Summer R Workshop 1

Matthew D. Turner

mturner46@gsu.edu

Department of Psychology

Georgia State University

R at GSU

- This past spring we started a new R Users Group
  - It began as a regular meeting in psychology convened by Chris Goode
  - I basically hijacked it to start a GSU-wide users group
  - If you are not on the email list, please add it or contact me (mturner46@gsu.edu) to be added
  - We were overwhelmed with the response
  - Biggest Need at GSU: Training

What is R?

- R is a *variant* of the **S programming language**
- The S language was developed in 1976 at Bell Labs
  - S was designed for data analysis and statistical modeling
  - S Developers: **Rick Becker**, **Allan Wilks**, **John Chambers**
  - S was developed by people working on particularly hard data analysis problems where standard solutions did not usually apply
- S Goal: *"to turn ideas into software, quickly and faithfully"* (Chambers)

What is R?

S → S-Plus → R

- R was developed by **Ross Ihaka and Robert Gentleman** at University of Auckland, New Zealand
- It is an open source version of the commercial language S-Plus
  - S-Plus was an earlier derivative of S developed by Statistical Sciences, Inc. (SF, CA)
  - S-Plus introduces many features not in the original S (much of the OOP system)
  - S-Plus was basically made non-viable as a commercial product due to R
- R is now the **dominant dialect** of this family of S based languages

What is R?

- R is now an international project
  - R language is controlled by the R Core Team
  - Managed by the R Consortium:

    www.r-consortium.org

  - Shared/Delivered by CRAN: cran.r-project.org
  - Package development is shared by the global R community

What is R?

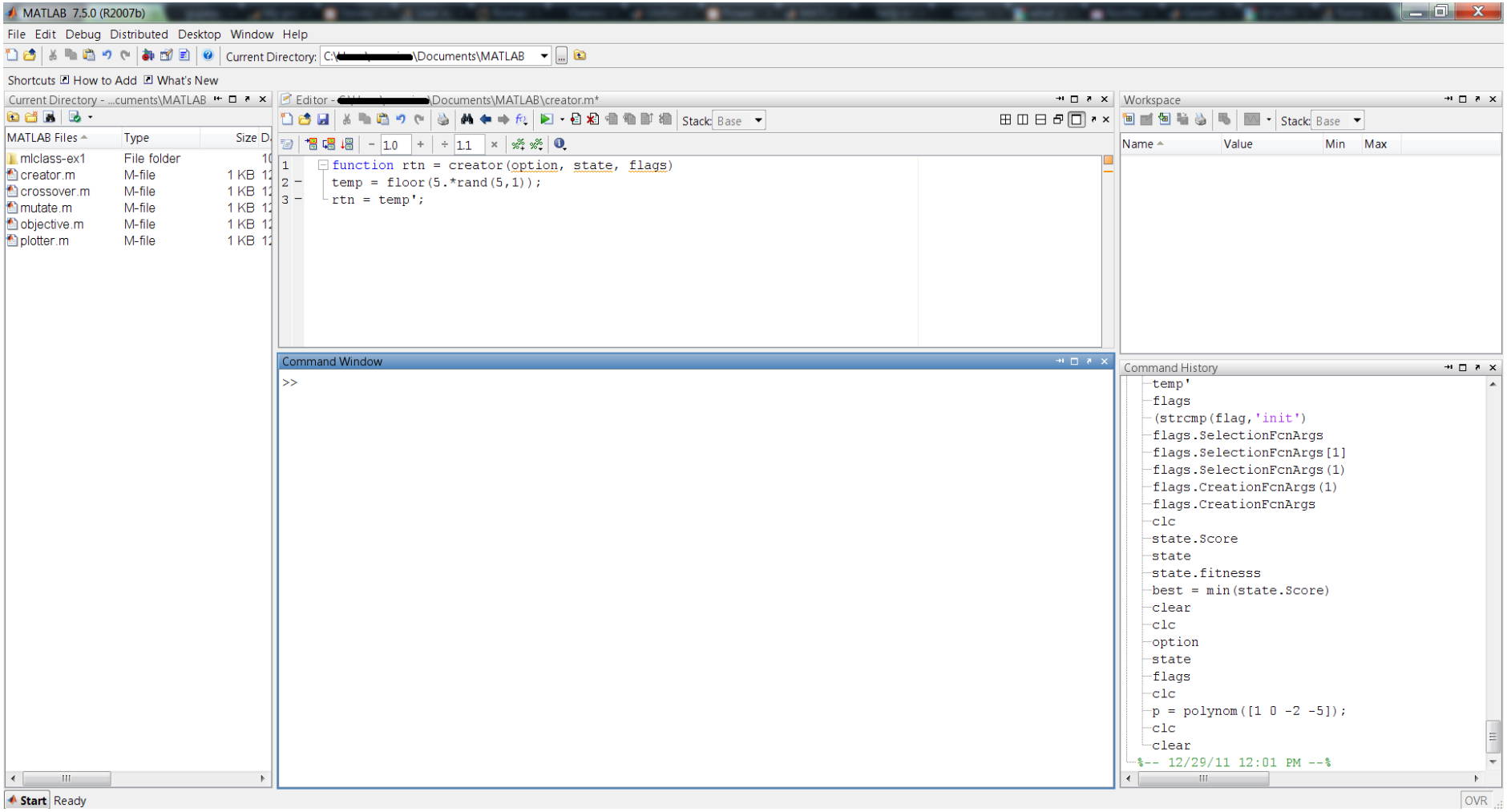
- S/R is an **opinionated language**, it emphasizes:
  - "**Exploratory data analysis**" (Tukey's ill-conceived term)
  - **Quick/Exploratory Graphics**
  - Easy and direct mathematical manipulations in models
  - Programmatic interface to analysis (functions)
- R is an evolutionary product
  - Many of its features only make sense if you know the history
  - "Vestigial" components abound!

