

GEOS F436/636 Beyond the Mouse

Christine (Chris) Waigl

University of Alaska Fairbanks – Fall 2018

Week 12: Your own data analysis environment

Topics for week 12 + 13

- Simple web presentation: GitHub Pages and Markdown
- Where to go from here: Options for plotting, mapping and data analysis
 - If you want your own MATLAB, but don't have a license: Octave
 - If you want an all-around useful language with strong data analysis and mapping: Python with Matplotlib, Basemap and Cartopy
 - If you want a language that is great for statisticians and statistical modeling: R with RStudio
 - The General Mapping Tool (GMT): a Unix command line based mapping tool

ALSO: We will be working on getting you all started on your projects

Your BtM project on GitHub

You have received a document setting out the requirements for your project. The project must be submitted in the form of a GitHub repository, and the write-up must be presented on the web.

You have options for that! One is to use the built-in GitHub Pages

In your Repository, click on Settings, then scroll down and enable Pages. You can write the content in Markdown and then apply a template.

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Source

GitHub Pages is currently disabled. Select a source below to enable GitHub Pages for this repository. [Learn more.](#)

None ▾

Save

Automatic page generator

Create a beautiful site for your project with our automatic [GitHub Pages](#) generator. Author your content in our Markdown editor, select a theme, then publish.

Launch automatic page generator

Custom and Jekyll-based sites

To publish a page manually, push an HTML or [Jekyll](#) site to a branch named `gh-pages`. Read the [Pages help article](#) for more information.

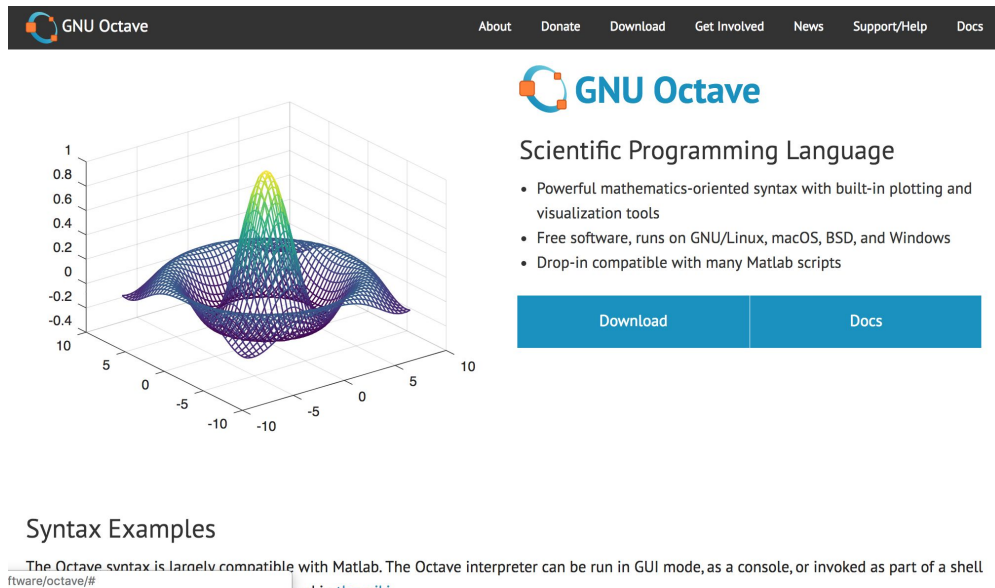
Markdown is a markup language (pun!)

Markup languages are computer languages in which *tags* are inserted in the content of a *document* to mark certain parts and elements of the document for presentation or logical structure. A popular markup language for data transmission is XML.

Many markup languages are *human readable*. Popular markup languages for text-based documents are RST (ReStructured Text) and Markdown. HTML is the markup language that web pages are written in. It was supposed to be human readable, but these days is often generated by computers and quite complex ... sometimes! More about Markdown: <https://markdown-it.github.io/> and <https://guides.github.com/features/mastering-markdown/> .

