# Initial Report

Last Modified: 01/29/2018

## 1. Name (first and last)

| Text Response |
| --- |
| Yiqi Tang |

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

## 2. Email

| Text Response |
| --- |
| ytang22@ncsu.edu |

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

## 3. Contact Phone

| Text Response |
| --- |
| 910-584-7772 |

| 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 |
|---|
| Master's, Statistics, NC 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**

I seek to gain practical experience in data science to complement the foundation in theoretical statistics I have begun to develop as a graduate student by working on a project that is highly impactful to society. This would combine all of my passions – statistics, big data, and affecting a tangible impact on the lives of others. I seek to understand the world in the form of data because of how it can be leveraged in the search for objective truths. I chose statistics, and before graduate school, math and economics, because of my love for problem solving and quantitative analysis. I have enjoyed the theoretical courses I have taken since they allowed me to understand the fundamentals of data analysis. I hope to use my statistics knowledge to enhance popular data science methods. I am also enthusiastic about big data and machine learning. I was highly involved in a local machine learning meetup group in the Charlotte area before moving to Raleigh. Being one of the managers of the group, I helped organize presentations, and actively participated in hackathons and data competitions with other group members. I want to apply these same motives within a formal setting like DSPG, which will offer me an invaluable learning opportunity, but more importantly, provide the proper venue to maximize my impact on society. In addition to my passions for data and statistics, I have a strong desire to better our world. A quote from Kai-fu Lee’s autobiography summarizes my motivations: “Imagine two worlds, one with you and one without you. Maximizing the difference between the two is the meaning of your life.” I have thought of it regularly since reading it over ten years ago, and embody it in my dedication to community service. As a Bonner Scholar at Davidson College, I successfully completed over 1500 hours of volunteering over four years. Bonner Scholars are a group of students committed to creating positive change to the community by service, research and action. I volunteered at many different organizations, such as Legal Aid of North Carolina and the Council for Children’s Rights in Charlotte. At the Council for Children’s Rights, I worked with the director of Education Initiatives, and pushed forward the most important agenda for the council: becoming an umbrella organization for Out-of-School Time (OST) programs in the Charlotte area. To that end, I helped created an interactive map of OST programs so that parents can easily find programs suitable for their children. These experiences helped crystallize my ultimate goal: using data science to better our world. I have always been passionate about helping others, and now, I would like to use the empirical toolkit I have built to solve real human problems. Spending the summer working with other talented people on data projects that have a high social impact would be a dream come true. I hope to gain practical experience doing something that I am passionate about, something that could affect a real impact on a human problem.

| 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**

I have worked on several empirical research projects that dealt with large amounts of data, such as consulting work with Dr. Timothy Chartier for Dunham’s Sports and the NBA league office. For the NBA project, we pulled data from SportsVU cameras to analyze. The datasets were flat files each containing data on a single NBA game, with information on measures of time (game clock, shot clock, etc.) and coordinates (x, y, z) of the basketball and the players at all times. We worked to analyze tracking data to answer some of their most pressing questions. I also worked with Dunham's Sports to analyze their sales data and how well their advertisements increased sales. I was able to clean up and merge very messy datasets from the sporting goods company, and dig into the most important pieces of information. The project involved quite a bit of feature engineering, where I created many different variables relevant to sales and the questions we were trying to answer, such as the effect of weather on sales. Before beginning my graduate studies, I worked for two years as an analyst at the OANDA Corporation, a fin-tech company. I worked on building a variety of models to answer business questions for the company. I built a random forest model to predict high value clients, built a model to determine sales team compensations, and built a regression model to project what our revenue would look like in an upcoming year. I also worked on competitive analysis projects for several lines of businesses for the company, as well as building visualization cards in Domo for management reporting. My work was highly valued by the company and the executives. My model is used directly in determining compensations, and the competitive analysis insights and Domo cards are used frequently by management. In the project to predict high value clients, I took data from our very large SQL database, taking only the relevant fields, particularly the relevant fields clients fill out during registration and subsequently their trading activities. After a lot of data cleaning using the Pandas and Numpy libraries in Python, I was able to obtain a fairly nice dataset. From there on, I worked on feature engineering and started building models, seeing what models worked best. I eventually decided on a random forest model, but soon realized that my dataset is very unbalanced. I mitigated the problem by balancing the dataset and adding in a penalty to the model to balance it out. I believe that my experiences as a research assistant and in independent projects, as well as in my previous job have all prepared me well for this highly data-driven opportunity.

