Proposal for addressing the necessity of students in the Life Sciences pursuing additional Computing background

For:

Vanessa Auld,

Head of Department of Zoology

And

Prospective and Current Undergraduate Students in the Life Sciences

University of British Columbia

Vancouver, British Columbia

By:

Aleem Tariq

10417153

December 17, 2020

Table of Contents

ABSTRACT

Preparing students to be up-to-date with the demands and expectations of their respective academic field is a critical aspect of every University, and the University of British Columbia has a responsibility to ensure this occurs. This report aims to address the gap between the computing skillsets expected from researchers in Life Science fields and the skillsets that students in this field possess, through recommending measures both departmental heads and students can take action on. Data collection through surveys of research professors were conducted, in fields ranging from Neuroscience to Molecular Biology, Evolution, Systematics and Physiology.

Findings suggest that not only do these respective fields heavily involve computing techniques, but that the emphasis on this technical background will only increase as time passes. In addition, researchers prefer students with strong backgrounds in programming, data analysis, and statistics, with additional preference for those familiar with Machine Learning.

As a solution, this report found that in addition to students seeking out extra resources and taking elective courses in the Computer Science faculty, the department could:

- Add additional Computer Science, Mathematics, and Statistics pre-requisites.
- Create courses for students in the Life Sciences that focus on the computational and analytical techniques used in the field of research.
- Revamp the current courses, so that a Life Sciences stream has a computational undercurrent that is developed and enriched throughout the degree.

I. INTRODUCTION

A. BACKGROUND OF UBC AND ITS ROLE AND OBLIGATIONS

The University of British Columbia (UBC) is a global leader in teaching, learning, and research, having been consistently ranked among the top 20 public universities in the world. With regards to the Life Sciences UBC features the Life Sciences Institute (LSI); the largest of any such institute in Canada, and among the best by many metrics (1), whose research mission is to perform innovative, interdisciplinary science.

As per the institutions role in being a global leader and possessing the largest Life Science Institute of its kind in Canada, the University also bears a set of responsibilities and obligations towards its students. Most notably, in cultivating the next generation of academics, employees and employers, which are prepared to face the rapid pace of change seen in the professional world.

Therefore, it stands to reason that the Life Sciences should sufficiently prepare the next generation of students for what is expected in the working world of research and industry related to the field. Both the scale and scope of applications of computation is rapidly increasing and having a background in such an area confers the ability to both work with data in novel ways, and the critical analytic skills to interpret such data (2). Given the nature and complexity of the fields in the Life Sciences, a large-scale revolution that involves extensive use of computation is occurring and will continue to accelerate (3).

B. OVERVIEW OF PROBLEM AND PURPOSE OF REPORT

Given the nature and complexity of the Life Sciences, working in research currently requires and will increasingly demand computationally intensive skillsets, in areas ranging from Neuroscience to Genomics to Immunology and Ecology, as data-analysis has started to take precedence over data collection. (4)

These skillsets are not covered by the pre-requisite demands of multiple departments in the UBC Life Sciences. The Life Sciences, in general, provide insufficient preparation for the changes which are emerging in both academia and industry.

The purpose of my report is to:

1. Explore the importance of Computer Science in the field of Life Sciences; primarily research.
2. Communication with faculty in the Life Sciences, collecting data on the computational tools they both use in research, and demand from prospective researchers.
3. Develop a strong case to sufficiently motivate the intended audience to acquire additional computational background.
4. Offer recommendations that may address this problem to the departmental head of Zoology.

C. METHODS AND SCOPE

My primary data sources (see Appendix A) will consist of surveys of current UBC professors conducting research; all of which are aimed towards assessing the necessity of recommending that students in the Life Sciences pursue obtaining computational skills. Eleven professors who conduct research in areas ranging from Cognitive Neuroscience to Ecology, Biomathematics and Molecular Biology responded to the survey.

To assess the necessity of such a recommendation, the following questions were addressed in survey, and will be expanded upon within this report:

1. What are the perspectives of professors on the role of computing in their research?
2. What sorts of computational skills do professors seek out from prospective students interested in research?
3. What changes do professors anticipate will occur in their field with time, and will such changes involve increased use of computational skills and techniques?

