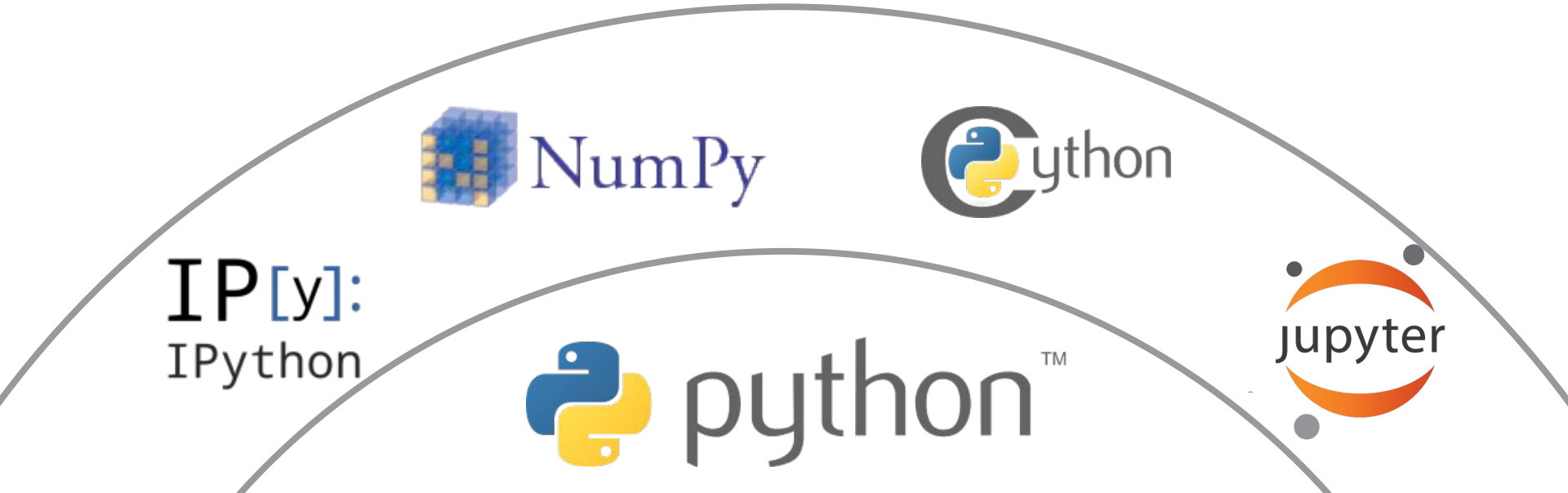
The Python Scientific Stack

Jake VanderPlas @jakevdp

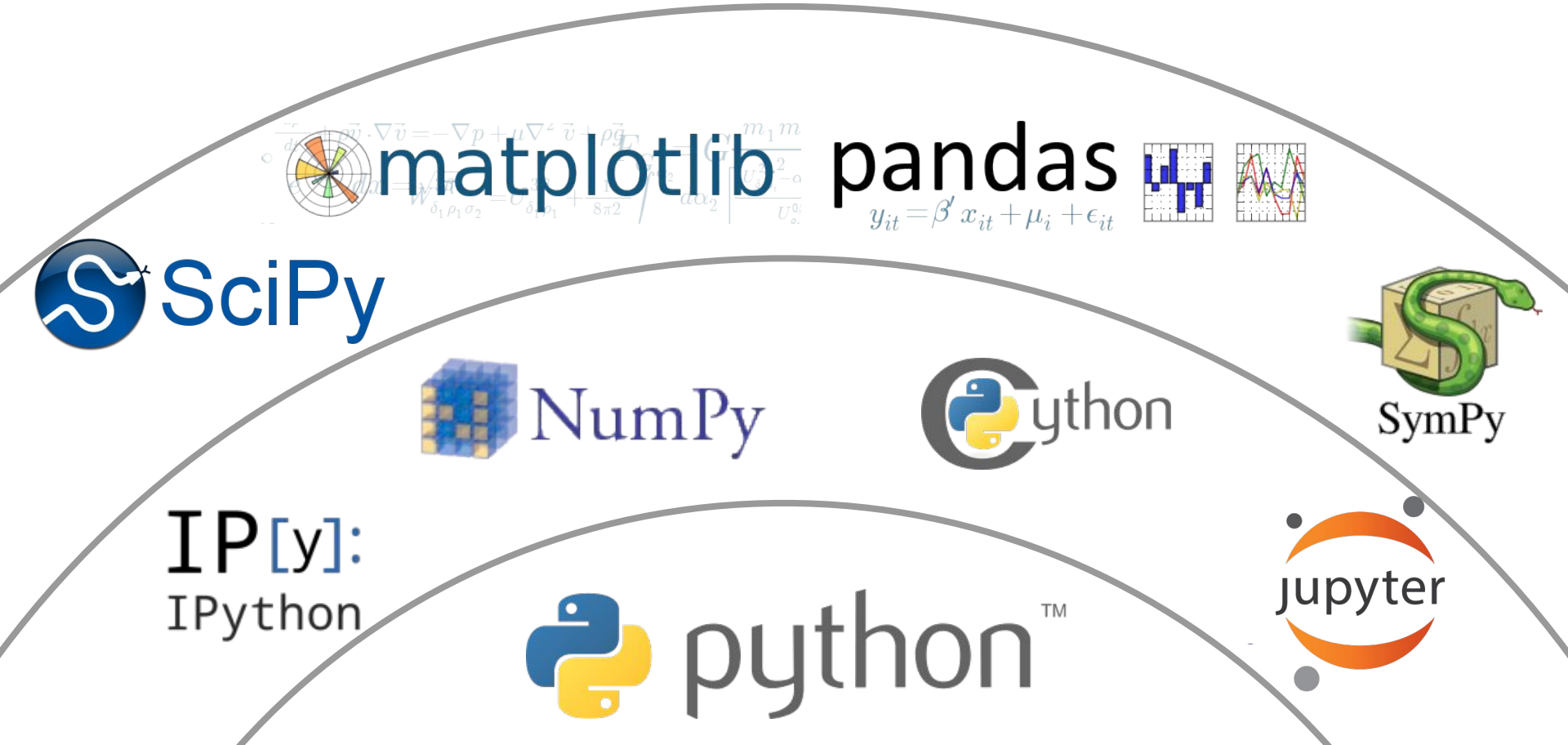
Dec 1, 2015

Python's Scientific Ecosystem

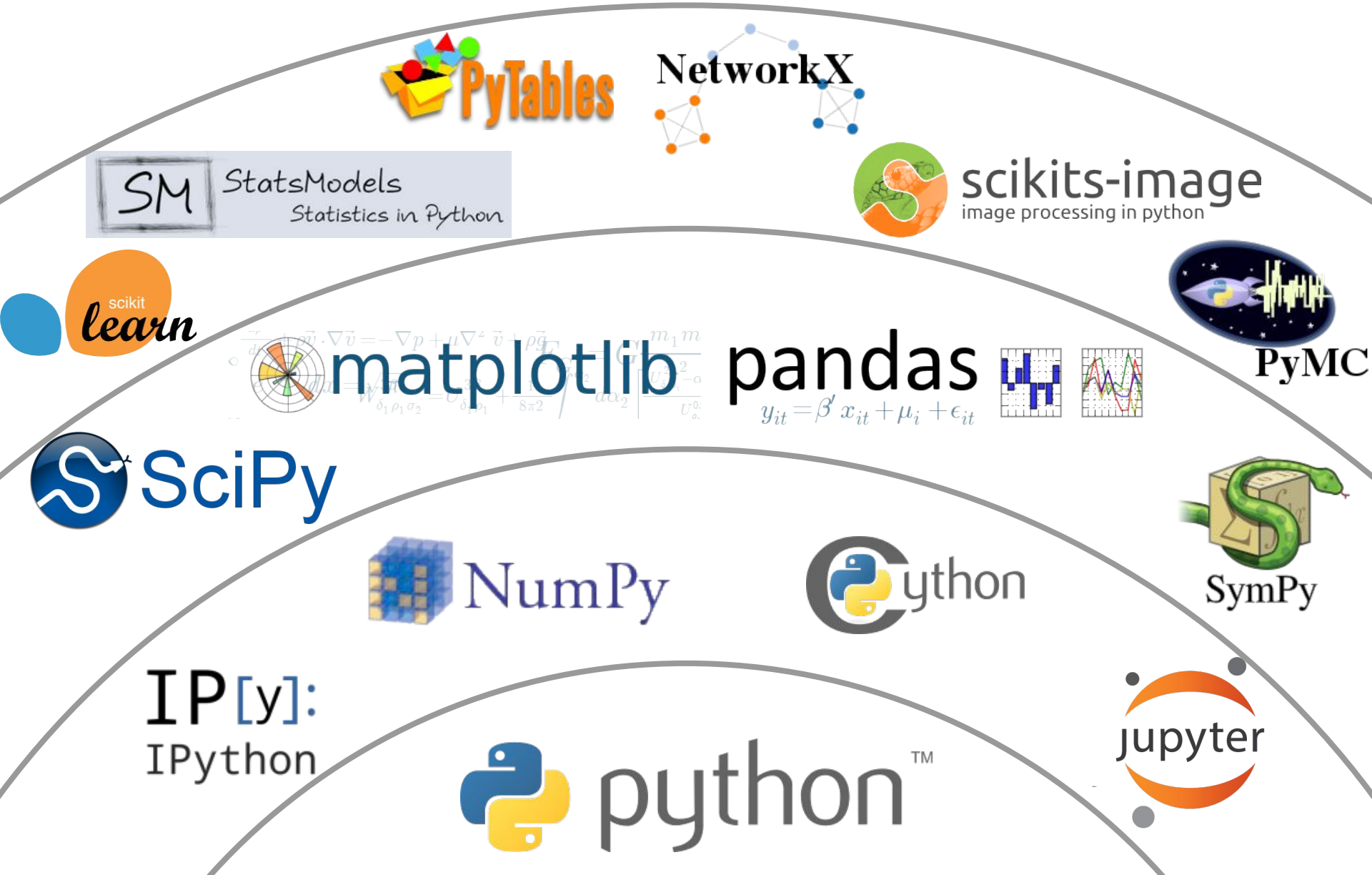
Python's Scientific Ecosystem



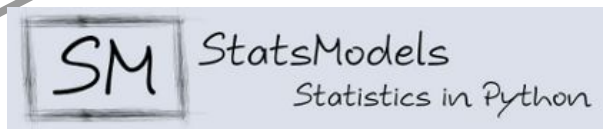
Python's Scientific Ecosystem



Python's Scientific Ecosystem



