Application for a Tenure-Track Instructor Position in Statistics at UBC

Vincenzo Coia

2019-12-13

Contents

1 Cover Letter

Dear members of the search committee,

I am writing to enthusiastically apply for a tenure-track instructor position in Statistics (Job #35876). Ever since joining the department of Statistics full time in early 2017, largely to advance the Master of Data Science (MDS) program at UBC, I've come to realize just how ideal of a fit the educational leadership stream is for me. My aim with this application is to indicate why this is so, as well as why the department and UBC (and external to UBC!) would benefit by promoting me to the educational leadership (EL) stream. An electronic version of this application can be found at coia-ubc-application.netlify.com.

Next to this cover letter, you'll first encounter a detailed Curriculum Vitae, where I list my career accomplishments such as the courses I've been involved with and my industry experience. This background has provided me with the necessary data science skills for the EL stream at UBC Statistics: ample use of advanced R and modern tools such as the tidyverse and tidymodels, including a working knowledge of python; other computational tools such as bash scripting, continuous integration, web hosting, and authoring tools such as various markdown flavours and LaTeX; data visualization skills through `ggplot2`, R `shiny`, `plotly`, and others; data formats such as csv, json, and xml; version control through git and collaboration through GitHub; and an advanced knowledge of Statistics, especially when it comes to the practicalities and the "why" of Statistics. I also have the leadership and teaching skills required to keep Data Science education thriving at UBC, as well as the beginnings of contributions (and promises for more) that I believe satisfy the definition of EL.

I hope these skills speak for themselves throughout this application, in the three statements that follow the CV:

- The "Statement of vision for education in statistics and data science" describes my take on Statistics for Data Science. There, I describe the importance and meaning of taking a "problem-first" approach, instead of the "framework-first" approach that should be reserved for Statistical Science. You'll also find my plans for EL, such as a book and software I'm writing.
- The "Statement of teaching and training philosophy" describes my take on teaching in general, along with how I go about leading groups. There, you can find a discussion of things such as learning objectives being a "true north" when developing a course, and allowing for some spontaneity while teaching, as well as how I lead a group using clarity and equality.
- The "Diversity Statement" explains my sensitivity to discrimination, and how simple things such as having pride in who I am, along with calling out discrimination, can promote an inclusive and diverse culture. You just might also find stories about discrimination that I myself have experienced.

I'm applying for the EL position because I want to take my current position one step further, where I can have a bigger impact on Data Science in general. Also, the allure of developing a new Minor program in Data Science is a new and exciting challenge that I can make valuable contributions to, because I know this program is going to be world-class.

I do admit, however, that I would be torn to leave MDS, because I'm invested in the many projects that are ongoing. But, I imagine maintaining a connection between MDS and the Minor program (as well as STAT 545A/547M, which I've been teaching since 2017), whereby all entities grow together through collaboration and discussions, ensuring that we support each other and provide a service to the community in our own unique ways. I will have been involved in all three data science entities at UBC, so am energized to lead collaboration efforts.

I'd like to reassure you that, although I've only ever officially taught *graduate* level courses as opposed to undergraduate, I expect the transition to be smooth. Courses in the MDS program are structured like undergraduate courses, and it's not unusual for content to also be at the undergraduate level. Also, I've been involved in the teaching teams of many undergraduate courses as a teaching assistant (TA).

Here are a few links to examples of teaching material that I've created. Note that I don't see any of these as being "finished", but rather a work-in-progress.

- For STAT 545A/547M (Exploratory Data Analysis), I made the lecture notes and course website publicly available. Less relevant are Lectures 11+, taught by Dr. Firas Moosvi, and Lecture 6, a guest lecture taught by my TA, Victor Yuan.
- For DSCI 551 (Probability for Data Science), I wrote most of the lecture notes by using Dr. Mike Gelbart's notes from the previous year as scaffolding.
- You can find the DSCI 531 notes here from the 2018/19 academic year, reflecting an older way I would structure course material.
- You can find the DSCI 562 notes here from last year. This still needs lots of help, but the foundations are there.
- I'm writing an open-source book on regression analysis for Data Science, called Interpreting Regression. It's in its early stages.

You may also find it useful to obtain a list of MDS courses that I would feel comfortable (or uncomfortable) teaching. You can find that list in the Course Involvement section of my CV.

I'd be remiss to forego mentioning just why this position fits my career goals. I've identified four pillars of an ideal career. It took me a while to figure out that the EL position nicely satisfies all four, which is why I didn't apply for the position back in 2017. These pillars are, in order of culmination:

1. **Application**: addressing real problems with data science. I will continue to actively seek to "sharpen the saw" by putting myself in front of real problems, as I mention in the Educational Leadership of my vision for education in Statistics.
2. **Research**: discovering better ways to do data science. To me, this means reframing Statistics from a problem-first approach, and bringing useful yet under-utilized tools to the table. You can read about this in the Statistics for Data Science section in my vision for education in Statistics.
3. **Development**: turning new research into programs and tools, which are ideally accessible to the public. To me, this means developing products under the umbrella of EL, such as books and software. You can read

about this in the Educational Leadership section in my vision for education in Statistics.

4. **Teaching**: spreading data literacy and enthusiasm to the public. Teaching university courses is a dream come true for me, and so is being a leader by training in a group setting. You can read about this in my Statement of teaching and training philosophy.