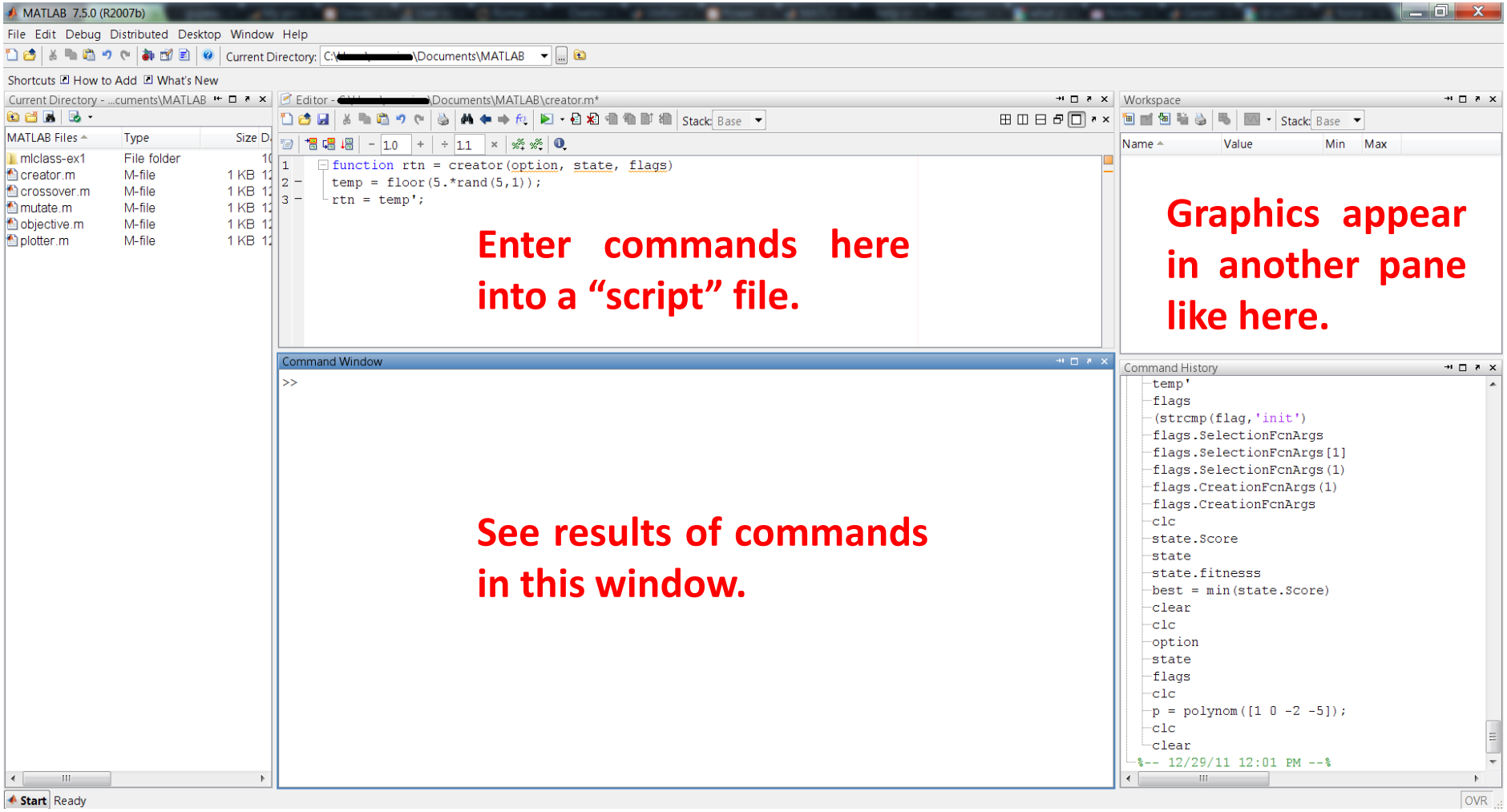
What is R?