| 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**

Throughout my undergraduate and graduate education, I have taken many courses in mathematics and statistics. I double majored in mathematics and economics in undergraduate, and am currently pursuing a Master's degree in statistics. I have taken calculus III, linear algebra, differential equations, sets and proofs, complex analysis, numerical analysis, dynamical systems, data structures, mathematical modeling, topology, and real analysis. As a current graduate student in statistics, I took statistical theory I, applied time series, and experimental statistics last semester. I am currently taking statistical theory II, linear model and variance components, and applied multivariate and longitudinal statistics. I have done a number of research projects in mathematics and statistics. An example of a project I have done is mapping out an efficient and effective classroom for a local nonprofit, Ada Jenkins Center. The Ada Jenkins Center is non-profit providing community-based education. They needed help designing a classroom that would optimize learning outcomes and efficiently utilize limited space. After researching and reviewing relevant literature, I learned that the best practice is to maximize space between students for independent tutoring, while making sure that each student remains relatively close to one other student for easy pairing. After plotting the classrooms on the computer, I used two methods to find an optimal solution: k-means cluster analysis and binary integer programming. I presented the results of my work to the non-profit with a blueprint of the best layout, which they ultimately adopted. I have also done a research project in theoretical mathematics, where another student and I researched the basins of attraction for the fixed points and periodic points in function iterations as a final project for our dynamical systems class. We browsed relevant literature and learned about the basins of attractions and their associated boundaries through reading papers and understanding theorems. We wrote a final paper on the subject. One project I took on was organizing a fun event that involves mathematics. My senior year of college, we were one of 20 sites around the world to participate in the MegaMenger project, a fractal building project where people come together to build a level-4 Menger sponge by uniting 20 level-3 Menger sponges from the 20 sites. A Menger sponge is a self-similar fractal that is the same at every scale. We built our level-3 Menger sponge with old business cards collected from local businesses. It took a total of over 1 million business cards! I organized the entire build with math professor Dr. Donna Molinek, planning out every detail from securing grant money, to getting students, faculty, and other members of the community to come be a part of the fun. We received coverage from news outlets such as the Huffington Post (https://www.huffingtonpost.com/tim-chartier/a-million-business-cards_b_6128880.html) and the Charlotte Observer (http://www.charlotteobserver.com/news/local/community/lake-norman-mooresville/article9228377.html). The whole build took several weeks, but it was a very rewarding experience, getting people involved and maybe even influencing people to see the fun in math.

| 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**

As a mathematics and economics double major, I took a variety of courses in economics at Davidson College. I have taken introductory economics, intermediate microeconomics, intermediate macroeconomics, introduction to accounting, economic growth and sustainable development, international trade, economics statistics, econometrics, sports economics, and senior session, which is the capstone course for the economics major. I have done quite a few independent research projects in economics. I have looked at whether single-sex education is beneficial to student academic performance, and explored high school GPA and SAT scores as an indication of college success. I have also researched brief country study on international trade in Mainland China, as well as written an empirical paper looking at the effect of Arena Football League (AFL) franchise relocations on Major League Baseball (MLB) attendance. I am especially proud of the project I did for my senior capstone research where I examined retention at Teach For America. I tested a novel approach to a classic problem in education research: factors affecting teacher retention. Scraping data from LinkedIn, I obtained a raw dataset of more than 10,000 observations, and ultimately used a random sample of 1,000. My approach produced a variety of TFA participants, but also introduced both sampling bias due to limitations of Internet data and omitted variable bias due to constraints on the available inputs. I eventually settled on five dummy variables: return to teaching, prior urban-living experience, prior education experience, urban school district during TFA, and top college. I regressed return to teaching on the other four variables using a linear probability model with robust standard errors. I found that on average, corps members in urban areas are 10.2% more likely to return to teaching, ceteris paribus. This past year, I participated in a competition called the Ignite NC challenge. The challenge encourages participants to come up with a proposal that uses gigabit fiber and smart city technologies to solve one of four problems: improved video streaming and collaboration, digital infrastructure for the home, safer road intersections, and regional sharing. My other two teammates and I decided to come up with a solution proposal for safer road intersections. We focused on using machine learning algorithms to "learn" traffic patterns and determine how safe a specific intersection is at different times of the day. Our machine can absorb real-time streaming video data from traffic cameras, and make our intersections smarter. We researched existing practices on building machine learning models using video streaming data and how people utilize traffic camera data. Our innovative solution was selected as one of 10 finalists in the state, and we eventually took home 3rd place at the final pitch, winning 120K in IBM Watson IOT/Bluemix Cloud credits. I learned a lot from the experience, researching social science methods in the form of city planning and road intersections, as well as how to perfect pitching in front of judges.