Here are the names of three references who will submit letters by the end of December 13, 2019:

1. Dr. Tiffany Timbers, UBC (current supervisor)
2. Dr. Michael Gelbart, UBC (current supervisor)
3. Dr. Harry Joe, UBC (PhD supervisor)

I am eager to begin the position on the indicated July 1, 2020 start date, but am ultimately flexible. I'm looking forward to talking more about my fit in this EL position in person.

Sincerely yours,

Vincenzo Coia, PhD

https://vincenzocoia.com/

2 Curriculum Vitae

2.1 Work Experience

2017/01 - present
Lecturer of Data Science
(Initially: Postdoctoral Teaching and Learning Fellow)
Masters of Data Science Program, Department of Statistics
The University of British Columbia
Vancouver, BC

2019/05 - 2019/08
Part-time Flood Forecasting Consultant
BGC Engineering
Vancouver, BC

2009/05 - 2014/05
Short-term statistical consulting (6 projects)
UBC and private

2.2 Education

PhD in Statistics

2012/09 - 2017/02

The University of British Columbia

Conferred May 29, 2017

MSc in Mathematics and Statistics (Statistics)

2011/09 - 2012/08

Brock University

Conferred on October 13, 2012

BSc (3-year) in Biological Sciences

Minor in Earth Sciences

2005/09 - 2011/04

Brock University

Conferred "With Distinction" on October 22, 2011

BSc (Honours) Mathematics Integrated with Computers and Applications

Concentration in Statistics

2005/09 - 2011/04

Brock University

Conferred "With First-Class Standing" on June 7, 2011

2.3 Teaching

2.3.1 Course Involvement

The following is a table of courses that I've been either lecture or lab instructor for. They all took place at UBC.

I also list a rating as to how comfortable I am to teach the lectures of the course (perhaps a little bit of enthusiasm is mixed into the scale, too). The ratings are:

5: Very comfortable

4: Comfortable

3: Neutral

2: Uncomfortable

1: Very uncomfortable

Course	Comfort Level	Title	Involvement
BAIT 509	4	Business Applications of Machine Learning	Instructor, 2 years
STAT 545A	5	Exploratory Data Analysis, Part I	Instructor, 3 years
STAT 547M	5	Exploratory Data Analysis, Part II	Instructor, 2 years
DSCI 511	5	Programming for Data Science	Lecture and Lab instructor, 2 years
DSCI 531	4	Data Visualization I	Lecture instructor, 2 years; lab instructor, 1 year
DSCI 562	5	Regression II	Lecture Instructor, 1 year; Lab Instructor, 3 years
DSCI 551	5	Probability for Data Science	Lecture Instructor, 1 year
DSCI 561	5	Regression I	Lab Instructor, 1 year
DSCI 571	4	Supervised Learning I	Lab Instructor, 1 year
DSCI 532	2	Data Visualization II	Lab Instructor, 2 years
DSCI 573	4	Feature and Model Selection	Lab Instructor, 1 year
DSCI 563	2	Unsupervised Learning	Lab Instructor, 1 year
DSCI 553	4	Statistical Inference and Computation II (Bayesian Statistics)	Lab Instructor, 1 year

Course	Comfort Level	Title	Involvement
DSCI 574	5	Spatial and Temporal Models	Lab Instructor, 2 years
DSCI 554	4	Experimentation and Causal Inference	Lab Instructor, 2 years
DSCI 591	5	MDS Capstone Project	Supervised 8 student teams over 3 years

Courses that I have not been part of the teaching team for, but have a comfort level of at least 4:

Course	Comfort Level	Title
DSCI 521	5	Computing Platforms for Data Science
DSCI 542	3-4	Communication and Argumentation
DSCI 523	5 (R); 2 (python)	Data Wrangling
DSCI 552	5	Statistical Inference and Computation I

Courses that have a comfort level of 1 that I have not been a part of:

Course	Comfort Level	Title
DSCI 513	1	Databases and Data Retrieval
DSCI 572	1	Supervised Learning II
DSCI 541	1	Privacy, Ethics, and Security
DSCI 525	1	Web and Cloud Computing
DSCI 575	1	Advanced Machine Learning

2.3.2 Course (Re-) Developments

I've been involved with the redevelopment of the following courses:

- DSCI 561: Here, I lead a team and supervised one graduate student to redevelop the labs.
- DSCI 562: Both lectures and labs
- DSCI 531: The lectures one year, then the labs the following year.
- DSCI 532: Here, I lead a team and supervised two TA's to redevelop the labs.

- DSCI 574: The labs.

I've been involved with the development of the following courses:

- DSCI 591: I helped develop the capstone course with my colleagues.
- BAIT 509: I used DSCI 571 as inspiration for developing this course.

Some courses have been partially redeveloped.

- STAT 545A/547M: Some redevelopment of content; much redevelopment on course website and infrastructure.
- DSCI 551: Some redevelopment of lectures and lab to accomodate a problem-first approach.
- DSCI 554: The labs, after some shuffling of content to and from other parts of MDS.

2.3.3 Teaching Assistantships

Duration: From the latter part of my undergrad, to the end of my PhD.

