From Student Site
< User:Thomas Hollis

Thomas Hollis' insights!

(Thomas Hollis' insights! page)

Contents

Top 25 Shortlist

For sake of concision, the following constitutes my top 25 bioinformatics insights by chronological order (effort was made to create rhymes but alas Shakespeare's wit is a challenge mere bioinformaticians can only strive toward).

Insight #1 - Standing On The Shoulders of Giants

In 1675, Isaac Newton is famously quoted to have said: "If I have seen further it is by standing on the shoulders of Giants." [1]. In a similar vein, my first few days spent on this course made me realise the following insight that I have phrased into a poem:

The future will always be better than the past

From the shoulders of bigger, better and more numerous giants

We may now use new tools to learn increasingly fast

Allowing us to visit more science and become its ultimate clients

16 September 2018

Insight #2 - You Know What a Good Insight Is?

You know what a good insight is? Rhetorical questions. After having read "What do you think I think of that? Exactly." on the Netiquette page (http://steipe.biochemistry.utoronto.ca/abc/index.php/FND-Netiquette) , I had not only a wholesome chuckle but also an insight. As an aspiring educator, the following advice will undoubtably be useful to future-me:

Never allow technology to make your educational delivery boring. The age-old tip of engaging with your audience is just as valid for online content!

16 September 2018

Insight #3 - Always Follow the Breadcrumbs

Sometimes I wonder if I follow too many breadcrumbs. Today, I was not disappointed when I followed the URL to Prof. Steipe's GitHub (https://github.com/hyginn) . The profile picture being a QR code was too enigmatic for me to ignore. Like a child on an easter egg hunt, I scanned it and spent the next hour getting lost in Wikipedia's various pages on the art of Hieronymus Bosch (https://en.wikipedia.org/wiki/Hieronymus_Bosch) .

Always follow the breadcrumbs

As it may lead to fascinating outcomes

You never know what you might stumble across

Just make sure that you minimise time loss

17 September 2018

Insight #4 - Never Forget the Big Picture

I have always believed that maintaining a high level view of the situation helps to motivate me to work in an efficient and targeted manner. It always me to know what I am doing and to know where I want to get to.

Therefore, I have decided to gradually check off each unit once I have finished going through it in a graphical high level way.
(see screenshot of progress aside)

A screenshot of my current progress

17 September 2018

Insight #5 - Bluelight Is Your Sleep Cycle's Worst Enemy

I have known for many years now that blue light (light closer to the UV side of the EM spectrum) was particularly good at suppressing the secretion of melatonin. I have however always assumed that, if I was *really* tired, my lack of melatonin would not be an issue. However, after much late night coding on a monitor with fullscreen RStudio I find myself in disagreement with my past-self.

RStudio has a night mode that can be activated in the global settings!

18 September 2018

Insight #6 - Why You Should Never Trust Someone Who Says "Trust Me"

When Prof. Steipe says (without justification) that I should take the habit of always setting `stringsAsFactors = FALSE`, my eyebrows raised immediately. This made me think of people who make a statement or recommendation and when prompted to explain the rationale behind this, their reply is: "Trust me". I have found that, in most cases, "trust me" usually means "I have an idea why but I don't want to explain it to you". In many cases this is indicative of a weak idea where the supporting data is missing (e.g. "Trust me, X is a bad politician"). However, in other cases this is indicative of a highly knowledgeable person whose time is limited (e.g. "Trust me, that won't work"). I believe that in both cases the investigation is entirely worth it! If someone has a poor argument, take the most charitable interpretation of the argument and try to find data for or against it. In the case where someone doesn't have time to explain it to you, don't take their word for it - look it up yourself! You'll never know what kind of secrets could spill out. In this case the secrets behind `stringsAsFactors = FALSE` were definitely worth it (https://www.tutorialspoint.com/r/r_factors.htm) .

Never trust someone who says "Trust me".