(and many, many more)



Many more tools:

Performance:

Numba, Weave, Numexpr, Theano . . .

Visualization:

Bokeh, Seaborn, Plotly, Chaco, mpld3, ggplot, MayaVi, vincent, toyplot, HoloViews . . .

Data Structures & Computation:

Blaze, Dask, DistArray, XRay,

Graphlab, SciDBpy, pySpark . . .

Packaging & distribution:

pip/wheels, conda, EPD, Canopy, Anaconda ...

Recent-ish Developments

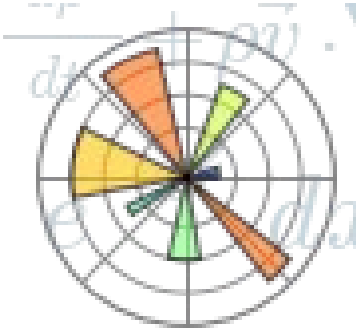
Recent Developments: Core Language

I'll just leave this right here . . .

If you haven't switched, it's time.

Recent Developments: Visualization

Matplotlib:
Evolving into a more Modern Package



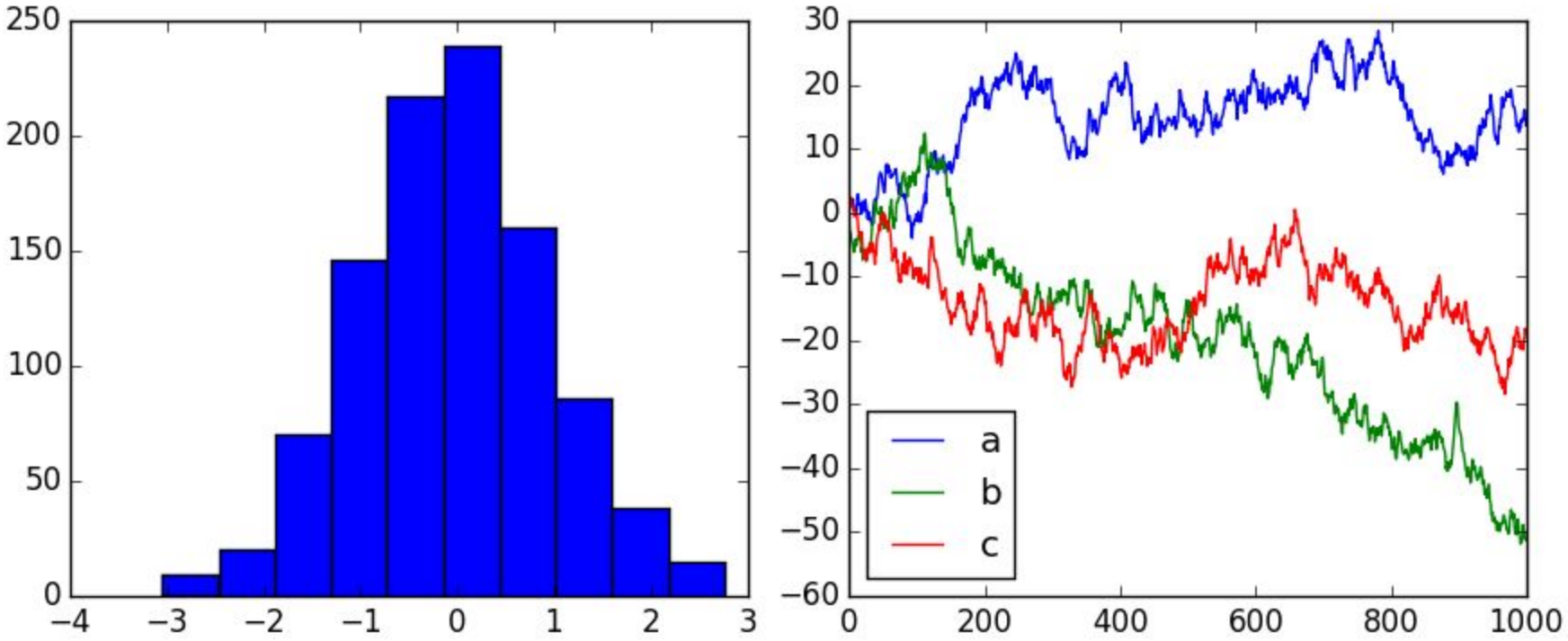
Matplotlib 1.4 features *stylesheets*,
with several very nice built-in styles.