UBC:

- **SCIE 300: Communicating Science** (5x)

Brock University:

- **MATH 4P82/5P82: Non-parametric Statistics**
- **MATH 3P82: Regression Analysis**
- **MATH 4P81/5P81: Sampling Theory**
- **MATH 3P81: Experimental Design** (2x)
- **MATH 2F40: Mathematics Integrated w/ Computers and Applications II**

2.4 Volunteer Positions

2016/09 - 2016/02
Science World at TELUS World of Science
Vancouver, BC
78.15 hours

2013/10 - 2014/05
Beaty Biodiversity Museum: Events Volunteer
Vancouver, BC
35.0 hours

2013/04 - 2013/09

UBC Farm

Vancouver, BC

102.5 hours

2011/06 - 2011/08

Project S.H.A.R.E. community garden

Niagara Falls, ON

15.0 hours

2.5   Research Assistantships

2013/05 - 2013/08

Robust penalized regression

Supervisor: Dr. Gabriela Cohen-Freue

Department of Statistics

The University of British Columbia

Vancouver, BC

2012/05 - 2012/08

2011/05 - 2011/08

2010/05 - 2010/08

Extreme value modelling

Supervisor: Dr. Mei Ling Huang

Department of Mathematics

Brock University

St. Catharines, ON

2010/09 - 2011/06

Quantum monte carlo simulations

Supervisors: Dr. Stuart Rothstein; Dr. Wai Kong (John) Yuen

Department of Chemistry and Department of Mathematics

Brock University

St. Catharines, ON

2.6 Publications and Talks

2.6.1 Articles in Preparation

- Coia, V., Joe, H., and Nolde, N. (2020?) Extreme Quantile Regression and the Copula-Marginal Composition Method. Status: Wrapping up the demonstration analysis

2.6.2 Articles Submitted to Refereed Journals

- Huang, M.L., Coia, V., and Brill, P.H. (2013) A cluster truncated Pareto distribution and its applications. ISRN Probability and Statistics 2013: Article ID 265373.

- Ayad, M., Coia, V., and Kihel, O. (2014) The number of relatively prime subsets of a finite union of sets of consecutive integers. Journal of Integer Sequences 17: Article 14.3.7

- Coia, V., and Huang, M.L. (2014) A sieve model for extreme values. Journal of Statistical Computation and Simulation. 84(8):16921710.

2.6.3 Articles Submitted to Conference Proceedings

- Huang, M. L., Coia, V., and Brill, P.H., A mixture truncated Pareto distribution, In JSM Proceedings 2012, Statistical Computing Section, Alexandria, VA: American Statistical Association, pp. 24882498.

2.6.4 Conference and Roundtable Contributions

- Coia, V. "The squared error has friends, too!" (Contributed Talk). SFU/UBC Joint Graduate Student Seminar (Winter). Spring 2019 at the SFU Harbour Centre, Vancouver, BC.

- Coia, V., Nolde, N., and Joe, H. Forecasting Extremes for Flooding (Invited Talk). The 44th Annual Meeting of the Statistical Society of Canada. May 29June 1, 2016 at Brock University, St. Catharines, ON.

- Coia, V., and Jeanniard du Dot, T. (Invited Demonstration) "Using the Grammar of Graphics and Interactivity to explore Biologging Data in R". May 6, 2015. Building a Bioanalytical Theory for Analysis of Marine Mammal Movements: A Peter Wall International Research Roundtable. The University of British Columbia, Vancouver, BC.

- Coia, V. "Flood Warning: An Application of High-Quantile Regression" (Contributed Talk). SFU/UBC Joint Graduate Student Seminar (Winter). February 28, 2015 at the SFU Harbour Centre, Vancouver, BC.

- Coia, V. "A New Sieve Model for Extreme Values" (Contributed Talk). SFU/UBC Joint Graduate Student Seminar (Fall). September 29, 2012 at the SFU Harbour Centre, Vancouver, BC.

- Coia, V., and Huang, M.L. "On Estimation of Heavy Tailed Distributions" (Contributed Talk). The 40th Annual Meeting of the Statistical Society of Canada. June 36, 2012 at the University of Guelph, Guelph, ON.

- Huang, M.L., Coia, V., and Brill, P.H. "A Mixture Truncated Pareto Distribution" (Contributed Talk). The 2012 Joint Statistical Meetings. July 28August 2, 2012 at San Diego, California

2.7 Awards

2.7.1 University Issued

- 2012/09 - 2016/08: Four-Year Fellowship
- 2012/09 - 2016/08: Faculty of Science Graduate Award
- 2011/09: Dean of Graduate Studies Excellence Scholarship
- 2011/06/07: Dean's Gold Medal
- 2011/06/07: Distinguished Undergraduate Student Award in Mathematics
- 2011/03: President's Surgite Award

2.7.2 Nationally Recognized

- 2013/06: Governor General of Canada's Gold Medal
- 2012/09 - 2015/08: NSERC Postgraduate Award (Doctoral, 3-year)
- 2011/09 - 2012/08: NSERC Alexander Graham Bell Canada Graduate Scholarship (Masters)
- 2010/05 - 2010/08: NSERC Undergraduate Student Research Award

2.7.3 Individual Donors

- 2012/04: Dr. Jack Lightstone & Dorothy Markiewicz Scholarship
- 2012/04: Dr. Raymond & Mrs. Sachi Moriyama Grad. Fellowship
- 2012/03: Tomlinson Entrance Scholarship for Excellence in Mathematics and Science
- 2011/06: John and Roslyn Reed Book Prize
- 2010/09: Art Bicknell Scholarship in Mathematics
- 2010/09: Ian D. Beddis Family Scholarship
- 2010/09: Terry and Sue White Mathematics and Science Scholarship
- 2007/09: M.J. ("Mel") Farquharson Scholarship
- 2006/09: Scholler Foundation Scholarship in Chemistry

