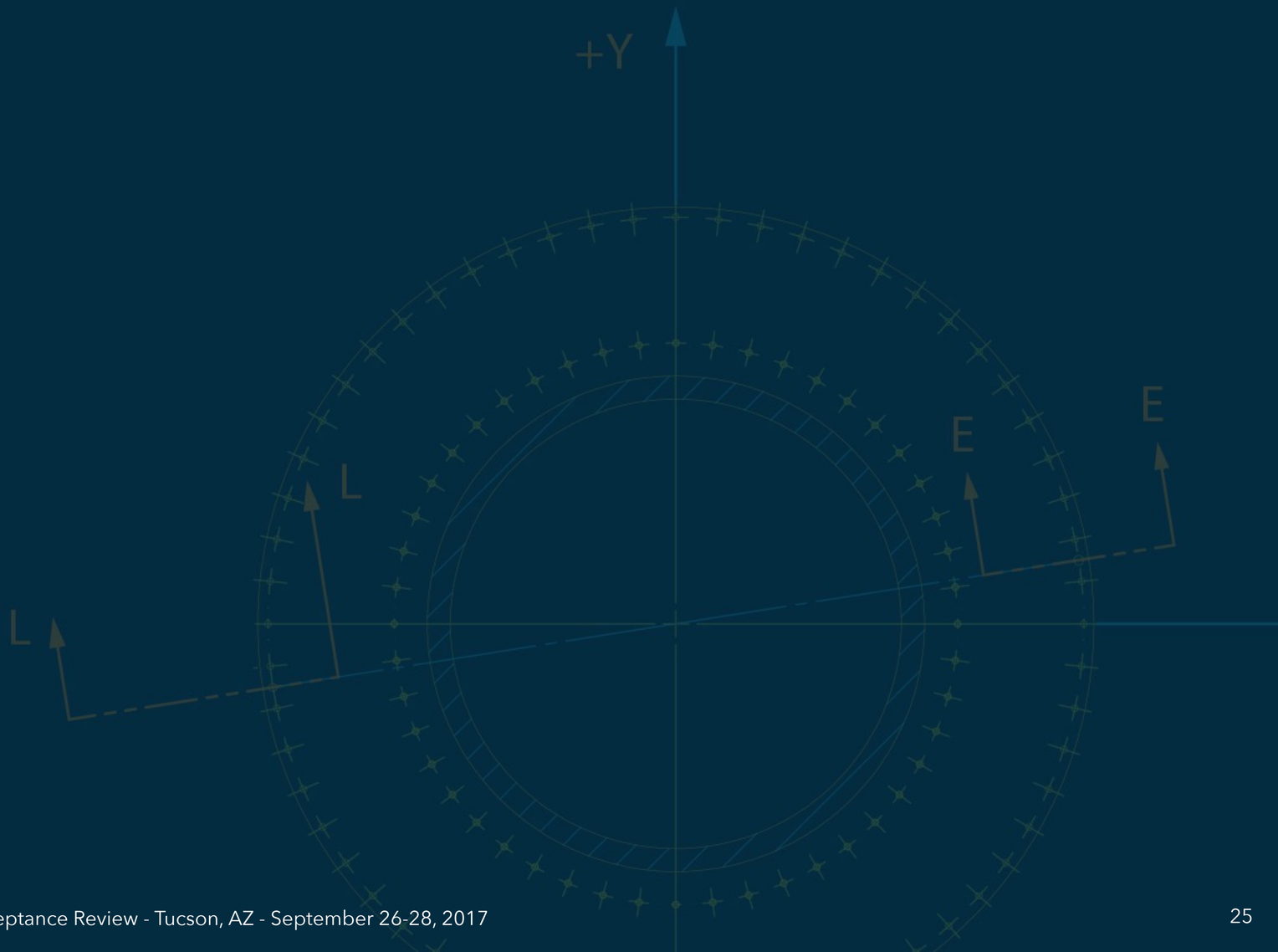
Education & Public Outreach Acceptance Review



September 26, 2017

Architecture

- Docker
- Kubernetes
- Hybrid cloud



Docker

Isolates And Unifies The Computing Environment.

- Docker is a specific container technology.
- "A **container** is an instance of an application that isolates its view of the operating system. It shares many of the advantages of Virtual Machines without the overheads and can be well orchestrated for scalability." DM Science Platform All Hands Meeting
- "A way to specify a computing environment to use with your software." Andrew Odewahn, CTO O'Reilly Media

Kubernetes

Manages Containers.

- Docker is a specific container technology.
- "**Kubernetes** is an open-source system for automating deployment, scaling, and management of containerized applications. It groups containers that make up an application into logical units for easy management and discovery." Kubernetes homepage
- "A way to specify a computing environment to use with your software." Andrew Odewahn, CTO O'Reilly Media

Hybrid Cloud

Benefits From Both Approaches.

- Use physical hardware for our most expensive compute resources. Such as storage.
- Use cloud services to scale, test and research. Such as caches, extending the Science Notebook platform temporarily. Performing large, one time tasks.
- Automate deployments so we can target available pro bono resources.

Resources

- http://jupyter.org/
- Jupyter Grant Details
- https://www.python.org/
- http://data8.org/
- https://www.oreilly.com/ideas/highlights-from-jupytercon-ny-2017