# Initial Report

Last Modified: 01/16/2018

## 1. Name (first and last)

| Text Response |
| --- |
| Alex Fout |

| Statistic | Value |
| --- | --- |
| Total Responses | 1 |

## 2. Email

| Text Response |
| --- |
| fout.alex@gmail.com |

| Statistic | Value |
| --- | --- |
| Total Responses | 1 |

## 3. Contact Phone

| Text Response |
| --- |
| 3037759445 |

| 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 |
|---|
| PhD in Statistics at Colorado 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 |
|---|
| Michael F. Baumann Award for Excellence in Analysis (US ARMY TRADOC Analysis Center, 2015), Superior Civilian Service Award (US Department of the Army, 2013), Research Experience for Undergraduates (University of Colorado Boulder Department of Physics, 2009). Eagle Scout (Boy Scouts of America, 2006) |

| Statistic | Value |
|---|---|
| Total Responses | 1 |

## 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 |
|---|
| In 2013, I had the privilege of deploying to Afghanistan as a civilian operations research analyst in support of the NATO mission. I was forewarned and quickly experienced firsthand that one of greatest challenges I would face on this deployment was that of self advocacy. Military leaders had little or no knowledge of what skills I possessed and how they were relevant, despite my desire to contribute. I suspect that my experience is not unique, and that many leaders in the public sector do not know how the broad set of skills encompassed by the term “data science” can be used in their particular domain. Perhaps the increasing adoption of data science in the private sector has demonstrated its promise, but moving beyond promise to implementation remains quite difficult and met with (perhaps healthy) skepticism. This excites and fascinates me. From my initial days working as an analyst for the US Army, I have developed a desire to apply sound quantitative techniques to real world problems, and work towards a common good. I think today’s decision makers have more information than they have time to effectively process, making the role of a data scientist crucial, and the potential impact great. I’m interested in this fellowship because I’d like to learn more about data science in the public sector. I would like to interface with leaders to better understand the challenges that they face and how data science can help. I look forward to the chance to meet with mentors to learn more about career opportunities in the public sector, whether local or national. Lastly, I hope to continue in my professional mission, which is to help improve the well being and quality of life for all people. |

| 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 |
|---|
| When I graduated from my undergraduate program, I worked as a civilian operations research analyst for the US Army. While there, I conducted studies to inform Army and DoD leadership decisions. Our studies varied in scope from a single focused question to a large, broad set of general goals, and we served customers at varying levels, from peer agencies to the Office of the Secretary of Defense. The analytic process began with a methodical decomposition of study questions into atomic elements that could be identified with appropriate methods and available data. This was critical to ensure that the study outcomes properly informed the decision maker. We often employed combat simulation, wargaming, field exercises, cost estimation, risk assessment, surveying and multi-attribute decision making (MADM) to adequately address a study’s needs. It was critical to effectively communicate both methodology and results to technical experts in detail as well as to decision makers at a higher level. I was privileged to support several studies in varying capacities, including calculating metrics and performing statistical tests on simulation output, developing a MADM framework and running sensitivity analysis on it, integrating artificial intelligence behaviors into entities in a combat simulation, and reducing, clustering, and sampling from a high dimensional design space to identify a representative subset of designs for in-depth analysis. I routinely participated in the study planning process, as well as the development and presentation of results. During a short internship with Numerica Corporation, I joined a small team which develops target tracking radar algorithms. I experimented with Long Short Term Memory (LSTM) neural networks for generating synthetic target data. I also contributed to the development of a predictive terrorist activity model. During a summer internship at the National Security Agency, I developed a machine learning pipeline to perform anomaly detection on Hadoop Distributed File System namenode log events. This involved parsing log files, producing thousands of features, generating an efficient low dimensional representation, clustering groups, and generating empirical p-values to identify anomalous events. As a Computer Science graduate student at Colorado State University, I joined a research group which focuses on bioinformatics problems. My MS thesis investigated the use of deep learning methods to the problem of protein interface prediction. The goal was to exploit the irregular 3D structure inherent in proteins using a convolutional neural network. We ended up representing proteins as graphs, which each amino-acid residue as a node in the graph, and relationships between residues captured on the edges. We then designed convolution operations which operate on graphs, and trained specialized neural networks using known protein interfaces to identify which residues of proteins in the test set were part of the interface. Data consisted of raw 3D atom coordinates. I wrote code to generate features from the coordinates, build neural networks, and run experiments on them using python and tensorflow. A paper based on this research was accepted to the Neural Information Processing Systems 2017 conference, where myself and a coauthor were able to present a poster. |

