Visualization using Jupyter Notebooks

Robert Sinkovits, PhD

Director of Education

San Diego Supercomputer Center

Overview

- Today we'll briefly explore visualization in Jupyter notebooks with Python (Matplotlib) and R (ggplot2)
- Knowledge of Python and/or R programming is not necessary to execute the notebooks and you should be able to modify the examples with very minimal programming experience
- I'm partial to Python and will spend most of my time there, but we'll also touch on R since the ggplot2 package is extremely powerful and popular with vis experts

Motivation

- Doing your visualization in Jupyter notebooks makes it easy to interactively tweak parameters to get plots just the way your want
- Having code in a notebook allows you to automate repeated visualization tasks
- Using notebooks also makes it possible to integrate your visualization with your computations, especially if you're working in R or Python

GitHub repos

Python

https://github.com/sinkovit/PythonSeries

Matplotlib.ipynb

R

https://github.com/sinkovit/Rseries

ggplot2.ipynb

NOTE – may need to install gridExtra to do the final example
install.packages("gridExtra)

Running the notebooks

If you use Anaconda or Galyleo, you'll probably have all the packages you need. The one exception was that I had to manually install the gridExtra R package when using engagelively Galyleo implementation (https://galyleo.engagelively.com/)

If using Galyleo, I suggest choosing the Datascience Environment since that allowed me to run all Python and R examples. I also found that the SciPy Environment worked for the Python notebooks (but not R)