

Intro to Informatics for the Brain Resilience Study

Part 2: Data Wrangling with Python

Session Details

- Interactive
 - you'll need to be on a computer to follow along
- Make sure you have your login info for Digital Research Alliance (Compute Canada)
 - Should be the same credentials you use to log into <https://ccdb.alliancecan.ca/security/login>
 - Also have your MFA device ready
- Feel free to interrupt or leave a message in the chat if you have questions, need something repeated, or are having trouble with any steps
- This session will be recorded – but only for internal use, within the BRS

Outline

- Using JupyterHub on cedar
- Working with data on cedar
 - Extracting subsets of data (e.g. participants with BAI/BDI)
 - Loading in tabular data
 - Visualizing tabular data (e.g. printing tables, quick plots)
 - Simple analyses (e.g., correlations)
 - Copying subsets of data from projects to scratch
 - Generating a new summary score for a given participant
- Running scripts from command line

What is JupyterHub?

- **JupyterHub** is a **web-based interface** that lets you access Cedar and run code in a browser
- JupyterLab and notebooks are meant for **short** interactive tasks such as testing, debugging or quickly visualize data (few minutes).
- Notebooks also allow you to:
 - include both code and documentation/notes in one place
 - create and visualize outputs easily
- Running longer analyses must be done in a **non-interactive job (sbatch)**

Getting started: JupyterHub

JupyterHub for Cedar:







<https://jupyterhub.cedar.alliancecan.ca/>

Documentation:

<https://docs.alliancecan.ca/wiki/JupyterHub>

JupyterHub on clusters

On the following clusters[†], use your Alliance username and password to connect to JupyterHub:

JupyterHub	Comments
Béluga 	Provides access to JupyterLab servers spawned through jobs on the Béluga cluster.
Cedar 	Provides access to JupyterLab servers spawned through jobs on the Cedar cluster.
Narval 	Provides access to JupyterLab servers spawned through jobs on the Narval cluster.
Niagara 	Provides access to JupyterLab as one of the applications of the SciNet Open OnDemand portal. To learn more, see the wiki page  .
Graham 	Provides access to JupyterLab servers spawned through jobs on the Graham cluster.

Server Options

Reservation

None

Account

def-_cpu

Time (hours)

2.0

Number of cores

1

Memory (MB)

2375

☐ **Enable core oversubscription?** Recommended for interactive usage

GPU configuration

None

User interface

JupyterLab

Start

Now let's work in the notebook!