Initial Report

Last Modified: 01/29/2018

1. Name (first and last)

Text Response
Kenneth Flagg

Statistic	Value
Total Responses	1

2. Email

Text Response
kenneth.flagg@msu.montana.edu

Statistic	Value
Total Responses	1

3. Contact Phone

Text Response
9512889172

Statistic	Value
Total Responses	1

4. In Fall 2018 you will consider yourself to be a:

#	Answer	Bar	Response	%
1	VT Freshmen Undergraduate Student		0	0%
2	VT Sophomore Undergraduate Student		0	0%
3	VT Junior Undergraduate Student		0	0%
4	VT Senior Undergraduate Student		0	0%
5	Between Undergrad and Graduate school		0	0%
6	Graduate school		1	100%
	Total		1	

Statistic	Value
Min Value	6
Max Value	6
Mean	6.00
Variance	0.00
Standard Deviation	0.00
Total Responses	1

5. What degree(s) and major(s) are you pursuing along with institution?

Text Response
Ph.D. in Statistics at Montana State University

Statistic	Value
Total Responses	1

6. List any Minors you are pursuing.

Text Response

Statistic	Value
Total Responses	0

7. List any honors and/or awards received:

Text Response

Statistic	Value
Total Responses	0

8. Currently we have 1 program accepting applications which includes:

#	Answer		Total Responses
15	Data Science for the Public Good	0	0
	Total	0	-

Statistic	Data Science for the Public Good
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	-

9. Essay (up to 500 words): “What do you want to get out of this experience?”

Text Response

The DSPG fellowship offers a unique opportunity to connect data and statistical methods to real problems. As a statistics graduate student, my studies have emphasized methods but provided limited exposure to the broader decision making context in which many statistical analyses take place. Thus, I look forward to experiencing novel applications of statistics used by real-world policymakers while engaging in mutual sharing of ideas with other researchers from diverse backgrounds. One of the biggest things missing from statistics coursework is the opportunity to communicate with the individuals who use statistical information. I can write reports for my classes, but commentary from professors on how accurately I explain the conclusions of contrived examples is not the same as feedback from people who actually to use the results in their own work. I expect that working with representatives from government and local DC-area organizations will provide insightful lessons about presenting results concisely without the statistics becoming a distraction. I would also gain much from meeting the other DSPG fellows and researchers from around the country. The people I would work would have learned different methods and in different contexts than I did. Discussing different approaches to the same problem presented by different individuals would help all of us solidify our understanding of statistics, social and behavioral science, and inference. In this respect, I would not only benefit from the experience but would be conscious of my contributions to the program. I see the DSPG program as a way to grow as an applied statistician through helping real-world decision makers and sharing background and experiences with the other fellow. As I explain in the other essays below, I have a comfortable background in mathematics, environmental statistics, and computing. I will eagerly hone these skills on the projects that come up this summer and learn from other fellows with complementary skill sets.

Statistic	Value
Total Responses	1

10. Essay (up to 500 words): "Please describe any previous research experience and/or work experience you may have.”

Text Response

Since beginning graduate school, I have had extensive experience with applied statistics work and research. My work experience established an interest in policy and regulatory issues; I hope to grow this interest through further work because most of my research has been purely academic. I previously had a summer internship at Neptune and Company, a Colorado-based environmental statistics firm. This was my first taste of the intersection between statistics and policy. I worked on a variety of environmental stewardship projects, including monitoring of several Superfund sites. The typical use of statistical results was to decide among three alternatives: stop monitoring because the site was no longer contaminated, continue monitoring, or remediate because the site was unsafe. For some Neptune was a contractor for the site owners, while on others we worked for the EPA. Thus, I saw how data are used on both sides of this regulatory decision-making process. Unfortunately, I also saw how inappropriate statistical procedures proscribed by too-broad regulations can lead to incorrect decisions, for example, when naive substitutions for non-detects and excessive conservatism lead to overestimation of the mean contaminant concentration. This experience convinced me that statisticians should be outspoken on matters of policy and regulation. Since August of 2016, I have held a research assistantship with my university’s statistical consulting group. The clients are, by and large, graduate students working on their theses. Hence, I routinely provide design recommendations and analyses for agricultural experiments and biomedical studies. Occasionally, projects verge upon statistics research. One project began with an ecology student observing “suspiciously narrow” confidence intervals in papers using a new abundance estimation method. This method was introduced in a conservation biology journal without including any mathematical derivations or justification, so the project had me deriving the variance estimator and investigating the effects of assumption violations. This project in particular furthered my interest in having a voice in matters of statistical guidance, but my role as a consultant is primarily to help clients justify their own conclusions. My research focus is spatial point processes with environmental applications. In an application of my dissertation work, I am using nonparametric smoothing to estimate the spatial point intensity of Poisson processes. The motivation behind my model is to find unexploded ordnance from noisy metal detector data; this is an important public safety issue both around the world and in rural America where former military training sites are being transferred to private ownership or public use. Beginning the dissertation has been a slow and transformative process where I have learned to balance reading, writing, programming, and the feeling that everything should have been done yesterday. I am no stranger to working with statistics on real problems. However, my interactions have so far been with other statisticians or with consulting clients who use my results to form their own policy arguments. Thus my work has always been a step removed from its end use. Through DSPG, I would hope to see my work have direct benefits on the public well-being.