Matplotlib 2.0 will break backward compatibility
to provide *new plot style defaults!*

(See *State of Matplotlib* talk for more details)

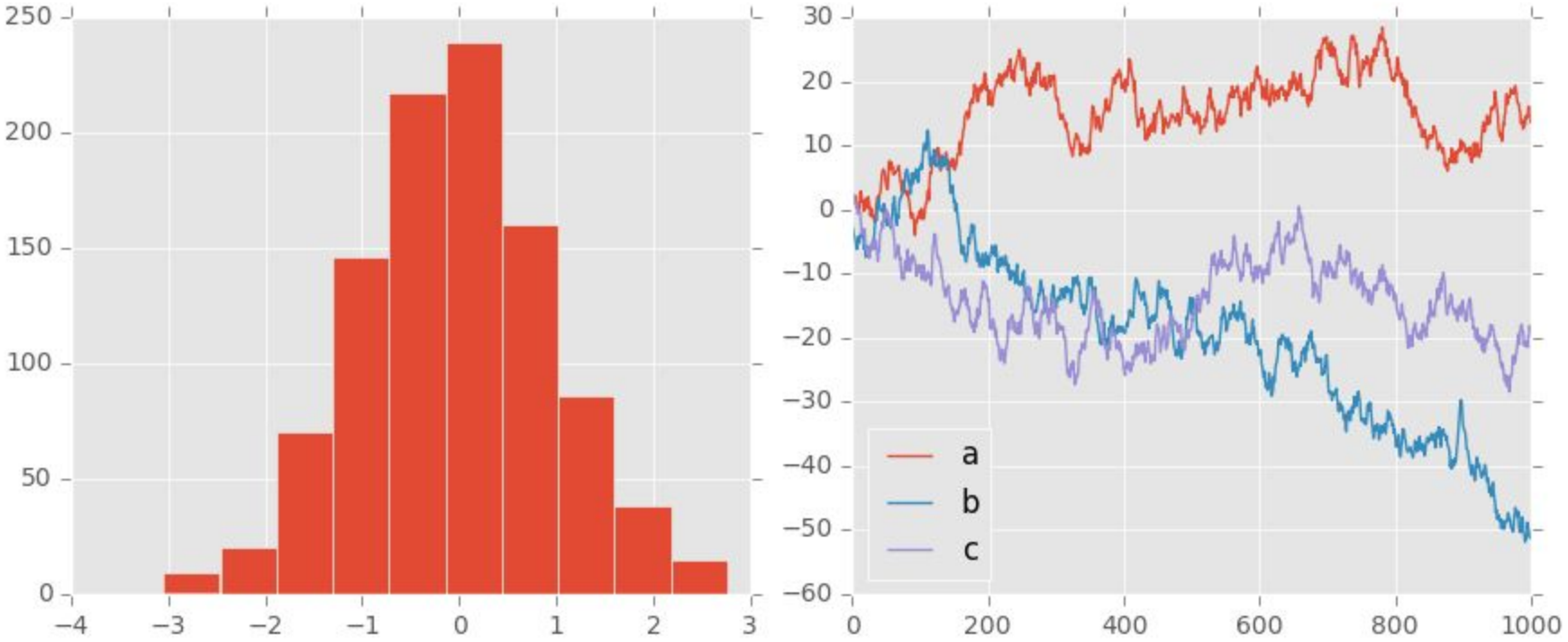
New Matplotlib Styles:

In [3]:
```
make_plots()
```



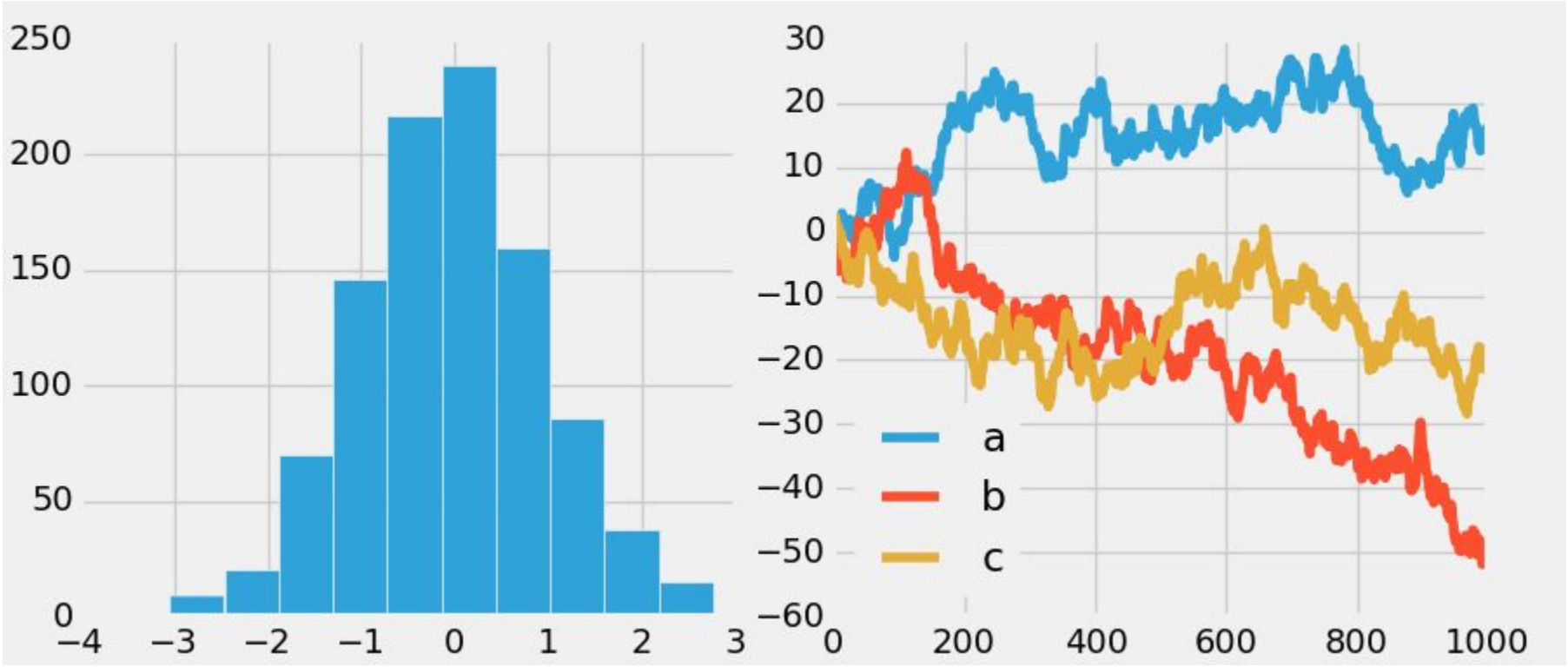
New Matplotlib Styles:

In [4]:
```
plt.style.use('ggplot')
make_plots()
```



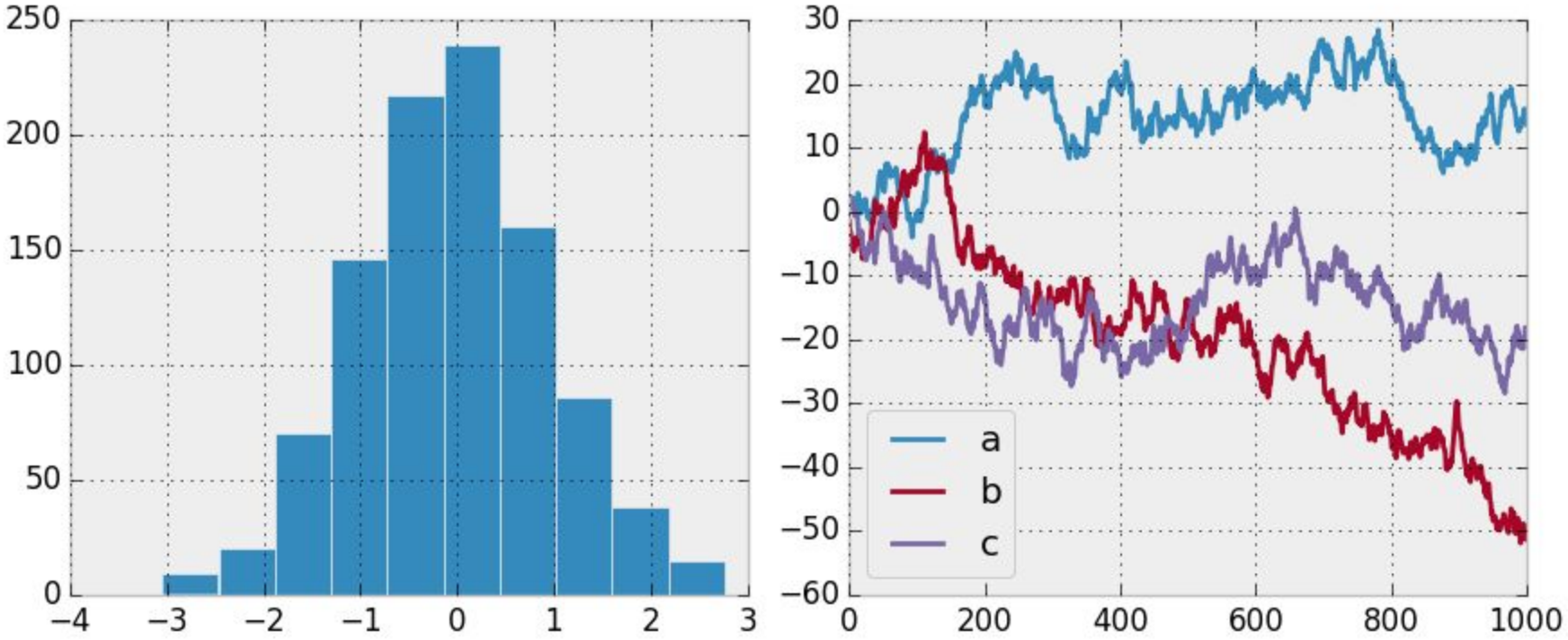
New Matplotlib Styles:

In [5]:
```
plt.style.use('fivethirtyeight')
make_plots()
```

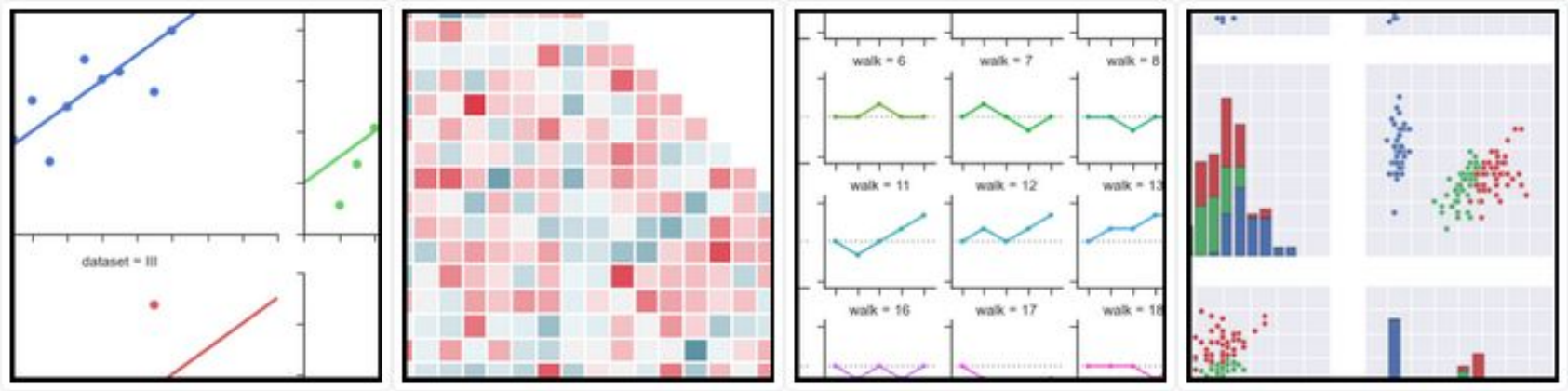


New Matplotlib Styles:

In [6]:
```
plt.style.use('bmh')
make_plots()
```



Seaborn: Matplotlib + Pandas + Statistical Visualization

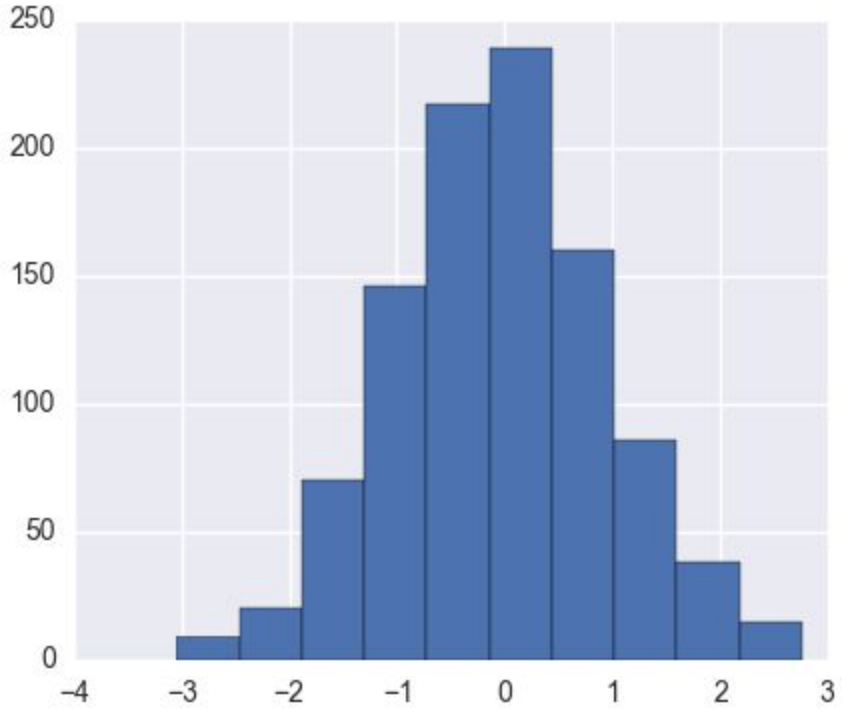


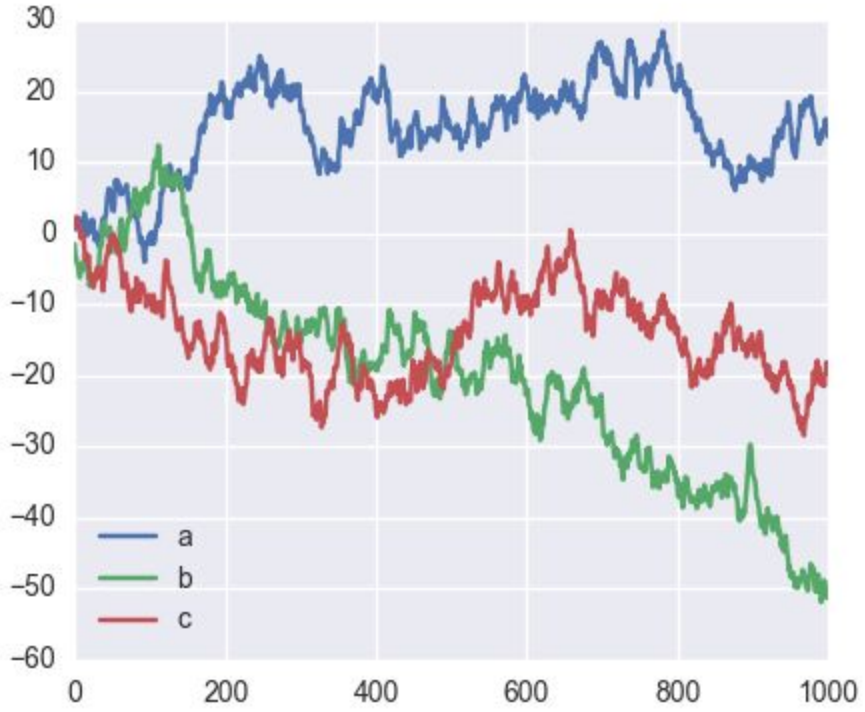
- built on top of **matplotlib**: able to use any of its backends & output formats
- **pandas**-aware: quick plotting of labeled data
- provides beautiful, well-thought-out default plot styles

http://stanford.edu/~mwaskom/software/seaborn/

Seaborn's Matplotlib Style:

```
In [7]: import seaborn; seaborn.set()
make_plots()
```





(style available natively in next matplotlib release)

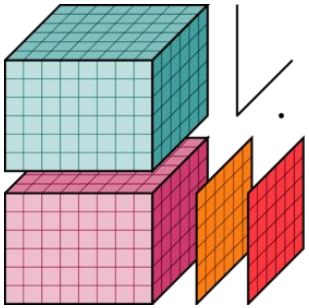
Bokeh: Powerful Interactive Viz

- HTML5 output, both server and client-side
- Flexible in-browser interactivity
- Fundamentally a **Javascript library** with Python bindings

http://bokeh.pydata.org/

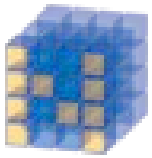
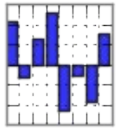
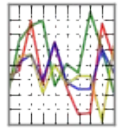
Recent Developments:
Arrays & Data Structures

Arrays and Data Structures



xray

`xray` implements numpy-style ND arrays with Pandas-style labels & indices.

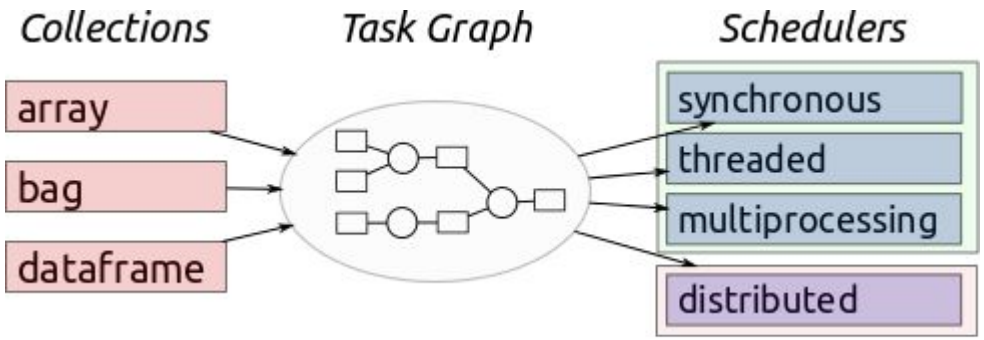
( NumPy + pandas $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$  )

Modern data is heterogeneous, noisy, and complicated. Anonymous dense arrays are no longer enough!

http://xray.readthedocs.org/

Arrays and Data Structures

Dask: a lightweight tool for general parallelized array storage and computation.



The project is still young, but the possibilities are very exciting!

http://dask.pydata.org/

Recent Developments: Computation & Performance

Computation & Performance:

Numba: with a simple decorator, Python JIT compiles to LLVM and executes at near C/Fortran speed!

```
def fib(n):
    a, b = 0, 1
    for i in range(n):
        a, b = b, a + b
    return a

%timeit fib(50)
```

```
100000 loops, best of 3: 3.83 µs per loop
```

Still some features missing, but very promising (see my blog posts for some examples).

http://numba.pydata.org/

Computation & Performance:

Numba: with a simple decorator, Python JIT compiles to LLVM and executes at near C/Fortran speed!

```
@numba.jit
def fib(n):
    a, b = 0, 1
    for i in range(n):
        a, b = b, a + b
    return a

%timeit(fib(50))
```

```
1 loops, best of 3: 468 ns per loop
```

20x speedup!

