Initial Reactions to the TiaToolbox: README.md

By David Epstein

TiaToolBox is an excellent idea, to which I am happy to devote time.

The target user groups of TiaToolBox—I will refer to them simply as *users*—mentioned to me by Shan, are 1) beginning PhD students and 2) pathologists. This covers a vast range of initial user knowledge and skill. This vast range needs to be kept in mind when writing documents for the toolbox. Perhaps there should be a separate document saying what background is assumed—I will call this not-yet-existing document, the Tialab Software Welcome or TSW. TSW should be short. Each topic should have a single link (or at most two links) to a very suitable webpage. Too many links to one topic are confusing and should be avoided. There are probably many examples of Welcome Packages on the Internet that can be used as models.

The first item of the README file

<https://github.com/TIA-Lab/tiatoolbox/blob/develop/README.md>

should be a link to TSW. Some topics that I think should be included in TSW:

1. Github.
2. Git.
3. Anaconda
4. conda.
5. conda virtual environments. Do not discuss other varieties of virtual environment.

I may want to extend this list when I start looking at the Jupyter Notebooks in TiaToolBox.

Since I have a mac laptop, many of my remarks will be relevant to macos only.

I use red to indicate snippets from the README file.

* OpenSlide [here](https://openslide.org/download/).
* OpenJPEG [here](https://www.openjpeg.org/), using conda install -c conda-forge openjpeg>=2.3.0.

This is too terse. Perhaps give one-line explanations of why OpenSlide and OpenJPEG are needed.

The links to OpenSlide and OpenJPEG should not be daunting, technical lists of possible downloads as at present. A possible consequence of the current README document is that some users will spend a couple of weeks struggling unsuccessfully to get source code (perhaps extracted from tar files) to compile—the opposite of what we are trying to achieve, and bad for morale. Instead the links should be to a chatty non-technical introductions to their benefits, and an explanation of why openslide and openjpg are important components of tiatoolbox?

It seems to me that the above snippet from README confuses two different uses of the word *install*. Users will probably be familiar with installing a package that makes certain commands accessible to the OS shell. But *conda install* is somewhat different, and makes sense only when conda is operating within a certain virtual environment. I’m not sure what the *conda install* command would do if a user was following README carefully. I wouldn’t be surprised in the command would hang, because no environment had been activated. Or else openjpeg gets installed in the base. Since tiatoolbox doesn’t yet exist at that point in the README sequence, and openjpeg is not mentioned in the requirements file, the user won’t have it in the relevant virtual environment.

Commands should be given in the following order.

1. Create a directory in which to work. Change to that directory.
2. Clone the TIAToolBox into that directory
3. Locate the requirements file in the clone.
4. Create the virtual environment tiatoolbox using a requirements file. Explain where the environment has been placed. Explain that the name of the environment is hidden inside the requirements file. (Consistently and throughout, try to demystify. An expert may prefer terseness and brevity, but few of the users are experts.)
5. *conda activate tiatoolbox*
6. *conda list*---this is where one might make a few remarks about some of the packages used by tiatoolbox and why they are helpful.

Omission of some of these six commands from the README file is the underlying cause for the errors I have found..

If openslide and openjpeg are listed in the requirements file, then one should be able to avoid sending the user on a needless battle with the unnecessary plethora of choices in openslide.org and openjpeg.org, One should be able avoid failure of the *conda-env create* command in Step 4 above by listing the right channels, or by some other method.

One advantage of my recommended approach as compared with

conda install -c conda-forge openjpeg>=2.3.0

is that the *conda install* command is silent and naïve users will think it has failed. In fact users who think that the command has failed are likely to be correct: using the current README file, since *conda install* will not know into which environment to install openjpeg.

The line

pip install -r requirements.txt

should be deleted. It’s confusing and unnecessary to give alternative methods of doing the same thing.

Do you want users to execute the following commands one after the other? If not, how would users know that they are only supposed to do some of them?

python -m tiatoolbox -h

pip install tiatoolbox

conda env create -n tiatoolbox python=3.{6,7,8}

conda activate tiatoolbox

pip install tiatoolbox

Alternatives are confusing and should be avoided unless there is good reason (like the difference between Macos and Windows). Since README is already using conda, just stick to that, and don’t give any alternatives. Perhaps you should say what to do if the commands fail.

Is the README file the right place for the help command? The best way for someone to learn how to use a command is not to have a formal list like this, but to see lots of examples, with explanations of what the commands achieve. Once the user understands the command and its options/flags, the formal description as in the README file can be an efficient way to recall details.