What's in a Name? An Introduction to Data Exploration in Python (Teen Track)
This tutorial introduces SciPy analysis and visualization tools by exploring the Social Security Administration's baby names data set within a Jupyter notebook environment. We will cover basics of Python and NumPy syntax and matplotlib scatter plots. The day closes with a mini-project intended to increase participants' confidence in creating their own routines for data analysis.

Tentative Schedule
Note
The breaks included in this schedule are a minimum; I may include more as appropriate for the room. I will pace the class using Software Carpentry teaching methods, e.g., passing out green and red post-it notes to each student and suggesting that they put a green post-it on their computer if they are finished with a task, and a red post-it if they need help.

8am-9am (optional)
A free hour before the tutorial for installation assistance, for those who have had difficulty with installing Python or checking for correct dependencies.

9am-10:30am
Introduction to Ipython/Jupyter notebooks. Introduction to basic Python syntax: data types, if statements, for loops, indexing, and slicing. Introduction to importing packages, with some practice using the math package.

10:45am-12pm
Introduction to basics of NumPy: array creation, array indexing, and loading data from text files into arrays. We will practice these techniques using data from the SSA baby names data set (https://www.ssa.gov/OACT/babynames/limits.html). The data set will be pre-cleaned so that participants can work easily within a tidy, modified data set.

1:30pm-3pm
Introduction to the syntax of defining functions. Brief introduction to creating a scatter plot with matplotlib, through a continued investigation of the SSA baby names data set. Participants will use the tools introduced in the first part of the session to:
- plot the distribution of their name over time
- determine when was the largest greatest increase/decrease in their name's popularity
- (if time) compare the distributions of the most popular boys' names vs. girls' names over time

3:15pm-5:30pm
The final section is intended to build participants' confidence in creating their own routines by using one of the two data sets from prior in the tutorial to answer their own questions with the data, using statistical and visualization tools from the SciPy stack. The exact timeline will depend on the expected number of participants. I will provide a list of possible mini-projects, and participants will work in small groups to answer a question of their choosing. A few example questions: Where in the United States are your name(s) popular? How do the distributions of the most-popular names over time differ from those of the second most-popular names? If time is limited, I will guide all participants through the same question. No new concepts will be presented in this section.

Bio

Emily Quinn Finney is a newly-minted M.S. in astrophysics, who spent the past five years using tools of the SciPy stack to understand properties of galaxy clusters. She has extensive experience teaching workshops in summer camp contexts; she has designed and directed her own curricula for physics of music, environmental science, and computer programming camps for students ages 4-16. Most recently (April-June 2017), she taught science and math at The Peregrine School, an independent elementary school in Davis, CA.

Skills and Instructions

Skills used in this tutorial: Python, NumPy, matplotlib, Ipython/Jupyter notebook.

Background knowledge required: Introductory programming experience will be very helpful. Command line basics may be helpful but not necessary; I will avoid use of the command line in this tutorial. Some knowledge of basic statistics could be helpful but I will explain any statistical concepts used in the tutorial.

Installation instructions: Given the wide variety of operating systems, I recommend installing all dependencies using Anaconda (https://www.continuum.io/downloads). Download the Python 3.6 package appropriate to your operating system. If you have a Windows or Mac, double-click on the package name to start the graphical installer. If you have a Linux, type the command

```
bash Anaconda3-4.3.1-Linux-x86_64.sh
```

into your terminal window. Follow all instructions for the installer. For Linux or Mac, when prompted whether to add the Anaconda directory to your .profile or your .bashrc, type ‘yes’.