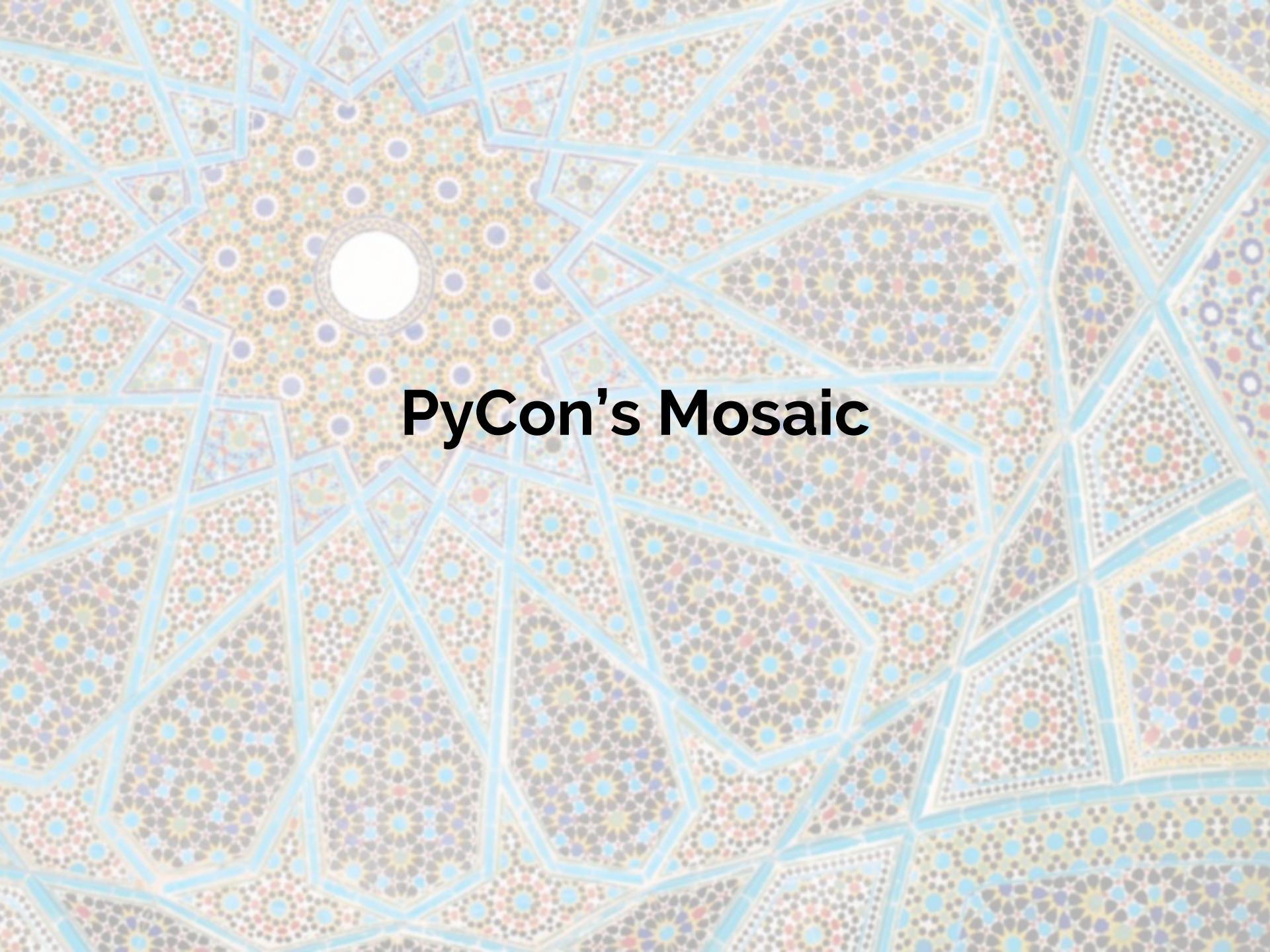




The Unexpected Effectiveness of Python in Science

Jake VanderPlas @jakevdp
PyCon 2017



PyCon's Mosaic

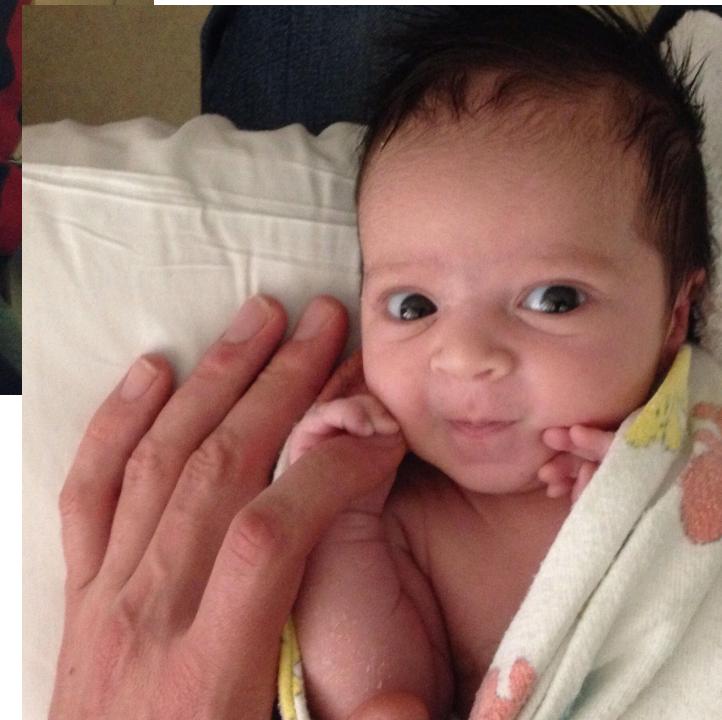


```
$ whoami  
jakevdp
```





\$ whoami
jakevdp





\$ whoami
jakevdp



Blog:

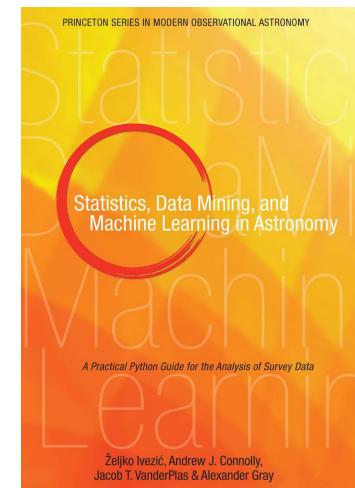
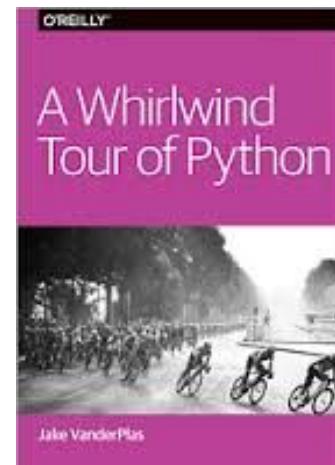
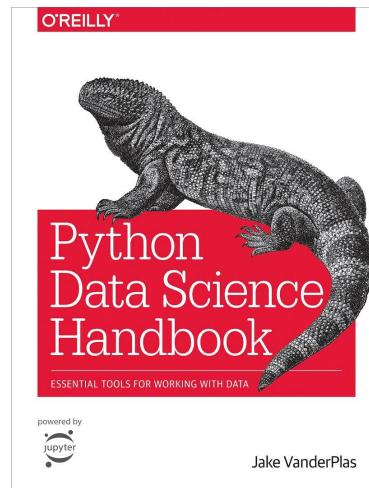


<http://jakevdp.github.io>

Code:



Books:





\$ whoami
jakevdp





\$ whoami
jakevdp

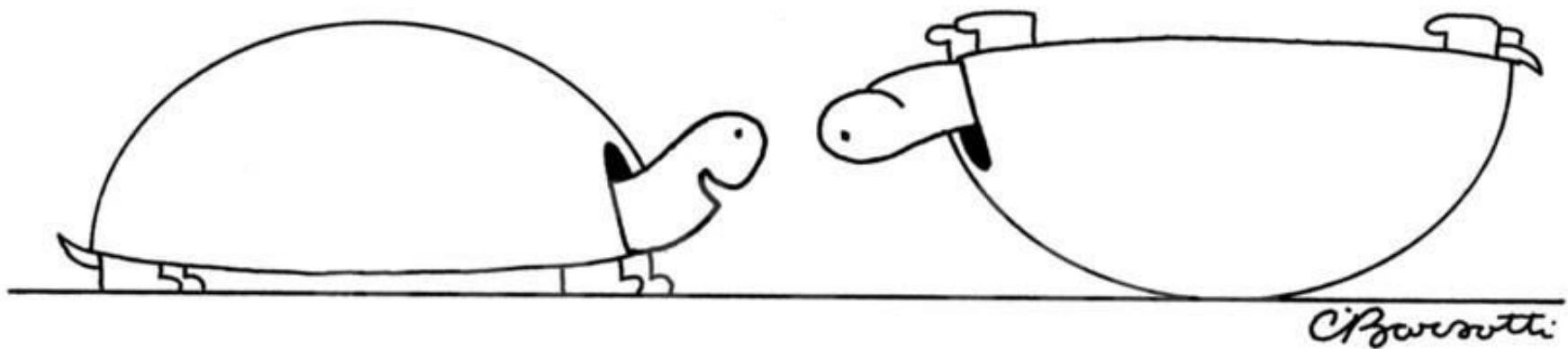


ALFRED P. SLOAN
FOUNDATION





\$ whoami
jakevdp



“Wow, I’ve never met an astronomer before.”

Charles Barsotti, New Yorker

Astronomy

Then . . .



Edwin Hubble at the 48" Schmidt Telescope,
Palomar Observatory, 1949. (credit: PNAS)

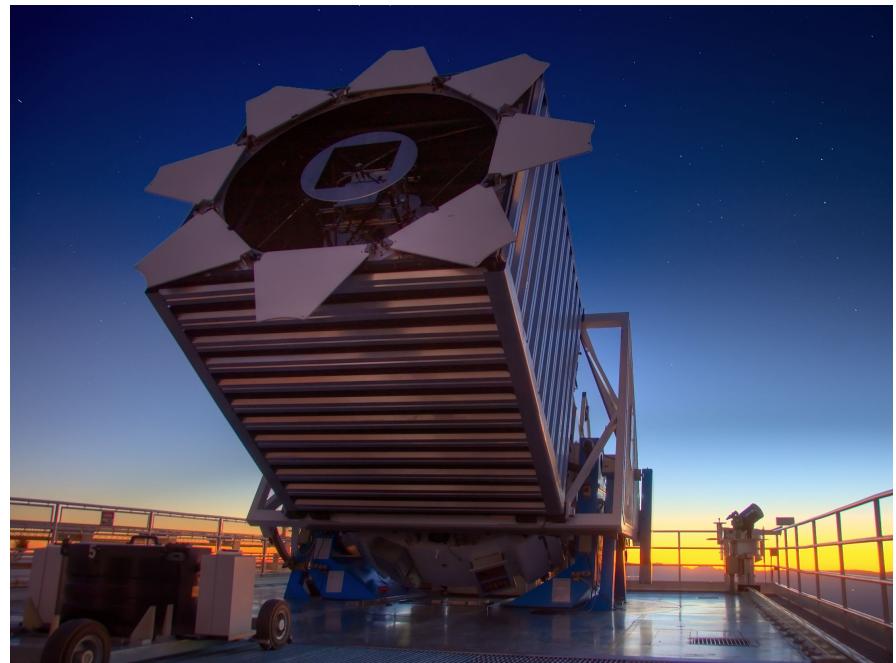
Astronomy Now . . .

Hubble Space Telescope



Source: <http://spacetelescope.org/>

Sloan Digital Sky Survey

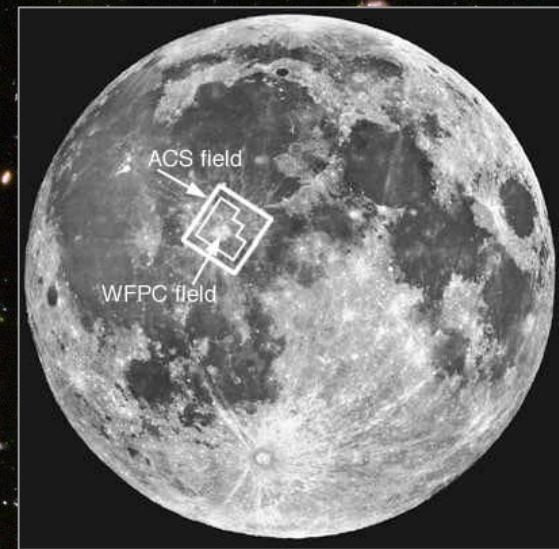


Source: <http://sdss.org/>

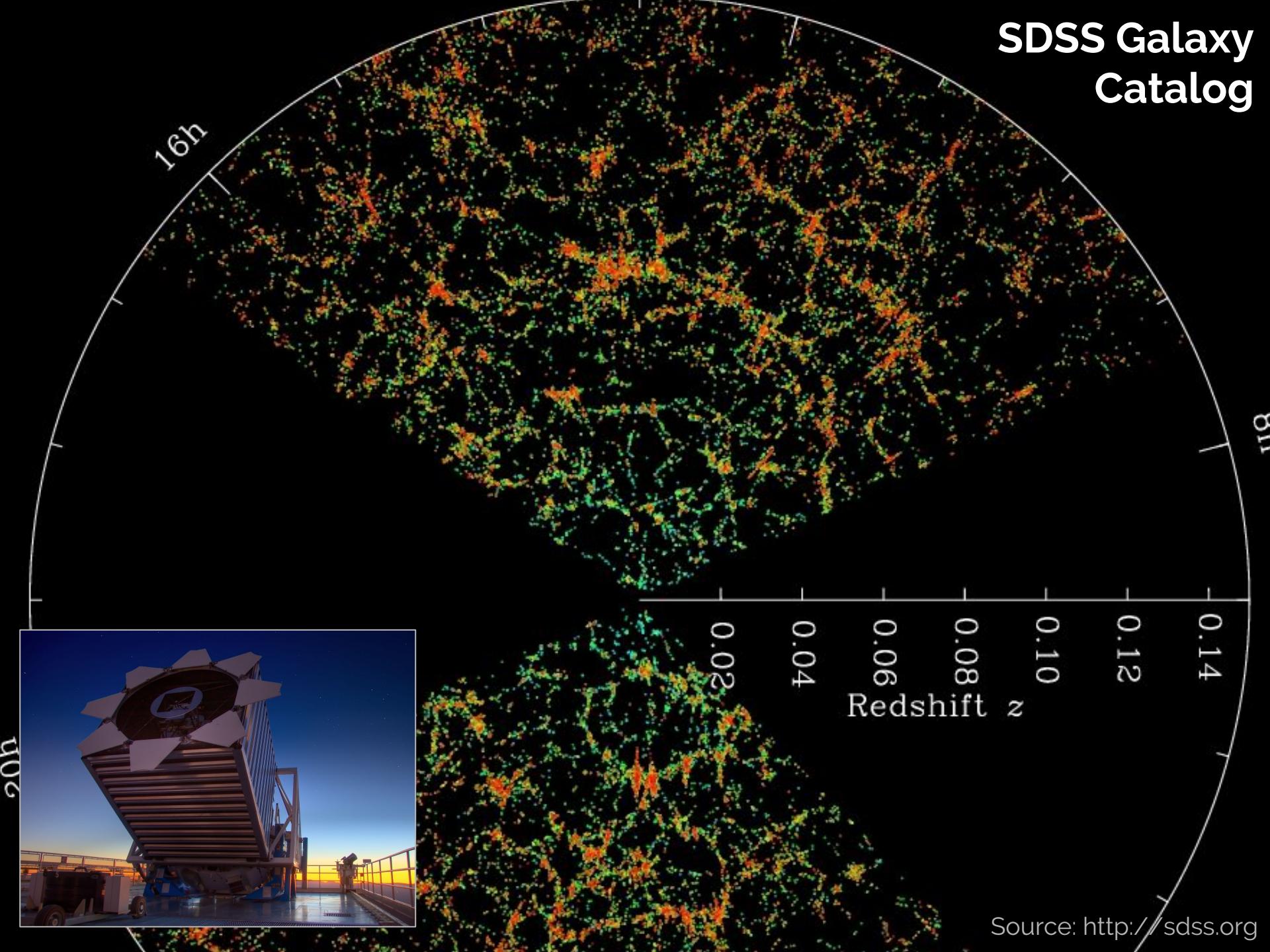
Hubble's "Ultra Deep Field"