2.7.4 Athletic Awards

- 2016/04/03: Sportsmanship Award, UBC Thunderbirds Sport Clubs (Fencing)
- 2016/02/28: Bronze Medal in Senior Mixed Epee (Intercollegiate Tournament, UBC)
- 2015/11/16: Bronze Medal in Senior Mixed Epee (Remembrance Day Tournament, UBC)
- 2012/03/28: RM Davis Surgite Award
- 2009/12/12: First Place Award Winter Epeedemic, Toronto Fencing Club
- 2009/03: Varsity Fencing Rookie of the Year Award

2.7.5 Declined Awards

- 2012/04: Ontario Graduate Scholarship (Doctoral)
- 2011/05: Ontario Graduate Scholarship (Masters)

2.8 Professional Activities

Dept. of Statistics, The University of British Columbia:

- 2016/05 - 2016/06: **Search committee member (for CRC 2 faculty position)**
- 2015/04/30: **Organizer of the "How to Write an Awesome Abstract" workshop**
- 2014/06 - 2015/05: **Organizer of the SFU/UBC Joint Seminar**
- 2014/06 - 2014/09: **Organizer of the Graduate Student Trip**
- 2014/04 - 2015/05: **Co-founder of the Graduate Writing Forums**
- 2013/06 - 2014/05: **Statistics Graduate Student Representative**

3 Statement of vision for education in statistics and data science

A major passion of mine is reframing Statistical Science from the lens of Data Science, by taking a "problem-first" approach. You can read about my vision in the first section, which is followed by a section describing how I go about executing this vision as educational leadership initiatives.

3.1 Statistics for Data Science

Why does Statistics need reframing for Data Science? Statistical Science tends to begin with assumptions / a modelling framework, then describing properties and methods that enable solutions to a problem. This is a "framework-first" approach. But Statistics for Data Science requires a "problem-first" approach, focussing on how to appropriately construct a modelling framework from real-world problems. These two approaches are not translatable,

because real problems don't neatly satisfy assumptions nor come with a description of what output is appropriate for addressing the problem. A real problem requires critical evaluation as to what outputs and what assumptions are useful and realistic. This isn't assumption-*checking*, it's assumption-*building*.

By the way, even Applied Statistics might sometimes seem to take a problem-first approach, because it starts with a motivating problem. But the focus is still on describing properties of the newly defined framework, falling squarely within the field of Statistical Science. On the other hand, Statistics for Data Science must focus on how to deal with a wide range of problems, not just one problem type. This is a challenge that relies on the properties already described by Statistical Science, hence why both approaches are important!

The way I see it, teaching a problem-first approach for Statistics for Data Science still has a long way to go. A central and powerful approach that I've been adopting is to create content focussing on the *type of problem* encountered in real-world problems, instead of the traditional Statistics topics. Perhaps the best example of this is DSCI 562 (Regression II), which I re-developed last year. Instead of teaching "Survival Analysis" and "Multiple Imputation", I teach "regression when data are censored" and "regression when data are missing". This changes the focus of the lessons to be about evaluating approaches for solving the type of problem, including evaluating the impact of ignoring the problem altogether (perhaps there's little to no impact!). This approach allows students to creatively come up with solutions to a variety of real world problems, which are sure to involve "combined" scenarios, such as a censored response with some missing data.

Sometimes, a traditional statistical topic is made up of more than one type of problem, resulting in a restricted way of thinking. For example, the topic of Generalized Linear Models (GLM's) can be split into two problems: regression when the range of the response is restricted, and using a distributional assumption to improve estimation performance. Both are valuable tools that expand beyond the scope of GLM's, but if just taught as GLM's, restricts students' thinking to the smaller range of problems that GLM's can handle. Determining the type of problem that a traditional statistical topic addresses is not always obvious. I found the GLM example to be tricky; here are some other tricky ones, from DSCI 551 (Probability for Data Science):

- "Probability Distributions" has the underlying problem of communicating an uncertain outcome.
- "Simulation" has the underlying problem of calculating probabilistic quantities (mean, probabilities, etc.) when the distributions or calculations are hard to come by.
- "Independence" is tricky because it's really "dependence" that is more interesting. Dependence, then, can have an underlying problem of garnering information from a dependent variable (or what I like to call "harvesting dependence").

This problem-first approach naturally promotes motivation in the classroom, because the topics are designed to be practical from the get-go. Sample real-world problems become important, and where possible, I draw from my capstone and industry experience to create compelling examples. Perhaps my favourite is the "ships making demands at the port" example in DSCI 551, which I encountered in one of last year's capstone projects, to demonstrate simulation under a more complicated situation. Another favourite is flood forecasting; another, determining the
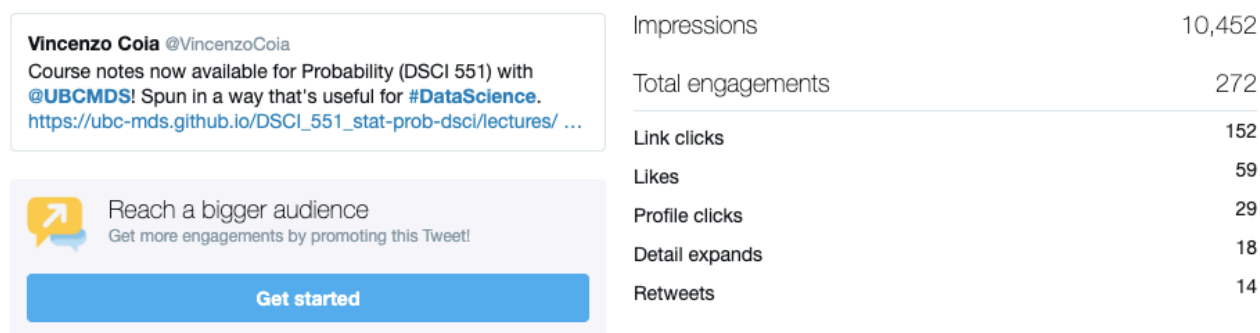
Figure 1: DSCI 551 tweet activity

effect of blue-light-blocking glasses on sleep.