II. DATA SUMMARY

A. IMPORTANCE OF COMPUTING IN RESPECTIVE FIELDS

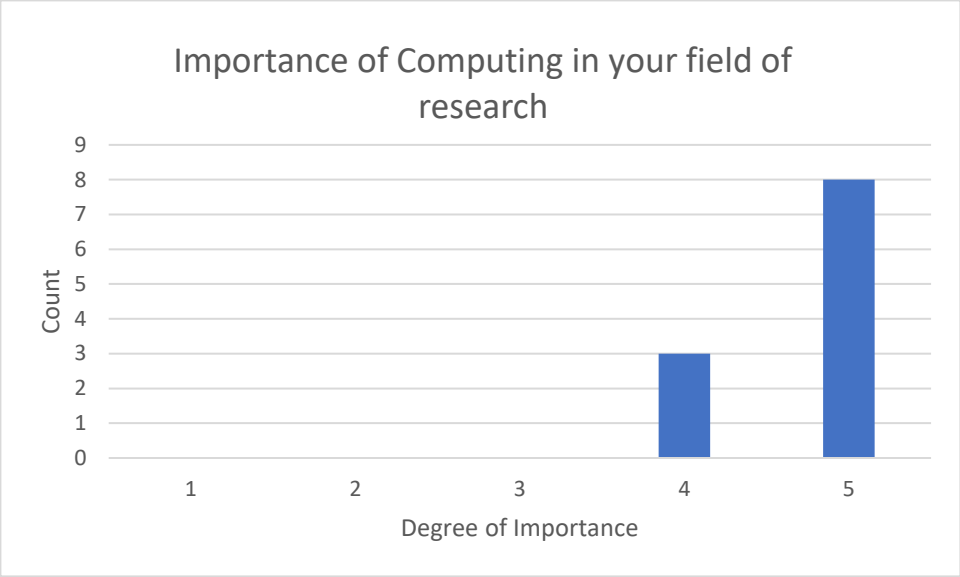

Fig 1. Bar chart of scores for importance of computing in one's field of research, ranging from low (1) to high (5), with number corresponding to count

Source: Tariq, Aleem, "Professor survey". Questionnaire. November 2020.

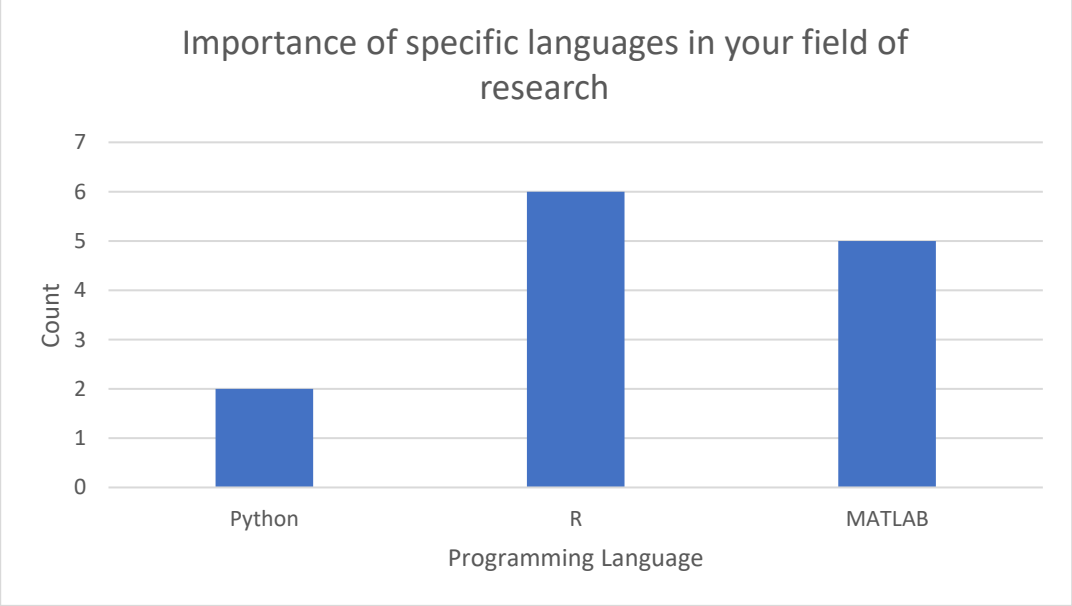

Fig 2. Bar chart of scores for importance of specific languages in one's field of research, all are scores of (4/5), with number corresponding to count

Source: Tariq, Aleem, "Professor survey". Questionnaire. November 2020.