- For most people R is:
  - A general purpose computer language for data
  - A statistics language
  - The *koiné* of the data science and statistics worlds
  - Most new statistics are implemented in R before they are elsewhere

Jupyter Notebooks

- There are two common styles of interface to scientific computing software
  - **Window-pane** GUI AKA Matlab-style
  - **Notebooks** AKA Mathematica-style


MATLAB 7.5.0 (R2007b)
File Edit Debug Distributed Desktop Window Help
Current Directory: C:\[redacted]\Documents\MATLAB
Shortcuts How to Add What's New
Current Directory - ...cuments\MATLAB
MATLAB Files Type Size
mlclass-ex1 File folder
creator.m M-file 1 KB
crossover.m M-file 1 KB
mutate.m M-file 1 KB
objective.m M-file 1 KB
plotter.m M-file 1 KB
Editor - C:\[redacted]\Documents\MATLAB\creator.m*
Stack: Base
1 function rtn = creator(option, state, flags)
2 temp = floor(5.*rand(5,1));
3 rtn = temp';
Workspace
Stack: Base
Name Value Min Max
Command Window
>>
Command History
temp'
flags
(strcmp(flag,'init')
flags.SelectionFcnArgs
flags.SelectionFcnArgs[1]
flags.SelectionFcnArgs(1)
flags.CreationFcnArgs(1)
flags.CreationFcnArgs
clc
state.Score
state
state.fitnesss
best = min(state.Score)
clear
clc
option
state
flags
clc
p = polynom([1 0 -2 -5]);
clc
clear
%-- 12/29/11 12:01 PM --%
Start Ready
OVR

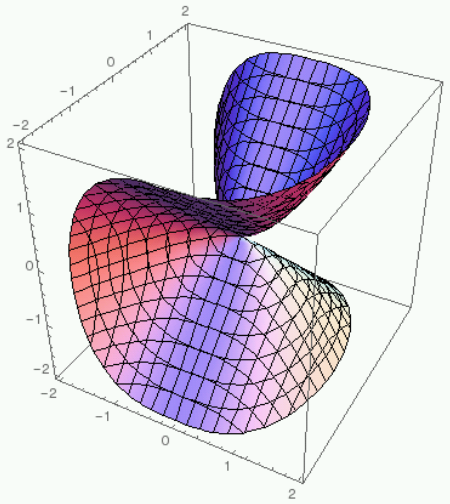

MATLAB 7.5.0 (R2007b)
File Edit Debug Distributed Desktop Window Help
Current Directory: C:\[redacted]\Documents\MATLAB
Shortcuts How to Add What's New
Current Directory - ...cuments\MATLAB
MATLAB Files Type Size
mlclass-ex1 File folder
creator.m M-file 1 KB
crossover.m M-file 1 KB
mutate.m M-file 1 KB
objective.m M-file 1 KB
plotter.m M-file 1 KB
Editor - [redacted]\Documents\MATLAB\creator.m*
1 function rtn = creator(option, state, flags)
2 temp = floor(5.*rand(5,1));
3 rtn = temp';
Enter commands here into a "script" file.
Command Window
>>
See results of commands in this window.
Workspace
Name Value Min Max
Graphics appear in another pane like here.
Command History
temp'
flags
(strcmp(flag,'init')
flags.SelectionFcnArgs
flags.SelectionFcnArgs[1]
flags.SelectionFcnArgs(1)
flags.CreationFcnArgs(1)
flags.CreationFcnArgs
clc
state.Score
state
state.fitnesss
best = min(state.Score)
clear
clc
option
state
flags
clc
p = polynom([1 0 -2 -5]);
clc
clear
%-- 12/29/11 12:01 PM --%
Start Ready
OVR

Untitled-5 * - Wolfram Mathematica

File Edit Insert Format Cell Evaluation Palettes Window Help

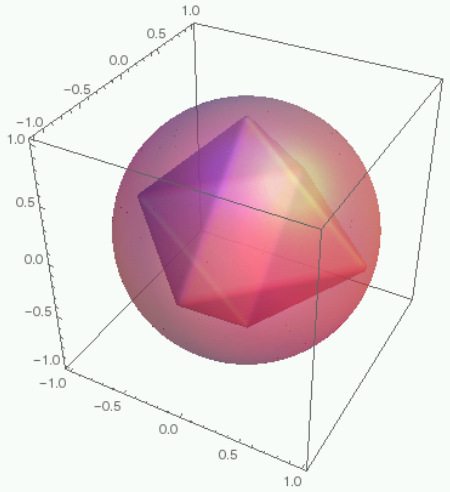
In[8]:=
```
ContourPlot3D[x^3 + y^2 - z^2 == 0, {x, -2, 2}, {y, -2, 2}, {z, -2, 2}]
```

Out[8]=



In[10]:=
```
RegionPlot3D[1 ≤ Norm[{x, y, z}, 1] && Norm[{x, y, z}] ≤ 1, {x, -1, 1},
 {y, -1, 1}, {z, -1, 1}, Mesh → None,
 PlotStyle → Directive[Opacity[0.5], Pink, Specularity[White, 20]],
 PlotPoints → 60, NormalsFunction → "Average"]
```