18 September 2018

Insight #7 - Keep a (Dream) Journal

I woke up in the middle of the night with an absolutely fantastic idea of something to write in my insights page. I had been visualising it in my dreams and I finally came up with something cool. Unfortunately when I woke up the next morning I had forgotten all about the idea. On the other hand this in itself provided me with an insight

Always keep a (dream) journal

For your memory is not eternal

Sleep or time can seize your thoughts to format or wipe

Keeping journals is not merely for Prof. Steipe

19 September 2018

Insight #8 - The N-Sided Dice Question

In lectures a few days ago, Prof. Steipe mentioned his dilemma of creating a physical fair *n*-sided dice for any *n* such that the dice has an axis of symmetry. I didn't think about this again for a while until my shower this morning where I asked myself what the most aesthetic solution would be. By the time I had dried myself off I would like to put forward my solution:

My solution is an n-gonal regular hosohedron (https://en.wikipedia.org/wiki/Hosohedron) but with each vertex thickened to prevent excessive rolling (think of an under-inflated beach ball with stiff sticks between each colour).

19 September 2018

Insight #9 - Humiliation Is Good Medicine

In lectures a few days ago, I had my incompetency revealed to myself in the same time as it was revealed to my classmates. I have never felt so silly before but it actually provided the invaluable lesson of never submitting work in a rush (or in my case very late at night)!

Humiliation is key, humiliation is learning, humiliation is good medicine

5 October 2018

Insight #10 - If It's Broke, Fix It

Unfortunately I was very attached to BABIN, my first MYSPE, as I had read a lot about information around it and wrote a wikipedia stub page for it. Unfortunately its sequence was incomplete and therefore BABIN had to be changed. I initially wanted to ignore this issue but I realised quite fast that in order to continue I really needed to fix this.

If it's broke, fix it

13 October 2018

Insight #11 - Disciplined Learning is Key to Avoid Swiss-Cheese Knowledge

I think that a lot of the learning units present us with brand new packages, explore them briefly without giving us practice or exercises and moves on. I often felt, when not being quizzed during the R tutorial script, that I was a lot more lenient on myself with my own learning when I was not implementing what I read. This came back to bite me as I figured out I was accumulating Swiss-cheese knowledge causing me to have to return to old units since I had forgotten many concepts. From now on I am more self disciplined and try to play and implement a lot more as I go through tutorials, even if not explicitly demanded.

Disciplined learning is key to avoid Swiss-cheese knowledge.

24 October 2018

Insight #12 - Understanding Your Own Biology is Understanding Yourself

In a spiritual moment I realised that the more I learn throughout this course the more I understand myself. When I started this course I got my genome sequenced. Since then I have been looking up papers on particular gene variants and their impact, tracking down the SNPs and checking my own SNPs. The more I understand how my body works the more my own actions and reactions make sense.

Understanding your own biology is understanding yourself.

26 October 2018

Insight #13 - Healthy Skepticism is Healthy ... But is it?

When writing R code it is all too easy to see the code run as expected without any bugs or warnings and produce the output desired. However, I have found that if something compiles and runs perfectly first time this is a miracle too good to be true. A bit of rubber duck debugging a day keeps the null pointers at bay.

Healthy skepticism is healthy ... but is it?

31 October 2018

Insight #14 - Do Not Expect Elegance in Biology, Expect Performance

Before taking the sequence alignment learning unit I expected to see a very neat method to align sequences. Maybe an elegant statistical method using maximum similarity or something of the like. I was bitterly disappointed when I found out how messy and imperfect the alignment process can be due to the biochemical underpinnings. This frustration kept me up at night as I wondered why such a lack of elegance was present in nature. In these moments I like to go back to basics and challenge my underlying assumptions. The critical question I asked myself which eventually put my mind at ease was: "Does evolution naturally select elegant solutions?". The simple answer is no. The complicated answer is not really. Some elegant solutions are naturally selected as they are more optimal but evolution has no reason to select a more elegant or simple method over an equally performant messy method. Upon realising that evolution was not in line with my desires I started questioning my own desires. Why do I desire elegant solutions? Is Occam's razor founded on strong evidence or intellectual laziness? Should I really write clean code if the most efficient computer on earth (human brain) is not trying to maximise elegance? Luckily, my knowledge in Machine Learning came to the rescue. If Occam's Razor was not correct, using regularisers would not make a difference on overfitting. I have seen countless model performances be increased by the use of a regulariser so I do think Occam's razor is founded on strong evidence. Indeed, it appears that human-designed algorithms have a slight edge over evolution when it comes to elegance!

