ASDS Code Camp

Day 4 - Good Practices

Martyn Egan (Teaching Fellow)

September 6 - 10, 2021

Good practices for learning data science

Learning data science is a continuous process. At times the learning curve can feel steep. Here are some tips to help you achieve your goals.

- Aim for steady, constant progress

  You will gain familiarity better by doing a little every day

- Keep a code notebook

  Remembering functions, and arguments to functions, can be a challenge. Making a note of useful functions can help in this

- Learn by doing

  Code is a language, and it's always easier to learn a language by using it - so find some data that interests you, and put your skills to work

- Take advantage of online learning resources

  Sites like Datacamp, Udemy and Kaggle can help improve your data science skills.

Good practices for doing data science

Organise your files

Use folders to organise your projects. Set a working directory for each new project. Keep all project files in that folder.

Setting a working directory in R/R Studio

We can set our working directory using the `setwd()` function. We can check our working directory using the `getwd()` function. In R Studio, we can use the **project** settings to assign a working directory to each project we are working on.

Setting a working directory in Python/Spyder

Changing the working directory in Python requires importing the os package (import os) and then calling the os.chdir() function. In Spyder, simply change the working directory using the file explorer bar in the top right of the screen.

Work *with* your IDE

IDEs (integrated development environments) can make your life much easier, and often contain various capabilities for improving your workflow. Makes sure you take the time to learn about the IDE you are using, and don't be afraid to experiment with new or different IDEs to find the one that suits you best.

- Learn your IDE's keyboard shortcuts - they will save you time and effort!
- Do not fall into bad habits - particularly regarding automatic settings such as preserving workspace.

In general...

Working **with** your IDE (as opposed to having it do everything **for** you) may involve turning off some automatic features. If you rely on your IDE automatically restoring your workspace, you may discover your script files stop working...

Use scripts

Working with scripts, rather than inputting commands directly into the console, is fundamental to a robust data science workflow. A good script will allow you to record and repeat your analysis, share your work with other analysts, and keep track of your progress. It also enables us to recycle our code, saving time and effort.

A good script...

- ...begins with the packages/libraries/modules you will require in your analysis.
- ...is appropriately commented to allow yourself and others to follow the workflow.
- ...is saved in the working directory of its project and points to that directory when creating/saving files.

Naming objects

Stick to a style!

In any data science project you will create numerous objects: datasets, variables, functions... Keeping track of everything can become difficult. It is therefore important to develop a good style for naming objects, and to stick with it. This will help both you, when looking back at old scripts, and others, when trying to interpret your work.

- this_is_snake_case
- THIS_IS_UPPER_SNAKE_CASE
- thisIsCamelCase
- ThisIsPascalCase
- this-is-kebab-case

  Remember: not all case styles will work in all languages
- dont.do.this.in.python