Hubble's "Ultra Deep Field"



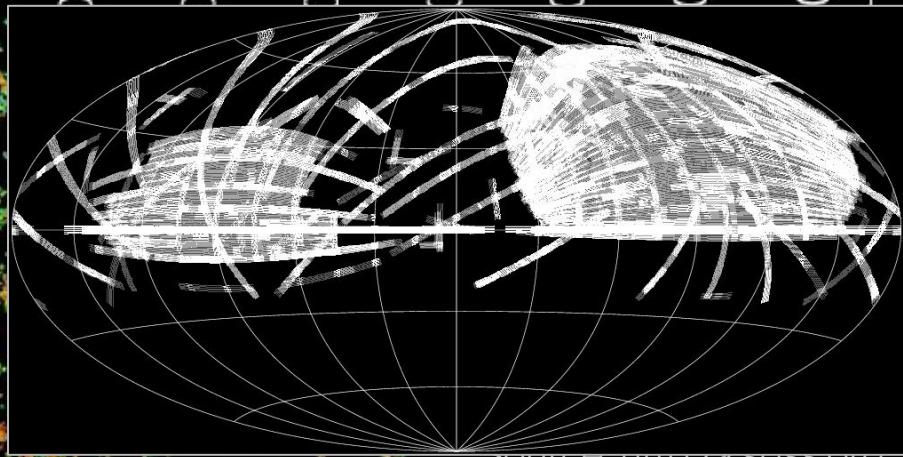
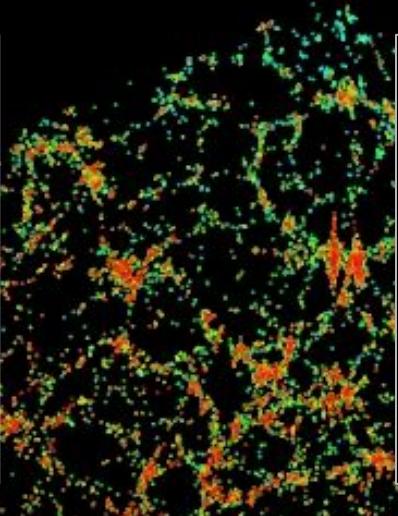
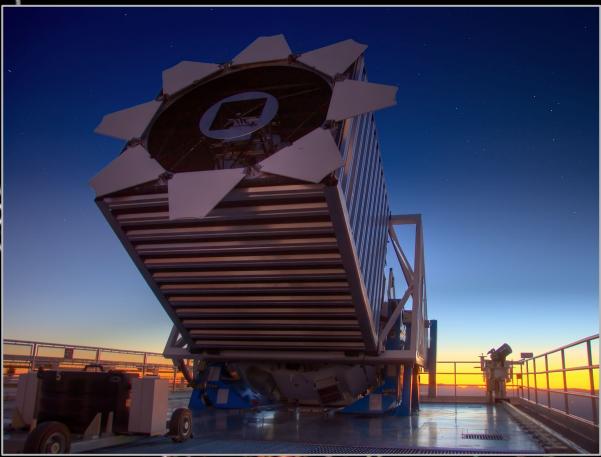
SDSS Galaxy Catalog



SDSS Galaxy Catalog

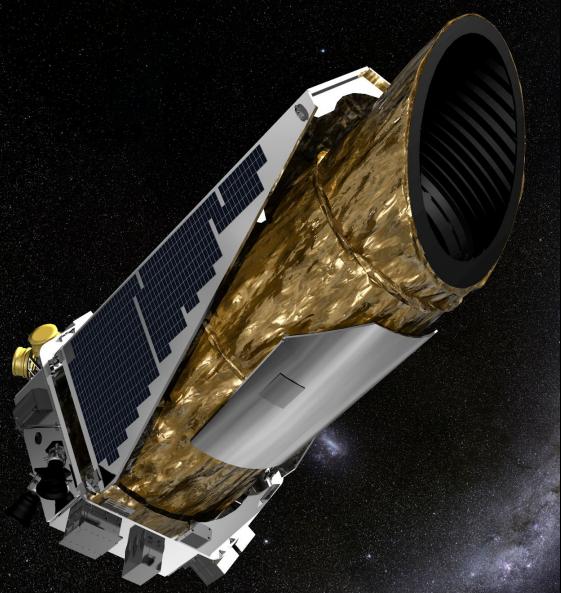
16h

118

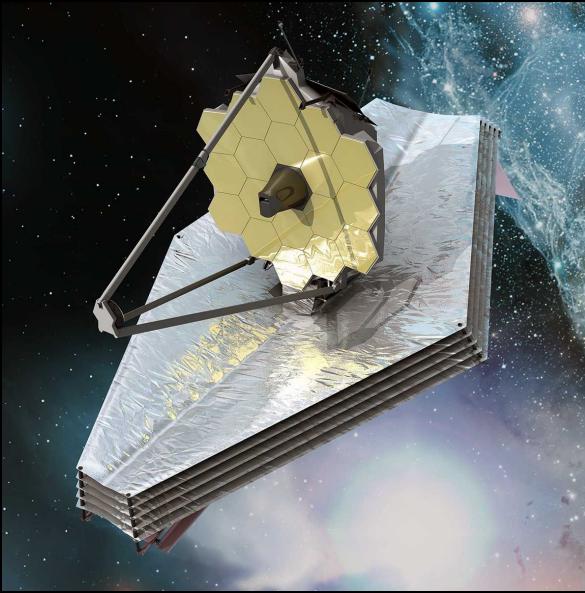


Source: <http://sdss.org>

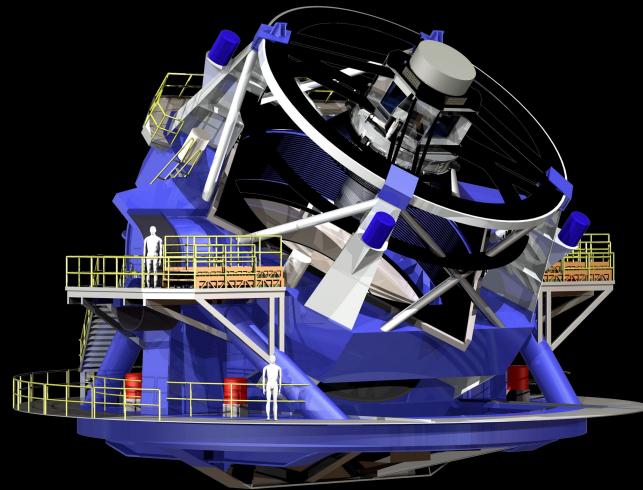
Astronomy in the 21st Century . . .



Kepler (2009)

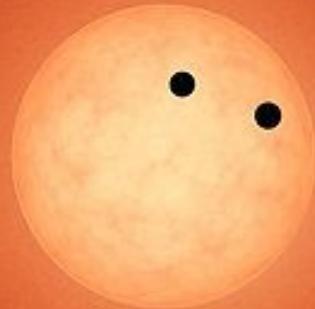


JWST (2018)

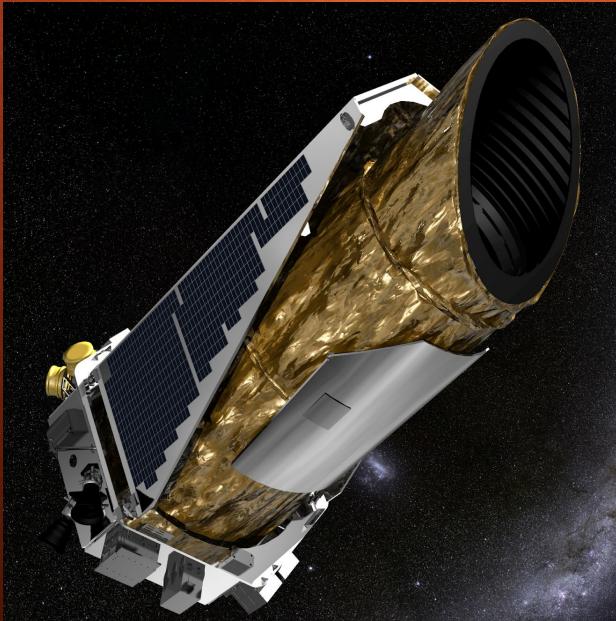
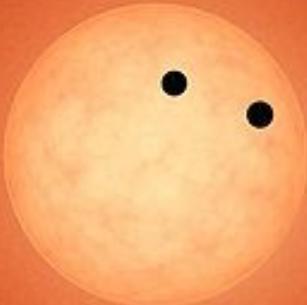


LSST (2020)

TRAPPIST-1 Exoplanetary System

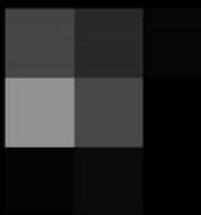


TRAPPIST-1 Exoplanetary System Kepler (K2) Observations



K2 View of TRAPPIST-1

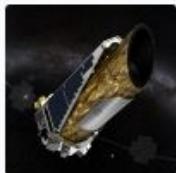
15 Dec 2016 20:55



[GitHub, Inc. \[US\]](#) <https://github.com/KeplerGO>

This organization Search Pull requests Issues Gist

Notifications + 



Kepler/K2 Guest Observer Office

<http://keplerscience.arc.nasa.gov>

KeplerGO@mail.arc.nasa.gov

 [Repositories](#)

 [People 0](#)

Search repositories...

Type: All ▾

Language: All ▾

PyKE

A suite of Python tools to analyze Kepler/K2 data

[kepler](#) [k2](#)

 Python  7  9 Updated 8 minutes ago



Top languages

-  Python
-  CSS
-  Makefile
-  HTML

KeplerScienceWebsite

Website for the Kepler/K2 Science Center.



People

0 >

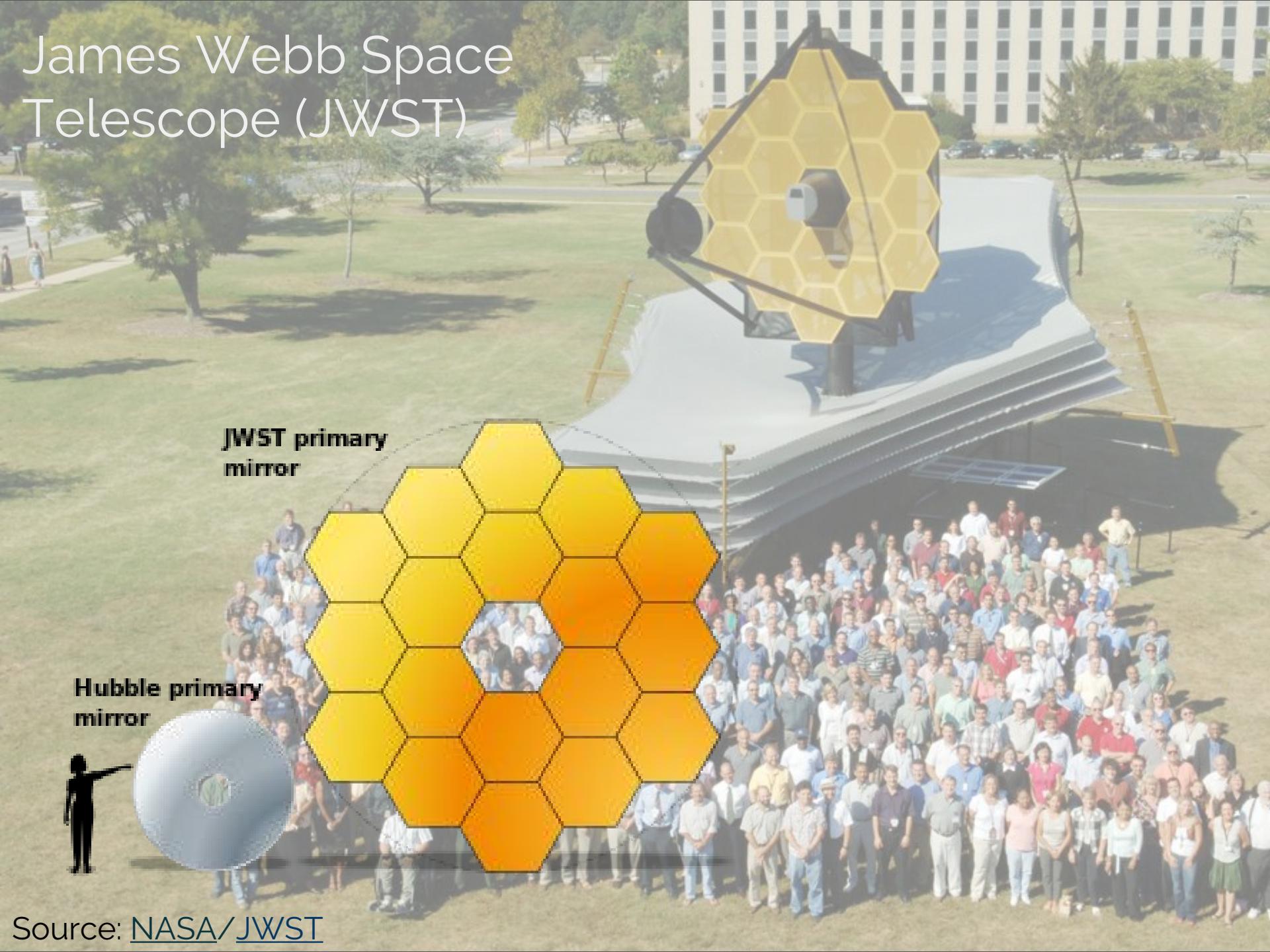
This organization has no public members. You must be a member to see who's a part of this organization.

