FOSE1025 — Scientific Computing

Week 10 Lecture 1: Towards Using Scripts for Reproducibility

Diego Mollá

FOSE1025 2020H1

Abstract

In this lecture we will have a very brief introduction to the use of scripts to store and manipulate data. We move away from Excel and enter the area of programming. The emphasis will be on how to use scripts for reproducibility, and we will focus in a particular environment: MATLAB.

Update May 12, 2020

Contents

Reading

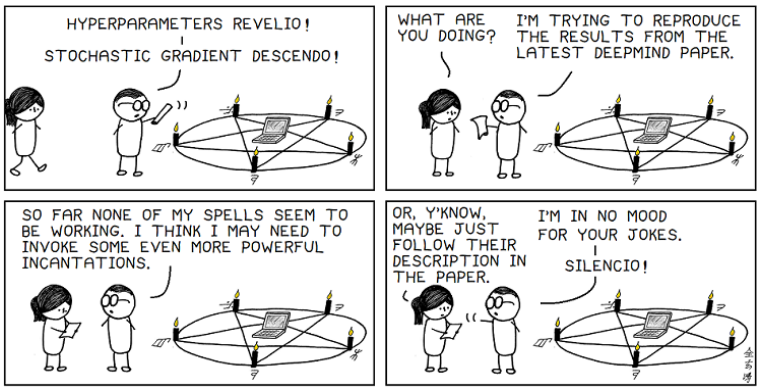
- These notes
- *https://au.mathworks.com/help/matlab/getting-started-with-matlab.html*
- *https://au.mathworks.com/videos/getting-started-with-matlab-1564521672719.html*
-

1 Scripts

The Problem with Reproducibility

It can be difficult to write clearly enough to allow reproducibility.

https://abstrusegoose.com/588



Scripting Languages

- Scripting languages are programming languages designed for *rapid prototyping*.

  $\Rightarrow$ These languages make it easy to quickly write and execute a program.

- Scripting languages are normally *interpreted languages*.

  $\Rightarrow$ This means that you can execute instructions one by one using a *run time environment*.

Example of Steps

1. Start the run time environment (e.g. MATLAB).
2. Type instructions (or load instructions stored in a file).
3. Run the instructions in the run time environment.

Top 10 Programming Languages for Data Science
https://www.analyticsinsight.net/top-10-data-science-programming-languages-for-2020/

1. Python (popular among programmers and web developers)
2. R (popular among statisticians)
3. SQL (designed for querying relational databases)
4. C (C++)
5. Java
6. JavaScript (originally designed to run in a browser)
7. MATLAB (the focus of this lecture)
8. Scala
9. Swift
10. Julia

Demonstration Using MATLAB Online

- In this demonstration, the runtime runs *in the cloud*.
- We use a web browser to interact with the runtime.
- Can be done with any computer as long as it has:
  - An internet connection.
  - A modern browser.
- There is no need to install additional software in your computer.

Scripting Languages and Reproducibility

- Instructions written in a scripting language ensure reproducibility . . . or does it not?
- While instructions written in a scripting language can be executed by a computer . . .
  - . . . instructions may not do what we intended them to do (e.g. because there are errors in the instructions).
  - Poorly-written scripts may not be understandable by people

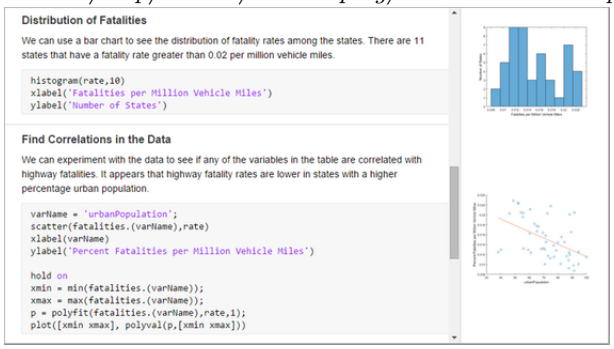
    $\Rightarrow$ and then we cannot tell if they are correct.
  - *Portability:* Scripts running in a computer might not run in another computer.

    $\Rightarrow$ often you need to provide instructions for installation of necessary software dependencies.
- Normally we want to supplement the instructions with comments and explanations.