Do not expect elegance in biology, expect performance

2 November 2018

Insight #15 - Compile High Level Summaries Early

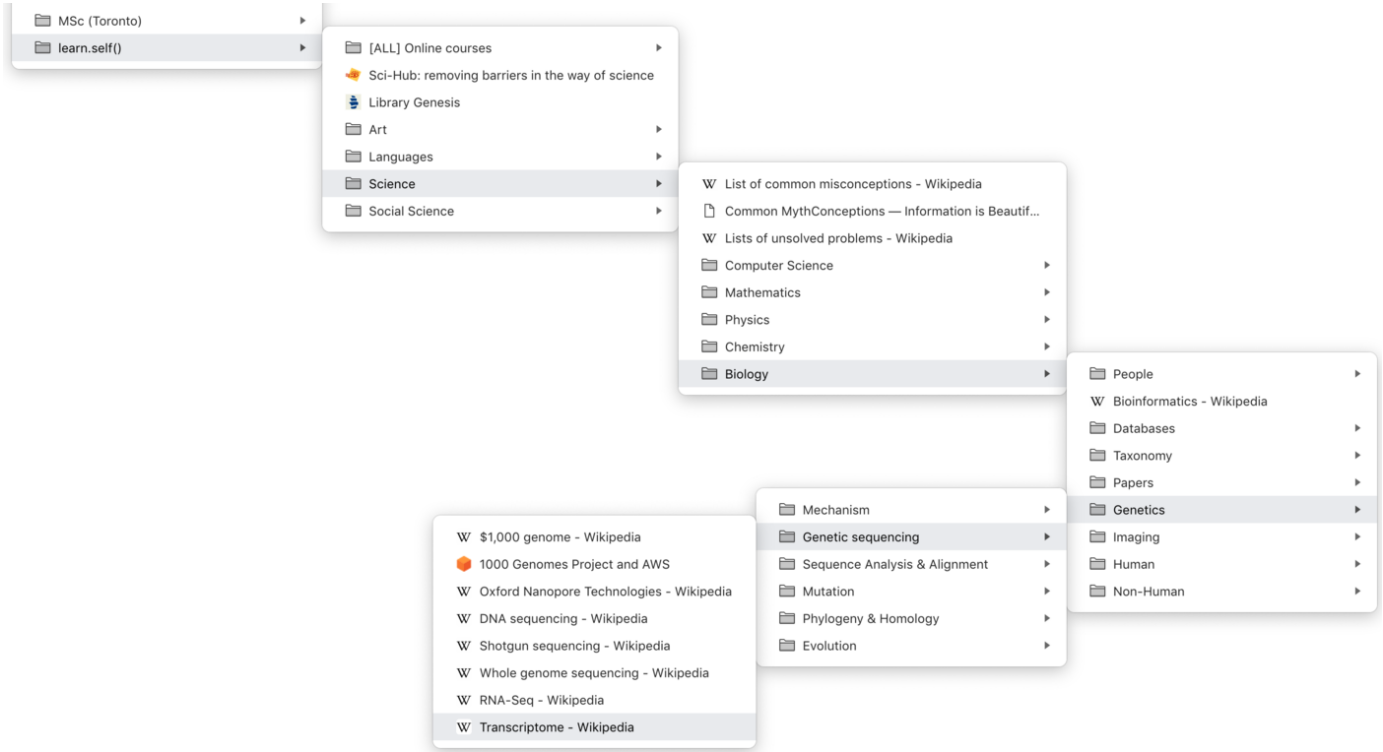
As I get quite advanced in this course I feel like I should have made a shortlist of my favourite databases in bioinformatics (a bit like I made my revision sheet of R syntax). I know that NAR publish a lot of material grouping all of the best resources together but my own personal list would be nice to have to help me compile all my tools in one place. I feel that now it is a bit late but I will backtrack through past units to compile this list. Perhaps I will make it available here or in my own unit design task.