Out[10]=



100%

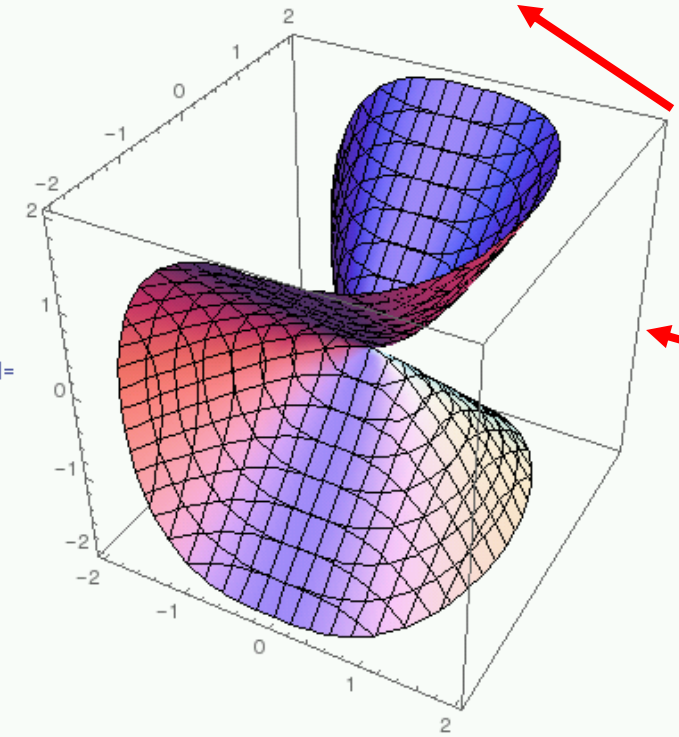
Everything appears in one window, in sequence.

Untitled-5 * - Wolfram Mathematica

File Edit Insert Format Cell Evaluation Palettes Window Help

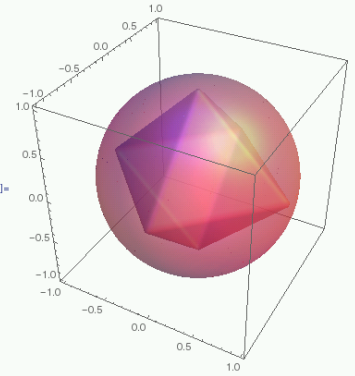
In[8]:= ContourPlot3D[x^3 + y^2 - z^2 == 0, {x, -2, 2}, {y, -2, 2}, {z, -2, 2}]

Out[8]=



In[10]:= RegionPlot3D[1 ≤ Norm[{x, y, z}, 1] && Norm[{x, y, z}, 1] < 2, {x, -1, 1}, {y, -1, 1}, {z, -1, 1}, Mesh → None, PlotStyle → Directive[Opacity[0.5], Pink, Specular], PlotPoints → 60, NormalsFunction → "Average"]

Out[10]=



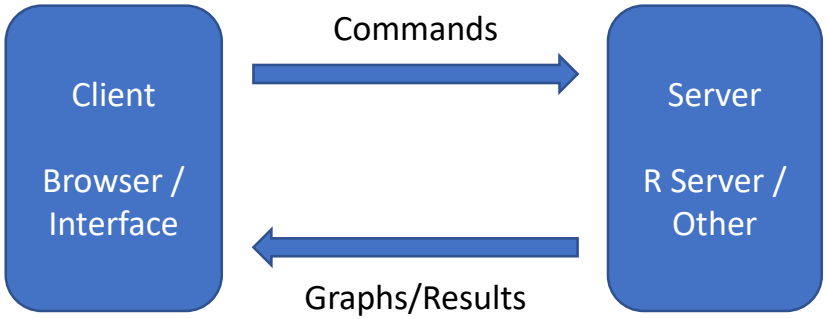
Each input is followed by…

Its corresponding output.

Jupyter Notebooks

- There are two common styles of interface to scientific computing software
  - Window-pane GUI AKA Matlab-style
  - Notebooks AKA Mathematica-style
- For R **both are available**
  - **RStudio** implements the Matlab-style interface
  - Jupyter implements the Mathematica-style interface
- **R is the SAME in both!**
  - Both RStudio and Jupyter use **exactly the same** R "server" underneath
  - Same language, same implementation, same everything

Scientific Software Architecture



One nice feature of this design is that it works cross-platform – that is, PC/Mac/Linux works the same.

- The server part of this can be located anywhere you are connected to:
- On the internet (like today): **`https://dice.gsu.edu:8000/hub/login`**
- On the same computer as the client: **`http://localhost:5000`**

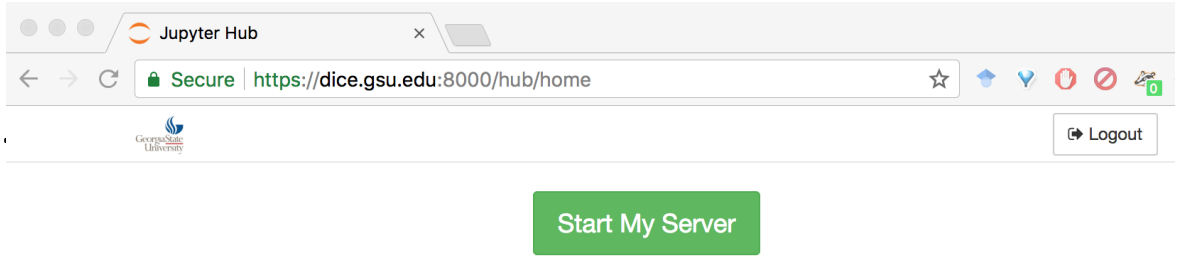
Jupyter Notebooks

- If you like Jupyter it supports most scientific programming languages:
  - Old timey: **FORTRAN** (coarray-fortran 2008)
  - Statistics programming: **SAS**, **R**, Incanter/Clojure
  - Math programming: **Matlab**, Julia (IJulia), Maxima,
  - Plotting/Graphics: **Gnuplot**, D3/Javascript
  - AI: prolog, smalltalk, Mathematica, scheme (LISP dialect)
  - Web languages: Ruby, Haskell, Javascript, Coffeescript
  - System languages: C, Go, Scala, Erlang, bash, Kotlin
- And many more

