NAME: OYEDAYO OYELOWO STUDENT NUMBER: 014717208 DATE: 05/11/2017

Read the paper Good Enough Practices in Scientific Computing by Wilson et al. (2017). Some of the advice in this paper is most useful for researchers, but much of it is applicable by students as well. List five recommendations from the paper that you think you should and could follow. Why? Would it be easy to follow this advice? Return your answers as a pdf-file. You can write in either Finnish or English.

1. **Ensure that raw data are backed up in more than one location**

This is perhaps the most important thing to do at first. This should be done, in case of any goes wrong with the computer or hard drive I use. Even cloud storage might be hacked, or inaccessible if one forgets the username or password. As unlikely, as most of these may seem, it is better to be safe than sorry and prepare for the worst-case scenario. With this, I can always access and return to the original data.

Considering the benefit of safety and now, that storage capacities and internet speeds are better, this does not seem difficult to follow. Although, one might be hesitant or procrastinative sometimes, repeated practices make it a habit and would encourage me to always have my raw data backed up both in my hard drive and in cloud.

2. **Create the data you wish to see in the world'**

This is crucial even if it is not a collaborative project. Data should be user-friendly because humans need to make meaning from it, except one intends to encrypt somethings. However, as science are leaning towards the open paradigm, it is important to make the data readable and unambiguous in the naming of variables. They should be well documented and immaculately collected for ease in amenability to computing too.

This should not be so difficult for me to follow because I recognise the usefulness, even for myself when I have to revisit or reuse the data in the future. Also, I would love other people to have access to data I have garnered, as better techniques can be developed, which is the whole essence of science.

3. **Give functions and variables meaningful names.**

I already put this into practise. While many people might think it is convenient to have shorter names or vaguely coded names, I ensure I name my variables, object and functions in a quite comprehensive, yet short and comprehensible ways. I also strive to explain what every line does and what the names stand for. Better to have more than less.

This also should not be difficult for me to put into practise as I already do it. When I return to my old scripts, I'm still able to understand everything even without having to read the comments deeply. This practise would especially be very useful in large collaborative projects. It is not nice to share your code and others have to ask many questions about the functions and the names. Also, it should be kept very simple.

4. **Record all the steps used to process data**

This should be done to understand the way the data cleaning and mining were done. When analyses are done and somethings seem odd, it becomes easy to diagnose and see if it could be from the data processing.

This does not appear herculean to me either. I like to make my life very easy when I revisit my data and analyses and I make sure I am very meticulous with the recording or documentation of every step. I do this because it is easier for me to crosscheck in the future and for others to understand my work when I share them. This eases reusability and repeatability.

5. **Create an overview of your project**

Similar to aforementioned, having an overview of every project gives a summary of what the project is about and other useful concise information. This is something that should already be a normal practise by every student and researchers. When others browse through your projects, they should be able to get a broad idea of what it is and if that might be interested in contributing or improving it. It is also useful for you too. Admittedly, as humans, we are tempted to believe we will always have a vivid memory of every project in the future but this is not necessarily so. When one creates an overview of every project, it becomes very useful, when one needs to reuse later.

I believe this should not take much time because it does not require much technical or mental exhaustion. This in fact, should be one of the primary things to adopt as it includes things like the title of the project and other updated information about it.