Octave is an open-source alternative to MATLAB

- You can download it for free at <https://www.gnu.org/software/octave/>
- GNU Octave supports most functions of MATLAB. It can be slower. It doesn't offer commercial support, and toolboxes. But it's not a bad solution to improve your MATLAB skills!



The screenshot shows the GNU Octave website. At the top is a dark navigation bar with the GNU Octave logo and links for About, Donate, Download, Get Involved, News, Support/Help, and Docs. Below the navigation bar is a large 3D plot of a function, likely a Gaussian or similar bell-shaped curve, rendered in a wireframe style with a color gradient from purple to yellow. To the right of the plot is the GNU Octave logo and the text "Scientific Programming Language". Below this are three bullet points: "Powerful mathematics-oriented syntax with built-in plotting and visualization tools", "Free software, runs on GNU/Linux, macOS, BSD, and Windows", and "Drop-in compatible with many Matlab scripts". At the bottom of this section are two blue buttons labeled "Download" and "Docs". Below the plot and text is a section titled "Syntax Examples" with a small text block that reads: "The Octave syntax is largely compatible with Matlab. The Octave interpreter can be run in GUI mode, as a console, or invoked as part of a shell".

Python and R are popular languages for scientific data processing and data visualization

- Both Python and R are **open-source** with very large and active user communities that contribute code and will help beginners
- Python is a general purpose language that is very widely used across all areas of science: on microcontrollers, on distributed systems, for web programming (interactive web sites), and for data processing. It is slower than compiled languages (FORTRAN, C), but easier to use.
- R started as a language specialized for statistics. It is a competitor of (expensive) commercial packages such as SAS and SPSS (used by economists). Like MATLAB, it works a lot with vectorized operations.

If you want to discover Python....

- Make sure you start with Python 3 (version 3.6 or later). The older version 2 (up to 2.7) is still in use, but now most people are moving to Python 3. If you have learned Python 3, working with Python 2 is not hard.
- If you are on a Linux machine, use the package manager (apt on Ubuntu) to install what you need:
<https://www.digitalocean.com/community/tutorials/how-to-install-python-3-and-set-up-a-local-programming-environment-on-ubuntu-16-04>
- If you are on Windows or Mac OS, I suggest you install the distribution called Anaconda: <https://www.anaconda.com/download/>
- You can install an IDE such as PyCharm or get used to a programming text editor. Good free options are Notepad++, Atom, VS Code, VIM and Bluefish.

If you want to discover R...

- Install the language itself from here: <https://www.r-project.org/>
- Install the IDE called RStudio from here: <https://www.rstudio.com/products/rstudio/> (commercial project, but the basic IDE is free)
- Install the libraries from the Tidyverse, which make data processing much much cleaner: <https://www.tidyverse.org/>

Python/R concepts that are different from MATLAB

- The concept of installing and importing packages (libraries)
- Defining functions is a lot easier than in MATLAB - do it often!
- The concept of *objects* that have *methods*

For example in Python, you can say:

```
>>> mystring = "hello, this is a string"  
>>> mystring.upper()  
'HELLO, THIS IS A STRING'  
>>> mystring.capitalize()  
'Hello, this is a string'  
>>> mystring.capitalize().lower()  
'hello, this is a string'
```

Which one is best? Up to you to choose!

You should make your selection based on your needs and the choices of your team, community or colleagues. Many have written tools for your use in seismology and other areas of earth science. R is widely used in ecology and general statistics. Python has fans in nearly all areas of science. More considerations: <https://towardsdatascience.com/r-vs-python-vs-matlab-vs-octave-c28cd059aa69>

	Pro	Con
MATLAB	Easy to get started with. Friendly IDE, easy plotting. Complete set of "Toolboxes".	Commercial (though Octave exists). Not easy to write modular code. Slowest.
Python	Open source. Fastest of the three. Widely used general purpose language. Large, friendly user community.	You need to give more thought about how to install a complete system, and which libraries you need.
R	Open source. Excellent for statistics and plotting. Widely used in data science. Friendly user community. Good tools.	Syntax can be confusing. Better for smaller-scale tasks.