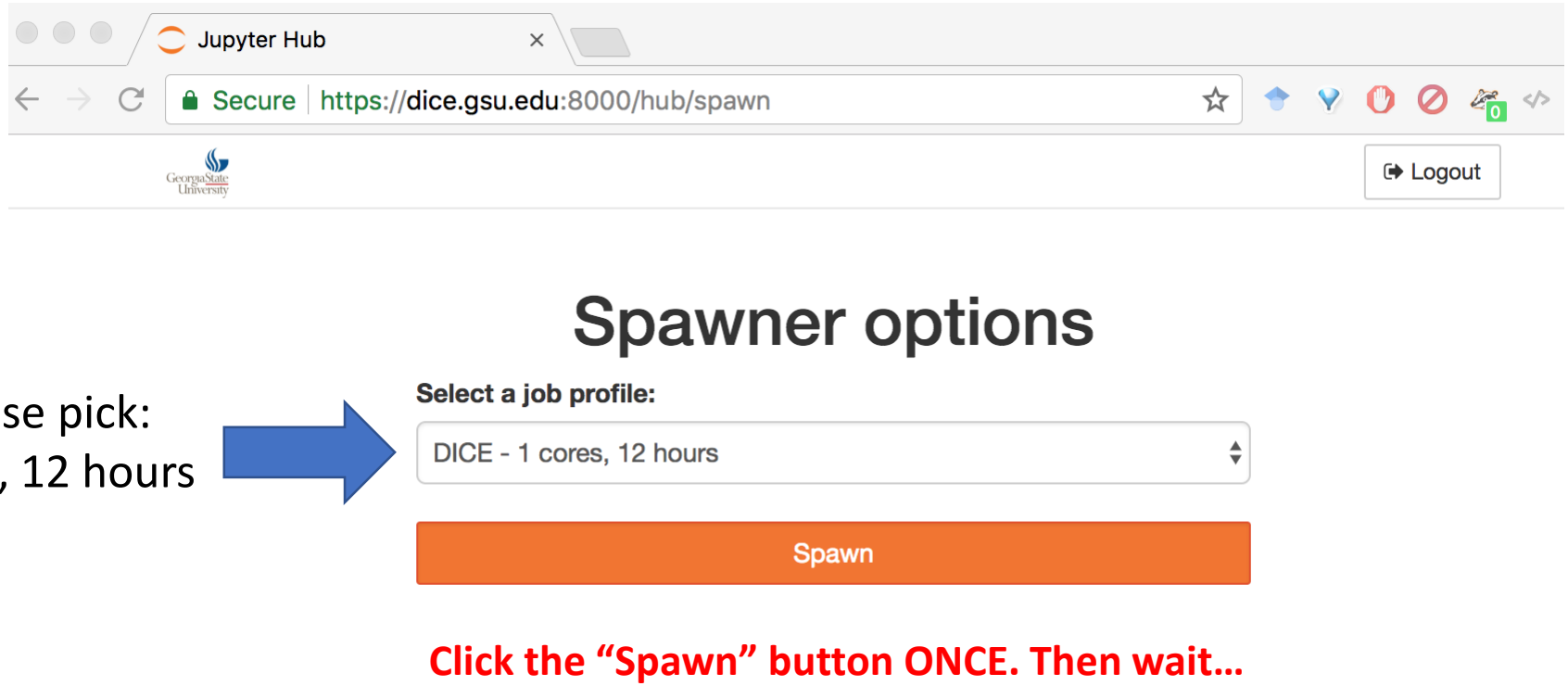
https://github.com/jupyter/jupyter/wiki/Jupyter-kernels

Workshop

- Log into DICE:
  - https://dice.gsu.edu:8000/hub/login
  - The minimal address: dice.gsu.edu:8000 should work, but sometimes does not
- Then you see