James Webb Space Telescope (JWST)

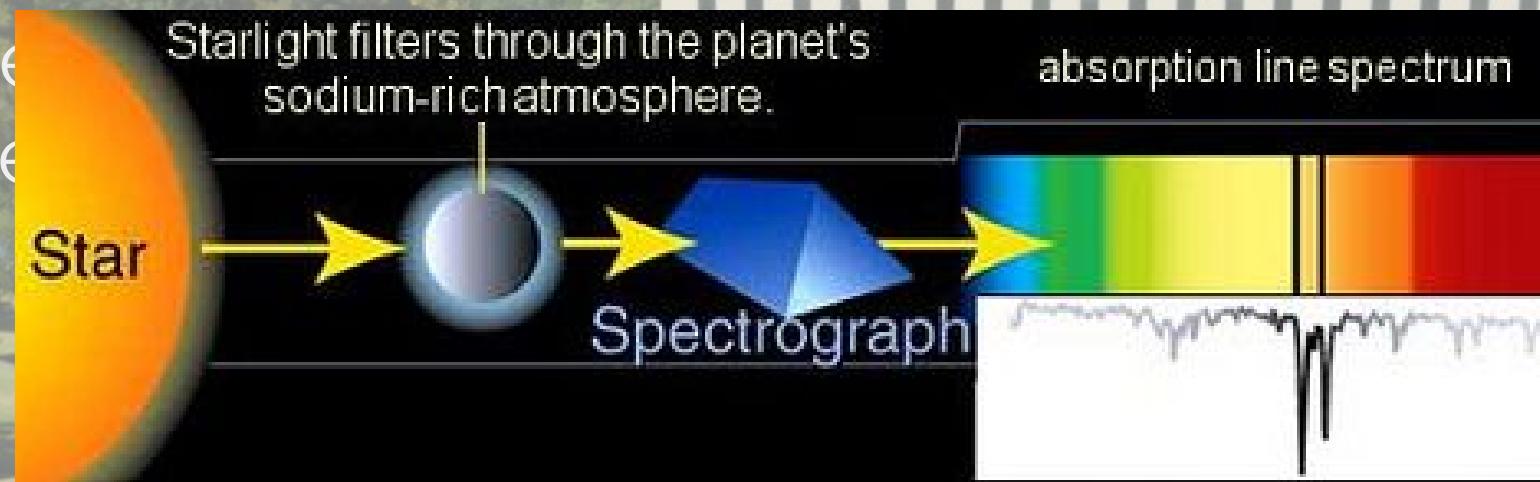


Source: [NASA](#)

James Webb Space Telescope (JWST)



James Webb Telescope





STScI-JWST 

Repositories

 People 2

Search repositories...

Type: All ▾

Language: All ▾

jwst

● Jupyter Notebook

wss_tools

Python tools for JWST Wavefront Sensing Software

[visualization](#) [astronomy](#) [jwst](#) [ginga](#) [wavefront-sensing](#)

Python 3 Updated on Mar 20



Top languages

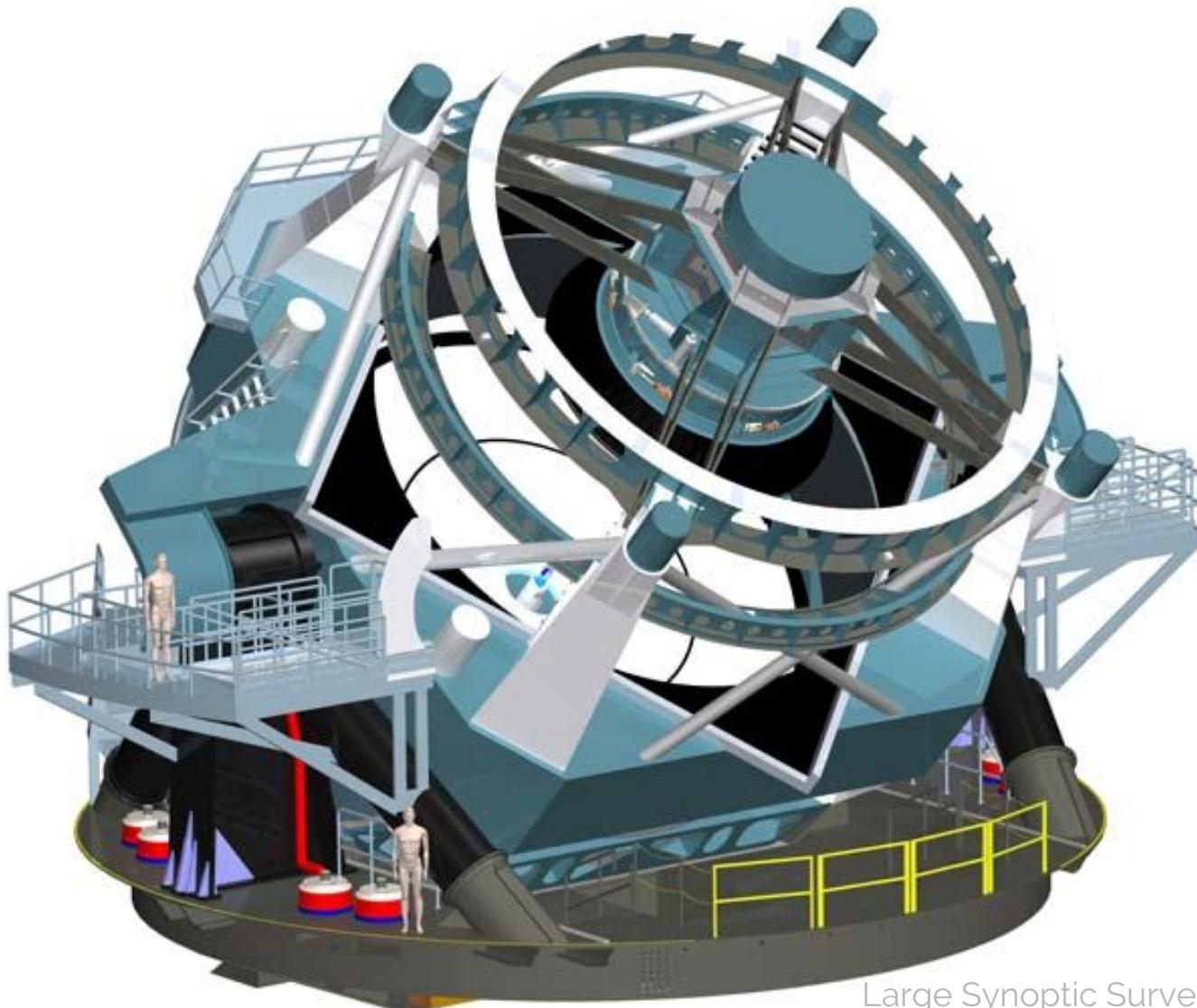
● Python ● Jupyter Notebook

People

 larrybradley
Larry Bradley

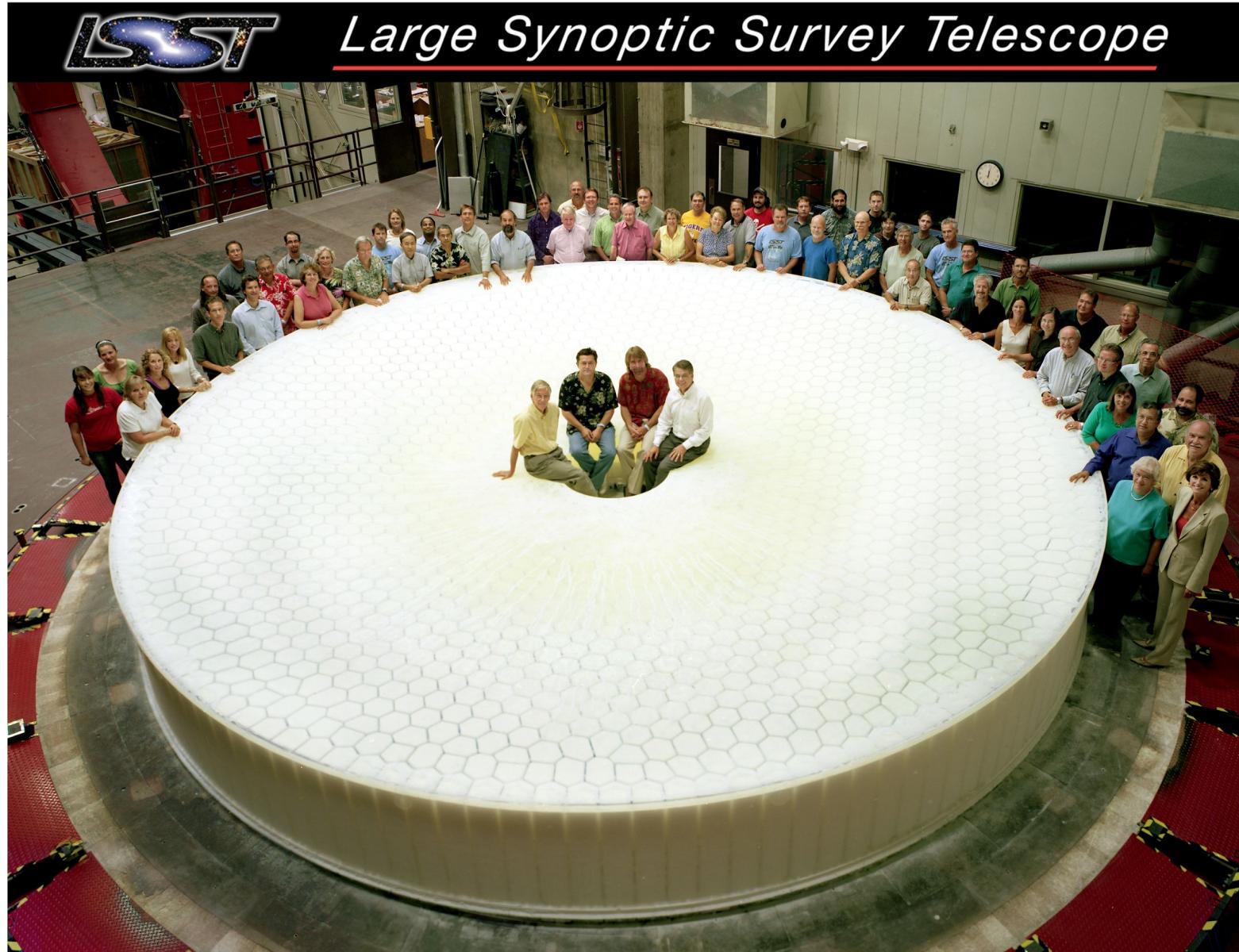
s0sev

Large Synoptic Survey Telescope

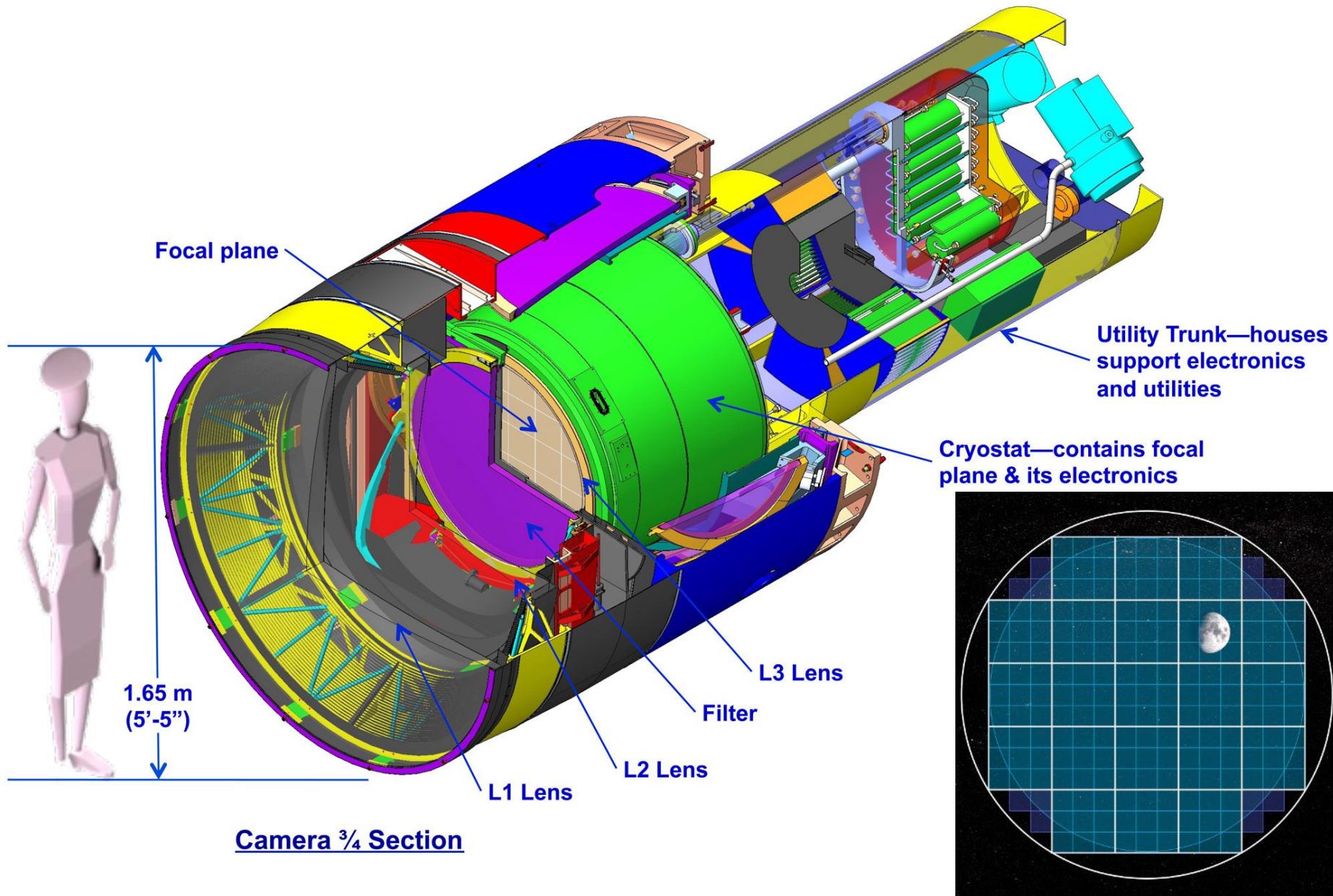


Large Synoptic Survey Telescope
(credit: LSST Corp)

