Initial Report

1. Name (first and last)

Text Response
Beomseok Seo

Statistic	Value
Total Responses	1

2. Email

Text Response
bzs32@psu.edu

Statistic	Value
Total Responses	1

3. Contact Phone

Text Response
4154639503

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. Statistics at Penn 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

Graduate Study Fellowships of The Bank of Korea(2016), Best Honors Scholarship in Korea University(2010), Honors Scholarship in Korea University(2006-2010), Ansan City Honors Scholarship from Ansan city in Korea(2005)

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

My dream is to make the world a better place by discovering information from data. For this reason, I started my first work at the Bank of Korea (BOK), which is the monetary policy apparatus of the South Korean government, and now am studying statistics in PhD. program at Penn State University. I have always believed that making a better world starts from knowing the world better. I could make my dreams partially come true by contributing my knowledge and experience in statistics on many projects that I worked on at the BOK and PhD program. My experience at the BOK was ultimately to help society by providing macro economic information to either policy makers or public. Now I would like to help society by dealing with not only macro economic data, but also micro data regarding social administration. Especially the project about emergency alert system, and early fire detection in SDAL attracted my interest. Those topics are definitely the parts that can be improved through data analysis and also are very important for the safety of society. Therefore, I believe that the internship position in SDAL will be the best opportunity to follow my dream that is to be a data scientist to help society. I expect that I will be able to learn more detailed hands-on technique that is helpful to deal with many social problems, and have the opportunity to see and deal with huge social data set that may be a key to improve the world. Moreover, the internship in SDAL will be a great time to enhance my research as well. Aa a PhD. candidate in statistics, I have worked on the field of data mining, and statistical learning for the last two years. Many problems that have seemed to be difficult to solve in the past are now settled through statistical learning and data analysis technique. One example is efficient logistics and transportation system by leading tech companies such as Amazon and Uber. Although those techniques have been developed mainly by technology companies in industry so far, I strongly believe that statistical learning and data mining are the key technology to solve many social problems as well. The experience of dealing with complex social problems will be a great opportunity for me to try my knowledge that I earned in academia to the real world.

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

Before I began my PhD. program at Penn State University (PSU), I have worked as a data scientist as well as a junior economist for six years at the Bank of Korea (BOK). My role for six years at the BOK was finding the significant patterns among various data sets. When I was a member of the Money Markets Team in the Financial Markets Department, I managed a model for risk factors in short-term interest rates in Korea using various time series methods, such as VAR and GARCH. Along with this, one of my most memorable accomplishments was constructing a model for risk exposers between markets using network analysis. I believe this experience will assist me with analyzing network data types in SDAL. After I moved to the National Expenditure Team in the Economic Statistics Department, I worked mainly to compile and estimate the Gross Domestic Product (GDP) of South Korea. For this work, I used SQL and R on a daily basis to manage big data sets from the Bureau of Customs and the Public Finance. There, I used to write long SQL programs to construct the compiling process automatically. These hands-on experience about dealing with big data through programming languages, such as SQL, R, SAS and Python would be helpful when working with SDAL as well. In addition to my previous work experience, I have accumulated various research experience since I joined the PhD. program in statistics at PSU. In the last summer, I worked as a research assistant under Dr. Bumba Mukherjee in political science department at PSU. We worked on the historical conflict data on civil war duration in the US, and suggested a new estimation of the time duration of each local war based on the result of Bayesian survival analysis. Since 2017 fall when I chose my adviser, I have worked under Dr. Jia Li focusing on statistical learning, and data mining, specifically with clustering algorithms on various data types. The recent paper that I am now working with Drs. Jia Li and Lynn Lin is on a methodology to find the more robust clustering and the stability measures for clustering using the optimal transport approach. This paper is improving the accuracy of unsupervised learning algorithms that have been relatively neglected compared to supervised learning. I believe that I have improved my ability to analyze various types of data through my work and research experience, and my experience will be helpful to solve the problems that SDAL deals with.