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

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

**Text Response**

I have experience working with a variety of programming languages including SAS, Java, Stata, Matlab, Mathematica, and SQL with particular proficiency in Python, R, and Java. I self-taught introductory programming, and took Data Structures in undergraduate studies, where we used Java to implement various classic data structures algorithms. Since then, I have used Python regularly for work and for fun. Along with other members of the machine learning meetup group, we learned about different machine learning methods from online courses and worked on solving Kaggle challenges, mostly using Python. For my work with OANDA Corporation, I have done a lot of data cleaning, as well as built regression models, decision trees, random forest models, etc., using Python and Jupyter Notebook. I am currently enrolled in a research-based course in stochastic optimization, where we build experiments in R to empirically test how a number of solvers do in stochastic optimization problems, with objective functions both continuous and discrete. For the course, I am writing a lot of R code to wrap functions and run experiments. We are also using Latex each week to write up our findings, as well as write up our final manuscript. Last year, I worked on a project with an economics professor for an organization based in Raleigh called Rise Against Hunger, analyzing data they had to determine the most efficient way of holding meal packing events. We did feature engineering in Matlab, adding features that describe locations and distance with longitudinal and latitudinal coordinates. We implemented solutions by running regression models. This past March, I participated in the CLT Hackathon where I worked on data from Big Brothers Big Sisters. My team came up with an innovative solution by first finding out whether a big brother/sister is a good fit for the program based on the people who put Big Brothers Big Sisters on their LinkedIn profile, then, using a Random Forest model to find the most important features in big/little pairing for local optimization, finally, using Monte Carlo simulations to pair bigs and littles up to achieve global optimization. We were one of three finalists to present in front of all the participants and judges. I also participated in a data dive last year held on SAS campus in Cary, NC, where I got to work with data from a nonprofit organization Step Up. Together with a team of other people, I worked on cleaning the data, data visualization, feature engineering, and building a model to predict participant success in their program. My analysis was mostly done using Python and Jupyter Notebook.

| 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 am currently taking a research-based seminar course in the computer science department at NC State University called Stochastic Optimization: State of the Art and Beyond. Along with the professor Dr. Franc Brglez and five other students, I am empirically testing how stochastic optimization algorithms perform. We primarily use R to write code and implement experiments on stochastic optimization algorithms, to empirically test how well these algorithms perform in solving different types of problems, both continuous and discrete. So far, we have already investigated how different methods perform while finding the true value of π, and how fast these algorithms are when the number of significant digits required increases. We have also looked at different evolution algorithms, mainly the DEoptim R package and the DEopt function in the R package NMOF. We will be also looking into other algorithms and problems as well such as genetic algorithms, particle swarm algorithm, Markov chains and random walk problems. We will be publishing at least two papers on stochastic optimization solvers based on the work we do this semester.

| 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 |
| --- |
| Dr. Timothy Chartier, tichartier@davidson.edu; Phillipe Loher, phillipeloher@yahoo.com |

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