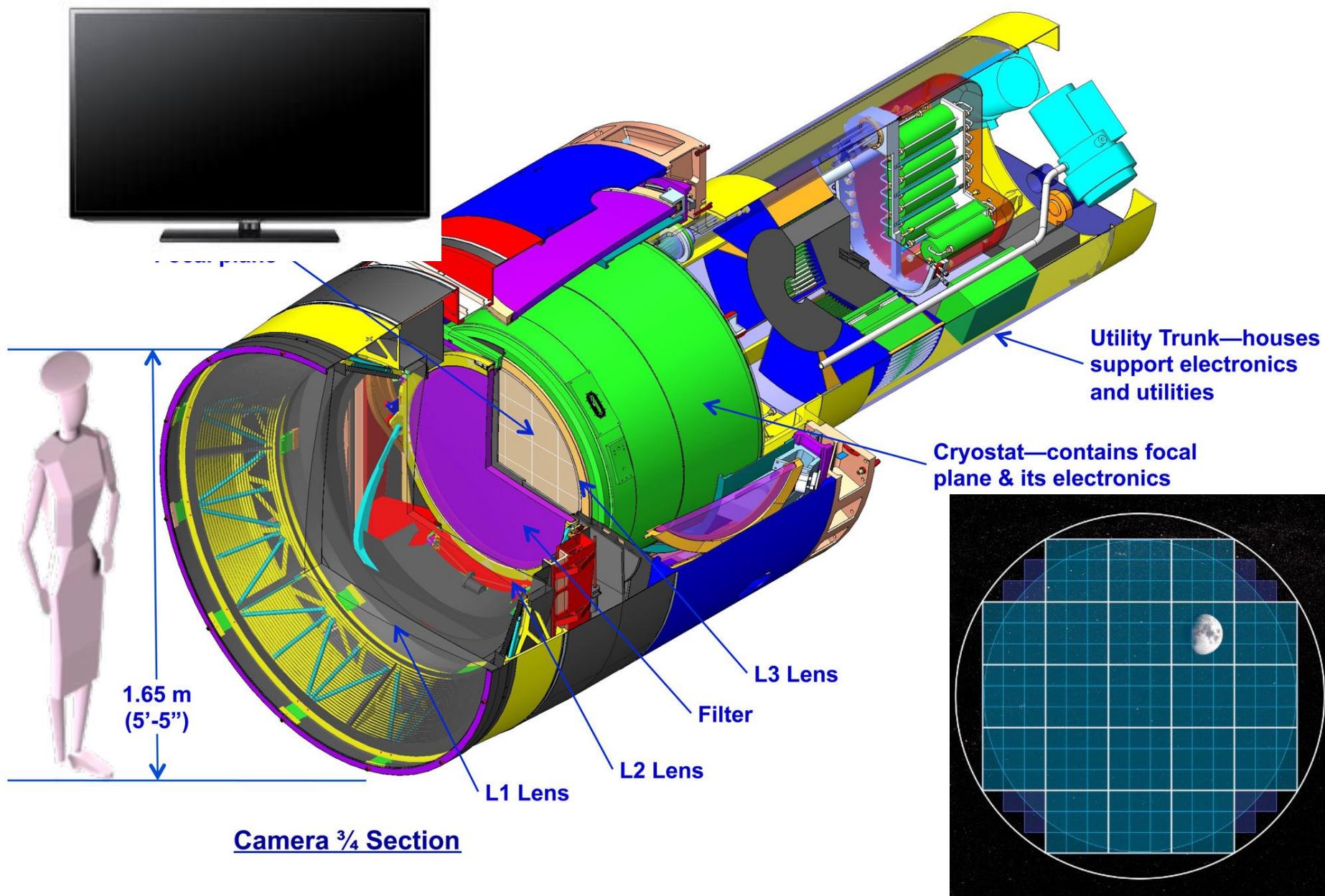
8.4-meter Primary Mirror



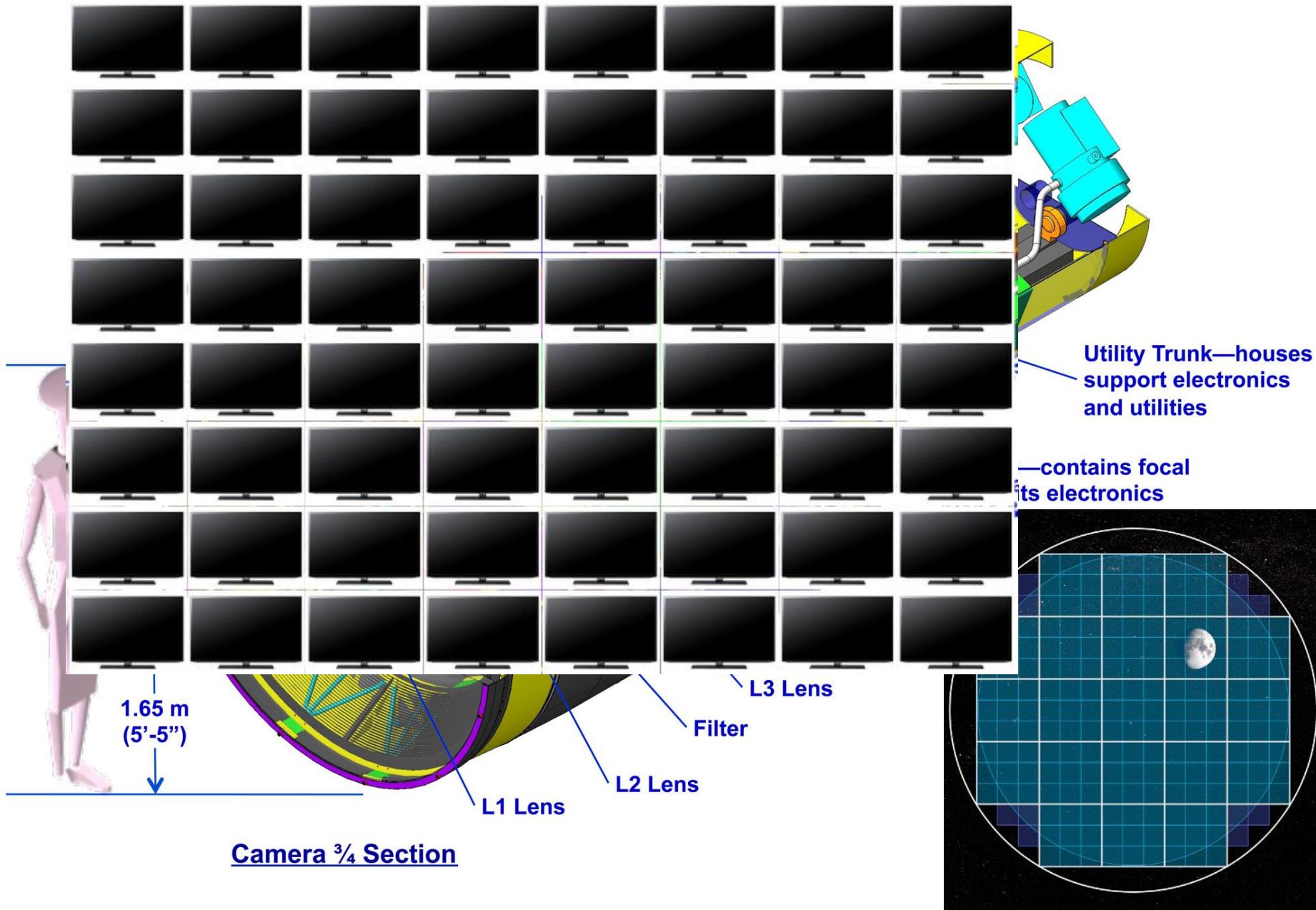
3 Gigapixel Camera



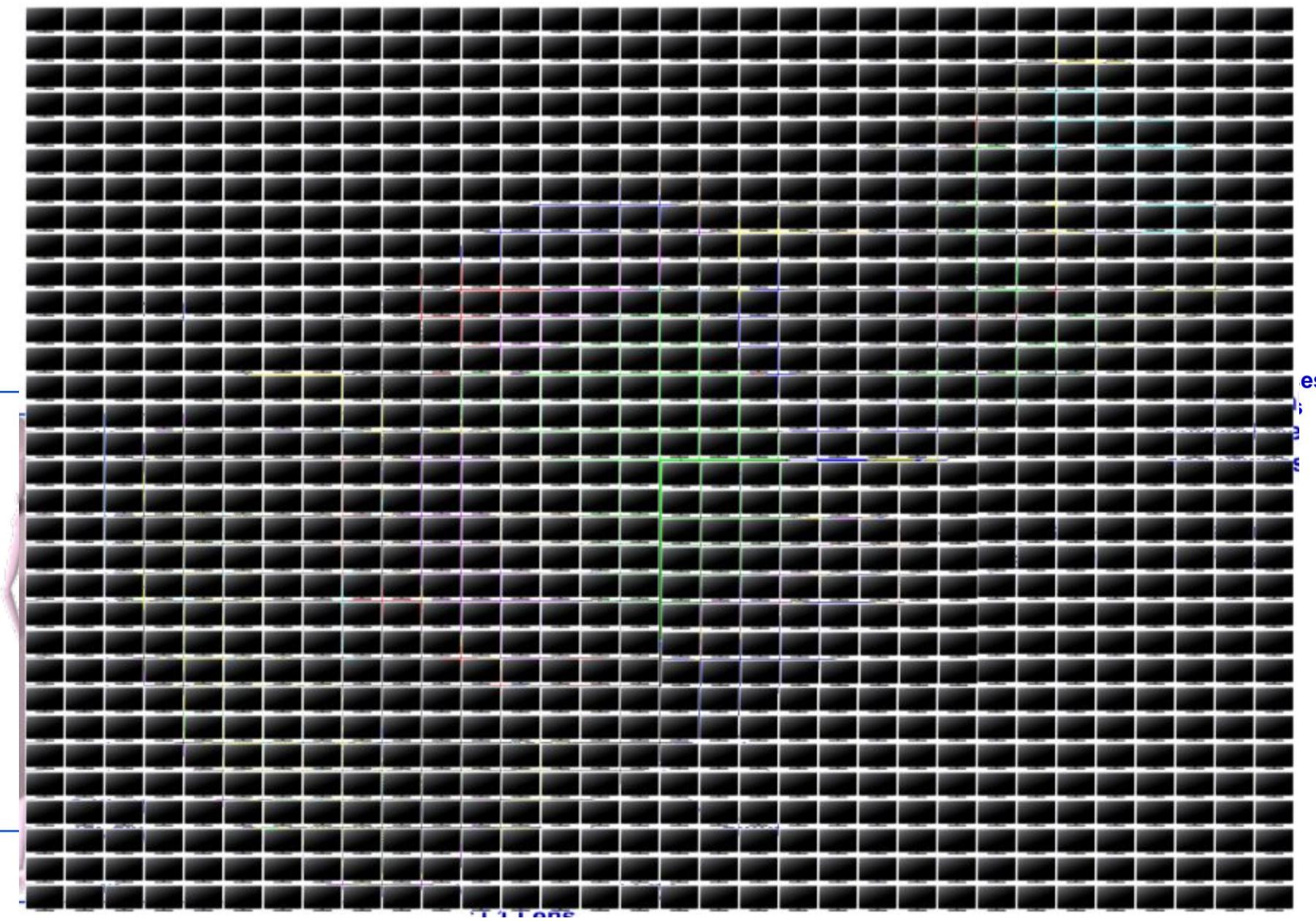
3 Gigapixel Camera

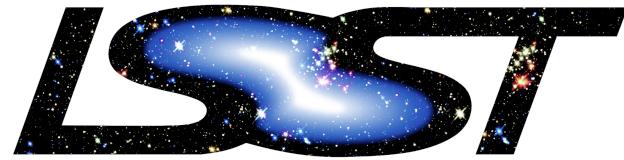


3 Gigapixel Camera



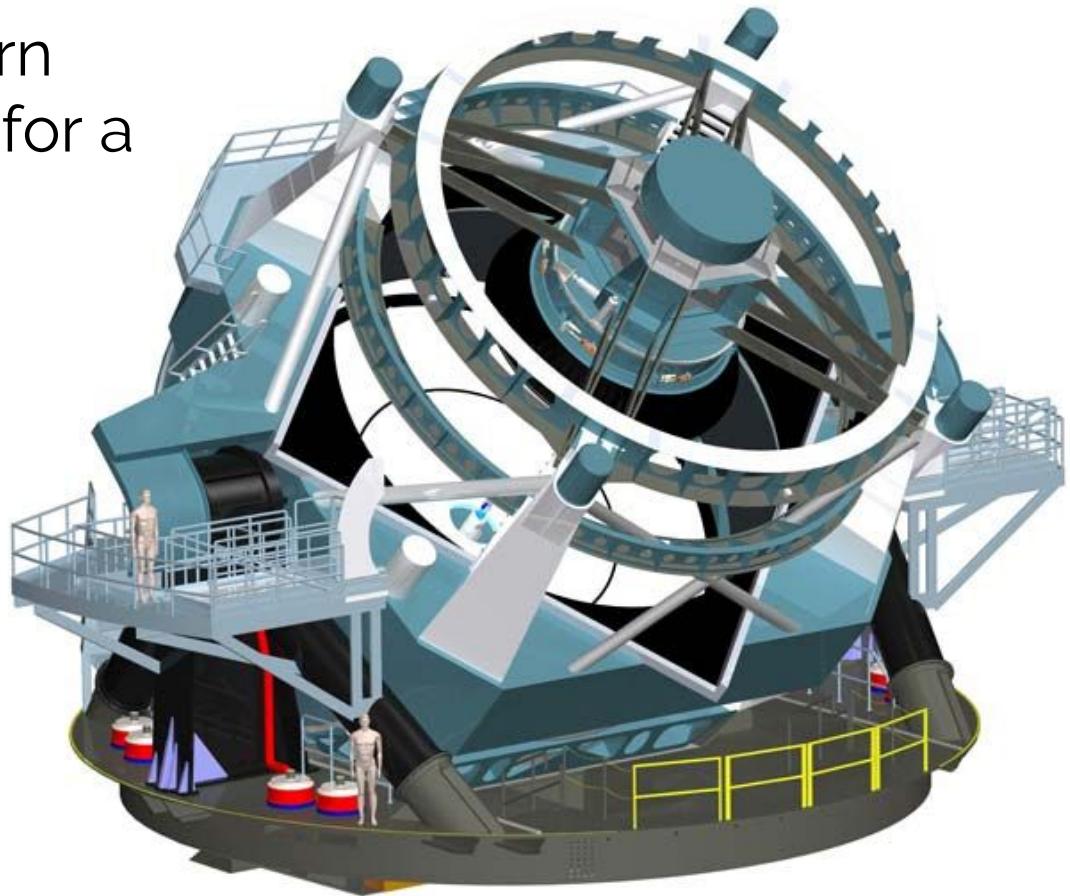
3 Gigapixel Camera = ~1500 HD TVs

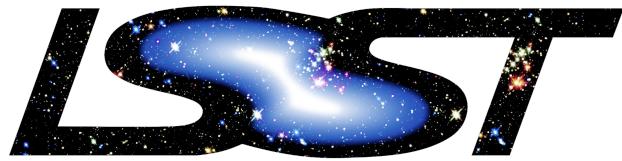




Large Synoptic Survey Telescope

- Survey mode: 2 exposures every ~30 seconds
- Images the full southern sky every three nights for a decade
- 15-30 TB/night!
- Final 10-year catalog: 100s of Petabytes