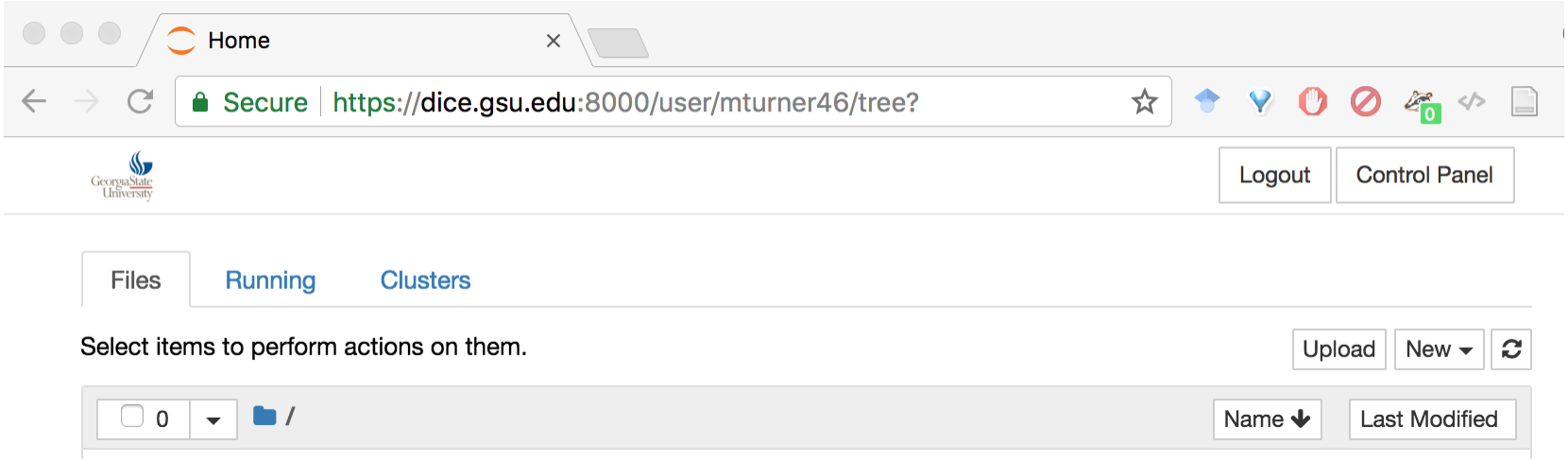


Click the "Start My Server" button.



Please pick:
1 core, 12 hours

Click the "Spawn" button ONCE. Then wait…



When you see a page where the TOP of the page looks like this, you are ready to start!