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

As a PhD candidate in statistics at PSU, I finished one and a half year of coursework in my program. The coursework covered mathematical background as well as applied topics in statistics including various course projects. To be specific, for mathematical courses at PSU, I took two statistical theory courses(513, 514) of PhD level, which covered basic probability and distribution theory, and inference theory including MLE, Bayesian estimation, and construction of confidence interval and test. Also, I took asymptotic tools(553), measure theory(517), stochastic process and Monte Carlo Markov chain(515), and now am taking more advanced inference course(561). For applied topics, I took not only the classical courses in statistics such as design and experiment(512), and regression models(511) but also cutting edge topic courses such as deep learning(math 597), spatial data analysis(597), and now I am more focusing on my research field by taking data mining(557) and data mining 2(ist 558) The advantage of taking applied topic courses was not only that I could learn new techniques, but also that I could apply what I learned to the course projects. I analyzed California temperature data and seismic faults data using various spatial data analysis such as point process and conditional autoregressive(CAR) model in spatial data analysis course. Moreover, I mainly used Python and Tensorflow to study convolutional neural network and recursive neural network in deep learning course. In college, I also focused on both the mathematical background and applied topics in the same time. I took mathematical statistics and topic course of mathematical statistics, matrix theory, introduction to probability, multivariate statistical analysis, econometric analysis, time series analysis, regression analysis, sampling theory, and so on. One of the most impressive classes in undergraduate courses was multivariate statistical analysis. As a final project, I, with my team members, analyzed the hierarchy of universities in Korea, using PCA, factor analysis, and hierarchical clustering and suggested a new index to provide a data-based ranking of universities by reducing the dimensions of large variable data sets. Lastly, in addition to college courses, I also took four more mathematics courses such as multivariate calculus, real analysis, stochastic process, and advanced linear algebra from 2012 to 2016, after I graduated college, at the Advanced Mathematics Program of the Bank of Korea Academy, which were lectured by professors in the Korea University College of Science, and had the same course hours and grading process as traditional university courses.

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 graduated Korea University with bachelor's degree in economics and statistics. For economics, I studied many courses providing intuitive idea about society. I took micro and macro economics, economics of strategy and information, economics of industrial organization, international finance, international trade, public finance and so on. In addition to the economics courses, I took a couple of additional courses that provide understanding of society such as introduction to public administration, introduction to law, statistical method in social science, and so on. The courses that I took in college helped me to understand how society works in a broad perspective of view. Whereas, the project that I worked on at the BOK provided the more detailed picture of specific parts in society. I worked on various topics regarding financial behavior across income quintiles at the BOK to provide relevant information to policy makers. One of the memorable work that I did at the Financial Markets Department in the BOK was that I analyzed the risk exposer of households to saving banks in Korea across income quintiles and estimated the impact of the business suspension of saving banks across income quintiles. I collected data from accounting documents from saving banks and used R, SQL tools to manage and express the data in an efficient way. Another memorable project that I worked on at the Economic Statistics Department in the BOK was an analysis project regarding the change of government expenditure in social welfare field in Korea. As the person in charge of general government expenditure statistics in national account of Korea, I managed public finance data and wrote regular reports including special topics about government expenditure. In order to capture meaningful patterns and impacts of government expenditure, I matched various data sets and estimated the effect of government expenditure using time series and regression models. The problems that I have worked on for the last ten years were to provide insightful pictures to society. I believe my experience will be helpful for the project that SDAL is working on as well.

Statistic	Value
Total Responses	1

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

Text Response

When it comes to programming, I studied C by myself when I was 12, and through the educational and work experience, I have enhanced my programming skills. I am now very confident in using R, SAS, SQL and have experience about Matlab, Python, and C/C++. In college, I took computational statistics I and II, introduction to statistical programing, and studied basic SAS, R and C programming. But, my programming experience was more improved through my work at the BOK. I used R, SQL on a daily basis in order to access and manipulate foreign exchange transaction data and the Bureau of Customs data. I used to write long SQL codes about deterministic process to automatize the compiling process of GDP in Korea. Moreover, to deal with missing data and predict the future values, I ran ARIMA models and various filtering techniques using R and SAS programming. Especially, the preprocessing work to manage incomplete data in real world improved my basic programming skills dramatically. My programming experience has been improved even more after I joined PhD program. I studied on various data types through different courses and managed those data using R. I analyzed spatial and time series data, image, geologic data, biological data, survey data, simulation data, and so on, and ran different R packages as well as worked on preprocessing for course projects. In addition to R, I also worked on Python codes to run convolutional neural network in the deep learning course. To study different coding experience, I attended various graduate students workshop where the senior and junior statistics PhD. students share their own coding experience and techniques together. One of the memorable specific experience regarding programming is that I wrote MCMC codes to estimate Bayesian survival models while working as a research assistant under Dr. Bumba Mukherjee. Writing codes according to theory is one of the main things that I have practiced in my PhD. program. I used to write codes to check theory by comparing it with simulation study. Moreover, my current research field is statistical learning and data mining, and I am now working on developing a new algorithm using C under Dr. Jia Li and a corresponding package that runs in R. I believe my experience in programming will be competitive in analyzing various data types in SDAL.

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

In addition to the regular courses in college and graduate school, I tried to attend in as many topic courses in various training organizations as possible. In July 2014, I attended summer school in the Barcelona Graduate School of Economics for one week and took two topic courses: empirical time series methods for macroeconomic analysis, and Bayesian vector autoregressions and small sample corrections in VARs. In addition, I also took two weeks training at the Singapore Regional Training Institute of IMF in December 2012. The training was mainly about basic statistical analysis for financial markets and new financial instruments.

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

Jia Li, Professor at Penn State, jiali@psu.edu, (814) 863-3074; Lynn Lin, Assistant professor at Penn State, llin@psu.edu, (814) 863-8677;

Statistic	Value
Total Responses	1