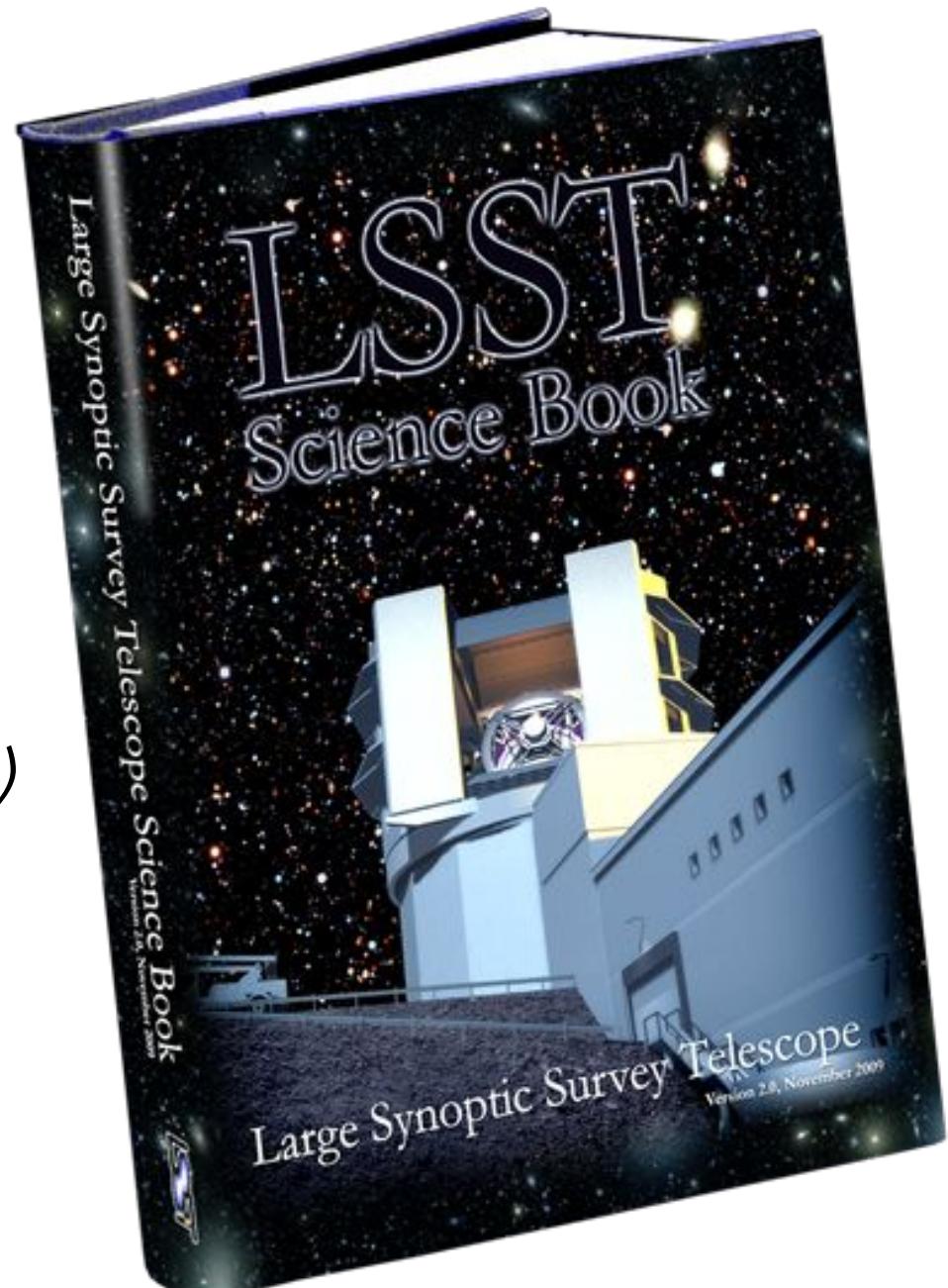




Large Synoptic Survey Telescope

What will we do with all this data?

*(Left as a 616-page
exercise for the reader)*



GitHub, Inc. [US] https://github.com/LSST

This organization Search Pull requests Issues Gist

Notifications + 



LSST



Large Synoptic Survey Telescope - Astronomy that's Wider, Faster, Deeper

Tucson, AZ

<http://lsst.org>

Repositories

People 18

Search repositories...

Type: All ▾

Language: All ▾

afw

LSST data management: astronomical framework, including images and tables

python c-plus-plus astronomy

● C++ ★ 5 ⚡ 8 Updated an hour ago



Top languages

- Python
- Shell
- C++
- TeX
- Makefile

pipe_tasks

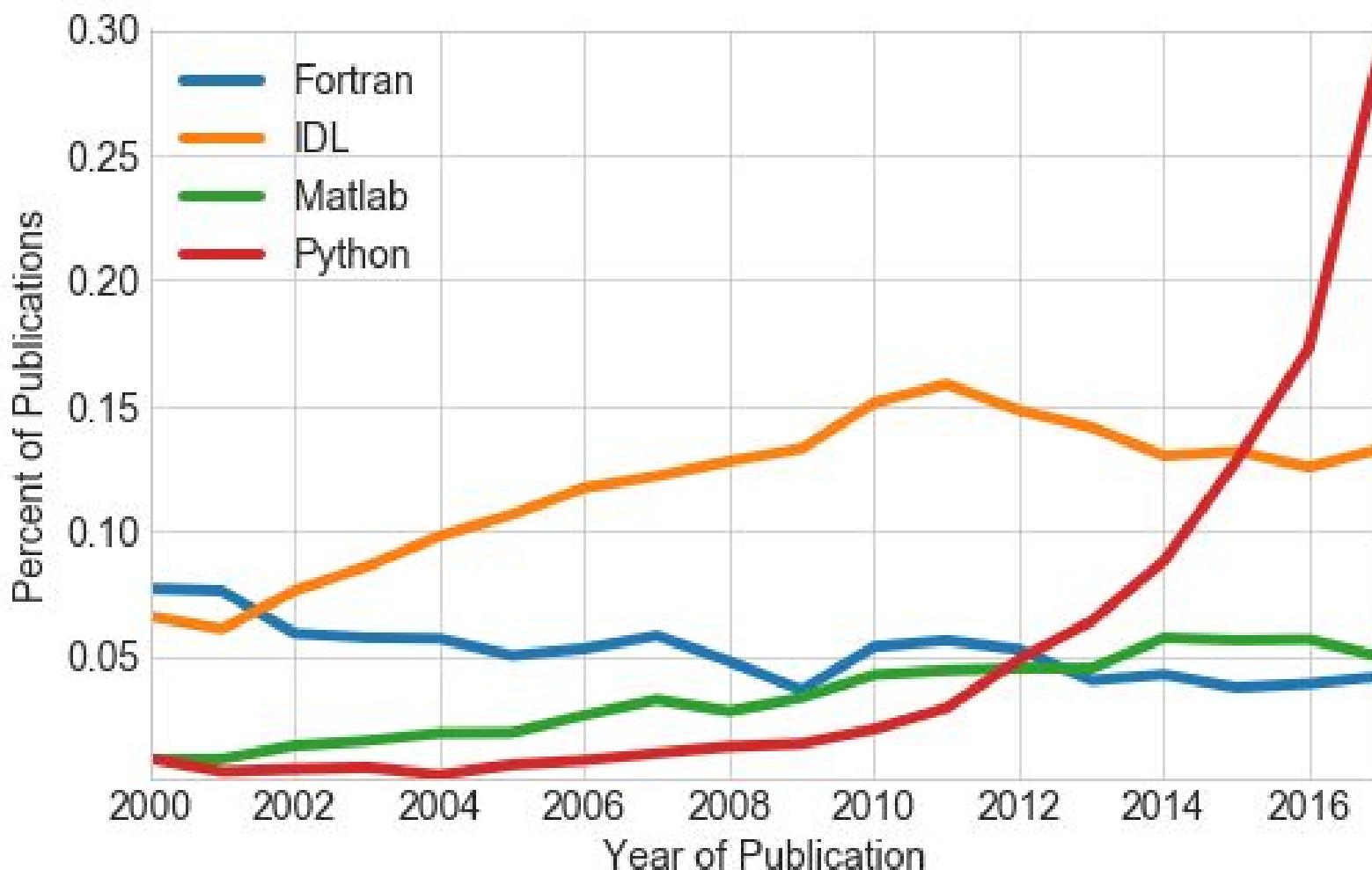
LSST Data Management: astronomical data processing tasks

People

18 >



Mentions of Software in Astronomy Publications:



Thanks to Juan Nunez-Iglesias,
Thomas P. Robitaille, and Chris Beaumont.



The Unexpected Effectiveness of Python in Science

But Why Python?

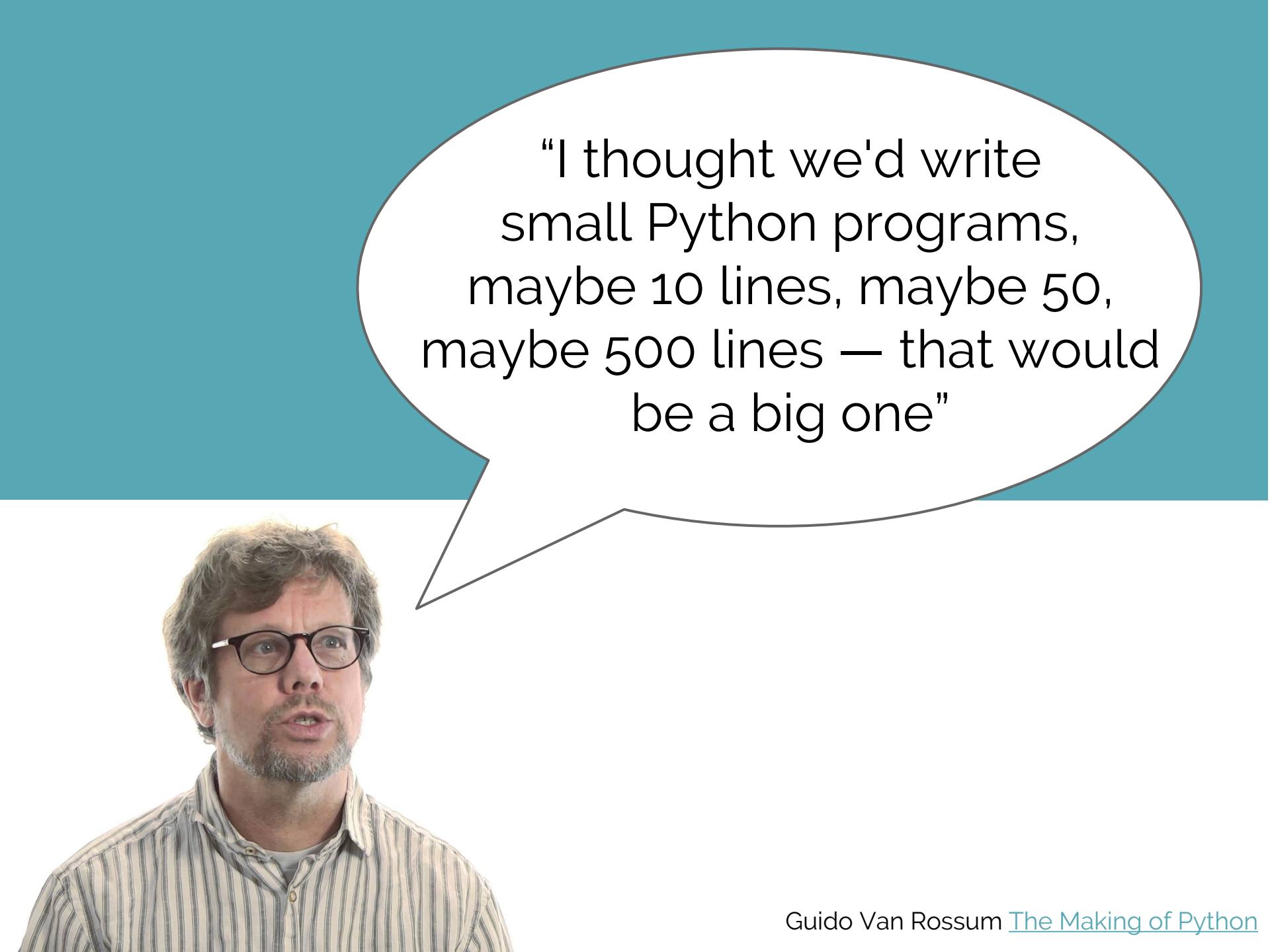
Python



Python is a “teaching language”

... created to “bridge the gap between the shell and C”

“never intended . . . to be the primary language for programmers.”



“I thought we'd write small Python programs, maybe 10 lines, maybe 50, maybe 500 lines – that would be a big one”

Why is Python such an effective tool in science?

Why is Python such an effective tool in science?

1. Interoperability with Other Languages

A portrait painting of Isaac Newton, an English polymath and a key figure in the scientific revolution. He is shown from the chest up, wearing a dark robe over a white shirt with a ruffled collar. His long, wavy hair is powdered white. He has a serious, thoughtful expression and is looking slightly to his left.

“If I have seen further, it is by standing on the shoulders of giants.”

- *Isaac Newton*

A portrait painting of Isaac Newton, an English polymath and a key figure in the scientific revolution. He is shown from the chest up, wearing a dark robe over a white shirt with a ruffled collar. His long, wavy hair is powdered white. He has a serious expression and is looking slightly to his left.

“If I have seen further, it is by
importing from the code of
giants.”

- *Definitely Not Isaac Newton*

Science Before Python . . .