Professors currently conducting cutting-edge research in a variety of fields have noted the significance of computing in their current research (Professor Survey, November 2020). In addition, various programming languages play a significant role in current research within numerous fields (Professor Survey, November 2020).

Based on my survey's results, computing is generally regarded as highly important in various fields of research within the Life Sciences. Virtually all research professors stress that computing is indeed very relevant to their field of research. 'Computing' as defined by methods such as data analysis, statistical tests, background in programming languages, and modelling and visualizing problem domains (Professor Survey, November 2020).

In particular, certain languages seem to be preferred over others. There is a significant preference for R and MATLAB. However, as of now there are a lack of courses which address this gap. Undergraduate students are not trained to be proficient in the use of these languages throughout their degree, as an underrunning theme of a 'technical skill' that is developed and expanded in every class. Rather, there are only two courses that involve the use of these languages, one of which is not mandatory (BIOL 300 & BIOL 301) (5).

B. NECESSITY OF ADDITIONAL COMPUTING BACKGROUND

Professors also near unanimously agreed that additional computing background would make students better researchers in their respective fields.

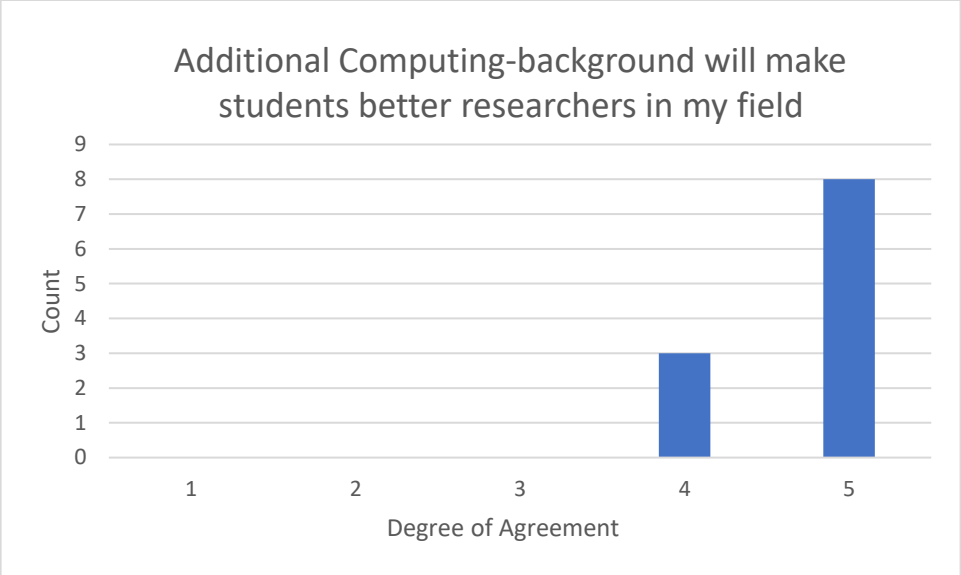

Fig 3. Bar chart of scores for additional computing making students better researchers, ranging from low (1) to high (5), with number corresponding to count

Source: Tariq, Aleem, "Professor survey". Questionnaire. November 2020.

With regards to specific skills that are highly sought after, professors tend to seek the following (Professor Survey, November 2020):

1. Data analysis and statistical tests.
2. Computational modelling skills, and an understanding of stochastic processes.
3. Basic coding (for loops, variables, data types, functions etc.) .
4. General programming and mathematical expertise, statistical knowledge.
5. Generating scripts to analyze large genome/proteome datasets in bioinformatics-based research.