An interesting thing that happens when taking a problem-first approach is that it brings to light areas of Statistics that are useful, yet are not receiving much attention. Two examples are quantile regression or distributional forecasting instead of mean regression. Another example is extreme value modelling (EVM), which becomes invaluable for problems that have heavy tailed data, such as at least one capstone project from Seahorse Strategies last year (the team should have used EVM, but did not, because we didn't teach it in MDS), as well as organizations tackling the issue of flood forecasting and flood frequency assessments. Interestingly, these topics tend to be deemed difficult, probably because of our community's obsession with the mean, as well as a general misunderstanding of the role of distributions outside of just being a "necessary statistical assumption". I believe these topics are more accessible than they're made out to be, if taught from a problem-first lens, because the mathematical details are inconsequential.

3.2 Educational Leadership

I'm finding that there's a general lack of problem-first education in Statistics, as well as a lack of good examples, but this presents opportunity. I envision being a leader in building Statistics for Data Science – it's actually a major component of my life goals. This section describes my strategy plan for having an impact, including some of the things that I've accomplished already.

Perhaps the simplest approach I've taken is making my course material public, which I intend to continue doing. People are starting to notice. See the figure for activity from my most recent tweet of my most recent release, for DSCI 551. Also, the STAT 545A/547M content is still receiving a lot of global attention, largely thanks to Dr. Jenny Bryan's original efforts with the course (although the content needs more upgrading than I've had capacity to supply).

I like to think beyond the structure of a classroom, by developing material intended to be used by a wider range of people than those who refer to course notes. So far, this means writing a book focussing on regression analysis from

a problem-first perspective. It's far from done, but I'm also well into its development. I'm calling it "Interpreting Regression". For maximum return on investment, I try to write course notes that can be easily adapted to the wider audience of the book. I believe distributional forecasting deserves its own book as well, and would include approaches like extreme value modelling, and techniques like copula modelling. These topics are more practical and accessible than people think.

I also see the need for new R packages for making Statistics more accessible for Data Science. I made some S3 Object Oriented R packages during my PhD with many useful functions, but being my first R packages, they have many deficiencies, including a lack of vignettes, too many overly complicated functions, and a lack of tests. I have ideas for packages to replace these older ones, as well as altogether new ones. My goal for each package is to submit each to CRAN and ROpenSci, publishing a corresponding paper to the Journal of Statistical Software, as well as making helpful vignettes. Here are my ideas:

- The `distplyr` package (which is already underway) will make it easier to work with univariate distributions, including empirical ones, for the promotion of distribution-based analyses. This makes a distribution an S3 object, which can then be manipulated. This would, for example, aid in the teaching of mixture distributions, and aid in the creation of extreme value models using functions like `distplyr::right_graft()` or `distplyr::left_graft()`.
- The `coperate` package is similar to `distplyr` in that it makes it easier to work with copulas by making a copula an S3 object. This will replace the `copsupp` package, one of my deficient R packages from my PhD days, yet still has lots of useful functionality.
- Modelling packages to go alongside the above two.

In short, I'm aiming to lead the development of a suite of tools that have a "wow" quality, so that they become go-to resources for people needing help for Statistics for Data Science to satisfy their unique needs. Aside from books, public course content, and R packages that I've mentioned above, I'm curious as to what other medium would appeal to other aspects of learning. R shiny web apps is one such medium that is on my radar. For example, the impact of multicollinearity is much easier to grasp by visualizing a regression plane slicing a data cloud, allowing the user to change the correlation between the covariates. Other media are only in the space of my imagination, but that's where things begin! Maybe one day I will in fact start an engaging podcast with my colleagues, or curate a monthly "Statistics for Data Science Box" containing books or other resources for busy professionals wanting to grow and learn. The possibilities are endless.

A completely different area where I would like to provide leadership is in teaching educators how to go about teaching Statistics for Data Science, because the field is so new. I of course don't have all the answers, so perhaps the best medium for this is to host workshops on Data Science education. In fact, I've already helped kick off a vision for an international workshop by spearheading the creation of a vision for MDS. Related, I believe it's important to be stewards of Data Science by promoting responsible use, which is reflected in our material, but also in this workshop / conference.

Because a problem-first approach is so critical for Statistics for Data Science, I believe a final area of leadership is to exercise my data science skills, or as Stephen Covey says in his book "The 7 Habits of Highly Effective People", to "sharpen the saw". Just as it's important for us to take time to stay healthy, it's also important as an educator to keep my Data Science skills sharpened, as well as up to pace with a changing field. Whether or not this is considered EL, it's important so that I can give students the most relevant and best education I can possibly give. Simple ways of remaining sharp, or at least up to date, is to subscribe to relevant newsletters such as RStudio's and ROpenSci's, and stay involved with the Data Science community on Twitter. A bigger way to remain sharp is to actually do a bit of work for external organizations to help them with their data science needs. I did this last summer with BGC Engineering by dedicating a small amount of time each week, and this opened my eyes to their values and their data science problems, which I can then bring back to the classroom.