“Scientists... work with a wide variety of systems ranging from simulation codes, data analysis packages, databases, visualization tools, and home-grown software-each of which presents the user with a different set of interfaces and file formats. As a result, a scientist may spend a considerable amount of time simply trying to get all of these components to work together in some manner...”



- **David Beazley**
Pythonista Extraordinaire
Scientific Computing with Python
(ACM vol. 216, 2000)

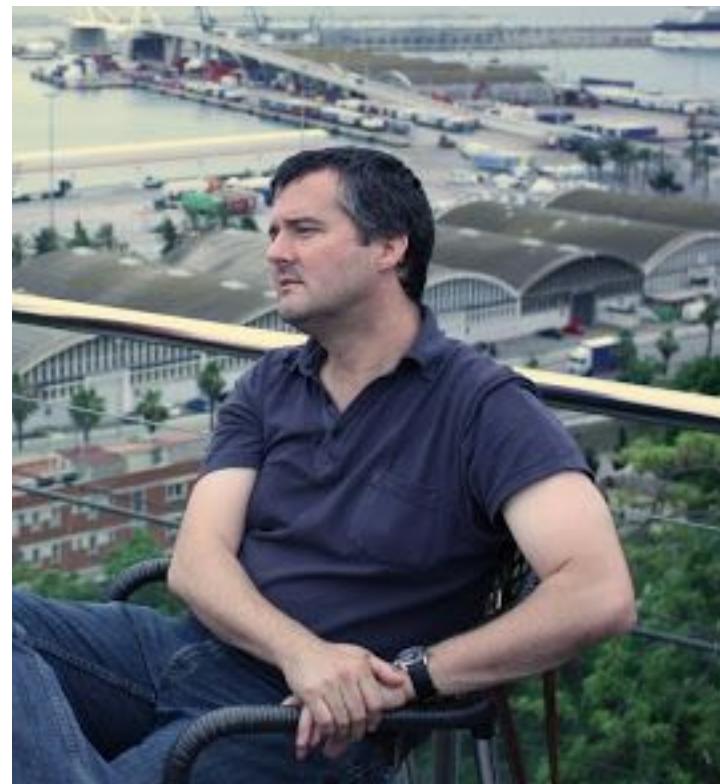
Science Before Python . . .

“I had a hodge-podge of work processes. I would have Perl scripts that called C++ numerical routines that would dump data files, and I would load them up into MatLab to plot them. After a while I got tired of the MatLab dependency . . . so I started loading them up in GnuPlot.”

-John Hunter

creator of Matplotlib

SciPy 2012 Keynote



Science Before Python . . .

“My advisor had a heavily customized awk/sed/bash workflow to manage job submissions and postprocessing of C codes for supercomputing runs... So I used her scripts to run my jobs, and on top of that had added my own layer of Perl, plus a hefty amount of Gnuplot, IDL and Mathematica.”

- **Fernando Perez**
creator of IPython
via email



Python is Glue.



Python glues together this hodge-podge of scientific tools.

High-level syntax wraps low-level C/Fortran libraries, which is (mostly) where the computation happens.

Python is Glue.



Why is Python such an effective tool in science?

1. Interoperability with Other Languages
2. “Batteries Included” + Third-Party Modules

Python has built-in libraries
for *nearly* everything . . .

. . . and there are third-party
libraries for everything else.

The Genesis of Scientific Python

“Prior to Python, I used Perl (for a year) and then Matlab and shell scripts & Fortran & C/C++ libraries. When I discovered Python, I really liked the language... But, it was very nascent and lacked a lot of libraries. I felt like I could add value to the world by connecting low-level libraries to high-level usage in Python.”

- **Travis Oliphant**

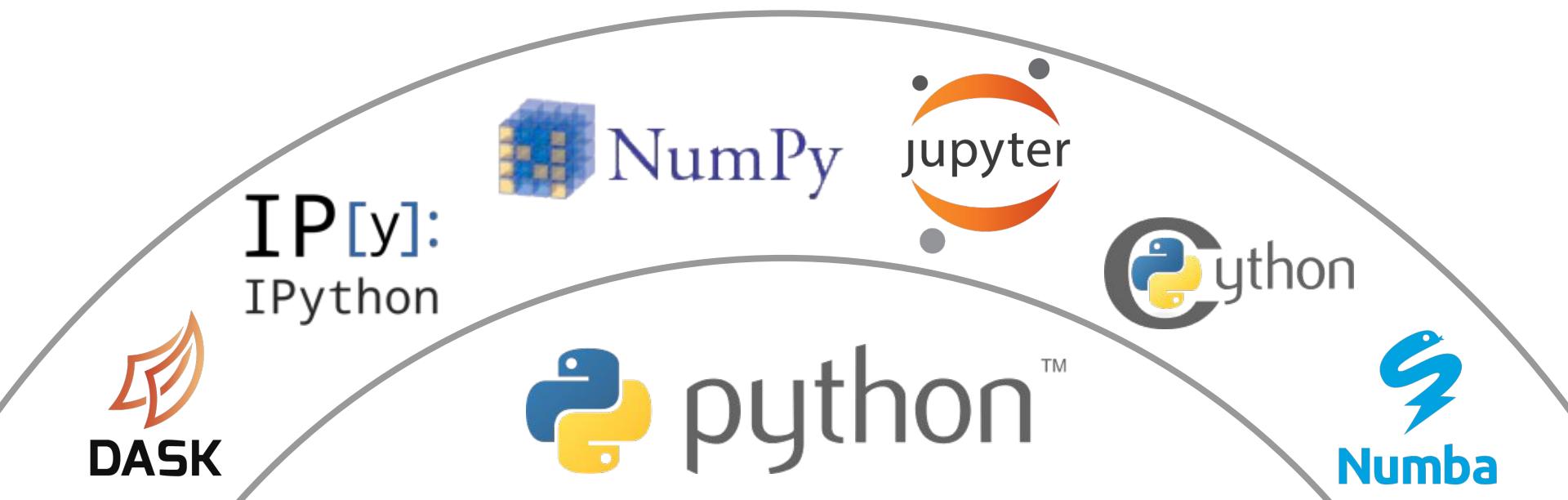
creator of NumPy & SciPy
via email



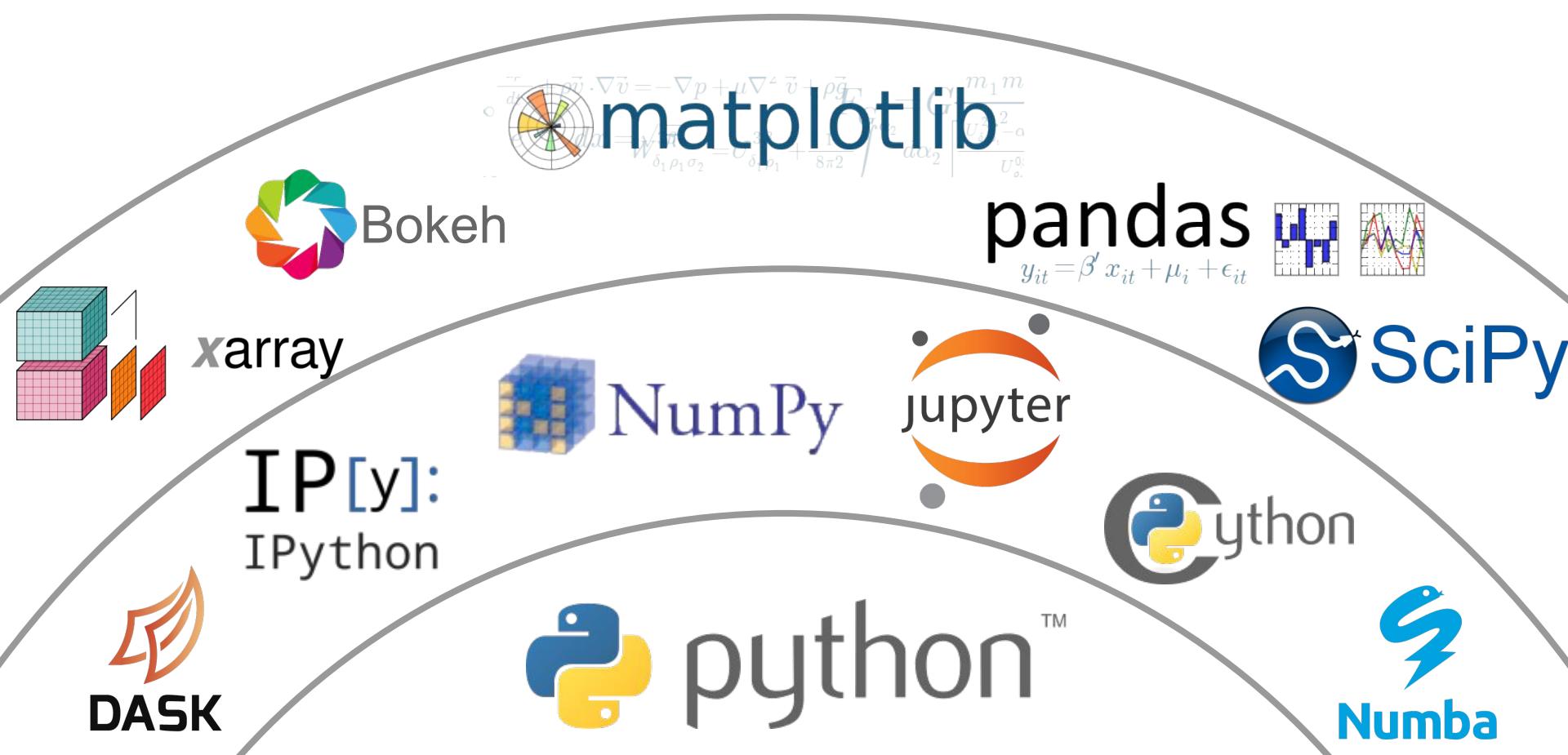
Python's Scientific Stack



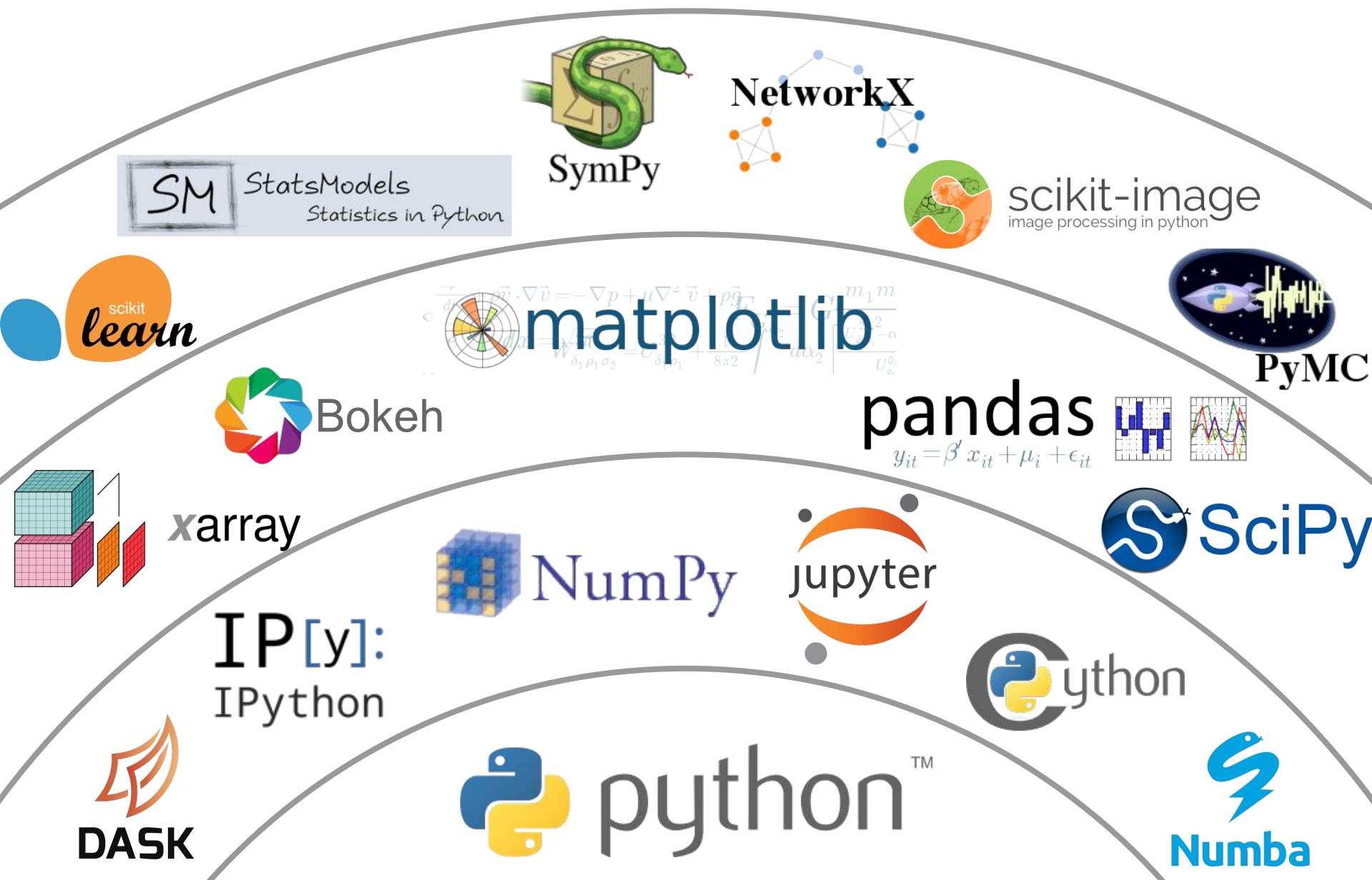
Python's Scientific Stack



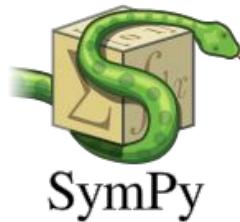
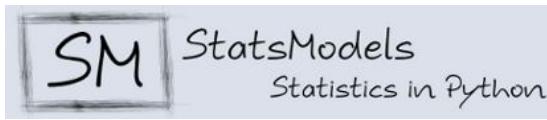
Python's Scientific Stack



Python's Scientific Stack



(and
many,
many
more)



scikit-image
image processing in python

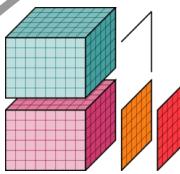
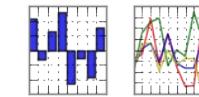


