

## Installation Instructions (for saasbook VM users)

(borrowed heavily from the [installation instructions](#) in CS194-16: Introduction to Data Science)

**Note: While you should also be able to set up a similar environment on your Mac or PC without needing a Virtual Machine, the course staff will not support such configurations - so you're on your own if you choose to go that route!**

### Python/IPython

1. Follow [the instructions available here](#) to install VirtualBox, and the virtual machine image. The virtual machine image already includes most of the software necessary to run the code. We will install extra packages below. *Note: If you have troubles installing Virtualbox on your Operating System and/or booting up the virtual machine, Google will be your best friend. We only provide official support after you have been able to get past the login screen and are inside the virtual machine*
2. Start up the machine. Enter **saasbook** as the password.
3. Launch a terminal. (Third icon in the launcher on the left.)
4. Grab a copy of `setup.bash`, for example, from [this Dropbox link](#), and save it on the virtual machine. You can either do this via Firefox, or in the terminal: `wget https://www.dropbox.com/s/4qrbkw7joebygxb/setup.bash`
5. Run `sudo bash setup.bash`. Make sure you are in the same directory as `setup.bash` before executing this command. Enter the same password again to install lots of packages.
6. Grab a coffee or something - it will take a few minutes to build/install these components. Also if you see warnings etc. on the screen, don't worry, that is expected.
7. **STOP:** If you are only interested in working in Python, you have now installed everything you need. If you are interested in using Julia, skip the rest of this section and move on to the Julia section. 8. To test that everything works correctly, run `ipython notebook`.

In the browser window that pops up, create a new notebook, and enter the following in the first cell:

```
%pylab inline
x = np.random.randn(5000)
plt.hist(x, 50)
```

You should see a beautiful normal distribution in the output, as seen below. To exit, simply hit Ctrl+C, 'y', and close Firefox.

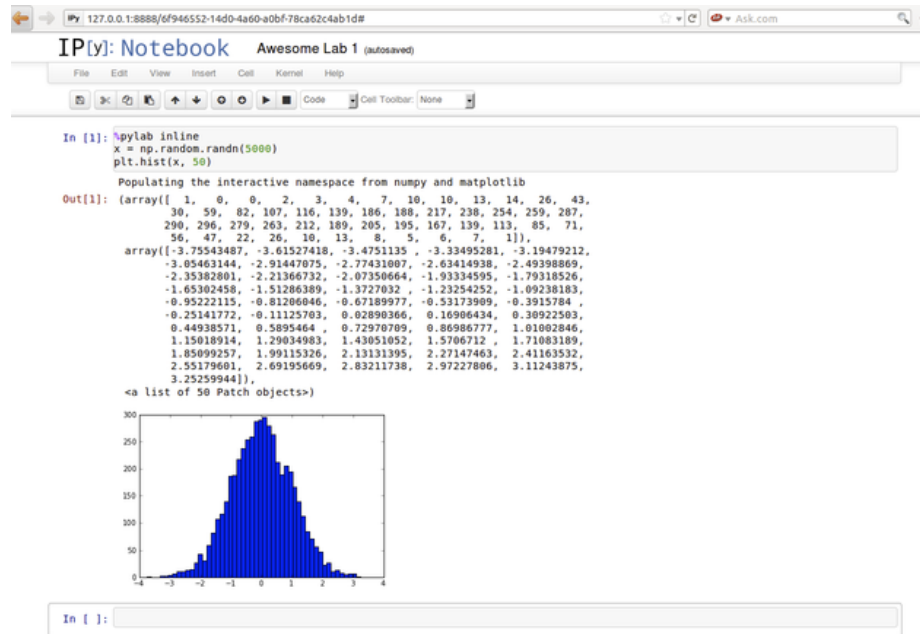


Figure 1: normal

## Julia/IJulia

Continue on if you are interested in working in Julia and IJulia.

9. Type `julia` in the command line to start the Julia REPL, and install IJulia with: `Pkg.add("IJulia")`.
10. Continue the installation with `Pkg.add("PyPlot")` to install PyPlot, which is a plotting package for Julia based on Python's Matplotlib.
11. If the commands above returns an error, you may need to run `Pkg.update()`, then retry it. To exit the Julia REPL, hit Ctrl+D
12. To start the IJulia interface, run `ipython notebook --profile julia`. In the browser window that pops up, create a new notebook, and enter the following in the first cell:

using PyPlot

```
x = randn(5000)
plt.hist(x, 50);
```

You should see a beautiful normal distribution in the output, just like the picture above.

