

# Jupyter Notebooks for Performing & Sharing Bioinformatics Analyses

Jonathan Dursi, OICR

[jonathan@dursi.ca](mailto:jonathan@dursi.ca)

<https://github.com/ljdursi/glbio-jupyter-workshop>

# The plan for this morning

- Run Jupyter notebooks on your laptop, in R and Python
- How to perform interactive analyses in a web browser using Jupiter
- How does it all work?
- Using markdown and latex to format notebooks nicely
- “Port” an R bioinformatics workflow from scripts into a Jupyter notebook
- Sharing a Jupyter notebook online:
  - SageMathCloud
  - GitHub and git; nbviewer
  - [mybinder.org](https://mybinder.org)

# The plan for this morning

Perfectly reasonable and interesting things to discuss that we *won't* be talking about today:

- Combining R and Python: cool and possible but requires a little setup
- IPython magics (probably)

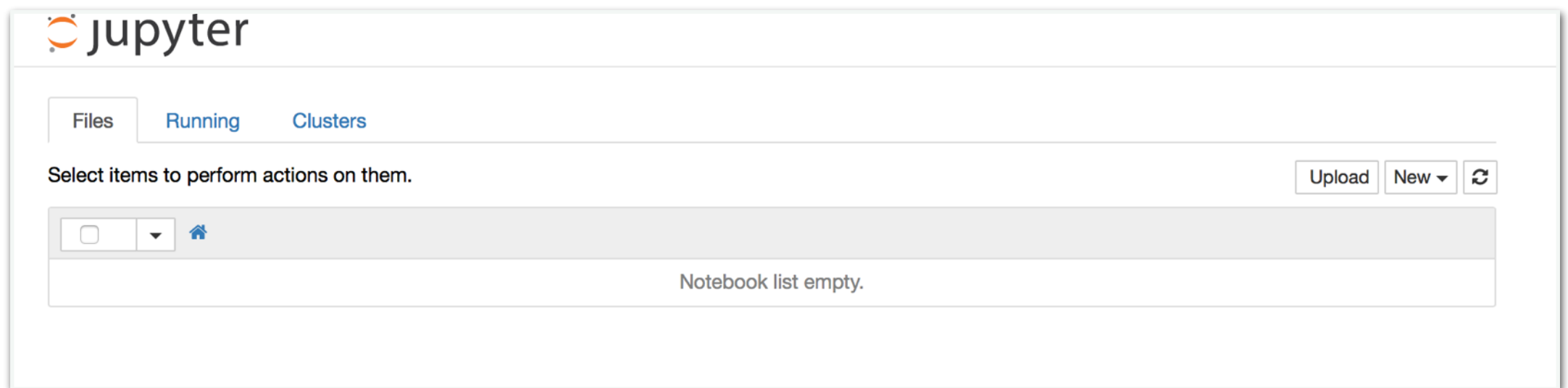
# Let's start up a notebook

- Make a Notebooks directory and start from there
- Command Line: Linux, Mac

```
1. python3.5
coredump:~ ljdursi$ mkdir ~/Desktop/Notebooks
coredump:~ ljdursi$ cd ~/Desktop/Notebooks/
coredump:Notebooks ljdursi$ jupyter notebook
[I 12:56:20.381 NotebookApp] The port 8888 is already in use, trying another random port.
[I 12:56:20.390 NotebookApp] Serving notebooks from local directory: /Users/ljdursi/Desktop/Notebooks
[I 12:56:20.391 NotebookApp] 0 active kernels
[I 12:56:20.391 NotebookApp] The Jupyter Notebook is running at: http://localhost:8889/
[I 12:56:20.391 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
```

# Let's start up a notebook

- Make a Notebooks directory and start from there
- Command Line: Linux, Mac
- Browser should open (or go to localhost:8888)



# Let's start up a notebook

- Or you can start from home (eg, launch from start menu:Win) and navigate folders, and create new folders:

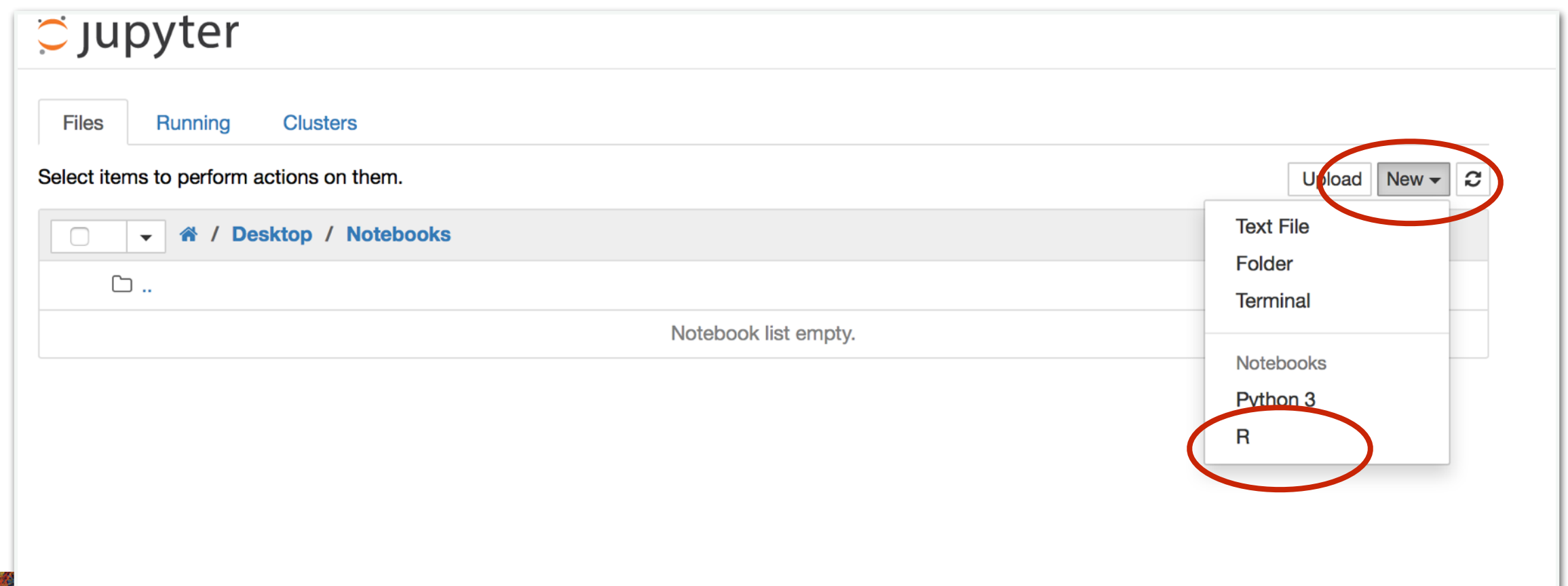
 jupyter





# Let's start up a notebook

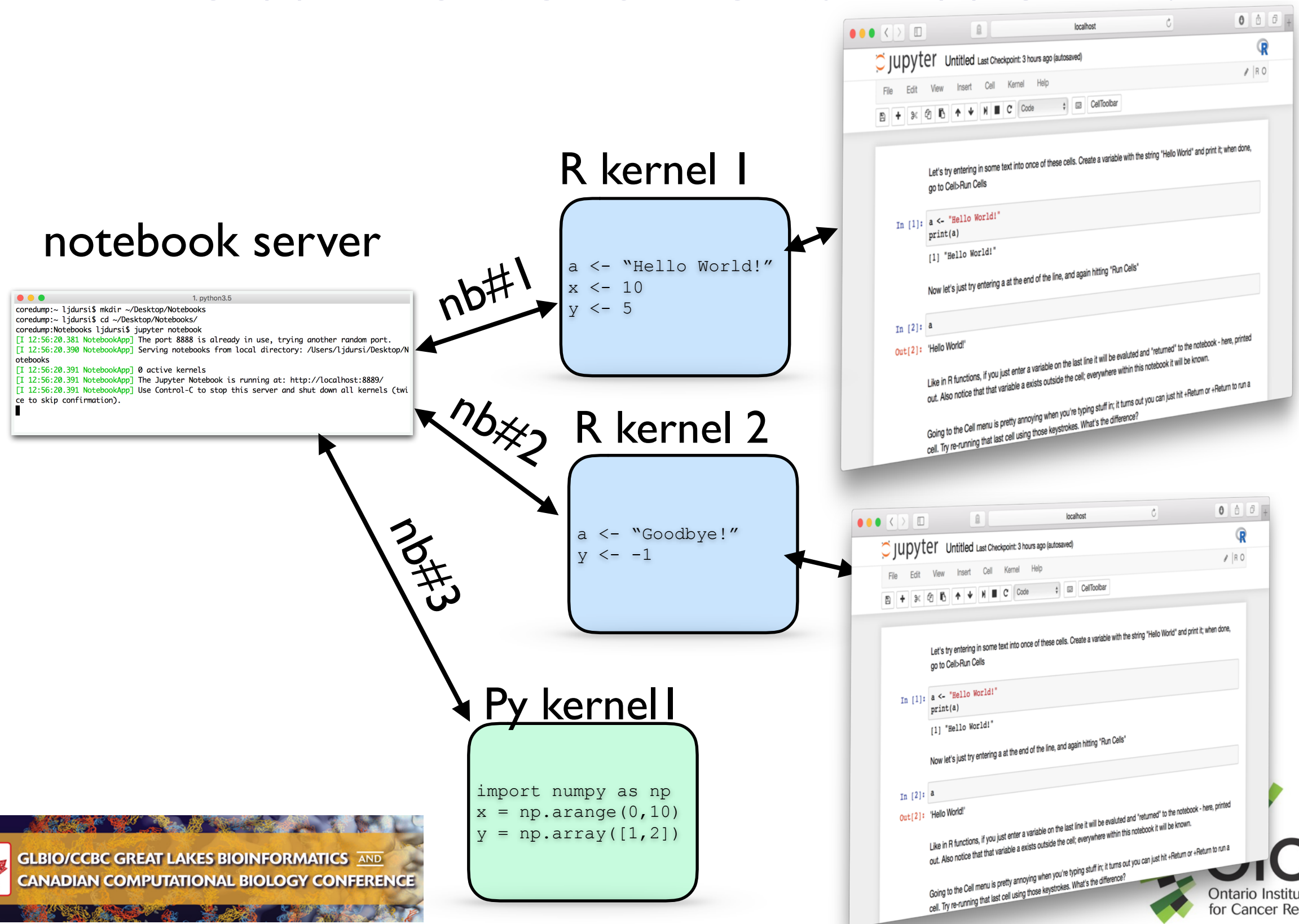