PyMC



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



xarray

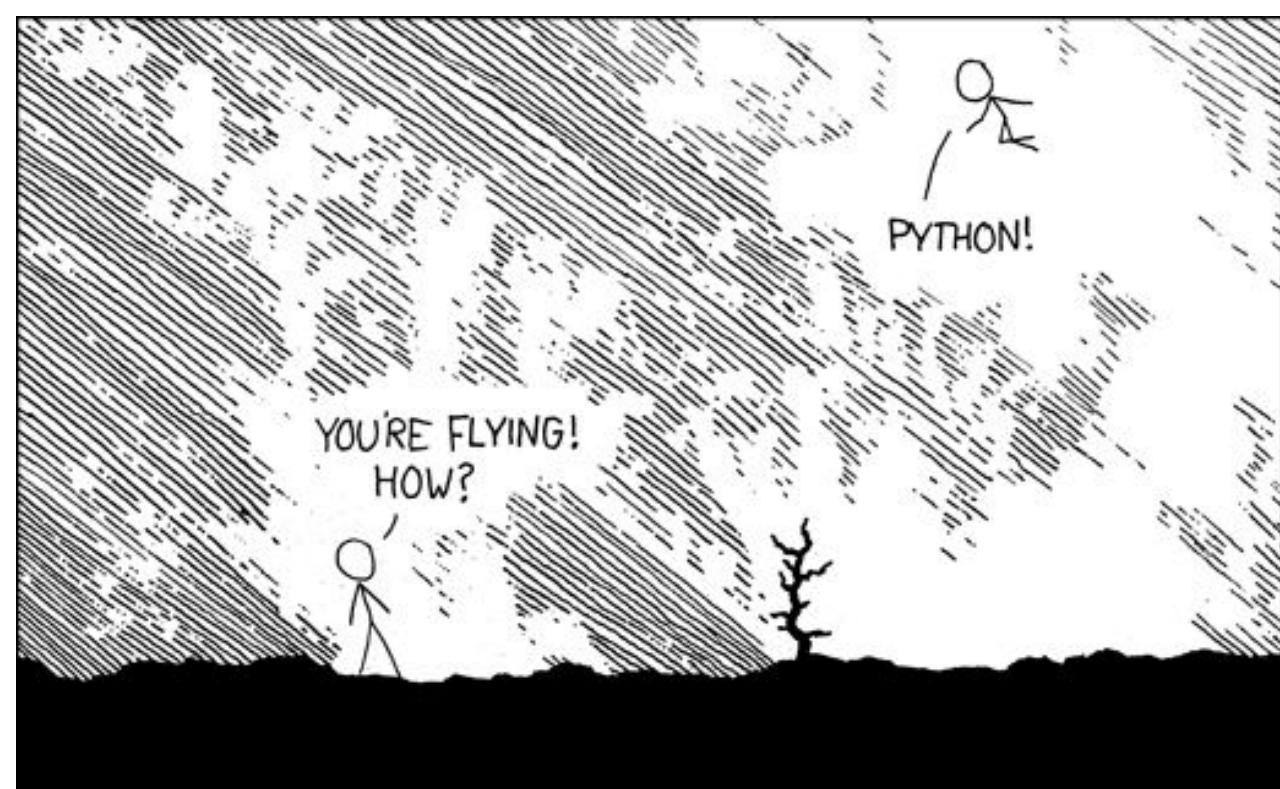
IP[y]:
IPython



python™

Why is Python such an effective tool in science?

1. Interoperability with Other Languages
2. “Batteries Included” + Third-Party Modules
3. Simplicity & Dynamic Nature



I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!

HELLO WORLD IS JUST

print "Hello, world!"

I DUNNO...
DYNAMIC TYPING?
WHITESPACE?

COME JOIN US!
PROGRAMMING IS FUN AGAIN!
IT'S A WHOLE NEW WORLD UP HERE!
BUT HOW ARE YOU FLYING?

I JUST TYPED
import antigravity
THAT'S IT?

... I ALSO SAMPLED
EVERYTHING IN THE
MEDICINE CABINET
FOR COMPARISON.

BUT I THINK THIS
IS THE PYTHON.

Python Enters Science:

“Python is a language that is very powerful for developers, but is also accessible to Astronomers. Getting those two classes of people using the same tools, I think, provides a huge benefit that's not always noticed or mentioned.”

- **Perry Greenfield**
Space Telescope
Science Institute
PyAstro 2015



Often-overlooked fact . . .

For day-to-day scientific data exploration,
speed of development is primary, and
speed of execution is often secondary.

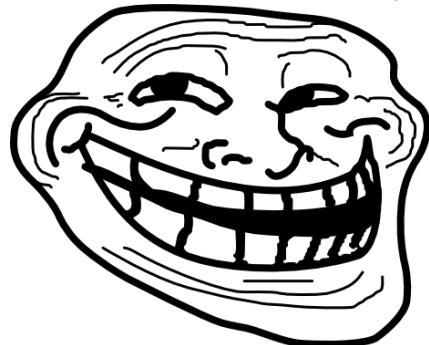


Why don't you use C instead
of Python? It's so much faster!



Why don't you use C instead
of Python? It's so much faster!

Why don't you commute by
airplane instead of by car? It's
so much faster!



Scientific Coding is Nonlinear and Exploratory

Ada Marie did what scientists do:
She asked a small question,
and then she asked two.
And each of those led her
to three questions more,
And some of those questions
resulted in four.





Jupyter notebooks embody this kind of quick, nonlinear exploration:

A screenshot of a Jupyter notebook interface. At the top, there's a menu bar with File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Python 3.7 kernel icon. Below the menu is a toolbar with various icons. The main area contains four code cells. The first cell shows the command `from altair import Chart`. The second cell contains the command `chart(data).mark_point().encode()` with a cursor at the end of the line. The third cell is empty. The fourth cell is empty. Above the code cells, there's a data table with four rows and seven columns. The columns are labeled with numbers 1 through 6 and a final column. The data is as follows:

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|------|---|-------|-------|------|----------|-----------|
| 1 | 11.5 | 8 | 350.0 | 165.0 | 19.0 | buck | skylane |
| 2 | 11.0 | 8 | 318.0 | 159.0 | 18.0 | plymouth | satellite |
| 3 | 12.0 | 8 | 304.0 | 150.0 | 16.0 | amc | rebel |
| 4 | 10.5 | 8 | 302.0 | 140.0 | 17.0 | ford | torino |

Why is Python such an effective tool in science?

1. Interoperability with Other Languages
2. “Batteries Included” + Third-Party Modules
3. Simplicity & Dynamic Nature
4. Open ethos well-fit to science



Sections ≡

The Washington Post

Monkey Cage

Does social science have a replication crisis?

By **Joshua Tucker** March 9 





Sections

Monkey Cage

Does so replicat

By Joshua Tucker Ma

Psychology's Replication Crisis Has a Silver Lining

It's an opportunity for the field to lead.

684



TEXT SIZE



PAUL BLOOM

FEB 19, 2016

SCIENCE

There is a crisis in psychology. It's not those rare cases of outright fraud, as when

the social psychologist Diederik Stapel simply made up the results of dozens of



Sections

Monkey Cage

Psychology's Replication

ing

The replication crisis has engulfed economics

November 2, 2015 7:31pm EST

No two alike? Image sourced from Shutterstock.com

Email

Twitter

Facebook

LinkedIn

A sense of crisis is developing in economics after [two Federal Reserve economists](#) came to the alarming conclusion that economics research is usually not replicable.

[The economists took 67 empirical papers](#) from 12 reputable academic

t fraud, as when
of dozens of



Sections ≡

Monkey Cage

The replication crisis has engulfed even cancer research

November 2, 2015 7:31pm EST

No two alike? Image sourced from ShutterStock

[Email](#)[Twitter](#)[Facebook](#)[LinkedIn](#)

A sense

[economist](#)

usually

[The economist](#)

By Daniel Engber



Psychology's Replication Crisis Is Engulfing Cancer Research

Cancer Research Is Broken

There's a replication crisis in biomedicine—and no one even knows how deep it runs.

as when
ans of



Sections

Big Science is broken



Pascal-Emmanuel *Gobry*



Novel

No tw

Ema

Twe

Facebook

68

usually

LinkedIn

26

The eco

as when
ans of





Sections

Big Science is broken



Pascal-Emmanuel Gobry



The replication crisis in science has just begun. It will be big.

[24 Replies](#)

Summary: After a decade of slow growth beneath public view, the replication crisis in science begins breaking into public view. First psychology and biomedical studies, now spreading to many other fields — overturning what we were told is settled science, the foundations of our personal behavior and public policy. Here is an introduction to the conflict (there is pushback).



Sections

Big Science is broken



Dascal Emmanuel Cohen

BBC

Sign in

News

Sport

Weather

Shop

Earth

Travel

More

NEWS

[Home](#)[Video](#)[World](#)[US & Canada](#)[UK](#)[Business](#)[Tech](#)[Science](#)[Magazine](#)[Entertainment](#)

Science & Environment

The replication
will be big.

[24 Replies](#)

Summary: After a decade of
beginning to break into public
many other fields — over
personal behavior and pub

Most scientists 'can't replicate studies by their peers'

By Tom Feilden

Science correspondent, Today programme

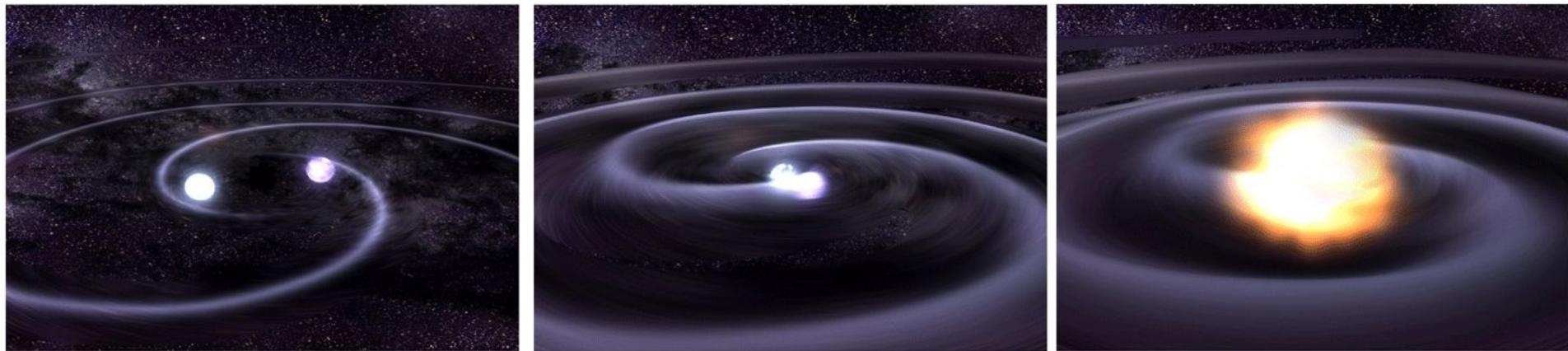
⌚ 22 February 2017 | [Science & Environment](#)

[Share](#)

“An article about computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.”

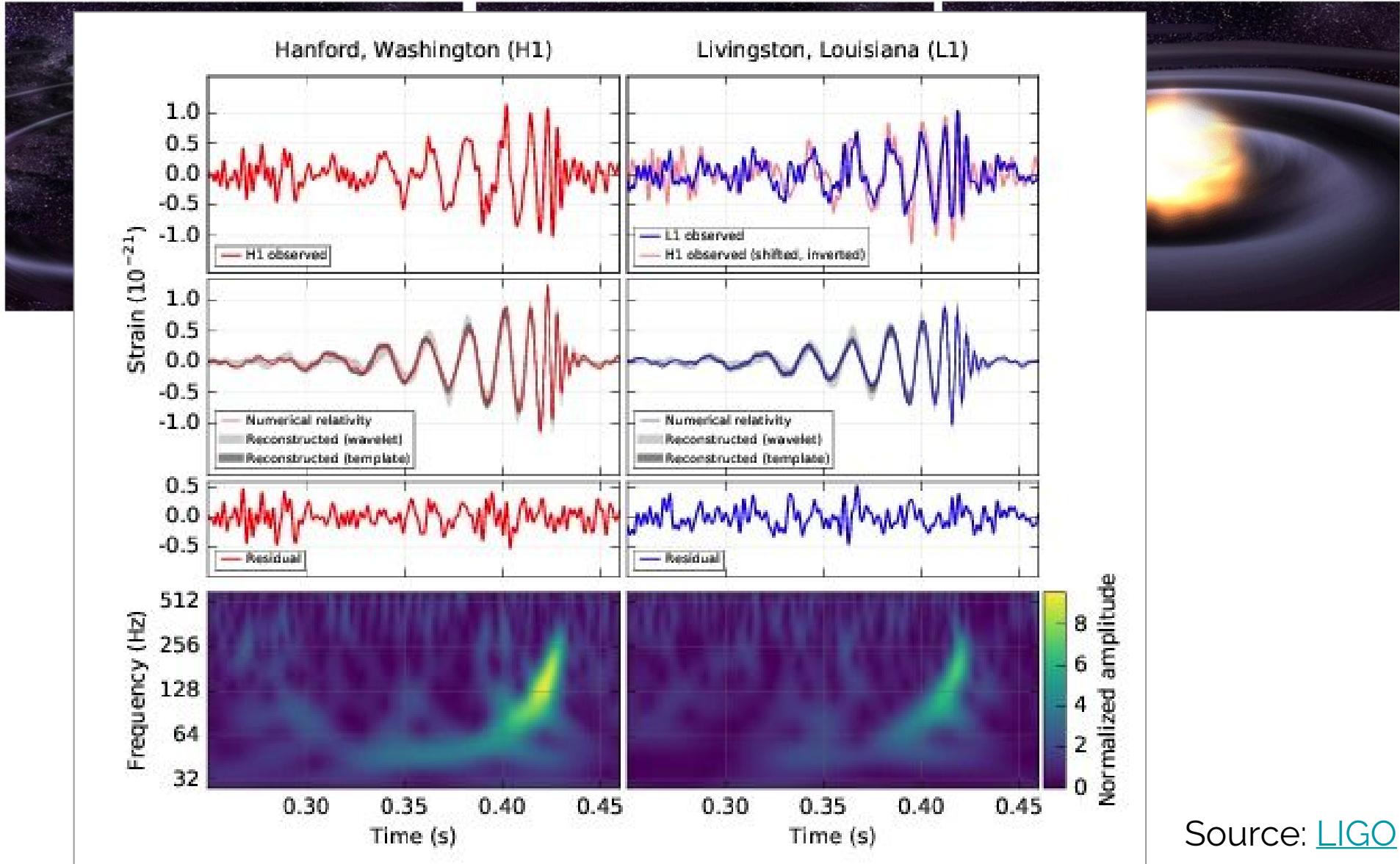
–Buckheit and Donoho (1995)

LIGO Gravitational Wave Event (GW150914)



Source: [LIGO](#)

LIGO Gravitational Wave Event (GW150914)



Source: [LIGO](#)

LIGO Gravitational Wave Event (GW150914)

← → C nbviewer.jupyter.org/urls/losc.ligo.org/s/events/GW150914/GW150914_tutorial.ipynb

jupyter nbviewer JUPYTER FAQ </> ⚡ ⏪ ⏴ ⋮

SIGNAL PROCESSING WITH GW150914 OPEN DATA

Welcome! This ipython notebook (or associated python script GW150914_tutorial.py) will go through some typical signal processing tasks on strain time-series data associated with the LIGO GW150914 data release from the LIGO Open Science Center (LOSC):

- <https://losc.ligo.org/events/GW150914/>
- View the tutorial as a web page - [https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html/](https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html)
- Download the tutorial as a python script - [https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.py/](https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.py)
- Download the tutorial as IPython Notebook - [https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.ipynb/](https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.ipynb)

To begin, download the ipython notebook, readligo.py, and the data files listed below, into a directory / folder, then run it. Or you can run the python script GW150914_tutorial.py. You will need the python packages: numpy, scipy, matplotlib, h5py.