4 Statement of teaching and training philosophy

4.1 Teaching Philosophy

My philosophy on teaching can best be broken down into four categories: **appropriate course structure**, **responsive teaching**, **motivation**, and **self-improvement**.

4.1.1 Appropriate Structure

You and I don't really know English. We only know enough for it to be a sufficiently useful tool for our own daily activities. I take this concept to heart as I design and deliver my courses. It's fallacious to design a course around *topics*, but rather on building *skills* for a certain purpose. To me, these skills formulated as a list of learning objectives (LO's) act as a "true north" that all components of a course point towards, and are the most important structure to the course. For example, course content on the topic "heteroskedasticity" could be developed in many directions. The students, thinking they have to "know heteroskedasticity", get stuck studying its never-ending scope. Instead, a learning objective such as "modify a linear model to appropriately incorporate heteroskedasticity based on data diagnostics" provides direction for the teaching team to create material, and clarity for the students. For concrete examples, take a look at the DSCI 551 notes, or my other course materials.

Clarity about the course structure and expectations is important for equipping students with a map of what they need to do to succeed. In addition to discussing how students should use the LO's, describing the summative assessments and a course syllabus is also important, as well as making it easy to navigate course material (see, for example, the STAT 545A/547M syllabus website – although the "Learning Outcomes" listed need updating from pre-2017). Informing the students of any changes or of any mistakes, no matter how painful, is also critical for maintaining clarity.

4.1.2 Responsive teaching

Sticking too close to course structure can stymie a course. This is because, while preparing course material, the instructor is not truly equipped with the clarity of how to most appropriately deliver material until the material has been delivered. For example, a question for the class might have seemed insightful when it was being prepared, but maybe during class the question doesn't end up being insightful. Or, maybe the lecture delivery time was underestimated. The answer to addressing this issue is responsive teaching.

Responsive teaching to me is about staying in touch with the class to get a feel for areas that need more or less attention. For example, it means: taking a break when the energy of the class is low; responding to class confusion by explaining something in a different way on the whiteboard; or spending more time on something after realizing its importance mid-class. In this way, teaching is more akin to improv comedy, where actors must respond to random cues from the audience (I've taken lessons). I believe in embracing a mindset of letting go and trusting in yourself to respond appropriately the moment things go off course, and to also be humble and admit when you don't know something.

This type of spontaneity sometimes also requires realizing and appropriately acting on your agency in adapting a course on the fly. Understanding that a course is flexible allows you to be nimble during contact hours with students, as opposed to feeling stuck in delivering the course exactly as it was originally laid out. However, it's important to be strategic when making changes so that things remain orderly and as seemless as possible for the students, and to be clear about any changes made. I like using the start of class to check in and return to a roadmap of the course so that everyone remains on board a moving bus.

Contact time with students is critical for adapting teaching to the students, and provides tremendous value for students to actually enrol in a program as opposed to taking an online course. Perhaps the most effective method for doing this is making a point to engage with each student in lab. Equally as powerful are discussion-style office hours. I no longer hold my office hours in my office, because the office hour model whereby students just drop by to ask a question is just not effective. There ends up being a queue of students, usually asking common questions, and these students feel pressured to leave the office so that others can get a turn. Instead, I hold my office hours in a lab-like room suitable for collaboration. I end up leading a whole-group discussion prompted by student questions. This gives me even more insight as to how things are going in class, and allows me the opportunity to modify the course moving forward or make clarifications. Plus, the students express appreciation for this approach.

Some other methods for connecting with the students are:

- being present on the course Slack channel (and sometimes even other course channels) to provide more insight,
- sending out a 1-minute long early survey about how the course is going,
- taking the time to talk to students who approach me during the mid-class break or after class,
- pausing to ask questions and check for insight in the classroom,
- applying active learning strategies such as think-pair-share or live coding, and

- checking in to see how things are going at the start of each class, by asking questions like "how are we feeling about the quiz coming up next week", or "how are we feeling this week".

4.1.3 Motivation

Neither the presence of structure nor responsive teaching will suffice if the students aren't excited or motivated about the course material, because the ultimately, learning happens outside the classroom, as the student engages with the material.

Taking a problem-first approach, as I mention in the Statistics for Data Science section, helps with this. But it also requires some level of entertainment and enthusiasm from the instructor to keep students engaged. It means being aware of the lengthy (usually 80-minute) time frame that students are present for (and that's just for one class), by taking a break mid-class, returning to the big-picture, and providing additional instructions for exercises for people who may have "fallen off the bus". It means telling stories. I'm pleased that I regularly receive ample praise on my instructor evaluations, and I encourage you to take a look.

4.1.4 Self-Improvement

What's better than effective teaching? Teaching that continually becomes more effective.

Community engagement is one method to become a better teacher. This means keeping an open dialogue and sharing experiences with other teachers, especially my colleagues. Aside from simple acts like sharing ideas and experiences through Slack and gatherings, I'm proud that my team gives and receives formative feedback on our teaching by visiting each others' lectures. I also take the opportunity to partake in workshops, such as UBC's Instructional Skills Workshop, and MDS's annual workshop on Data Science education.

An after-action review is another effective method, involving capturing the insight you gain after teaching. There are three ways I engage in an after-action review. First, I capture regular insight throughout the delivery of the course as GitHub Issues, so that the insight can easily be referred to in the future and by any of my colleagues. Secondly, I find keeping a teaching journal that's not tied to a specific course is useful for becoming a better teacher in general – and it's even easier now that my team has our lectures recorded. Thirdly, my colleagues and I engage in a "retreat" at the end of each term, to discuss our insight on our courses and the MDS program as a whole.