Compile high level summaries early

4 November 2018

Insight #16 - Build A Tree Of Knowledge

After reading about the Tree of Life in the Phylogeny Concepts learning unit, I decided it was time for me to share one of my most valuable learning tools discovered to date: bookmarks. Now this may sound silly, but used correctly, bookmarks can become an incredible way to memorise and absorb a high level overview of all the different nuggets of information fed to us throughout our life. Everytime that I look something up of value on Wikipedia (or elsewhere) I bookmark it in my Tree of Knowledge (rigorously organised bookmark structure):



Note: The above also has the added benefit of caching all these terms so that if you need to look something up again, the second time you search it will come up pre-loaded instantaneously in the Chrome Omnibar (similar behaviour will occur on Safari, Firefox and other browsers).

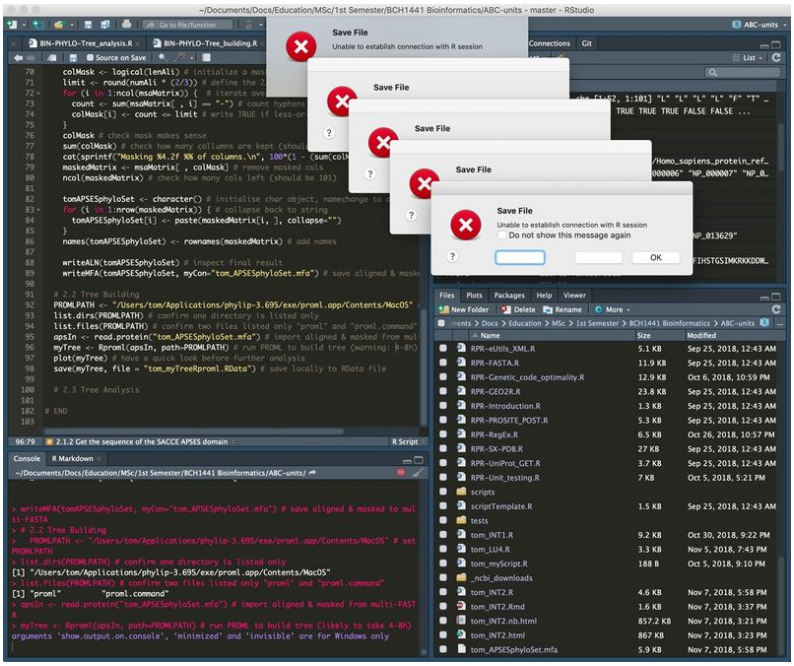
Build A Tree Of Knowledge

5 November 2018

Insight #17 - Take Error Message Documentation Seriously

I should preface this insight with a confession that this is going to be more of a rant than an insight.

RStudio, while running an 8 hour long method call by PROML started to repeatedly issue the following error messages:



Now don't get me wrong, I love a bit of constructive criticism. But this is not helping. Who decided they would give me the option to chose between " ", " " or "OK" in such a critically long method call! How do I know this prompt hasn't interrupted the call? I would know if RStudio implemented progress bars but I guess that's too much to ask so if the developpers of a package don't write their own then the users are left in the dark... Sad!

Anyway, I know what will be able to help me out: official R error discussion forums! Let's have a look what I found:

Fantastic, right? Believe it or not this was actually a **design choice**! The head of department of R forums decided that for old posts he would simply replace all the usernames and pictures by his own. Why? Because John Malkovich (https://www.youtube.com/watch?v=Q6Fuxkinhug) ...

Take Error Message Documentation Seriously

6 Nov 2018

Insight #18 - The Power of Literate Programming

When I first took the Literate Programming unit (bonus unit), I thought my burst of motivation and keen energy would only help improve my grade in this subject and I did not really believe I would end up embracing the concepts of Literate Programming. A few months later here I am, writing every single one of my machine learning assignments in a Jupyter Notebook... In fact, I have started writing some of my R scripts into .Rmd files and knitting them in HTML into a notebook! I love this new method of programming! It is so useful for teaching or running code in chunks! Fun fact: Kaggle (the biggest Machine Learning Competition platform) encourages literate programming via its user of publicly-accessible Kernels! I am sure this is no coincidence...