On Windows, or if you prefer, you can use a python development environment such as Anaconda (<https://www.continuum.io/why-anaconda>) or Enthought Canopy (<https://www.enthought.com/products/canopy/>).

Questions, comments, suggestions, corrections, etc: email losc@ligo.caltech.edu

v20160208b

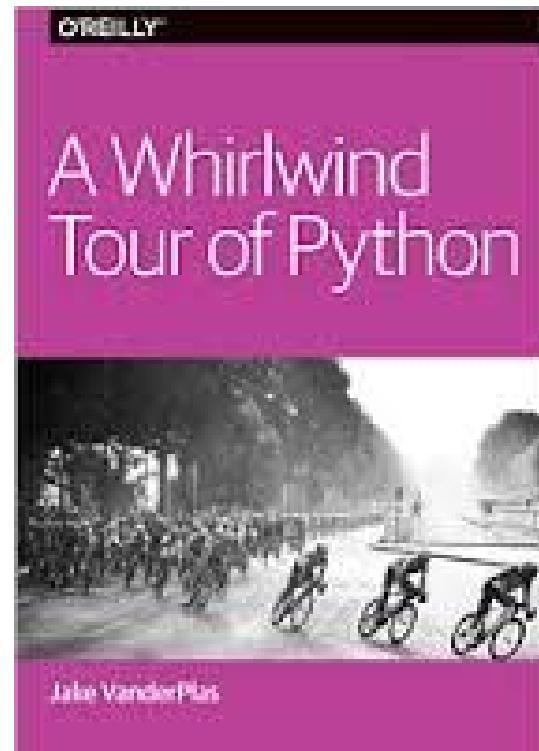
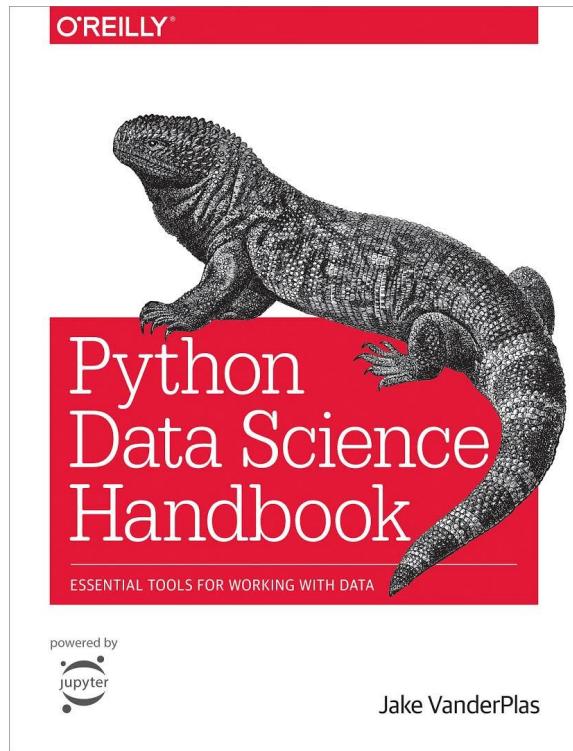
Intro to signal processing

This tutorial assumes that you know python well enough.

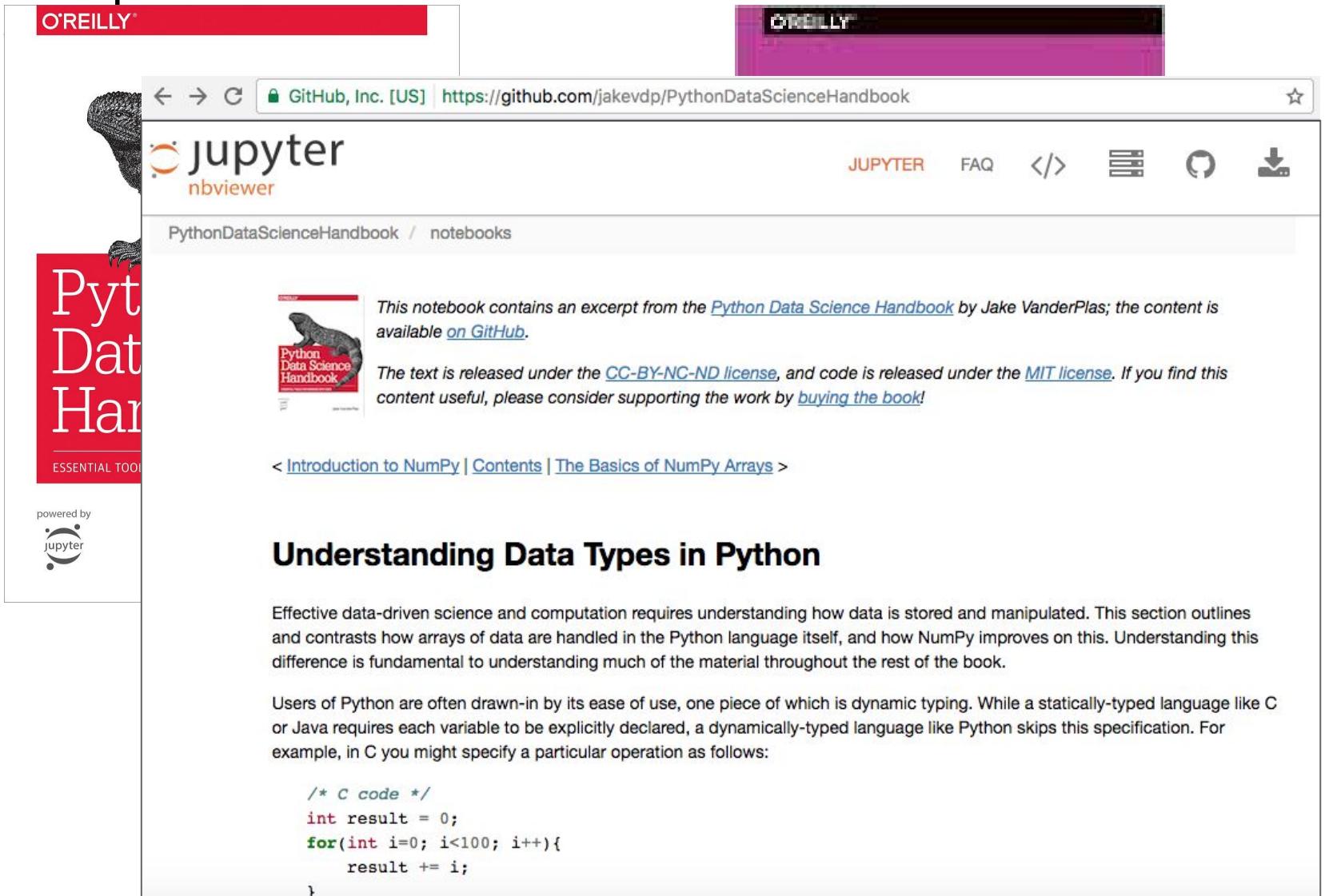


Source: [LIGO](#)

My Projects: Same Open Philosophy



My Projects: Same Open Philosophy



The screenshot shows a Jupyter notebook viewer interface. At the top, there's a red header bar with the O'REILLY logo. Below it is a navigation bar with back, forward, and search icons, followed by the GitHub URL <https://github.com/jakevdp/PythonDataScienceHandbook>. The main content area features the "jupyter nbviewer" logo and the title "PythonDataScienceHandbook / notebooks". On the left, a vertical sidebar displays the book cover of "Python Data Science Handbook" with the subtitle "ESSENTIAL TOOLS FOR DATA SCIENCE" and the text "powered by jupyter". The main content area contains a snippet from the book, mentioning the "Python Data Science Handbook" by Jake VanderPlas and its availability on GitHub. It also notes the CC-BY-NC-ND license for text and MIT license for code, with a link to support the work. Below this, there's a navigation bar with links to "Introduction to NumPy", "Contents", and "The Basics of NumPy Arrays". The main article title is "Understanding Data Types in Python". The text discusses the difference between Python's dynamic typing and statically-typed languages like C or Java. It includes a code snippet in C:

```
/* C code */
int result = 0;
for(int i=0; i<100; i++){
    result += i;
}
```

Entire content available on GitHub as Jupyter Notebooks

Python World Influencing Science . . .

Scientists are increasingly hosting research code on Github & similar services to aid in reproducibility.



Screenshot of a GitHub search results page for "Astronomy Python Library".

The search bar at the top shows "Search GitHub" and filters for "Pull requests" and "Issues".

The main title is "Astronomy Python Library" with a link to <http://astropy.org>. It shows 22 people associated with the repository.

Three repositories are listed:

- astroplan**: Observation planning package for astronomers. Updated 4 minutes ago.
- astropy**: Repository for the Astropy core package. Updated an hour ago.
- astropy-helpers**

Each repository entry includes a thumbnail image, the repository name, a brief description, and the last update time.

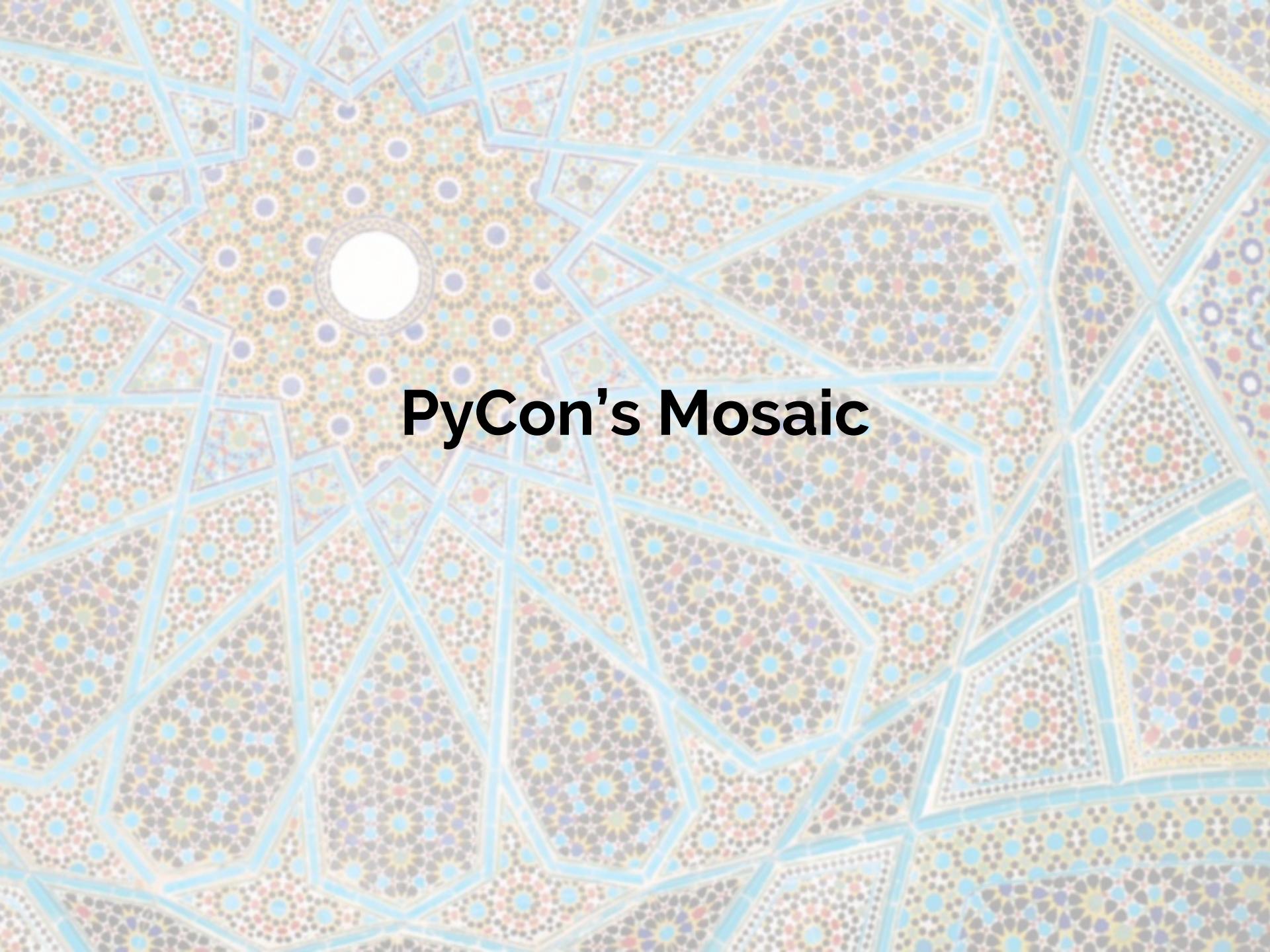
Python World Influencing Science . . .

Python's software practices increasingly adopted by academia

| Traditional Astronomy Software | Python & Open Source |
|----------------------------------|-----------------------------------|
| Possessive/non-sharing | Cooperative/sharing |
| Fragmented & Overlapping efforts | Build on common projects |
| Top-down planning | Bottom-up/Loose organization |
| Committee-oriented design | Design by “doers” |
| Endless analysis & argument | Action-oriented & experimentation |
| Unwilling to discard old tech | Good at replacing old tech |
| No leader to resolve conflicts | BDFL resolves conflicts |

Why is Python such an effective tool in science?

1. Interoperability with Other Languages
2. “Batteries Included” + Third-Party Modules
3. Simplicity & Dynamic Nature
4. Open ethos well-fit to science



PyCon's Mosaic

Thank You!



Email: jakevdp@uw.edu



Twitter: [@jakevdp](https://twitter.com/jakevdp)



Github: [jakevdp](https://github.com/jakevdp)



Web: <http://vanderplas.com/>



Blog: <http://jakevdp.github.io/>