DICE Demo

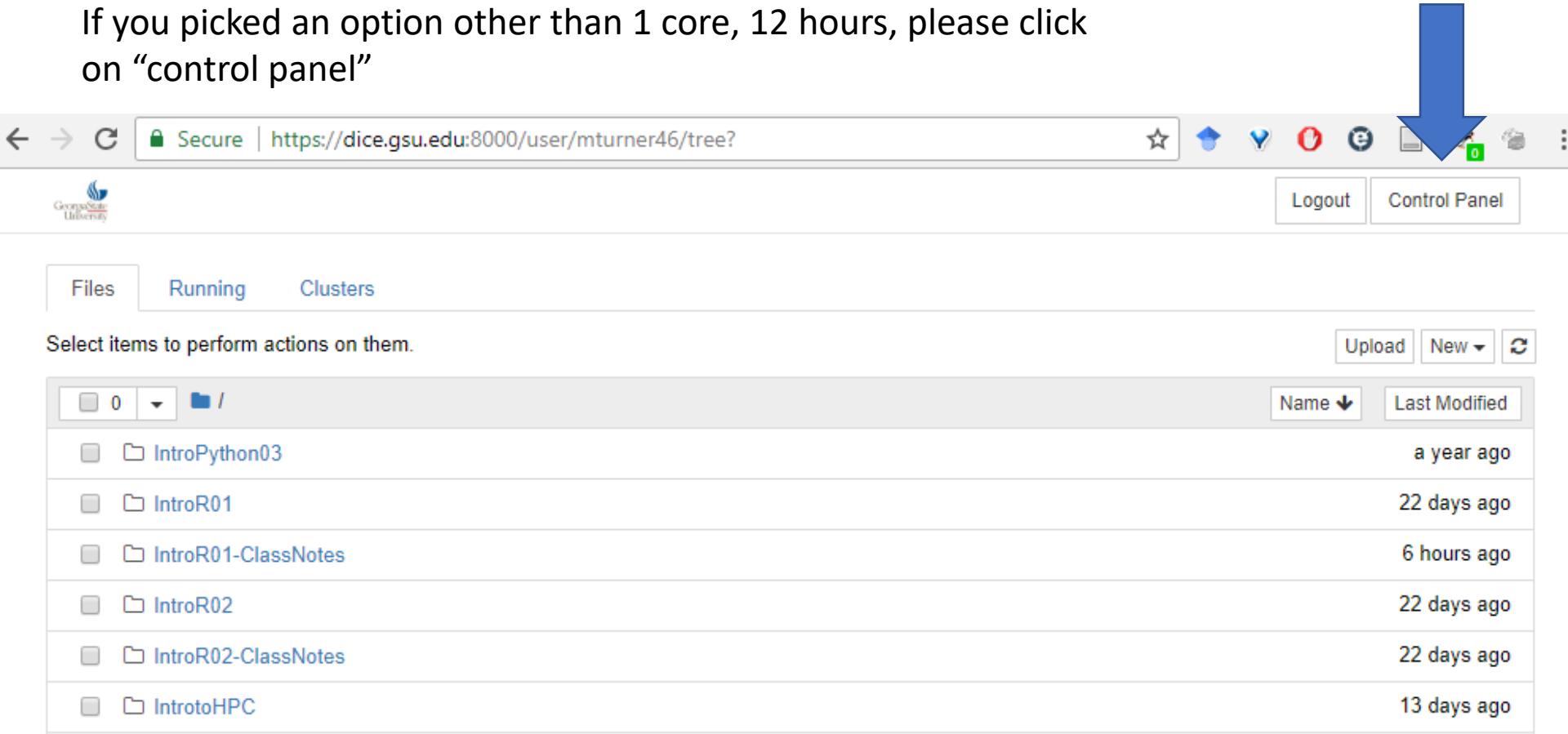
Today's Files

- Click on the folder: IntroR01
- And within that folder: ClassNotes

- The Jupyter notebooks for today are numbered from 01 to 05

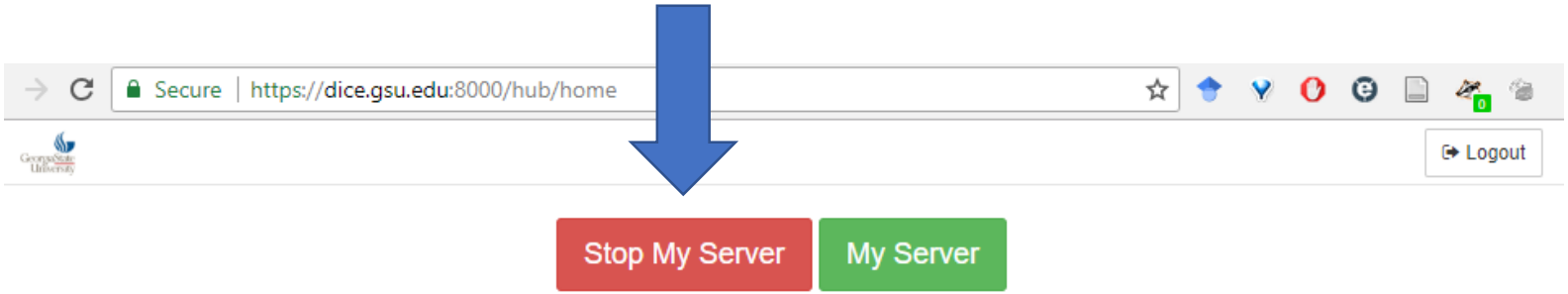
- Launch the first and go!

If you picked an option other than 1 core, 12 hours, please click on “control panel”



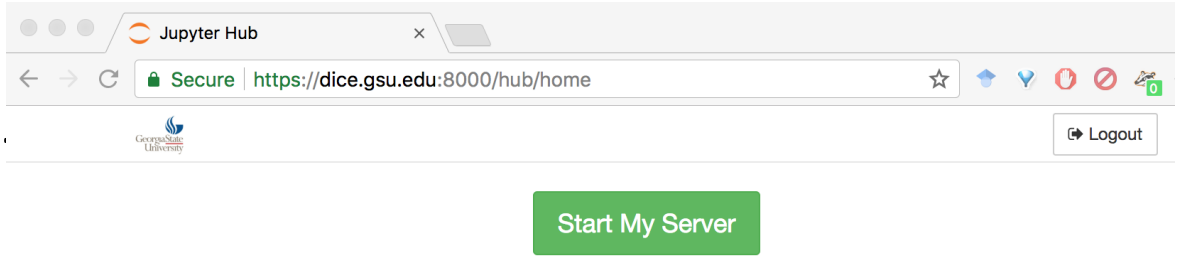
If you picked an option other than 1 core, 12 hours, please click on “control panel”

Then click on “Stop My Server.” This will return you to the spawner page where you can choose the correct option.



Workshop

- Log into DICE:
  - https://dice.gsu.edu:8000/hub/login
  - The minimal address: dice.gsu.edu:8000 should work, but sometimes does not
- Then you see



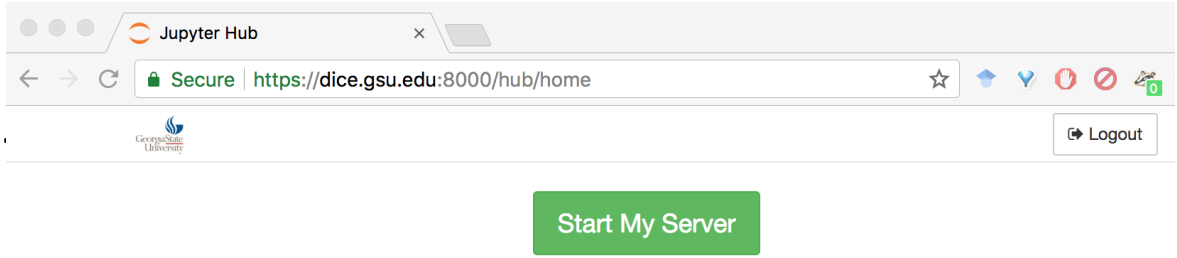
Click the "Start My Server" button.