## TeX and nbconvert

13. If you want to use the `nbconvert` utility to convert your IPython/IJulia Notebook to pdf format, you will need to install some flavor of TeX. One way to do so is via the `texlive` package. Be aware that the following command will download about 600 MB of software, which expands to over 1 GB. In other words, it will take a **very** long time to download and configure.

```
sudo apt-get install texlive-latex-extra
```

You can always right click from the web browser and choose to print the notebook as PDF, which is much easier, though the formatting may be messed up. You can also look into other convert options [here](#)

## Don't like working in the VM?

*(Credit goes to Yan Zhao, Victor Lyman, and Sridatta Thatipamala - current CS194-16 students)*

Good news - you can still work on your local machine using your own browser! Follow the instructions below.

## Setting up SSH

14. Select your VM inside VirtualBox Manager.
15. Go to *Settings > Network > Advanced > Port Forwarding*.
16. Add a new rule with “**name => ssh, protocol => TCP, Host Port => 3022, Guest Port => 22**”, and everything else BLANK.
17. INSIDE your VM, install the ssh server with `sudo apt-get install openssh-server`
18. SSH Command: `ssh -p 3022 saasbook@127.0.0.1`.  
SCP command: `scp -P 3022 saasbook@127.0.0.1:~/VM_DIRECTORY_NAME LOCAL_DIRECTORY_NAME`

19. If you didn't change, the password is **saasbook**
20. Optionally, you can look into SSH-ing without a password. There are many tutorials on the web, one of which can be found [here](#).

## Accessing notebook from host system

21. Add another rule to Port Forwarding with guest port as 8888, and host port as 8888.

To help double check that you get everything correctly, here's a screenshot of what you suppose to see after adding the above rule



22. To start the notebook, run `ipython notebook --no-browser --ip='0.0.0.0'` instead since the guest system won't expose 127.0.0.1 to other machine in the LAN.
23. Open **localhost:8888** on your host browser, you will see the notebook page. Congrats, you can now work in your local browser!

## Headless VM

Now you may wonder, if I can do all of the above tricks, why do I need to open a window of the guest system. The answer is you don't have to :)

24. Run the following command in terminal to start a headless VM:

```
VBoxManage startvm "name_of_vm" --type headless
```

where "name\_of\_vm" should be the name of the virtual machine as in VirtualBox Manager

25. To turn off the vm, just run `sudo poweroff` in the ssh session.

## ssh config file

If the ssh command is too long to type, put everything in a config file.

26. The config file is `~/.ssh/config`, if it is not there, just create one.
27. Put the ssh host in the config file in the following format, notice the blank before User, Hostname, port is a tab:

```
Host vm70
  User saasbook
  Hostname 127.0.0.1
  Port 3022
```

28. Just type `ssh vm70` to ssh into the vm.

## Frequently-asked Questions

- Q: Do I have to use Python or Julia? Do I have to use the VM?  
A: **No**, you are free to use any software/library/package on any OS, as long as you complete the lab and what you do is reproducible (please always document the steps).
- Q: I have an error installing Virtual Box / I can't get my VM to boot.  
A: Sorry, but this is out of our reach. We do not own nor host the image. Your best bet is to consult Google.
- Q: I'm following the instructions, but I get an error at step X. Help!  
A: In this case, please also try to Google first. Chances are the problem have been solved (many times) in the past. If you find a fix, please post it on Piazza and share it with everyone. You are a wonderful human being for helping your classmates and instructors.  
If Googling doesn't help, please post it here, along with the error message and at which step you encountered the error, and we'll try to look into it.
- Q: I see that you guys provide solutions in Python, Julia, and Matlab, but there's no instruction for installing Matlab. What gives?  
A: Unfortunately, Matlab is a proprietary software, meaning you will have to pay (or use the school computers) to use it. It is possible to install Matlab's free, open source brother Octave, as well as R, but we will not provide official support for other languages or software packages.
- Q: If I use the VM to complete the lab, do I still have to document the steps?  
A: Yes, but it doesn't need to be detailed like this. Please also don't

copy paste this entire document into your writeup. At the very least, you should mention the packages that you use to complete the lab (NumPy, Matplotlib, etc.) and link to the VM download page or this instruction page, so that anyone who read your homework can reproduce the exact steps.

- Q: How do I copy the lab and/or writeup back to my local machine for submission?

A: There are many ways. An easy solution would be to email it to yourself or upload it somewhere (Dropbox, Github) but please never make it public. If you want, you can set up SSH/SCP (as seen in the **Setting up SSH** section), and scp your files back and forth.