- Let's start up an R notebook



# Demo - starting notebook

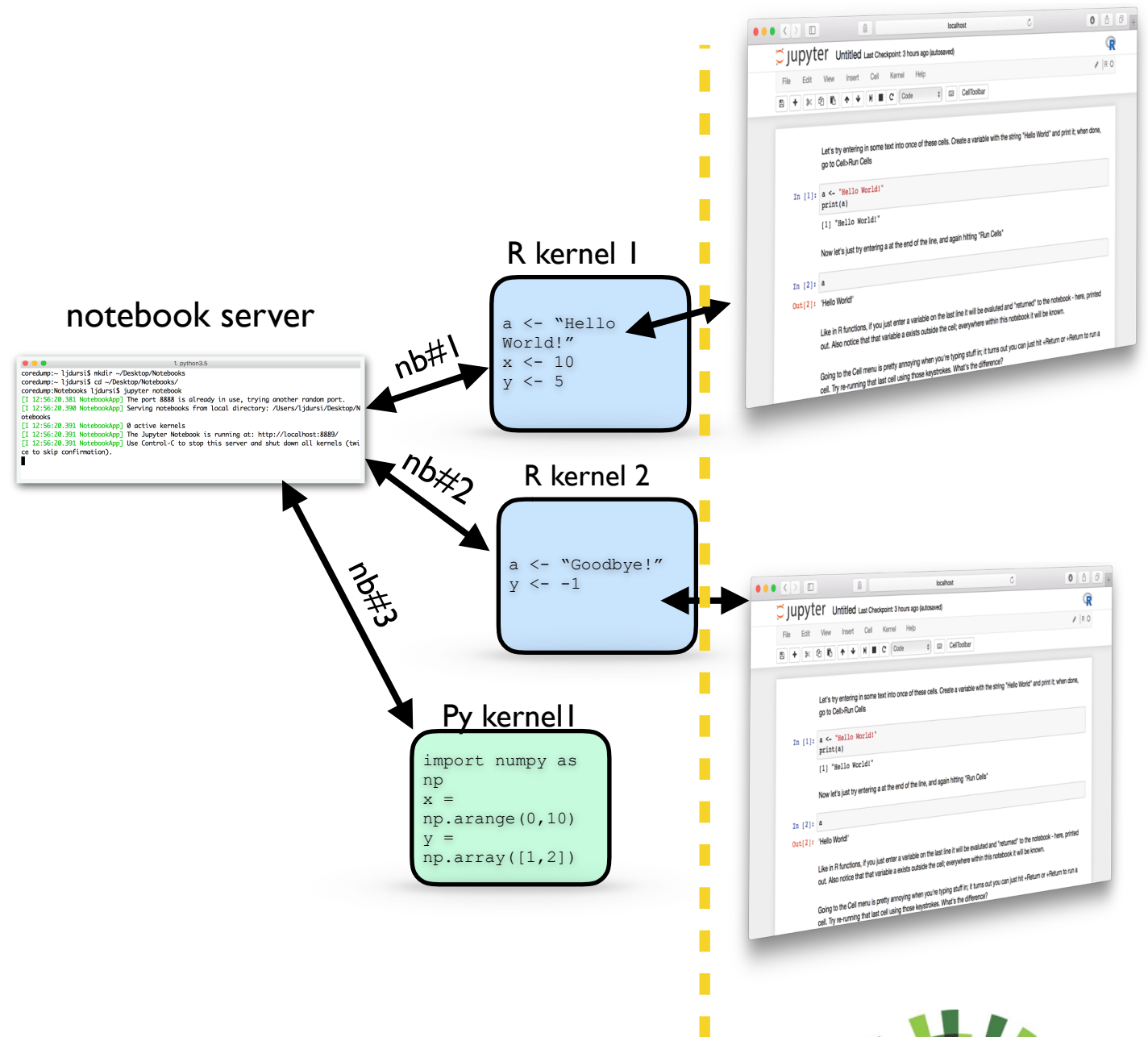