Notebooks for Reproducibility

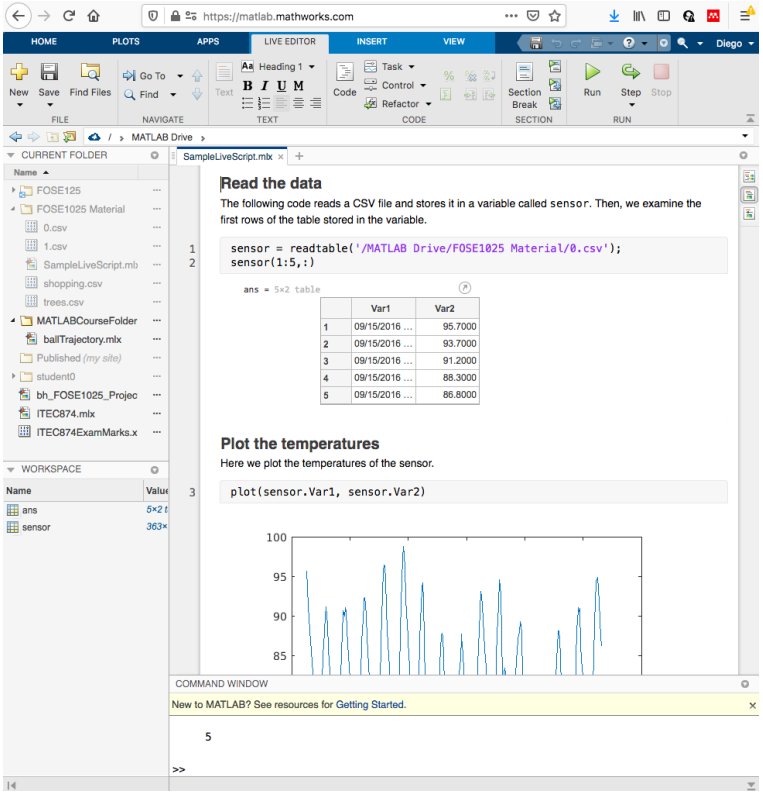
- Some run time environments allow the creation of notebooks.
  - Called *live scripts* in MATLAB.
- These notebooks are the digital equivalent of lab notebooks.
- Notebooks contain sections that can be executed.
- The results of execution appear in the notebook.
- Notebooks also contain formatted text for documentation and explanations.

https://au.mathworks.com/help/matlab/matlab_prog/what-is-a-live-script-or-function.html



Demonstration of a Live Script

SampleLiveScript.mlx



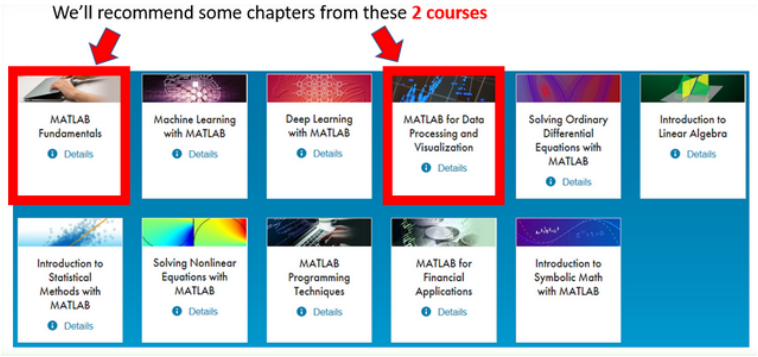
2 MATLAB

What is MATLAB?

- MATLAB is a scripting language.
- Includes types designed to store and manipulate data.
  - Matrices (MATLAB = MATrix LABoratory)
  - Tables
- Includes a large library of functions for data analysis, manipulation, and visualisation.
- Has extensive documentation and on-line courses.
- Easy to use
- Others programming languages have attempted to integrate some of MATLAB's features.
  - Matrices, tables
  - Plots
  - Interactive notebooks

