Homework 8. Due by 11:59pm on Sunday 11/9.

Linux, the command line, and the open source software movement.

Linux is the dominant environment for scientific computing. For example, supercomputers generally run some variant of Linux (`https://en.wikipedia.org/wiki/Linux`). As another example, most cloud servers are built on Linux, and Linux is therefore dominant for data science applications that involve cloud computing. At University of Michigan, the main resource for high performance computing is the Great Lakes Linux cluster. It should be apparent that Linux skills are useful for a research statistician, as soon as your data analysis or simulation study is too large for a laptop.

Linux expertise in this class ranges from novice to expert. Our goal is to advance our understanding and share knowledge.

Write brief answers to the following questions, by editing the tex file available at `https://github.com/ionides/810f25`, and submit the resulting pdf file via Canvas.

1. Linux, R and Python are all open source and free.

   (a) How do you think these projects led to high quality products given that the usual financial incentives for building, coordinating and running a development team are missing?

   Well, my though could be naive. But there are always warm hearted people. Suppose those people are 1% of the whole population, and people who will work for money count for 50%. However, a close enviroment makes the whole population much smaller. For example, if there are 1 million people in the community, but only 100 people are in the team, we still have $1000000000 \times 1\% \geq 100 \times 50\%$.

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

   Yes, they can. They can get known by people, and earn great opportunities to sell courses or other products. They can find collaborators to found a new company more easily too.

2. To what extent do you agree or disagree with the opinions at `https://valohai.com/blog/command-line-for-data-science/`? Specifically,

   (i) "most data scientists are on UNIX-based systems these days."

   (ii) Five reasons why command line use is critical for data science: speed, agnosticism, automation, extensibility, lack of other options.

   I totally agree with (i), it seems that every developer is using linux or other unix-based system instead of Windows. For (ii), I do not agree that lack of options is a reason for developers to choose CLI. I would also add that it is less costly(both in power or in money), and also more straightforward.

3. Unless you vigorously disagree with the assertions in Question 2, if you have not yet had much experience working with the command line, you should be motivated to change that! The instructor of this course endorses the value of working with the command line, so even if you are deeply skeptical about the value of learning the ancient art of command line computing, please put that to one side for the rest of the course.

If you are new to Linux, or if your Linux skills are limited to a handful of commands, read the introduction to command line Linux at

`https://tutorials.ubuntu.com/tutorial/command-line-for-beginners`

If you have a Mac, try out the commands on a Terminal app, which runs a version of Unix that works identically to Linux for many everyday purposes. If you run Windows, try out some commands on the Windows subsystem for Linux

`https://docs.microsoft.com/en-us/windows/wsl/`

You can also log in to the UM Linux login service, using SSH Secure Shell. For example from a Mac terminal, type

`ssh your_uniqname@login.itd.umich.edu`

Optionally, if you find it an effective way to practice, play the terminus adventure at

`https://web.mit.edu/mprat/Public/web/Terminus/Web/main.html`

Report briefly on what you have learned.

Unfortunately, my SSH connection did not work, it said permission denied after I inout my password. The last website is a fun tool to learn CLI commands.

4. If you are a relatively experienced Linux user, share some words of advice for beginners. How and why did you get started with Linux?

Well, I would not call my self "experienced", which is too much for me. But I would advise that using a virtual machine to study Linux. There are multiple advantages:

- It can make it up if they are using other operating system like Windows;
- They can try more "dangerous" commands like "rm -rf" without worrying about destroying their own machine;
- They can choose any version of Linux to learn based on their own preference.

Optionally, you can also use this opportunity to learn some more Linux-related skills, e.g., from the Linux intermediate tutorials at

`https://www.linux.org/forums/linux-intermediate-tutorials.124/`