# How does this all work?



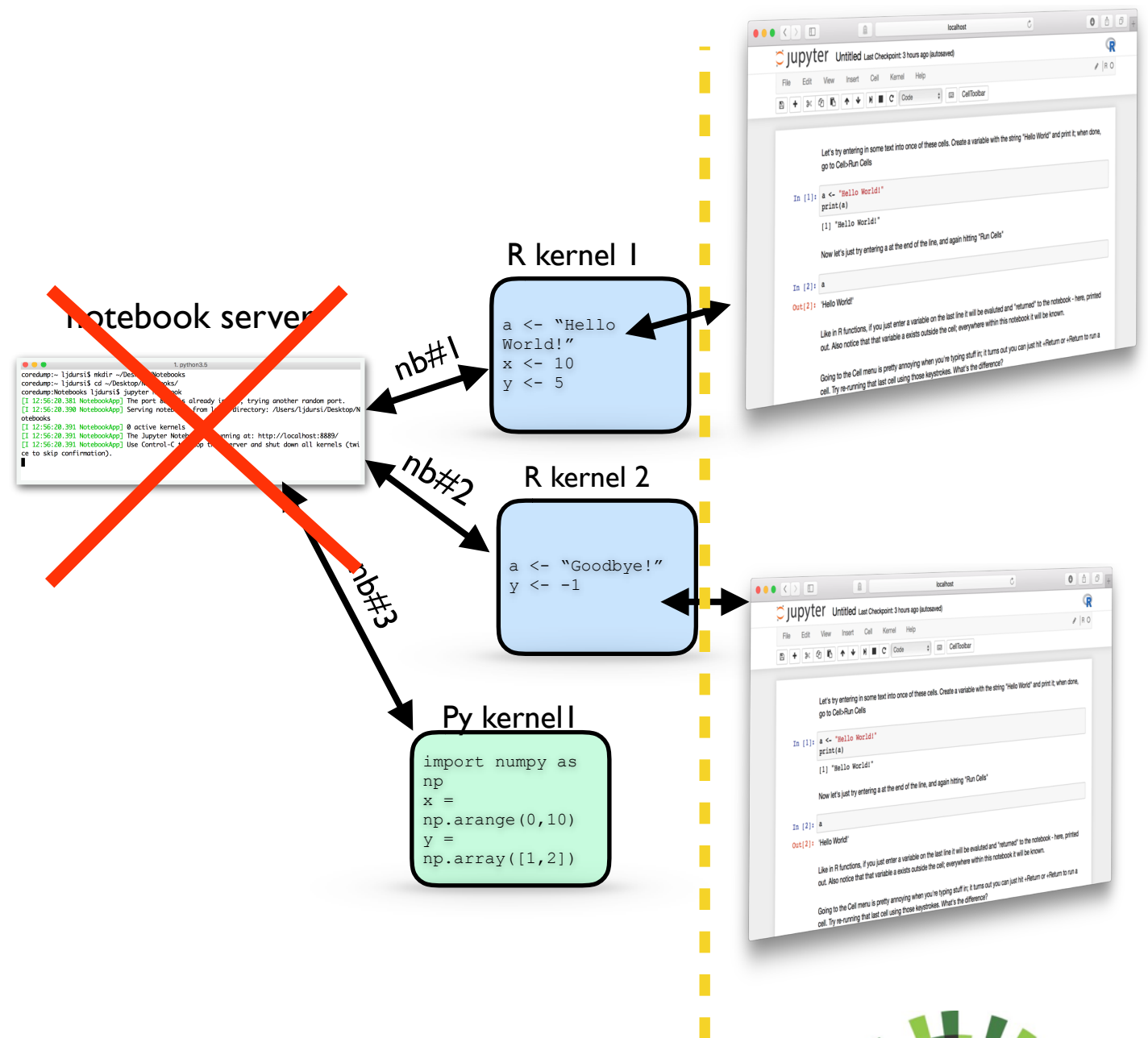
# How does this all work?

- There's no rule that says the notebook server/kernels has to be on same machine as browser!
- Can be running on remote server (or cluster)
- AWS demo