6. Develop programs to enable faster and more complex data extraction.

There is general lack of these being covered, or if they are, only in highly restricted courses. An exception rather than the rule in terms of how the field is taught at an institutional level.

C. FUTURE IMPLICATIONS

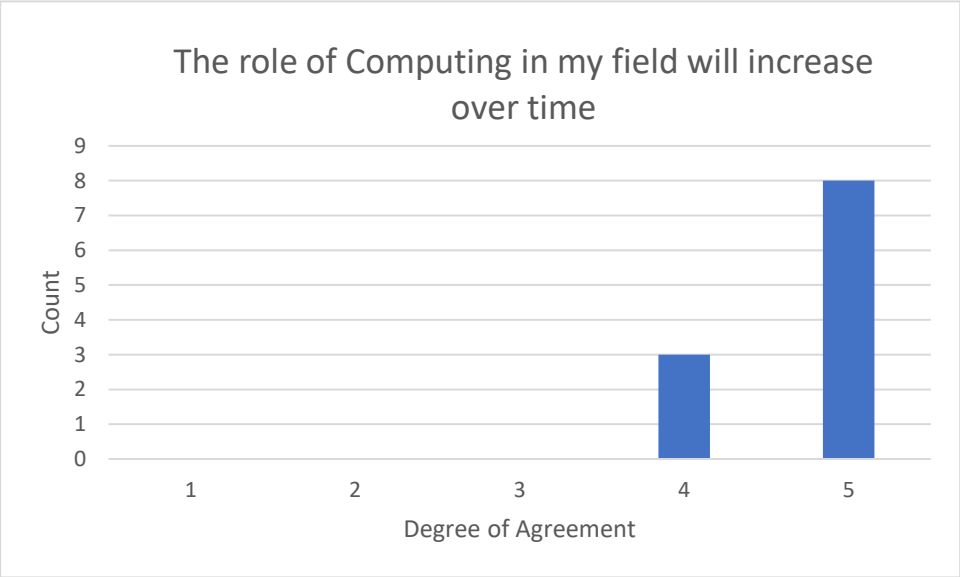

Fig 4. Bar chart of scores for role of computing increasing with time, ranging from low (1) to high (5), with number corresponding to count

Source: Tariq, Aleem, "Professor survey". Questionnaire. November 2020.

The research professors surveyed almost unanimously agree that the role of Computing in their respective fields will increase over time. One professor in particular noted that "being able to program will become only more important with time. It was an optional extra when I was in training, but it will be a necessity for the next generation of scientists. It gives you so much more freedom to ask the questions you want to ask and increases the impact of your publications." (Professor Survey, November 2020).

Their respective perspectives on the nature of such a change are highlighted below (Professor Survey, November 2020):

1. **Cog. Neuroscience** → The field is moving towards questions of how activity in populations of neurons, over time, reflects experience and controls behavior. This requires advanced methods for detecting and quantifying signals present across large numbers of neurons.
2. **Phylogenetics, Systematics** → Much of the research in systematics is going towards 'big data' – looking at thousands of nodes across many different lineages in the evolutionary framework. This requires increased understanding of computing languages and scripts to analyze these data, and also how to draw out relevant data for specific questions, like major trends or patterns across these different lineages.
3. **Neuroscience** → Increasingly large data sets measuring different parameters (gene expression, neuron activity, etc.), which will require more computational power to analyze.
4. **Cognitive Neuroscience** → Behavioral data can be analyzed and modelled with ML algorithms, which is a particularly exciting area for future research.
5. **Molecular Biology, Biological Sciences** → With bigger and bigger data sets (genomes, transcriptomes, etc.) computing becomes even more important. Big Data and -omics will become more prevalent, to capture the many variables that influence cell processes.

II. CONCLUSION

A. SUMMARY OF FINDINGS

Professors in general agree that not only is computing relevant in their field of research as of present, but will be increasingly so as time passes, and that they preferably seek out students with these skills to take on as research assistants.