Statistic	Value
Total Responses	1

11. Essay (up to 500 words): "Please describe your background (courses taken, research projects, etc.) in statistics and mathematics."

Text Response

My mathematical and statistical background includes undergraduate and graduate coursework in both math and statistics, graduate-level statistics courses, theoretical research at both levels, and three years of statistical consulting as a graduate student. During my undergraduate studies, I developed a passion for the mathematical and logical foundations behind the inferential techniques of statistics. This began in my final semester of junior college while I was taking introductory linear algebra and differential equations courses in preparation to transfer as a statistics major. These were the first courses where I was exposed to proofs, and I was fascinated by the concepts of finding general cases and exploring the limits of deduction. I still wanted to work on concrete real-world problems, but I gained a desire to understand the why in addition to the how. Therefore, I changed my major to mathematics. I then transferred to the University of California, Riverside, to pursue a Bachelor of Science in mathematics with a concentration in applied mathematics and statistics. My program of study included rigorous courses from both the mathematics and statistics departments. The core mathematics courses consisted of linear algebra, a one-year sequence in probability and mathematical statistics, and another one-year sequence in ordinary and partial differential equations. For fun, I took set theory and abstract algebra as electives. I applied the math I learned to my statistics courses, which included mathematically formal one-quarter courses in stochastic processes and generalized linear models, as well as a quarter of advanced SAS programming. I cannot say I had a single favorite undergraduate course, but I enjoyed seeing the common mathematical themes running through all of my courses. This universality of math inspired me to work hard and ultimately graduate summa cum laude. For graduate studies, I came to Montana State University, Bozeman, where the statistics faculty work hard to support ecological, agricultural, and biomedical research on campus. Therefore, my courses emphasized clear communication so that statistical models and results can be easily understood by collaborators from other disciplines. I earned a Master of Science in statistics after taking courses in probability and mathematical statistics, regression, sampling, experimental design, Bayesian statistics, and other topics. I then continued on as a doctoral student, taking more courses, including spatial statistics and generalized linear models, before switching my focus to my own dissertation research. Other essays adequately describe my graduate-level statistics research and consulting work, so I will conclude by briefly introducing the applied math research I did as an undergraduate. During my first year at UC Riverside, a professor invited myself and two other undergrads to shadow his work developing clustering algorithms. We implemented spectral clustering in Matlab and experimented with different distance metrics. In my second year, I studied computer vision in a small group under the mentorship of two graduate students. We created a simple rectangle-detection algorithm and implemented it in Python. These projects, as well as my whole educational journey, showed me that there is still a lot more out there for me to learn.

Statistic	Value
Total Responses	1

12. Essay (up to 500 words): "Please describe your background (courses taken, research projects, etc.) in social and behavioral sciences."

Text Response

I have some experience with social and behavioral research that I gained through working as a statistical consultant. From these consulting projects, I learned about the statistical aspects of survey design and about the ethics of studying sensitive topics. I will focus on two specific projects. My first project that could be considered social or behavioral research involved an annual survey of my institution's freshmen. The research goal was to associate attendance at several first-semester events with changes in belonging, engagement, preparedness, and self-efficacy as measured by an instrument that the clients had developed themselves. The results were inconclusive, with no evidence that either the total number of events attended or attendance at many of the individual events was related to changes in any of the four traits of interest. Perhaps the most interesting finding was that the correlations among the traits increased as the semester progressed. This project was largely a lesson in the behavior of the clients – a team of three researchers with different stakes in the project who did not communicate well with each other. One of the single most educational consulting projects I've worked on is a sequence of studies about interventions to reduce stigma around suicide and help-seeking behavior in Montana high-school and college students. The client is a professor with extensive qualitative research experience but who has struggled to design studies to quantitatively very her conclusions. We have been collaborating for over a year, through several rounds of pilot studies (with a variety of data quality hiccups). My role progressed from churning out cookie-cutter regression analyses to helping clarify possible mechanisms of action as I learned more about the theoretical framework of the client's work. The overarching hypothesis is that Witte's Extended Parallel Process Model (EPPM) is at work, with individuals engaging in either fear control (rejecting the messages of the intervention and acting to remove fear of the threat) or danger control (accepting the messages and acting to remove the danger, i.e. the threat itself) depending on the balance between perceived threat and perceived efficacy. As such, I am now reasonably well-read in the scale design literature as we need reliable ways to measure these concepts. We are currently building the EPPM and other possible mechanisms into an SEM for an NIMH grant proposal. My role in that is to specify the mathematical relationships among the various traits while the client balances the ethical issues of screening and helping subjects who may be distressed by the study despite those individuals being the ones we could learn the most from. Even though my experience with the social and behavioral sciences is limited and informal, these consulting projects gave me some background to build upon. I anticipate learning much more as I continue to work on statistical applications in this area.