DICE Spawn

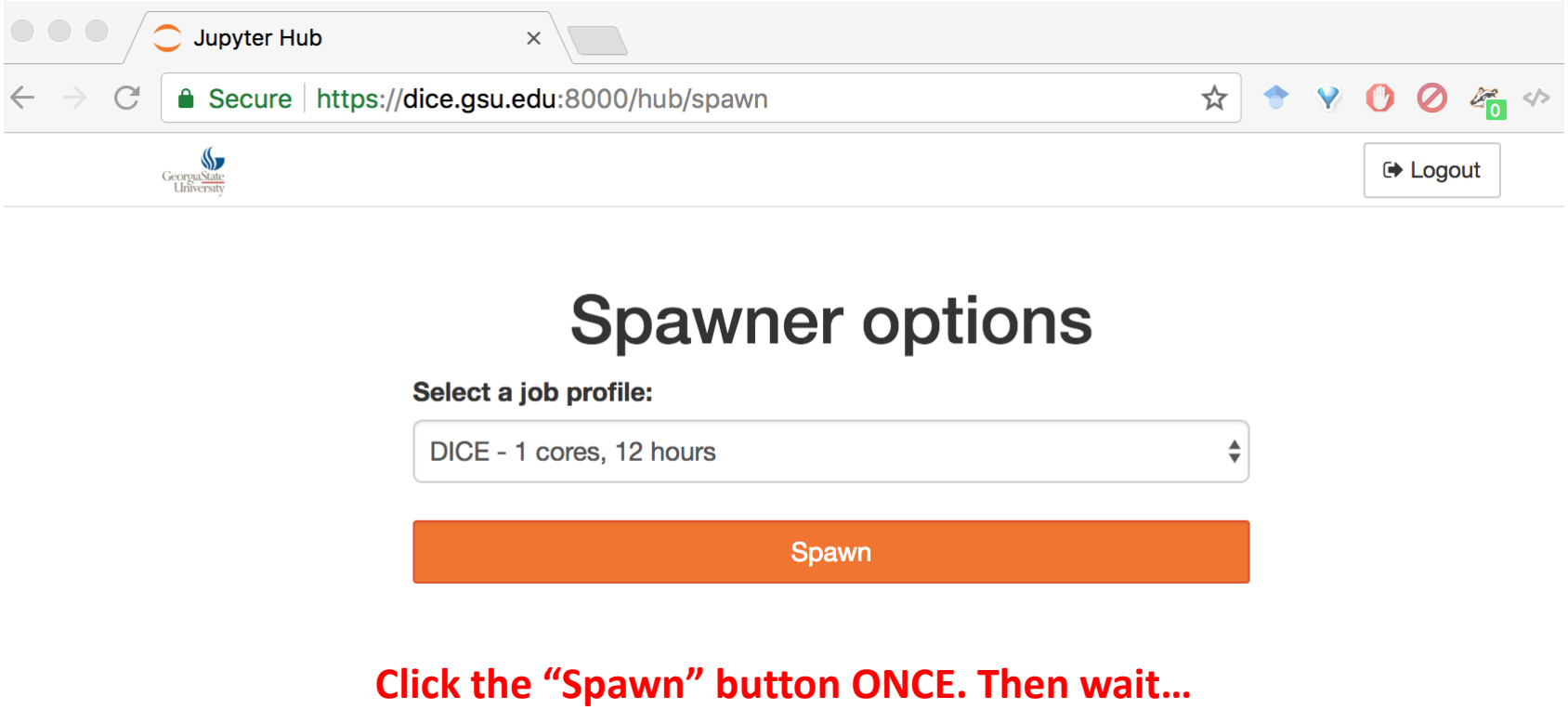
- Please choose the 1 core, 12 hour option. If you choose the others the system will be slow for everyone!

Workshop

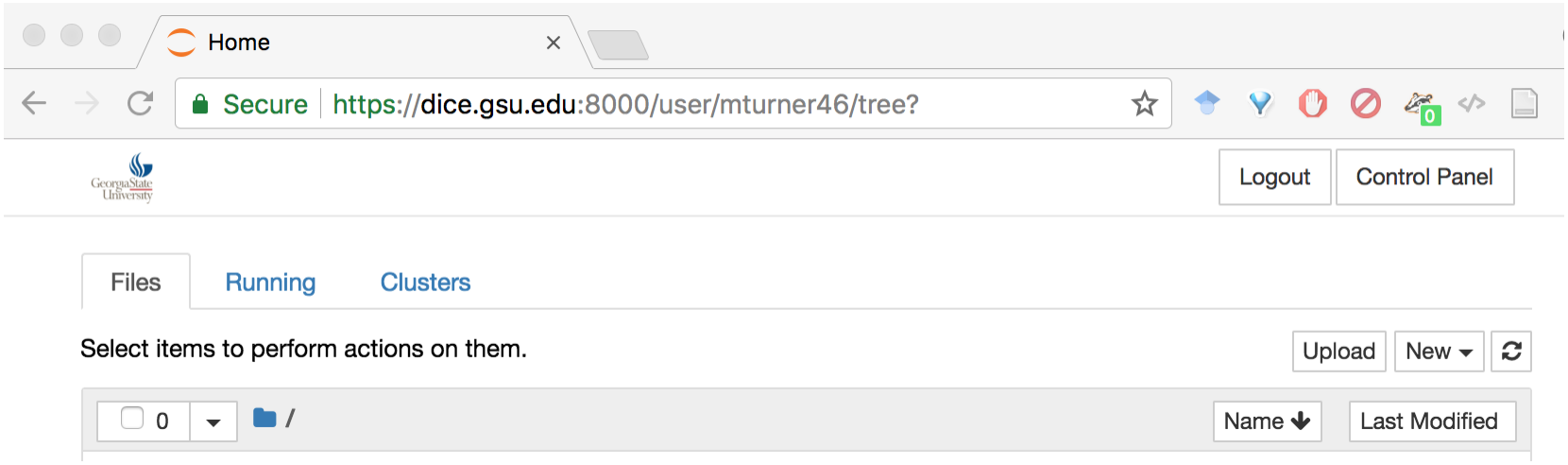
- Log into DICE:
  - https://dice.gsu.edu:8000/hub/login
  - The minimal address: dice.gsu.edu:8000 should work, but sometimes does not
- Then you see



Click the "Start My Server" button.



Click the "Spawn" button ONCE. Then wait…



When you see a page where the TOP of the page looks like this, you are ready to start!