Computer coding in science

November 29 2016

To be a successful scientist in academia it is no longer sufficient to be good at science. In addition to expertise in experimental methods and data analysis, scientists must also excel in public speaking and writing. Furthermore, scientists must be able to successfully network to form and maintain beneficial inter-institutional collaborations. Proficiency in computer coding has arguably become one of the most important skills a researcher needs in science today. There are a number of reasons for this, chief among them is, as a consequence of advances in tools and technology, researchers are now collecting and working with larger datasets. These datasets require computing coding and machine learning steps for unbiased, large-scale analysis.

The statistical package R is often a researcher's first exposure to a coding language. R has become one of the most popular statistical packages in scientific research, largely due to the fact it is free, highly customizable, and analytical packages are available in online repositories. R operates within an environment using its native coding language to perform powerful data manipulation and visualization. Graduate students, therefore, find themselves learning how to code and use R early on in their careers. From here, R becomes a gateway to further coding proficiency.

Furthermore, many analysis programs have downloadable versions that operate via command line. Interacting with a program or computer via command line refers to instances where commands are given in the form of text, rather than the use of menus or a graphical interface. These are often open source, free, and allow more customization and tailored analysis. For example, there is a command line version of BLAST (Basic Local Alignment Search Tool) available for download. BLAST is a tool that searches a user-provided DNA or protein sequence against a named database to identify similar sequences or regions of sequence similarity. It is often used to identify similar genes or proteins in disparate species. The command line version allows a great deal more flexibility than the web based tool. For example, users can create their own database and customize how and what information is given in the output. The exponential increases the amount of data researchers can now collect makes computer coding an invaluable analytical tool. Coding also allows for the automation of many repetitive analytical tasks. The increased importance of coding is far from specific to biology but widespread across all science and engineering disciplines.

What ways are there to learn coding?

There are many ways to learn how to code and it is important to identify the method that works best for you. First you have to decide which computer language you are going to learn. A few of the bigger and recommended languages for science include Python, Perl, and Scala. Many academic institutions offer introductory coding classes that can be a great starting point to learn the basics of coding.n some cases these classes can be progressive, for example starting with Python for beginners and progressing to intermediate. It is worth noting that these courses are geared towards students from all disciplines so there may be portions of the class that are not applicable to your own work in an obvious way.

Another method of learning code is through free online courses, tutorials, and dedicated blogs. A quick online search for "learn code online" will give you seemingly innumerable options. It is best to look across sites to identify the one that best fits your level of experience, and teaches material that is relevant to you personally. An advantage to online courses and tutorials is that they allow you to work at your own pace and cherry pick classes that you will find useful. Some students take self-controlled learning to the extreme and learn code using textbooks. There are many textbooks available, some of which are tailored towards particular disciplines. These textbooks outline the basics of the coding language, and then move onto relevant examples and practice exercises using provided datasets.

A further way to learn practical coding is to identify your own Yoda of coding who is willing to answer questions, and offer advice, without complaint. However, such characters are few and far between and should generally be considered supplemental to your own learning, with them providing more practical guidance than foundational teaching.

How to start and where to get help

The best place to start is your institution's information technology services (ITS) department, particularly if there is a research-focused branch. ITS may be able to provide you with the physical resources required. For example, they can set up virtual machines (VM) that allow remote access to a maintained environment. These VMs are particularly useful if you are primarily a Windows user but your analytical tools require a Linux operating system. Because these VMs are set up and maintained by trained experts, they can help install programs, ensure they work, and troubleshoot any problems. If you do encounter errors (which is likely when you begin), the Internet is your friend. Although it can be hard to find the exact answer you are looking for, there are plenty of online forums and blogs full of people willing to help. Finally, many available packages are open source and available in online repositories, meaning they are developed by researchers and shared with the scientific community. As a consequence, whatever analysis you are attempting has likely not only been done before but a tool is available online, along with instructional documentation.

Proficiency in computer coding has become a definite advantage to researchers in all stages of their careers, and across disciplines. However, many find it a difficult skill to master, particularly when you are just starting out. I have definitely found there to be a steep learning curve full of frustration, particularly in applying textbook knowledge to my own data in a practical way. That being said, learning to code is worth the effort and will pay dividends in analytical power and scope. There is a multitude of ways to learn coding and it is important to choose the best method that suits you.

More information: R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL www.R-project.org/.

Wikipedia contributors. "Command-line interface." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 3 Nov. 2016. Web. 3 Nov. 2016.

Wikipedia contributors. "Graphical user interface." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 14 Nov. 2016. Web. 14 Nov. 2016.

Wikipedia contributors. "Python (programming language)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 20 Nov. 2016. Web. 20 Nov. 2016.

Wikipedia contributors. "Perl." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 20 Nov. 2016. Web. 20 Nov. 2016.

Wikipedia contributors. "Scala (programming language)." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 14 Nov. 2016. Web. 14 Nov. 2016.

Wikipedia contributors. "Operating system." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 18 Nov. 2016. Web. 18 Nov. 2016.

Provided by PLOS Blogs

Citation: Computer coding in science (2016, November 29) retrieved 26 June 2024 from https://phys.org/news/2016-11-coding-science.html