Statistic	Value
Total Responses	1

13. Essay (up to 500 words): "Please describe your background in programming."

Text Response

I am a trained computer programmer and longtime hobbyist who has made the transition to computational statistics. I work well with people of all programming skill levels, and have helped train some of my classmates. I firmly believe knowing how to read the documentation is the coder's most valuable skill and that by doing so I can learn any language or package as needed. My earliest programming experience was as a high school student writing games on my TI-83, an ability I acquired by spending hours studying its hefty user's manual. I ultimately earned an Associate of Science in computer programming from Riverside Community College. The curriculum provided a rigorous foundation in C++ and the theory of object-oriented programming, as well as including brief introductions to computer architecture, systems design, and relational databases. When I began studying statistics, I became familiar with SAS and R. Both of those languages are well-documented so I picked them up easily. Several of my classmates had more computational sophistication than I did, and they had learned their skills in different contexts. I enjoyed discussing parallelization and sharing efficient coding techniques with these more advanced colleagues. However, many of my classmates struggled because they had little training or experience with any programming languages or concepts. We found that collaboration was an effective way to mutually improve everyone's skill levels. We would talk about what we knew and did not know, with the more savvy coders guiding the group towards helpful resources so that we could solve computational problems together. The conclusion I drew from these experiences was that resourcefulness trumps natural programming skill because no one will know everything there is to know about modern computers. Computing is an area that is continually changing, providing opportunities to learn new algorithms and tools. For my current research in Bayesian spatial models, I wrote my own Monte Carlo samplers running in parallel, and I am presently learning the Armadillo linear algebra library and several GIS libraries. One tool that I would like more experience is the Python language. I used it for one project as an undergraduate, but I would eagerly and quickly learn more Python if the opportunity came up.

Statistic	Value
Total Responses	1

14. Essay (up to 500 words): "Please provide information about other significant courses you have taken within your field of study."

Text Response

Two graduate statistics courses stand out as being especially influential on my development as a statistician. The first was a course in generalized linear models (GLMs) with a surprising balance of fun math and straightforward interpretation of somewhat unintuitive models. The second was a course in Bayesian statistics that contained a refreshing amount of philosophy. My GLMs course was one of the most mathematically rigorous courses I have taken in grad school. (It included some matrix calculus!) Proving properties of the exponential dispersion family reminded me why I loved math so much as an undergrad, while working through the Fisher scoring and iterative weighted least squares algorithms gave me an appreciation for the power of modern computers. However, the course was also extremely practical. In class, the professor would routinely quiz us on giving precise interpretations of model coefficients. This was an intense task as we had to be able to explain interaction effects and backtransform them to the scale of the data. Thus, I became adept at making sense of these sometimes difficult models. These days, it seems that many researchers think of Bayesian statistics as a set of computational tricks to magically fit impossible models. I am grateful that I took a course taught by a true Bayesian who acknowledged that she could never fit all of Bayesian statistics into a single semester (but tried anyway). The course began with a philosophical debate about long-run frequencies, updating prior beliefs into posterior beliefs, and Fisher's subtle understanding of likelihood. This discussion turned me into a Bayesian when I concluded that Bayesian updating parallels the scientific method and that even "objective" frequentist researchers are influenced by their prior experiences. I was also convinced that Gelman's graphical ANOVA is much more sensible than the traditional sums of squares partitioning method. We spent several weeks thinking about the choices of priors for very simple models, and it was only after the students were comfortable justifying their prior beliefs that we moved on to studying complicated hierarchical models. We learned to use Stan and JAGS, and also picked up some general tricks for efficient MCMC. Throughout all of this, the professor trained us to make everything interpretable – e.g. if MCMC was simplest when working with the log of the variance, do that and then just transform it to get the posterior distribution of the standard deviation. This course had an enormous influence on how I think about statistics and I still think that an analysis should be understandable in order to be useful. Although I have taken quite a few different statistics courses, these two best summarize the insights I acquired in graduate school. Statistics can do amazing things with complicated models, but my approach as a data analyst is to understand the context and do what I can to make things simple.

Statistic	Value
Total Responses	1

15. Please list the name and contact information for 2 references (teachers, mentors, or employers) that we will contact for a letter of reference/brief survey. Please make sure you list the correct email and they know we will be contacting them soon (within the next week). Only 2 references will be contacted; do not list more than 2.

Text Response
Andrew Hoegh: andrew.hoegh@montana.edu, Lillian Lin: lillian.lin@montana.edu

Statistic	Value
Total Responses	1