

FOSE1025 — Scientific Computing

Week 10 Lecture 1: Towards Using Scripts for Reproducibility

Diego Mollá

Department of Computer Science
Macquarie University

FOSE1025 2020H1

Programme

- 1 Scripts
- 2 MATLAB

Reading

- Lecture notes

Programme

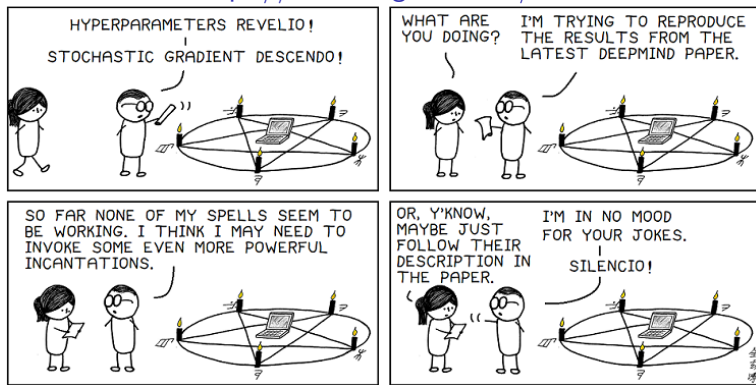
1 Scripts

2 MATLAB

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**.
 - ⇒ These languages make it easy to quickly write and execute a program.
- Scripting languages are normally **interpreted languages**.
 - ⇒ This means that you can execute instructions one by one using a **run time environment**.

Example of Steps

- ➊ Start the run time environment (e.g. MATLAB).
- ➋ Type instructions (or load instructions stored in a file).
- ➌ 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.

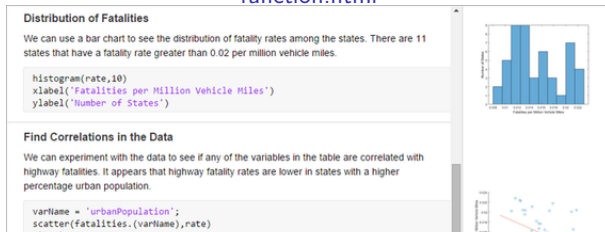
Scripting Languages and Reproducibility

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

The screenshot shows the MATLAB Live Editor interface. The top toolbar includes tabs for HOME, PLOTS, APPS, LIVE EDITOR, INSERT, and VIEW. The LIVE EDITOR tab is active, showing a script named `SampleLiveScript.mlx`. The script content is as follows:

```
1 Read the data
2 The following code reads a CSV file and stores it in a variable called sensor. Then, we examine the
   first rows of the table stored in the variable.
   sensor = readtable('/MATLAB Drive/FOSE1025 Material/0.csv');
   sensor(1:5,:)
   ans = 5x2 table
           Var1      Var2
   1  09/15/2016 ...  95.7000
   2  09/15/2016 ...  93.7000
   3  09/15/2016 ...  91.2000
   4  09/15/2016 ...  88.3000
   5  09/15/2016 ...  86.8000

3 Plot the temperatures
   Here we plot the temperatures of the sensor.
   plot(sensor.Var1, sensor.Var2)
```

The left sidebar shows the CURRENT FOLDER and WORKSPACE. The CURRENT FOLDER contains files like `FOSE1025`, `FOSE1025 Material`, `0.csv`, `1.csv`, `SampleLiveScript.mlx`, `shopping.csv`, `trees.csv`, `MATLABCourseFolder`, `ballTrajectory.mlx`, `Published (my site)`, `student0`, `bh_FOSE1025_Projec`, `ITEC874.mlx`, and `ITEC874ExamMarks.x`. The WORKSPACE shows a variable `ans` of type `5x2 table`.