Still some features missing, but very promising (see my blog posts for some examples).

http://numba.pydata.org/

Recent Developments: Distribution & Packaging

Distribution & Packaging:

`conda` distribution & packaging tool has changed the way many use, develop, & teach Python.

- like **pip**, but better management of Python & non-python dependencies
- like **virtualenv**, but allows different versions of compiled libraries
- Similar to **yum / apt / macports / brew**, but platform-independent!

http://conda.pydata.org/

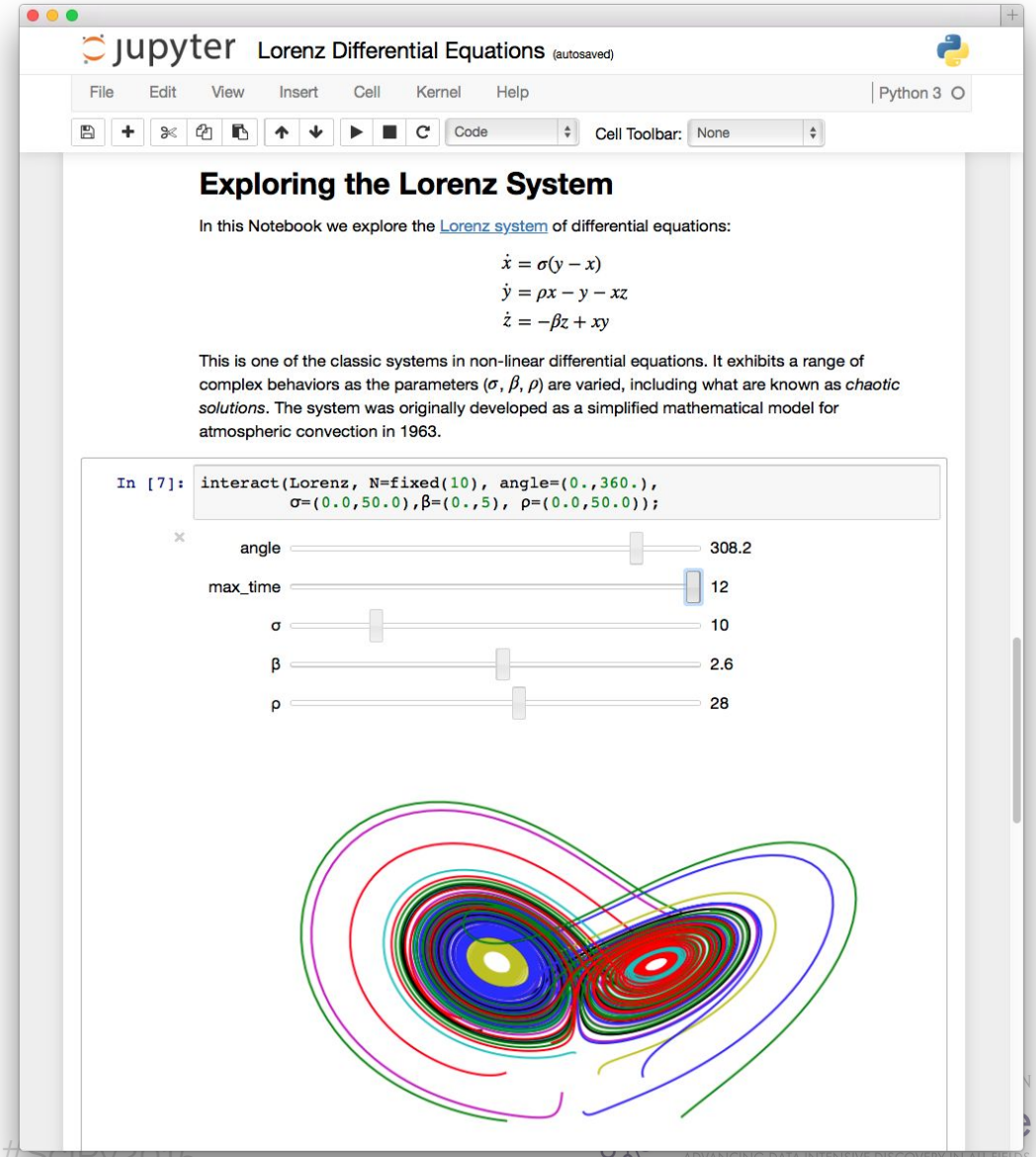
And of course . . .

IPython & Jupyter

So much happening . . .

- The IPython/Jupyter split
- Widgets = awesome
- Docker-based backends
- Jupyter Hub
- new $6M grant this week!

Python stack is branching out to benefit other languages!



And so much more . . .

Thank You!

Email: jakevdp@uw.edu

Twitter: @jakevdp

Github: jakevdp

Web: http://vanderplas.com

Blog: http://jakevdp.github.io