

WEB-BASED TOOLS FOR DATA ANALYSIS: JUPYTERLAB ENVIRONMENT AND WORKFLOW OPTIMIZATION

M Portela, November 3, 2020

The material is available in **GitHub**

https://github.com/reisportela/R_Training

1. Operating system

- Linux (e.g., Ubuntu 20.04), OSX Catalina, Windows 10

2. Packages

Windows: consider installing Chocolatey, a package manager for Windows (similar to `yum` in CentOS or `brew` in OSX)

- Python: install Anaconda – <https://www.anaconda.com>

Example, using Chocolatey: `choco install anaconda3`

or download and install

- **R:** <https://www.r-project.org>
- **Julia:** <https://julialang.org>
- **Stata:** <https://www.stata.com>

Recomendation: install RStudio

3. Jupyter

“The Jupyter Notebook is an open-source web application that allows you to **create and share** documents that contain *live code*, equations, visualizations and **narrative text**. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.”

3.1 Install jupyter

- Open a Terminal in either Linux or OSX
- Open Windows Powershell as Administrator

Run the following lines

- **jupyter notebook:** `pip install notebook` **or** `conda install -c conda-forge notebook`
- **jupyter lab:** `pip install jupyterlab` **or** `conda install -c conda-forge jupyterlab`

3.2 Install your kernels

- **Python:** this should be the first one installed ipykernel
- **R:** `irkernel`

Open an R console, e.g. within RStudio, and execute sequentially, `install.packages('IRkernel')`, `IRkernel::installspec()`

Add `Node.js` and `npm`

Visit [Nodejs.org](https://nodejs.org) to **install Node.js** and `npm`

- **Julia:** IJulia

Run Julia and execute sequentially, *using Pkg, Pkg.add("IJulia")*

- **Stata:** stata_kernel

for Stata see the instructions by Kyle Barron

Magics – “Magics are programs provided by stata_kernel that enhance the experience of working with Stata in Jupyter.”

3.3 Start ‘notebook’ or ‘lab’

Open a Terminal/Power shell, move to your working folder and type:

- **jupyter notebook:** *jupyter notebook*
- **jupyter lab:** *jupyter lab*

It should open your browser with the notebook and the installed kernels.

3.4 Remove a Kernel

```
jupyter kernelspec list
jupyter kernelspec uninstall unwanted-kernel
```

4. Binder

Running R Projects in MyBinder: Dockerfile Creation With Holepunch

- myBinder
- Gesis Notebooks

Check the following link

Configuration Files

mybinder allows you to create a linked icon to your interactive notebook



Check the example in GitHub with RStudio & R 3.6 + Python + Julia + Stata

reisportela/prjs

- or a setup where we can build a notebook with Python 3.0 or R (you can also run RStudio from this link)



Even better, use GESIS notebooks to launch your image

The concept using GESIS

MyBinder: EXAMPLES

5. A gallery of interesting Jupyter Notebooks

- Gallery
- Plotting and Programming in Python
- Exploratory data analysis in Python

6. Books

- Python Data Science Handbook
- Bookdown
- How to Hide all the code cells in Jupyter Notebook Python with single Click

7. Checks

- Binder Multi-language demo
- mybinder.io

8. SoS NOTEBOOK

- Local installation

pip installation

```
pip3 install sos
```

```
pip3 install sos-pbs
```

```
pip3 install sos-notebook
```

```
pip3 install sos-papermill
```

```
pip3 install sos-r
```

```
pip3 install sos-julia
```

```
pip3 install sos-stata
```

```
python3 -m sos_notebook.install
```

```
jupyter kernelspec list
```

```
jupyter notebook
```

9. Discussion on Julia

- use an environment julia-python
- a multi-language-demo
- Using Julia in Binder: interactive web environment for running your code

By default I will not activate a machine running Python, R and Julia as it takes too long to build the image. I recomend using the link.

10. GESIS Notebooks

- Create a login in GESIS Notebooks and add your machine (running RStudio)

Build and launch a new repository

GitHub repository name or URL

GitHub ▾

https://github.com/reisportela/prjs

Git branch, tag, or commit

Git branch, tag, or commit

URL to open (optional)

rstudio

URL ▾

launch

or a **Jupyter Lab**

GitHub repository name or URL

GitHub ▾

https://github.com/reisportela/R_Training

Git ref (branch, tag, or commit)

HEAD

URL to open (optional)

lab

URL ▾

launch

Copy the URL below and share your Binder with others:

https://mybinder.org/v2/gh/reisportela/R_Training/HEAD?urlpath=lab

📋

11. Further notes

11.1 Jupyter's extensions

```
conda install -c conda-forge jupyter_contrib_nbextensions
```

```
jupyter contrib nbextension install --user
```

11.2 Kaggle Kernels

Kaggle

11.3

How to Hide all the code cells in Jupyter Notebook Python with single Click

11.4 Pandas

Pandas cookbook

R and Dropbox

rdrop2

12. Usefull links

Binder

CODE OCEAN

GESIS Notebooks

Hypernet Labs

IBM Skills Network Lab

RStudio Cloud