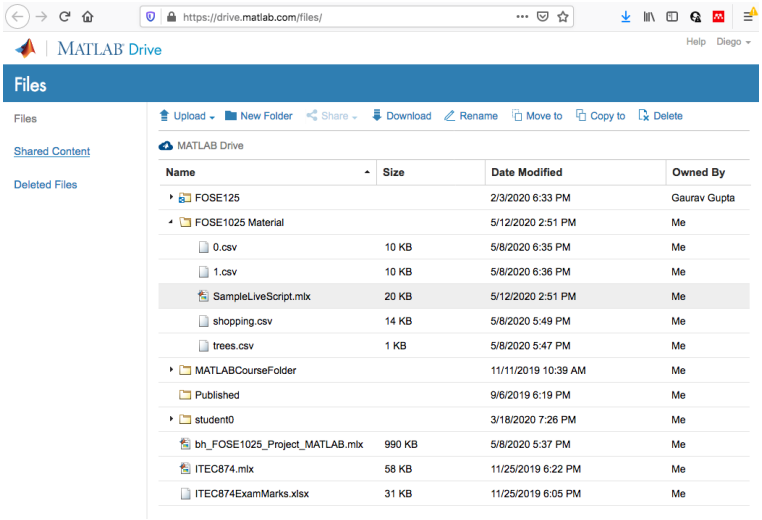
Accessing MATLAB and MATLAB Online

- Macquarie University has a license for students: *https://au.mathworks.com/academia/tah-portal/macquarie-university-916052.html*
- MATLAB Online here: *https://matlab.mathworks.com/*
- Getting started: *https://au.mathworks.com/help/matlab/getting-started-with-matlab.html*
- Self-paced courses: *https://matlabacademy.mathworks.com/*



MATLAB Online and MATLAB Drive

- MATLAB Online runs in the cloud.
- To upload files to the cloud you can use MATLAB Drive.
- You can use a browser to upload and download files.
- Or you can install software that integrates with your computer file system.
  - It looks and feels like MATLAB drive is a folder in your computer.



Loading data in MATLAB

- MATLAB Fundamentals, Chapter 10, "Tables of Data"
- *https://au.mathworks.com/help/releases/R2019b/matlab/matlab_prog/ create-a-table.html*
- MATLAB can store tables into variables.
- You can use the MATLAB "Import Data" wizard.
  - Looks like a more sophisticated version of Excel's Import tools.
- Or you can use the readtable instruction.
  - trees = readtable("trees.csv");

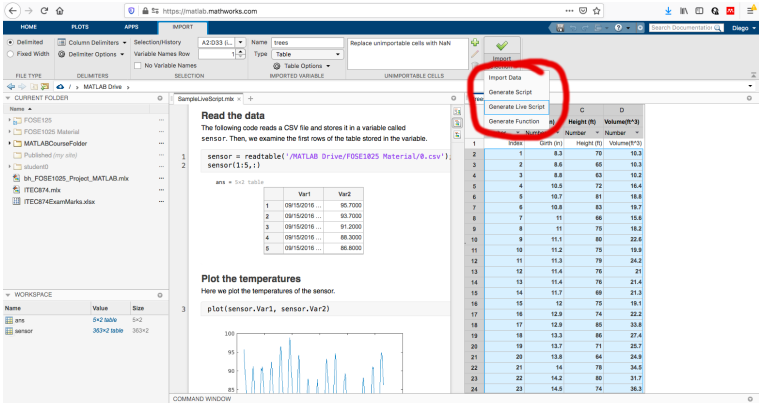
Processing data in MATLAB

- *https://au.mathworks.com/help/releases/R2019b/matlab/matlab_prog/ access-data-in-a-table.html*
- Accessing a column: girth = trees.("Girth (in)")
- Accessing a full row: sample = trees(5,:)
- Adding a column: trees.("Girth (cm)") = trees.("Girth (in)") * 2.54
- Concatenating tables:

```
table0 = readtable("0.csv");
table1 = readtable("1.csv");
table = [table0; table1];
```

Creating and Reusing MATLAB Scripts

- Many MATLAB wizards can generate scripts.
- You can write your own script.
- Then you can run it again later.



Take-home Messages

- Scripting languages are powerful means to allow reproducibility.
- Scripting languages can be executed by a computer.
- Some environments allow the use of interactive notebooks for better reproducibility.
- MATLAB is a powerful scripting language designed for data analysis.

What's Next

- Monday 18 May: Submit your report for reproducibility.
- Wednesday 20 May, 10am: In-class quiz 4.
- Friday 22 May: Submit Collaborator Module (hurdle, Turnitin).
- Friday 29 May: Complete peer review of report for reproducibility.
- Friday 29 May: Submit Professional Module (hurdle, Quiz).
- Friday 5 June: Problem Solver Module (hurdle, Quiz).