| 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**

One of my undergraduate majors was Applied Mathematics, which included courses in multivariate calculus, partial differential equations, linear algebra, real analysis, discrete mathematics, mathematical modeling, and applied statistics. In the mathematical modeling course I trained neural networks to recognize handwritten digits (MNIST data), simulated the effect of driver distraction on car accident rates, and performed Latent Semantic Analysis on the Greek text of the new testament. My second major was Engineering Physics, for which I took multiple courses in experimental physics. I would routinely have to model a physical system and fit data to it, being careful to propagate measurement and instrument uncertainty through all calculations to produce an uncertainty in the final result. During my Master's degree in Industrial Engineering, I took a multivariate statistics course. For this class, I completed a project in which I evaluated strategies for electrical transmission network expansion, using principal component analysis. I also took courses in discrete optimization, stochastic processes, and statistical quality control. During my Master's degree in Computer Science, I took a course in the analysis of algorithms, and another course in bioinformatics which covered biologically related algorithms like those used for gene sequence alignment and RNA folding. In my current degree program, I have taken graduate level courses in probability, regression, and time series.

| 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 don't have much formal training in Social and Behavioral Sciences, but I have had the opportunity to work on a few projects related to this field. I recently collaborated with a Research Psychologist who was interested in studying social networks in the context of an at-risk youth mentoring program. Her desire was to model changes in the social network over time, and understand how network growth mediates program outcomes. There were two experimental conditions, and she was also interested in how each condition affected network growth. The data collected for this program include several surveys taken throughout the semester which measured belongingness, mentorship alliance, and relationship strength. After examining the network data, we decided first to examine the end-state networks, recognizing that the network dynamics were quite noisy. We fit end state networks to a Stochastic Block Model and various other independent dyad models and evaluated their fit by comparing certain network metrics (e.g. betweenness centrality) to simulated networks randomly sampled from the models. At the conclusion of my involvement, the Psychologist was going to investigate more complex network models, such as Exponential Random Graph Models.

| Statistic | Value |
|---|---|
| Total Responses | 1 |

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

**Text Response**

I have experience programming in several different languages, but I am most comfortable with Python, Java, and R. I have used MATLAB, SQL, VBA, Mathematica, C, HTML/CSS, PHP, and Javascript. I was a graduate teaching assistant for an object oriented programming course where I would teach Java programming to students with one semester of Java experience. We covered classes, inheritance, polymorphism, interfaces, recursion, native Java data structures, and sorting algorithms, among other topics. I received generally positive feedback from my students, with one remarking "Alex is very knowledgeable of everything taught and wants to make sure his students learn it. He helps by getting you to think about problems and guides you to your own solution which is great for a class like this". I programmed exclusively in Java during two Big Data courses, where we designed and implemented MapReduce algorithms and ran them on a Hadoop Distributed File System, which we set up and configured ourselves. We utilized Amazon Web Services for this as well. I also implemented a version of Craig Reynolds' steering behaviors in an entity level discrete event combat simulation written in Java. My Python experience is closely linked to machine learning, as I learned the language during a machine learning course and used it extensively throughout my Master's research. With two other graduate students, I wrote an extensive application which loaded data, produced features, constructed neural networks, and ran training/testing/validation experiments both on a single machine and distributed over multiple machines. For this project we made use of several python libraries, including numpy, scikit-learn, and tensorflow. I contributed over 80% (roughly 8500 lines of code) of the code for this application. I have also written a web crawler in Python to extract information about Jeopardy games and contestants from j-archive.com. I used R during a graduate Regression course and also as a consultant to a Psychologist who wanted to analyze social network data. For that project I wrote R code to load and clean the data, create various visualizations of the social network and individual relationships, fit network models to the data, and run Monte Carlo simulations to identify model lack-of-fit. This code was made available to the researcher at the conclusion of my involvement. In addition to these three languages, I have varying levels of experience in a number of other languages. I am very comfortable with git and have used it in many projects during the past three years. I wrote my Master's thesis in LaTeX, although I have not created my own templates, just modified existing ones. I created my own website using a mix of HTML, CSS, Javascript, and PHP. And several years ago, when I only knew how to program in MATLAB, I wrote a script to read the Greek text of the New Testament, tokenize it, apply TF-IDF weighting, and perform Latent Semantic Analysis to assess inter-chapter similarity. I have written basic SQL queries to extract and summarize results from a combat simulation.

| 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**

I have no other significant courses that have not been noted elsewhere.

| 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**

Asa Ben-Hur (asa@cs.colostate.edu), Bailey Fosdick (bailey@stat.colostate.edu)

| Statistic | Value |
|---|---|
| Total Responses | 1 |