Therefore, I have decided to embrace the Power of Literate Programming.

7 Nov 2018

Insight #19 - Bug Hunting as a Way of Learning

There are many ways to learn how to write code. Bottom up: learn intricate details of a language from documentation and write carefully and incrementally more challenging code. Top down: throw yourself in the deep end of complex code and figure out how it works and how to change it to suit your needs. However, I feel that bug hunting is more of Middle-Out learning approach. You need to a have a strong understand of the language you are coding in but you also need to be going through code that is foreign and written by someone else. So you learn from both approaches at once. (If you, dear reader, were expecting a Silicon Valley joke, I am afraid this section will disappoint). I found this insight while I patched a few bugs in the mailing list and helped find a workaround for deprecated software tools.

If you want to learn a language while optimising for a strong baseline - go for bottom up
If you want to learn a language while optimising speed of discovery - go for top down
If you want a compromise of both and have some previous experience - go for middle out

7 Nov 2018

Insight #20 - Must All Good Things Come To An End?

I am getting close to the end now of BCH441. It almost feels sad as I can feel the day where I run out of units is just around the corner. It got me thinking if the quest of learning new things is one of those things in life that break the eternal saying "all good things must come to an end?". I believe so for now. Perhaps it is that saying itself which must come to an end (no mention of GEB will be made in this insight, do not fear).

The Quest For Knowledge Can Only Be Brought To An End By Death Or Dementia

8 Nov 2018

Insight #21 - Complete Means 100%

I decided I would check out all the optional learning units because I believe that something is not truly complete if all rocks have not been turned over. All the optional units are now complete so I know I will reach my goal of 100%!

Complete Means 100%

9 Nov 2018

Insight #22 - Version Control Saves Lives

Version control is absolutely fantastic. Git is God. Wiki history is fantastic. Writing code (and my CV) with version numbers and keeping all the old version is so useful I absolutely have to leave it here on my insight page!

Version Control Saves Lives

10 Nov 2018

Insight #23 - Do Not Spread Yourself Thin

I had around 3 units for evaluation currently being worked on at the same time (since I was stuck in two of them) and it got really confusing so I stopped and decided to complete them one after another, addressing each issue deliberately. I feel like this might be intuitive but it is all to often forgotten when maximising work output!

Do not spread yourself thin

10 Nov 2018

Insight #24 - Perfectionism: A Cautionary Tale

I am not sure if this is an insight but I am finding myself spending A LOT of time going through my journal and insights pages with a fine tooth comb trying to perfect all the small details. I am not convinced how productive this is but I feel like it is something I have to do for my own sanity...

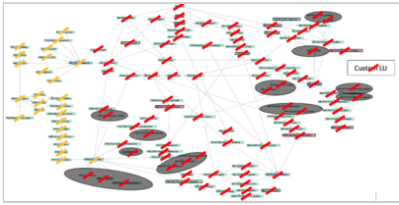
Perfectionism - A Cautionary Tale

11 Nov 2018

Insight #25 - Reflect on How Far You Have Come

As hindsight is settling in, it seems to me that looking back at the course map progression (shown in insight #4) allows me to see visually how much I have achieved. This has come to be hugely beneficial to me and it seems to others as well who have similar maps in their insights.

Therefore, I have decided to repost Insight #4 but updated showing the overall achievement (see screenshot of progress aside).



A screenshot of my final progress

05 Dec 2018

Honorable Mentions

Upon rereading my insights, I have found another insight: many of my insights use language such as "all", "never, "always" or "worst". This is bound to come back to bite me as I am sure I can find exceptions to many of these insights. Perhaps my final insight should have been:

The only strong opinion anyone is entitled to is the strong opinion that nobody is entitled to strong opinions.

More insights can be found on my GitHub (https://github.com/PsiPhiTheta) or by asking me in person.

References

1. ↑ Newton, I. Letter from Sir Isaac Newton to Robert Hooke. Historical Society of Pennsylvania. Retrieved 16 September 2018.
