

# Linux and the open source software movement

Linux, R and python are all open source and free.  
(a) How did high quality products emerge without financial incentives?

Note: Nowadays it may seem obvious that this is possible. Before Linux and R took off, 20 years ago, the consensus was that free products would remain marginal. Free products were obscure toys for computer scientists, and everyone else had to pay for Matlab, S-Plus, Windows, Unix, Mac, ...

(b) If developers are interested in making money, can they do this by writing free software? If so, how? If not, why do they do it?

A. *"Yes, they can make money because they are hired by IT giants to develop the software. The companies will pay the developer for their job to get done."*

B. *"They can write software of which partial code is open source and ensure most common usage of the software. If people want full code or more complex function of the code, they have to pay the company."* This is called the "freemium" model.

C. *"Developers can make money writing free software if they find someone to pay them to do so (i.e. the government, universities, non-profits, etc.)."*

*“A successful open source software will help one build up reputation and get recognition for ones professionalism. This will link one to more valuable resources and thus provide many opportunities of making money. This really likes how researchers make money. One possible way is to provide consult service.”*

*“People initially just want to make something useful for themselves to use and are not thinking about making money.”*

examples: R, Hadoop, tensor flow, ...

*"They can make money by writing free software and make it influential, and get money from the user donation and advertisements."*

examples: facebook, wikipedia

# How much of the intro Linux tutorial is new to you?

entirely or almost entirely familiar (7)

intermediate (3)

mostly or entirely unfamiliar (4)

*"I installed dual boot Ubuntu on a computer that I used for my internship last summer. I gained some familiarity with basic commands like cd, mkdir, cat, and sudo. However, I didn't encounter any need to move or delete files from the command line or to work with hidden files."*



Do you agree or disagree with the opinions at  
[hub.packtpub.com/  
data-science-windows-big-no/](http://hub.packtpub.com/data-science-windows-big-no/)

*"[...] Although I have not (yet) encountered a large enough data science problem that the operating systems speed made a significant difference, I have found that Linux is essential to collaborating with other data scientists."*

*"The words I agree with the most are: "This is a moo point." To me, data science is really OS agnostic. Sure, there are instances when programming and computing that one OS is better than another, but not to a point where it will significantly hinder your abilities of becoming a data scientist."*

*"Though a very dedicated Windows user, I plan to install Linux onto my laptop over winter break to start to get used to the system."*

*"For researchers or people conducting projects with advanced topics of data science, certainly Linux would be a better choice and in the long run it is worthwhile to learn it. However, if people are just trying to make use of existing tools like visualization or just use SAS for data analysis, then Windows is enough."*

*"I don't agree with those opinions. Windows will be just fine for the average data scientist: using mainstream tools will always work; developing tools shouldn't be a problem in general. There may be some compatibility issues for some esoteric software and languages, but these relate to more advanced users which will most likely already be working on UNIX; parallel computing is slightly different (maybe less efficient?) on Windows, but then just use a cluster if you really need more computing power (and this can be done from Windows)."*

*"I have seen many people from CS department suggested using Linux for data science, so I guess I should agree. Most of my friends having Window computers say that they are using Window to play games only, when they need to work, they switch to Linux."*

## Recommend a Linux distribution (with a justification) and give some other helpful advice

*"I could suggest him to install a free Ubuntu and keep his Windows side by side by partitioning the hard disc suitable (usually called a dual boot). I used Ubuntu and Fedora as Linux distributors in my college computer (not much) and I do not really know how and why they are different. But they are free to download and use."*

*"Ubuntu is well-maintained as one of the most common Linux distributions, and many help resources are available online."*

Ubuntu was most popular, with variations...

*"I would personally recommend Fedora. It is one of the best Linux distributor out therein the market. Here are some reasons to be lured of Fedora. (a) Fast and Reliable Updates.(b) Better Package Management. (c) Reliable Multi-Level Security. (d) Greater Extent Hardware Support. (e) Very User Friendly.*

*<https://www.ubuntupit.com/>*

*is-fedora-linux-a-good-distro-best-reasons-to-use-fedora*

*"I would like to recommend Debian to him. I suggest him to read this [debian-handbook.info/browse/stable/sect.release-lifecycle.html](http://www.debian-handbook.info/browse/stable/sect.release-lifecycle.html). "*

*"I use Linux Mint which is light-weight and has that Windows feel but allows full UNIX capabilities."*



*“Oracle VM VirtualBox seems to be a good option. It’s free and fairly simple to use. There are over 100,000 registered users, so people who have experienced issues probably post about it online. It supports VirtualBox, something I’ve used a lot, so I guess I might be biased here.”*

and how about WSL?

*"I have tried Ubuntu and it is OK, also on Windows 10 now you are able to install a Linux subsystem (like Ubuntu) which I only tried few times but felt OK. MacOS is what I am using now, it is also well developed and really handy if you also have an iPad or other Apple devices. So probably Mac would be a good choice if you have enough funding. Otherwise, install a dual Linux system is the fastest solution."*