Intro to R

Fuzzy Rogers
Research Computing Administrator
Materials Research Laboratory (MRL)
Center for Scientific Computing (CSC)
fuz@mrl.ucsb.edu
MRL 2066B

Sharon Solis
Research Computing Consultant
Enterprise Technology Services (ETS)
Center for Scientific Computing (CSC)
swsolis@ucsb.edu
Elings Hall 3229

Paul Weakliem
CNSI Research Computing Support
California NanoSystems Institute (CNSI)
Center for Scientific Computing (CSC)
weakliem@cnsi.ucsb.edu
Elings Hall 3231

Other Research IT

Ted Cabeen
Life Sciences
Chemistry
ted.cabeen@lscg.ucsb.edu

Michael Colee
Earth Research Institute (ERI)
mtc@eri.ucsb.edu

Steve Miley
Bren School of Environmental Science & Management
smiley@bren.ucsb.edu

Glenn Schiferl
Physics
glenn@physics.ucsb.edu

Jim Woods
Marine Science Institute
jwoods@msi.ucsb.edu

Letters & Science IT
help@lsit.ucsb.edu
(805) 893-4357

Pre-class Instructions:

This lesson assumes you have the R software and the development environment RStudio installed on your computer.
R can be downloaded here.
RStudio can be downloaded here. You will need the Desktop version for your computer.

You also need to download some files to follow this lesson:

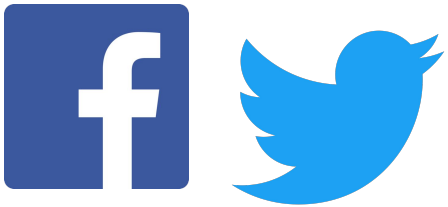
1. Make a new folder in your Desktop called r-novice-inflammation.
2. Download r-novice-inflammation-data.zip and move the file to this folder.
3. If it's not unzipped yet, double-click on it to unzip it. You should end up with a new folder called data.

What is R?

- R is a programming language and free software environment for statistical computing and graphics
- R was initially written by **Robert Gentleman** and **Ross Ihaka**—also known as “R & R” of the Statistics Department of the University of Auckland in 1992.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Who uses R?



https://www.listendata.com/2016/12/companies-using-r.html

Why Use R?

- Powerful, state-of-the-art
- Used by professional statisticians
- Lot of documentation (StackOverflow)
- Freely available for Unix, Windows & Mac
- Extendable, with numerous add-on packages available.
- R produces publication quality graphics.

http://thegrantlab.org/bio3d/79-static-content/85-why-use-r

What is RStudio?

Integrated development environment (IDE)

- Console
- Syntax-highlighting editor that supports direct code execution
- Tools for plotting, history, debugging and workspace management
- Pretty!

Post-It Notes

- **Red**: Help needed
- **Green**: Good to go

Data Set

- Data Set, Code and Lesson Material available here: http://swcarpentry.github.io/r-novice-inflammation/

What is a Script?

- How to run code
- Save yourself work!
- Don’t need to type over and over again
- Move easily between machines

Installing Packages

- What are packages
  - libraries
- How to install packages
- Fftw
- Ggplot2
- Other examples
- Swirl

Using R on a Cluster

- Use R not RStudio on the cluster
- Make sure your R code runs from start to end on your own machine
- Perform tests on your computer first
- A simple script (text file) can be used to submit to the queue:

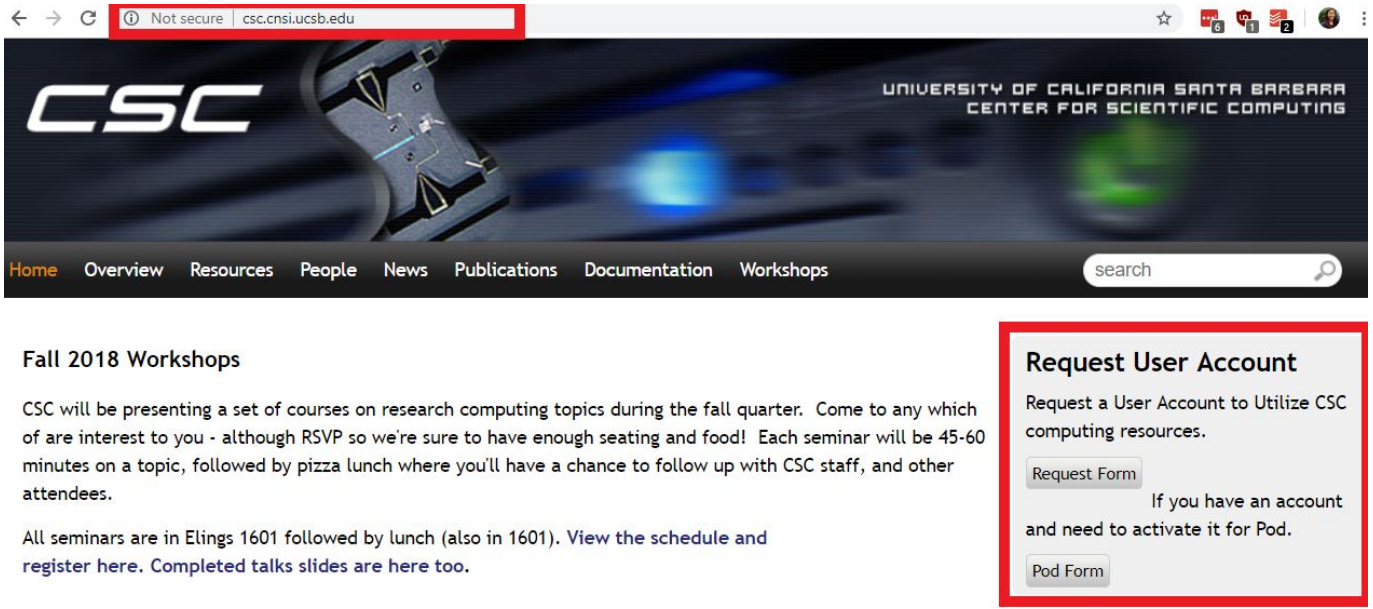
```
#!/bin/bash
#SBATCH --nodes=1 --ntasks-per-node=1
#SBATCH --time=1:00:00
#SBATCH --mail-user=user@ucsb.edu
#SBATCH --mail-type=start,end
cd $SLURM_SUBMIT_DIR
Rscript --vanilla example.R
```

Using R on the Cluster

```
#!/bin/bash
#SBATCH --nodes=1 --ntasks-per-node=1
#SBATCH --time=1:00:00
#SBATCH --mail-user=user@ucsb.edu
#SBATCH --mail-type=start,end
cd $SLURM_SUBMIT_DIR
Rscript --vanilla example.R
```

- The shell you are using
- Asking for one node and one task per node
- Walltime: 1 hour
- Mail to user
- Mail begin/end
- Change directory to the one where job is submitted from
- Run your R code, example.R

How to Request a User Account



What Else Can You Do with R?

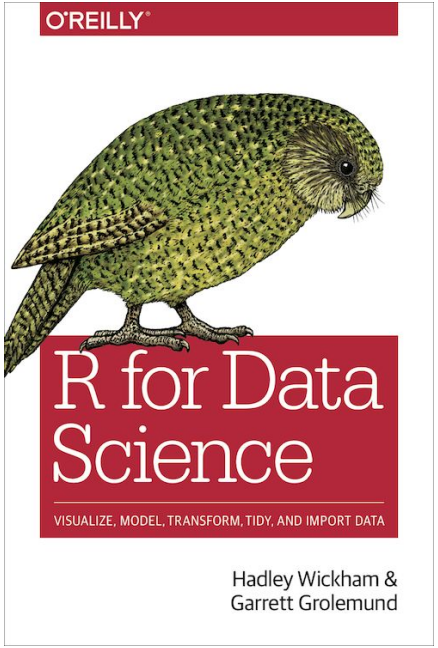
- Predictive modeling
- Machine Learning
- Statistical Analysis
- Economic forecasting
- Predict financial market changes
- Data visualization
- Semantic clustering

Swirl

- R package that leads you through an interactive tutorial to learn R
- Interactive within console

How to Learn More

- Online Tutorials:
  - Coursera, DataCamp, YouTube
  - Lynda.com (available to UCSB employees, including student employees)
- Swirl
  - (RStudio package, interactive tutorial within console)
- Stackoverflow
  - (a great forum of questions and answers about computer programming)
- One-on-One Consultation
  - Center for Scientific Computing (Elings Hall 3229)
  - Collaboratory
- Books
  - R for Data Science by Hadley Wickham and, Garrett Grolemund



Post-Workshop Survey

- Be sure to complete the pre-workshop survey. We use this to calibrate the pace of the workshop and, together with a post-workshop survey, to assess how it went.
  https://ucsbltsc.qualtrics.com/jfe/form/SV_e4YZt4Ubf3gc3tz

Contact Us

csc.cnsi.ucsb.edu

Sharon Solis
Research Computing Consultant
Enterprise Technology Services (ETS)
Center for Scientific Computing (CSC)
swsolis@ucsb.edu
Elings Hall 3229

Fuzzy Rogers
Research Computing Administrator
Materials Research Laboratory (MRL)
Center for Scientific Computing (CSC)
fuz@mrl.ucsb.edu
MRL 2066B

Paul Weakliem
CNSI Research Computing Support
California NanoSystems Institute (CNSI)
Center for Scientific Computing (CSC)
weakliem@cnsi.ucsb.edu
Elings Hall 3231