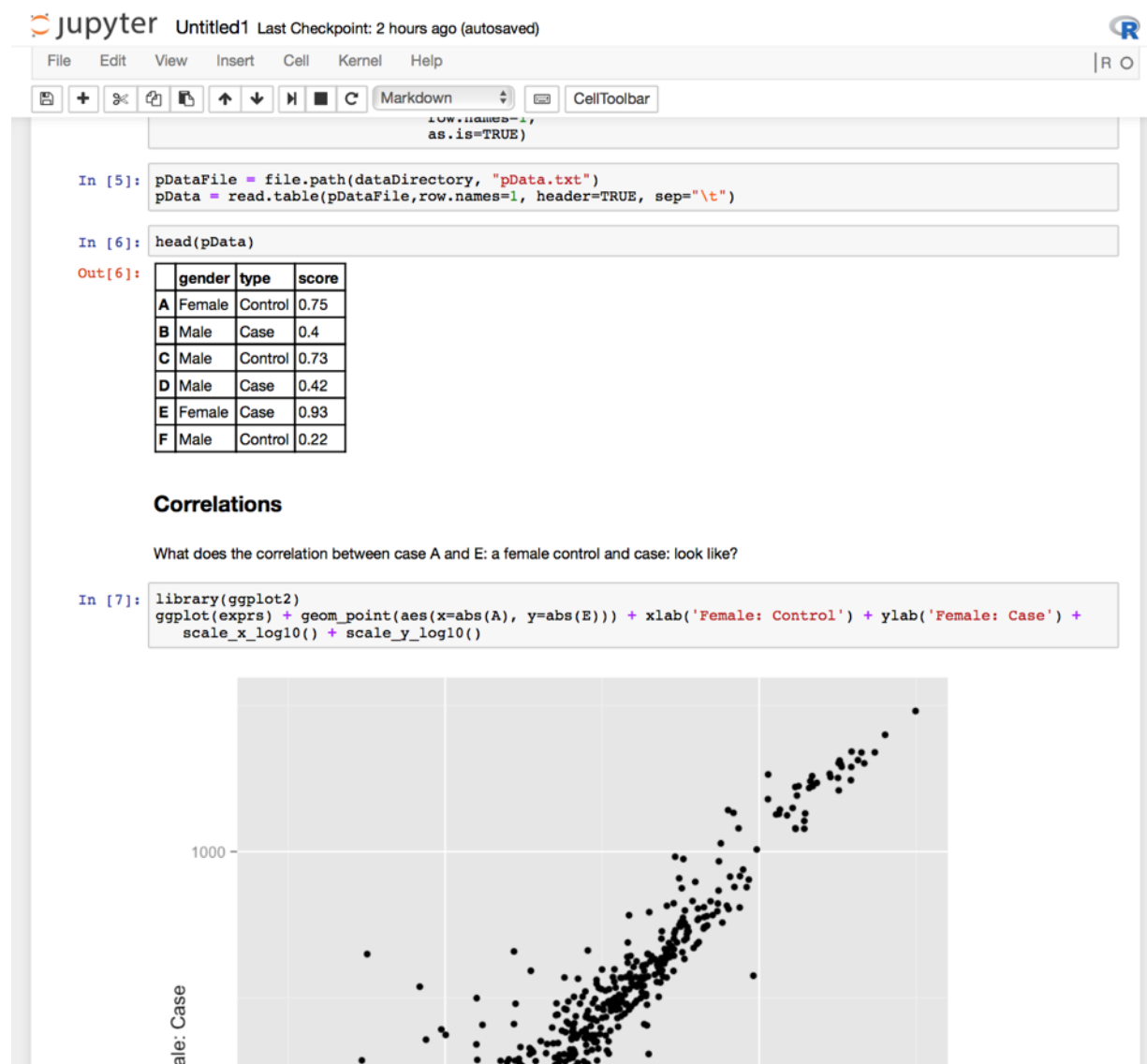


# How does this all work?

- AWS demo...
- What happens if the notebook server dies?
- Demo



# All input+output in ipynb file



```
{
  "cells": [
    ...
    {
      "outputs": [],
      "source": [
        "data <- read.csv('http://dursi.ca/content/images/GSE37704_featurecounts.csv')\n",
        "metadata <- read.csv('http://dursi.ca/content/images/GSE37704_metadata.csv')\n"
      ]
    },
    ...
    {
      "outputs": [
        {
          "name": "stdout",
          "output_type": "stream",
          "text": [
            "'data.frame':\t19808 obs. of  8 variables:\n",
            " $ ensgene  : Factor w/ 19808 levels \"\tENSG000000000003\",...: 15488 19676 19628 19554 19334 15792 16078 15
            867 15778 15793 ... \n",
            " $ length   : int   918 718 1982 939 939 3214 5539 3395 2833 3424 ... \n",
            " $ SRR493366: int    0  0 23  0  0 124 1637 120 24  4 ... \n",
            " $ SRR493367: int    0  0 28  0  0 123 1831 153 48  9 ... \n",
            " $ SRR493368: int    0  0 29  0  0 205 2383 180 65 16 ... \n",
            " $ SRR493369: int    0  0 29  0  0 207 1226 236 44 14 ... \n",
            " $ SRR493370: int    0  0 28  0  0 212 1326 255 48 16 ... \n",
            " $ SRR493371: int    0  0 46  0  0 258 1504 357 64 16 ... \n"
          ]
        }
      ]
    },
    ...
    {
      "outputs": [
        {
          "data": {
            "image/png":
              "iVBORw0KGgoAAAANSUgEUGAAA0gAAANICAYAAAD958/bAAAEDWLDQ1BJQ0MgUHJvZmlsZQAQAQ
              I2NVV1oHFUUPrtZyMkz1NsNIV0qD8NJQ2TVjShtLp/3d02bpZJNtoi6GT27s6Yyc44M7v9oU9
              FUHw6psUxL+3gCAo9Q/bPrQvLQol2tQgKD60+INQ6Ium65k7M5LpurHeZe58853vnnvuuWfvB
          "
        }
      ]
    }
  ]
}
```

- All the input+output is located in the .ipynb file