Realizing one's own struggles is also critical for self-improvement. For now, one thing I struggle with are student names. I feel embarrassed to use names because I'm sure to make mistakes, but calling students by their name is important because it shows that the instructor cares. I already review class lists, but have been stepping outside of my comfort zone by no longer avoiding using names. Another thing I struggle with is writing in an orderly fashion on the whiteboard. I use the whiteboard for spontaneity / responsive teaching, which means I don't know how I'll ultimately be using the board throughout the class. I end up running out of space, and end up using the board

in a non-linear fashion, which can sometimes be confusing. For now, I intend to remain mindful of this and keep practicing.

4.2   Training Philosophy

When it comes to training, I aim to foster a clear and highly communicative environment between me and my trainees.

Examples of group leadership include the development and delivery of courses with colleagues and teaching assistants, and the delivery of capstone projects together with partner organizations and student groups with MDS. I believe effective group leadership involves establishing everyone's roles, expectations, and time commitments; creating an empowering work culture for all group members; and defining success. When training is involved, delegation and feedback are key to promote the subordinate's growth.

Clarifying each team member's role, expectations, and level of authority for making decisions is the beginning step in defining a framework for which the group can operate effectively. Doing so allows everyone to respect each other's time and boundaries, so that group members know how to interact with each other. The MDS Capstone projects are one example, where I feel it is tremendously important to immediately clarify the roles of the student group, the partner organization, and myself, so that problems don't arise such as the partner organization treating the students like employees. Another example is having established roles at the beginning our DSCI 561 redevelopment kickoff meeting, involving a graduate assistant to be the main one putting "pen to paper", myself as her mentor, and colleagues for checking in with the high-level direction of the redevelopment.

When it comes to the actual nature of the interactions between group members and myself, I believe it's important to empower all group members to feel some ownership over the project that they can make a valuable contribution. The reality is that everyone has a unique set of skills that, when tapped, result in a better end product than I could have produced on my own. Doing this means to be deliberate about my interactions – it means bringing positive and friendly energy, being aware of my facial expression, and considering all input and ideas from others as if they are thoughtful.

When it comes to the actual teaching and training of my trainees, it's important to provide regular input both verbally and as written feedback. This involves taking the time to write detailed notes as to where they can improve or where they are exceeding expectations. For example, my capstone groups know that I will give them the courtesy of looking into the details of their work and give them detailed feedback. This is more time consuming than just skimming over their work and providing a few comments, but this is what promotes learning.

Another important aspect of training is to delegate, especially when it comes to those tasks that we as leaders feel more skilled at. This is difficult, because we feel ownership over our tasks. We feel that we are in the best position to do the job, and we gain satisfaction when they are done. But we must let go, and "gift" these tasks to our trainees who are less skilled at it, because the challenge it presents them is a powerful learning opportunity. For

example, I like to give my teaching assistants the opportunity to teach up to one lecture for most courses that I teach.

Finally, as in teaching, I always looking for ways to improve my leadership in training. This involves asking the uncomfortable question of how I'm doing as a leader. For me, as I learned from last year's capstone project, I need to be more mindful of group dynamics. Even if things between me and each team member or student is fine, the relationship between other team members might not be. This year, I'll be meeting with each student individually early on to see how things are going.

5   Diversity Statement

People deserve the same respect regardless of their identity. Any differential treatment is discrimination, and is problematic because it leaves people feeling disrespected, or worse, puts an impediment on one's life. As an educational leader, awareness of issues is critical, so I begin by expressing some of my views. I myself and no stranger to discrimination, and I'd like to share some stories in the second section. Lastly, I end with my views of how I promote inclusivity.

5.1   View

Our society has made some great progress on inclusivity. Racial and sexual discrimination are significantly less these days compared to several decades ago. But the battle for these is still not over, as is evidenced by graffitied rainbow sidewalks and defaced mosques. Some discrimination is still rife, but are actively gaining awareness. For example, women are still underrepresented and dismissed in the workforce. Indigenous communities are still prone to lower quality of life.

Yet other types of discrimination are still not even well known. This includes gender discrimination, whereby we are all judged more than we know based on our gender. There is even less awareness on the issue of discrimination against fathers.

Not only is it healthy for each individual to receive respectful and equal treatment, we have so much to gain by having a diverse group of people around us. There's strength in diversity.

5.1.1   What's the discrimination against dads?

It's likely you haven't heard of this one, so I think it's worth elaborating. There is an overarching issue of dads not being regarded as equally important figures as moms. Just about anything to do with parenthood refers to mom – for example:

- I can't find a book about parenthood that does not emphasize mom's role;

- Parenthood groups emphasize women over men, such as Facebook groups and community centers.
- Products are marketed largely for mom. Companies like 4Moms is akin to a business supply brand being called "4Businessmen".

The discrimination is even built into BC policies regarding birth registration. Three examples are:

1. A dad is not allowed to register his child's birth on his own, but a mom can.
2. A dad is not allowed to apply for MSP and Canada Child Benefits, but a mom can.
3. A dad is not allowed to be listed as primary caregiver without getting permission from the mom (if present), but a mom can list herself without permission of a present dad.

