前言

Python 科学可视化的领域非常广泛(见图1)。其涉及到形形色色的工具,从最通用、最广泛使用的,到更专业、保密的工具。这些工具一些是由社区开发,而另一些则是由公司开发。有些工具专门用于网络,有些则仅用于桌面,有些处理3D和大数据,而另一些则针对完美的2D渲染。

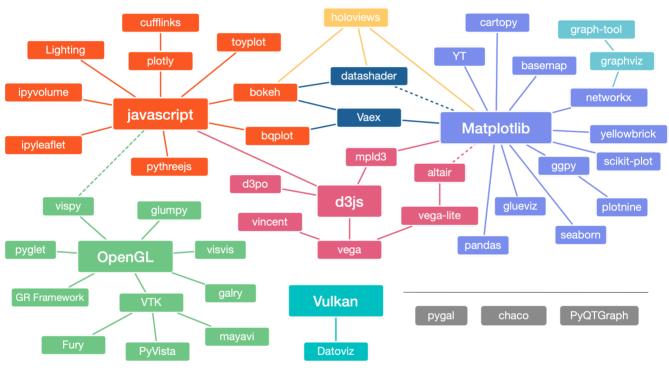


图1

2018年Python在科学可视化领域的概览 (部分内容)。该图基于Jake Vanderplas的原创进行调整。详细资源见: https://github.com/rougier/python-visualization-landscape

面对如此丰富的选择,您可能很难找到最适合需求的软件包,因为您甚至不知道某个软件包的存在。为了帮助您做出选择,您可以先问自己几个问题:

- 您的目标是桌面应用还是Web渲染?
- 您是否拥有复杂的3D渲染场景?
- 您是否对产品有质量要求?
- 您是否需要处理海量数据?
- 这些包的社区是否活跃?
- 这些包有文档和教程吗?

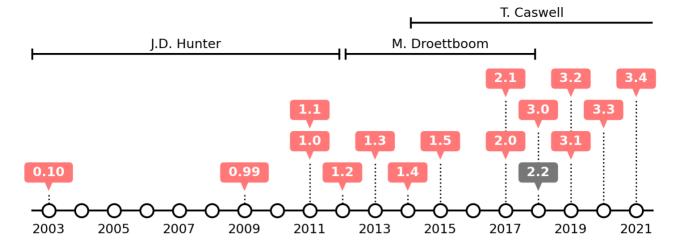


图2

Matplotlib最初由John D. Hunter编写,第一个公开版本发布于2003年。在John Hunter于2012年8月去世前不久,Michael Droettboom被提名为Matplotlib的首席开发人员,Thomas Caswell于2014年加入,并担任现任首席开发人员(截至2021年)。最新版本为3.4(撰写本文时),其仅支持Python 3,而版本2.2是兼容Python 2和Python 3的长期支持版本。详细资源见introduction/matplotlib-timeline.py.

根据您对上述的回答,您可以决定使用哪个软件包并投入一些时间学习它。例如,如果您需要在浏览器中进行交互式可视化,并与jupyter无缝集成,那么可以选择bokeh。如果您有非常大的数据并需要在桌面上进行3D可视化,那么vispy或mayavi是一个很好的选择。如果您对快速构建漂亮的图形方面非常感兴趣,那么seaborn和altair是你的得力助手。另外,如果你从事地球科学工作,那么你不能忽视cartopy等库。我在此无法列出所有库,我相信在撰写本章和实际出版书籍之间,会诞生出新的可视化库。一个很好的信息源是 pyviz 网站(用于数据可视化的Python工具),该网站收纳了很多可视化库的网站链接,并提供最新的活跃包列表。

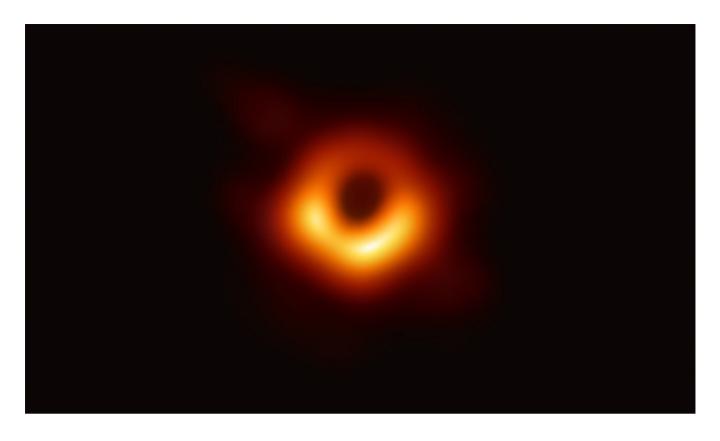


图3

2019年4月10日,事件视界望远镜发布的第一张图像描绘了超巨椭圆星系M87核心的超大质量黑洞,其质量约为太阳质量的70亿倍。详细资源见Wikipedia

在这个领域中,Matplotlib具有非常特殊的地位。它最初是由John D. Hunter于2003年创建的,目的是为了可视化脑皮层电图数据。以下是2003年5月23日发布在Python邮件列表上的官方公告¹。

¹非常感谢 Anthony Lee 给我指出这个公告。

Matplotlib

Matplotlib is a pure python plotting package for python and pygtk. My goal is to make high quality, publication quality plotting easy in python, with a syntax familiar to matlab users. matplotlib is young, and several things need to be done for this goal is achieved. But it works well enough to make nice, simple plots.

Requirements

python 2.2, GTK2, pygtk-1.99.x, and Numeric.

Download

See the homepage - nitace.bsd.uchicago.edu:8080/matplotlib

Here are some of the things that matplotlib tries to do well

- * Allow easy navigation of large data sets. Right click on figure window to bring up navigation tool bar for pan and zoom of x and y axes. This requires a wheel mouse. Place the wheel mouse over the navigation buttons and scroll away.
- * Handle very large data sets efficiently by making use of Numeric clipping. I have used matplotlib in an EEG plotting application with 128 channels and several minutes of data sampled at 400Hz, eg, plotting matrices with dimensions $120,000 \times 128$.
- * Choose tick marks and labels intelligently
- * make easy things easy (subplots, linestyles, colors)
- * make hard things possible (OO interface for full control)

Matplotlib is a class library that can be used to make plots in pygtk applications. But there is a matlab functional compatibility interface that you can get with, eg::

from matplotlib.matlab import plot, subplot, show, gca

Example scripts and screenshots available at http://nitace.bsd.uchicago.edu:8080/matplotlib

John Hunter