The usage of computational techniques in the field of geology has increased over the past decade and has seen effective applications in the field over time. Through this document, you will be familiarized with the basic tools and technologies that are easy to learn and understand and at the same time are sophisticated enough to handle your workflow problems. The topics covered below are exceptional starting points into the field of geoinformatics and adoption of the concepts of data science into your research.

A) R and R Studio:

- R: <https://www.r-project.org/>

When you don’t like to code or don’t understand coding that well, but absolutely have to perform certain statistical calculations on a large dataset or need to represent your data coherently in a single graphic, R comes to the rescue. The R language is easy to pick up and learn in a relatively less amount of time for domain scientists and use it for the data collected through experiments and research.

Documentation: <https://www.r-project.org/about.html>

Download: <https://cran.r-project.org/mirrors.html>

- R Studio: <https://rstudio.com/>

The optimal way to reap benefits of R is by using it in R Studio, which is basically an integrated development environment, or in other words an easy double click application on your machine in which you will try out writing a few lines of the code.

Download: <https://rstudio.com/products/rstudio/download/>

NOTE: R Studio needs R installed on your machine to run. So, make sure that you have R installed first and then R Studio to get started.

R Tutorials:

1. <https://www.edx.org/learn/r-programming>
2. <https://www.statmethods.net/r-tutorial/index.html>
3. <https://www.tutorialspoint.com/r/index.htm>

B) Jupyter Notebooks: <https://jupyter-notebook.readthedocs.io/en/stable/notebook.html>

A Jupyter Notebook is a web-based application which allows in browser writing, editing and executing of code, documentation and communicating results. It supports R as well as Python. Jupyter Notebooks provide an interactive console-based approach, which makes it easier for members you work with to understand your code, results and goals. Think of it as a living online notebook which invites you or your collaborator to explore and experiment with the code you have written to study the work in depth for instance, they could try to tweak with the parameters of your experiment and see how it affects the results.

<https://jupyterlab.readthedocs.io/en/stable/getting_started/installation.html>

The Deep Carbon Observatory, the 4D Initiative, and other collaborations with domain scientists are great examples of the adoption of these concepts and technologies:

1) <https://deepcarbon.net/dco-webinar-wednesdays-summer-data-science-series>

2) <https://4d.carnegiescience.edu/>

To follow up on any of the above topics covered, or any question related to the field of geoinformatics, feel free to reach out to us at the Virtual Data Help Desk, tweet using #DataHelpDesk and #GSA2020 or submit your question at <http://sgiz.mobi/s3/Data-Help-Desk-Questions>

Shweta Narkar

PhD Student

Tetherless World Constellation

Rensselaer Polytechnic Institute

 