



The
University
Of
Sheffield.

BMS 353

Bionformatics for Biomedical Sciences

Module coordinator: Dr. Marta Milo

Research Software Instructors: Dr Tania Allard, Dr. Mike Croucher

An introduction to the tools we will be using

R vs the old way



R is a free language and environment for statistical computing and graphics

- Very expensive
- Popular with industry

The logo for SPSS (Statistical Package for the Social Sciences), featuring the letters 'SPSS' in a bold, red, sans-serif font with a registered trademark symbol.

Not surprisingly R is far more popular now!

<http://r4stats.com/articles/popularity/>

The number of jobs requiring R has increased too!



Packages

These save time by using other people's code

- Comprehensive R archive Network (CRAN)
9662 packages (November 2016)
<https://cran.r-project.org>
- Bioconductor **1211** packages (November 2016)
<http://bioconductor.org/>

Packages

These save time by using other people's code

- Comprehensive R archive Network (CRAN)
11749 packages (November 2017)
<https://cran.r-project.org>
- Bioconductor **1476** packages (November 2017)
<http://bioconductor.org/>

Jupyter notebooks



The image shows a screenshot of the Jupyter website homepage. At the top left is the Jupyter logo, and at the top right are navigation links for INSTALL, PROJECT, DOCUMENTATION, BLOG, and DONATE. The main visual is a large orange Jupyter logo with the word 'jupyter' in lowercase, surrounded by various programming language icons including JS, PHP, Lua, VB, C#, F#, R, Spark, and C++. Below this, there is a section titled 'Jupyter Notebook' with a notebook icon. To the right of this section are two overlapping screenshots of a Jupyter Notebook interface. The front screenshot shows a notebook titled 'Exploring the Lorenz System' with a text area, a code cell, and a plot of the Lorenz attractor.

Jupyter Notebook

The Jupyter Notebook is a web-based interactive computing platform that allows users to author data- and code-driven narratives that combine live code, equations, narrative text, visualizations, interactive dashboards and other media.

Currently in use at

Google

Microsoft

IBM

Bloomberg

O'REILLY

ANACONDA.

rackspace
the #1 managed cloud company

SOUNDCLOUD

Quantopian

NetApp

software
carpentry

hhmi janelia
Research Campus

< CODE NEURO >

N-Site LLC

COCA.LC

BRYN
MAWR
COLLEGE

CAL POLY
SAN LUIS OBISPO

Berkeley
UNIVERSITY OF CALIFORNIA

The
University
Of
Sheffield.

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

CLEMSON
UNIVERSITY

MICHIGAN STATE
UNIVERSITY

Northwestern
University

NYU

NASA

AYASDI

The Data Incubator

Combine live computer code, data, text, and mathematics in an interactive document.



Academic papers only give a description of your analysis.

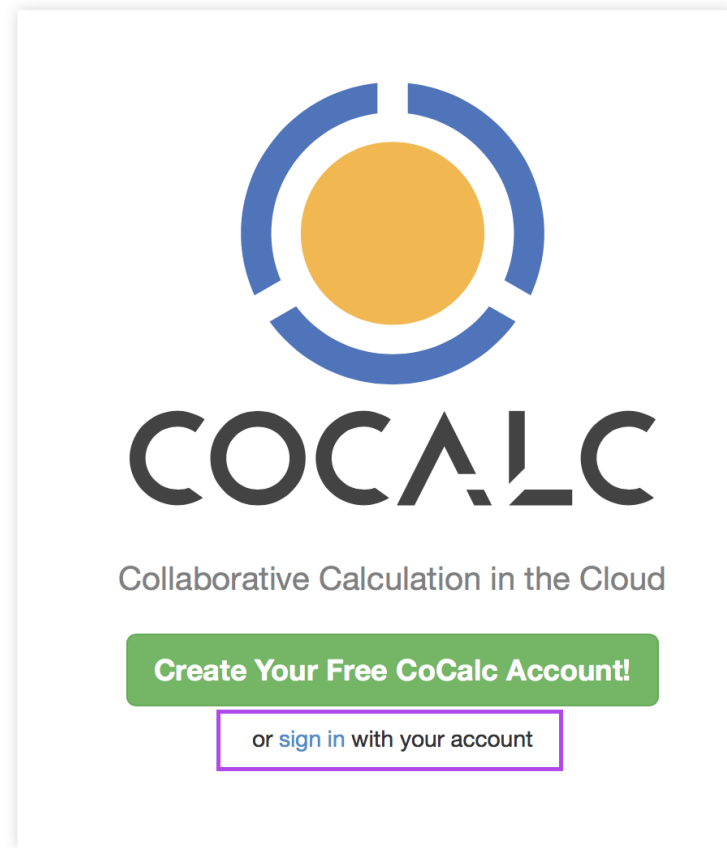
Jupyter notebooks contain both the description and the analysis.



- Collaborative computational mathematics
- All open source environment for running Jupyter notebooks (and others) in the cloud
- Dedicated Virtual Machine for this course on the Google Cloud Platform
- As powerful as a node on the Sheffield University Supercomputer: Iceberg

Log into CoCalc

<https://cocalc.com>



Handy tips for the notebooks

Ctrl/Cmd + y changes the cell to code mode

Ctrl/Cmd + m changes cell to markdown

Shift + Enter evaluates the cell

To add Latex in a markdown cell make sure to enclose in $$$$ or $$$

Also, we put together a **mini** CoCal tutorial that
you can check at
<https://tutorial.cocalc.com/students/>