Programme

1 Scripts

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

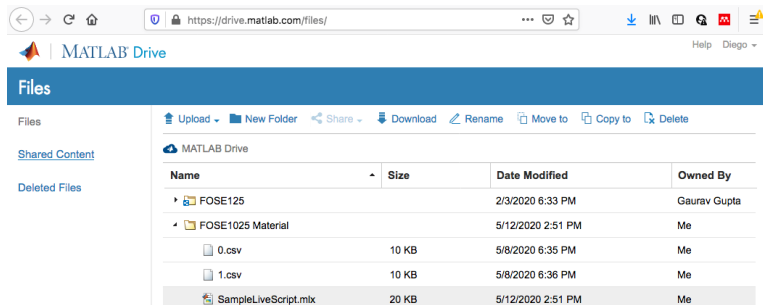
Accessing MATLAB and MATLAB Online

- 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.



The screenshot shows the MATLAB Drive web interface in a browser. The address bar displays the URL <https://drive.matlab.com/files/>. The page header includes the MATLAB logo and the text "MATLAB Drive". Below the header, there is a "Files" section with a table of files. The table has columns for Name, Size, Date Modified, and Owned By. The files listed are FOSE125, FOSE1025 Material, 0.csv, 1.csv, and SampleLiveScript.mlx. The interface also includes navigation links like "Shared Content" and "Deleted Files", and action buttons like "Upload", "New Folder", "Share", "Download", "Rename", "Move to", "Copy to", and "Delete".

Name	Size	Date Modified	Owned By
▶ FOSE125		2/3/2020 6:33 PM	Gaurav Gupta
▶ FOSE1025 Material		5/12/2020 2:51 PM	Me
0.csv	10 KB	5/8/2020 6:35 PM	Me
1.csv	10 KB	5/8/2020 6:36 PM	Me
SampleLiveScript.mlx	20 KB	5/12/2020 2:51 PM	Me

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");`

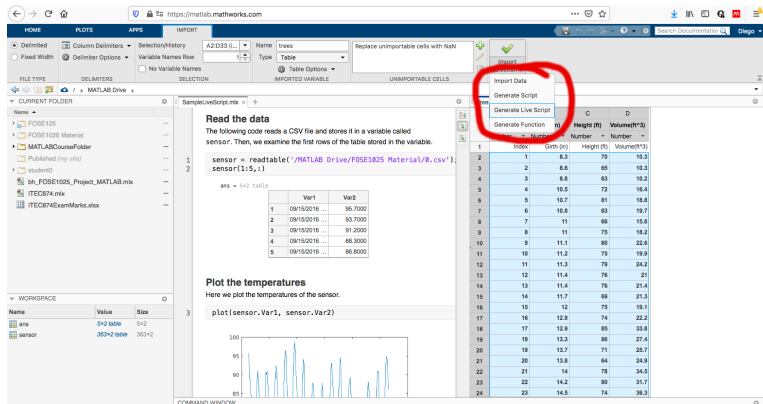
Processing data in MATLAB

- 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.



The screenshot shows the MATLAB Import Wizard interface. The 'Import' button is highlighted with a red circle. The 'Generate Script' button is also highlighted with a red circle. The 'Import Data' section shows a table with columns: Index, Girth (in), Height (ft), Volume (ft^3). The 'Generate Function' button is also visible.

Read the data
The following code reads a CSV file and stores it in a variable called `sensor`. Then, we examine the first rows of the table stored in the variable.

```
sensor = readtable('/MATLAB Drive/FOSE1025 Material/0.csv');  
sensor(1:5,:)
```

Plot the temperatures
Here we plot the temperatures of the sensor.

```
plot(sensor.Var1, sensor.Var2)
```

Workspace:

Name	Value	Size
ans	5x2 table	5x2
sensor	362x2 table	362x2

Command Window:

```
ans =  
    1    09/15/2016    86.7000  
    2    09/15/2016    83.7000  
    3    09/15/2016    91.2000  
    4    09/15/2016    88.3000  
    5    09/15/2016    86.8000
```

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).