This tells dads that they are only of secondary importance when it comes to child care, promoting a culture whereby dads don't connect as intimately to their children. Worse, it also harms women's ability to establish an equal presence in the workforce as men, due to the expectation that men need not expend as much energy on their child as women.

Even the notion of "primary caregiver" is discriminatory. Asking to identify such an individual is equivalent to asking "which parent is going to give less care to their child?", which is of course very presumptuous, but also promotes a culture of unequal parenting, putting more pressure on one individual.

All of this is compounded in the situation where there's no mom (i.e., two dads), or the situation where individuals don't squarely identify as either male or female.

5.2 My experience with discrimination

As a member of the LGBTQ+ community, I continue to experience discrimination. Me and my male partner holding hands in public still forms a spectacle to many, some even stopping to watch us after we've walked by. Wedding vendors still referring to a "bride" when mentioning our wedding. We've been invited to a wedding under the condition that we (specifically) show no affection ("for the sake of the children").

Things were worse in my adolescence, where homosexuality was "discouraged" in my environment, leaving me feeling socially out of place and fearful. Luckily, very few extended family members have a problem with my identity, and the rest embrace me.

Although I'm a cis male, I'm passionate about gender issues. I once used the women's washroom at the Whistler convention center, as an experiment and to test my confidence. The result were shrieks and gasps, gossip even outside of the washroom, and a closure of the women's washroom until I left.

As you may know, my partner and I recently had a child through surrogacy. But, on the delivery day, the hospital was not prepared to handle a situation with two dads. It's important for the parents to form a bond with their child, and to help the surrogate avoid attachment issues with the child – but the (well-intentioned) hospital staff overlooked both of these aspects. We (the parents) were supposed to be brought in as soon as the baby was born,

for initial skin-to-skin contact, but the hospital staff didn't bring us in until an hour after birth. Furthermore, they gave our surrogate first skin-to-skin contact – this promotes oxytocin in the surrogate, making it more difficult to part with the child, and denies the parents the opportunity for initial bonding. Worse, hospital staff kept referring to our surrogate as "mom", and referred to our son by her last name. Not only is this disrespectful for the parents, but it makes separation even more difficult for the surrogate.

5.3 Contribution

To me, contributing to inclusivity is about creating an inclusive environment, especially when it comes to making course content; calling out discrimination, even when the discrimination is unintentional; and being a role model for others by being comfortable about who I am.

5.3.1 Creating an inclusive environment

There are a host of things we can do to create an inclusive environment.

We can change our language to be less presumptious. This means saying things such as "pregnant people" instead of "pregnant women", referring to one's "partner" instead of saying "wife" or "husband" (assuming a gender), not referring to someone by their race if not relevant (such as "I was talking to a Latino man the other day, …"), etc.

We can post online content that suggest inclusivity. For example, not using data in the classroom that indicates gender as binary (because it's naive and ultimately untrue, even if gender is paraded as "sex"). Another example is not indicating female:male ratios and parading naive terms such as "gender balanced", but rather indicating percentage belonging to a minority group (such as "percent self-identifying as female") to show that we are sensitive to marginalized groups.

We can remove spaces that discriminate by gender. This means a complete decoupling of bathrooms and gender. UBC needs to first take low-hanging fruit by abolishing gender from single-person bathrooms (like we have in the stats dept). Then, the trickier issue of multi-person bathrooms can be addressed by finding the root problems (which at first seem gender-related, but are actually not), such as an overall lack of privacy in bathrooms.

We can include a "covenant" or "code of conduct" in collaborative online spaces. This explicitly states that discrimination won't be tolerated.

We can pay attention to how we treat everyone, to avoid unconsious discimrination. This is important, but challenging. For example, I was once accused of spending less energy on non-native English speakers on Slack. Although it's hard for us to recognize unconscious discrimination, I'm not convinced their claim was true – but regardless, this experience taught me that I need to be more deliberate about how I respond to students. I like to think, "how would I respond if this question or comment was made by someone else?". This question is useful for interactions in the classroom, too.

5.3.2 Calling out discrimination

Calling out discrimination means pointing out when non-inclusive language or behaviour is used, whether intentionally or not. Critically, this should be done with compassion, as opposed to accusation, because people may not know any better, and even if they do know better, this language can accidentally slip due to many years of belonging to a less-inclusive environment. Calling out discrimination can help educate people of the first type, as well as help re-program people of the second type. Examples include:

- Someone telling someone else that they're in the "wrong bathroom" needs to be informed that they don't know the person's gender, nor does it matter.
- A gay bar that has a "men only" policy needs to be told why it's wrong.
- A well-intentioned person celebrating "gender balance" needs to be informed as to why the term is non-inclusive to non-binary individuals.
- A government that doesn't allow dads to register their child needs to be written some letters as to why this is discriminatory.

By the way, all of the above are situations I've encountered (and yes, I responded in each case).

5.3.3 Being a Role Model

I hope to be a role model by telling people that I identify as a gay cis-male (and no, the term "coming out" is not appropriate). Some examples of how I've done this include:

- Posting my preferred personal pronouns (he/him/his) and a rainbow emoji online.
- Posting a Positive Space sticker outside of my office.
- Contributing my LGBTSTEM interview.

Doing this helps show that I'm proud of who I am, and indicates sensitivity to issues of gender and sexual identity. I hope this is empowering for others, as well as creating a safe space for students and others to talk about issues they may have with their identity. I also hope that it brings awareness to those who are not familiar with LGBTQ+ issues – for example, the more someone sees preferred personal pronouns being specified, the more likely they are to look up why more and more people are posting this and what this means.