Part of this is the nature of the field itself, in that there is a push towards 'big data' as being noted by professors as of present, which demands the technical skills necessary to analyze and work with this data. As per the findings, these skills are not only encapsulated within the realm of 'computing', but also include mathematical and statistical expertise. The issue is that modern life sciences are taught as if they are entirely 'technique driven'. Whereas in the past, most experiments one might conceive of were technically impossible, now with the numerous methods at present, the situation has reversed, and it is harder to conceive of an experiment that cannot be done. Consequently, the shelf is bare when it comes to drawing principles and patterns from the facts that have been revealed through the techniques often taught in the current institutions (4). There is no doubt that obtaining a skillset in Computing is a significant return on investment for a student in the Life Sciences. Such a student will not only be able to more proficiently work with data, but also approach existing problem domains in novel ways, and remain relevant given the anticipated changes expected in the field.

B. RECOMMENDATIONS

- Additional CS, Mathematics, and Statistics course requirements. The UBC Biology Department has already gone somewhat in this direction, with the inclusion of CPSC 103 recently as a mandatory course requirement. Additional course such as MATH 200, MATH 221, and CPSC 210 would partly address this gap.
- Expansion of available course options. In essence, more applied CS courses specifically for students in the Life Sciences. These would address the techniques currently used in the field, and rigorously involve problem-solving and working with data.
- Reformulation of current courses in the Life Sciences. The inclusion of computational techniques as an undercurrent to the entire stream. For instance, a laboratory course on Molecular Biology includes analysis of genomes and bioinformatics.

III. WORKS CITED

(1) *THE LIFE SCIENCES INSTITUTE*. lsi.ubc.ca/.

(2) Vincent, Robert D. "Applications of Computer Science in the Life Sciences." *McGill*, 29 Nov. 2011, www.cs.mcgill.ca/~jpineau/comp102/Lectures/26AIandMedicine.pdf.

(3) Brenner, Sydney. "The Revolution in the Life Sciences." *Science*, American Association for the Advancement of Science, 14 Dec. 2012, science.sciencemag.org/content/338/6113/1427.full.

(4) Sterling, Peter, and Simon Laughlin. *Principles of Neural Design*. The MIT Press, 2015.

(5) "The University of British Columbia." *Biology - Bachelor of Science - The Faculty of Science - Faculties, Colleges, and Schools - Vancouver Academic Calendar 2020/21 - UBC Student Services*, www.calendar.ubc.ca/vancouver/index.cfm?tree=12%2C215%2C410%2C418.

Tariq, Aleem, "Professor survey". Questionnaire. November 2020.

IV. APPENDICES

Appendix A: Professor Survey Questionnaire

1. Field of Research/Focus (e.g. Neuroscience, Immunology, Cognitive Neuroscience, Ecology, Biomathematics, etc.)

   Short answer text

2. Importance of computing in your field of research (e.g. data analysis, statistical tests, background in programming languages, modelling and visualizing problem domains)

   Low 1 2 3 4 5 High (circle)

3. Importance of specific languages in your field of research (e.g. in tasks such as data analysis, statistical tests, modelling problem domains, etc.)

Low	1	2	3	4	5	High	(circle)
R	1	2	3	4	5		
Python	1	2	3	4	5		
Java	1	2	3	4	5		
Ruby	1	2	3	4	5		
SQL	1	2	3	4	5		

4. Additional computing-background will make students better researchers in my field (e.g. skills related to programming languages, data analysis, data visualization techniques, etc.)

   Strongly Disagree 1 2 3 4 5 Strongly Agree (circle)

5. The role of computing in my field will increase with time (e.g. skills related to programming languages, data analysis, data visualization techniques, etc. will be required moreso in the future)

   Strongly Disagree 1 2 3 4 5 Strongly Agree (circle)

6. What are specific computational skills you seek out in students (undergraduate/graduate) interested in pursuing research in your lab?

   Long answer text

7. How do you anticipate your field of research will change with time? (In terms of not only skillsets required, but the types of research questions posed in the future, and what may be needed to answer them).

   Long answer text

8. Additional comments

   Long answer text