# Proceedings of the 5th International Big Data and Analytics Educational Conference

16th July 2017

Langkawi, Malaysia

E-Proceedings: http://waglobalacademicsnetwork.org/proceedings-of-the-5th-international-big-data-and-analytics-educational-conference/

**********

# Improving Student Learning with Experiential Learning in and with Data Analytics

Balasubramaniam Ramesh
Georgia State University, USA
bramesh@gsu.edu

Shuguang Hong
Georgia State University, USA
shong@gsu.edu

## Abstract

Experiential Learning Theory (ELT) (Kolb, 1984) is a prominent theory that has been widely tested and adapted in delivering experimental learning (McCarthy, 2016). ELT defines a four-stage, circular learning model that can be used to characterize the learning styles of learners. By recognizing the different learning styles, educators can develop suitable instructional design to improve teaching effectiveness, learning experience and outcomes.

ELT offers an excellent theoretical lens for introducing analytics in the curriculum. Since analytics often involves generation of hypotheses (rather than testing them) and is exploratory, the four stages of ELT can be readily adapted to courses on analytics. For example, in a course on big data analytics in the graduate program at GSU, students required to work on 'real-world' projects (e.g., insights from the presidential election). Students obtained structured and unstructured data from a variety of sources including census, data aggregators, twitter etc. to explore different facets of the election campaigns. Student teams presented intermediate findings periodically to the instructor, the class as well as experienced professionals during the semester. The feedback they were provided covered a variety of areas including tools, visualizations, analytical techniques, presentation of results, scope of the enquiry etc. Students were required to reflect on the feedback and emerging data from the field and incorporate this learning into refining their work. This iterative process helped significantly enhance student learning and performance. All the student groups submitted their final project presentations to a national competition on analytics. All the six teams in the class were selected as national finalists, and an all women team won the national title in its category.

Motivated by its success, the experiential learning cycle of think-act-reflect is being implemented in the field study courses. Student teams work on IT projects sponsored by partnering organizations. Students apply their knowledge and skill to solve real business problems. Our research examines the use of data analytics to improve the learning experience and outcomes in the courses. It collects and analyzes structured and unstructured data such as weekly project logs, project plans, risk analysis, working documents, email communications, text messages, group chats and social media postings. Our research seeks to generate insights by creating dashboards that provide targeted guidance and suggest interventions which can be used by students to assess their own performance as well as that of their teams. Reflective observations facilitated by this feedback is expected to enhance student learning and performance.

Keywords: Experiential learning, data analytics, reflective observation

REFERENCES

Kolb, D. A. (1984). Experiential learning: experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall, Inc.

McCarthy, M. (2016). Experiential Learning Theory: From Theory To Practice. Journal of Business & Economics Research, 14(3), 91-99.

# The Making of a Successful Analytics Master Degree Program: Experiences and Lessons Drawn for a Young and Small Asian University

Michelle Cheong
Singapore Management University, Singapore
michcheong@smu.edu.sg

## Abstract

Singapore Management University's School of Information Systems is a young school within a young and small university in Asia. Being young and small, establishing a successful analytics master degree program required extensive landscape research, assessment of its own strengths and weaknesses, having a committed team, and having a clear vision to meet the ever changing needs of the industry. The Master of IT in Business (Analytics) program, established since 2011, has grown from an annual intake of 16 to 128 students in six years. This paper attempts to describe the design process, challenges faced, decisions made, and the key actions taken, which resulted in an extremely successful analytics master program. The experiences and lessons drawn can become valuable references for other universities who are also planning to launch analytics master degree programs. The paper also summarizes the 11 key takeaways which can be used as a strategic guideline.

Keywords: Analytics master program, challenges and solutions, lessons learnt

Note: The full paper is available at the forthcoming issue of IJBIR: http://www.igi-global.com/journal/international-journal-business-intelligence-research/1168

# Reshaping Master in Data Analytics Curricula with the trends in the Big Data World

Elena Gortcheva
Department of Information and Technology Systems,
Graduate School, University of Maryland University College,
3501 University Blvd. East, Adelphi, MD 20783
elena.gortcheva@umuc.edu

**Abstract**

Any company has huge amount of structured and unstructured data and need data professionals to model, design, implement and maintain the organizational data structure and get value of Big Data. The main challenge in the academia is how to keep the curricula current with trends in the workplace and industry. In this work is presented the redevelopment work for the Master of Science in Data Analytics Program (MSDA) at the University of Maryland. The curricula have been redesigned based on the today's trends in the Big Data world and direct inputs from the industry.

Thesis: A high-quality academic program for Data Professionals has to address real workforce and industry needs in the realm of the emerging Big Data technology. Curricula can be kept current only having direct collaboration with industry.

**Big Data, Analytics, and Cognitive are coming together**

The program is developed with the following main competencies at the end of the program; student who completes the MSDA program will be able to:

 Apply statistical and machine learning techniques for data understanding, predictive modeling and interpret and communicate the results
 Transform large data sets into actionable information in an easy-to-understand format to support organizational decision making through the use of advanced analytical tools
 Apply big data analytics technology to specific area such as health care, marketing, insurances, cyber security, biological, medical and scientific applications
 Evaluate the appropriate methods and tools for data analysis in specific organizational contexts, including selecting a modeling approach, building a model using appropriate tools, validating the model, and deploying the model for prediction and analysis.

For students in MSDA the most significant technical skills to acquire on one hand are machine learning, data mining, and predictive modeling and on the other hand are the programming languages/packages and tools that enable them performing data cleaning, data mining and visualization. The most important programming languages/packages and tools currently are R, Python, SAS, Watson Analytics, Tableau, and of course distributed open source frameworks as Hadoop and Spark ML. It is very important to keep in mind that the technical skills in the second group are subject to rapid changes, so the graduate must possess the ability to learn fast and constantly new languages/packages and tools.

MSDA program is constantly updated based on industry input. There is a major need for faster, better, and improved ways for decision support, analytics, and BI systems. On the other hand cognitive computing has come to unlock the power of unstructured data in forms of text, audio and images. In the last couple of year we have observed natural confluence of

Big Data, Analytics, and Cognitive computing. More and more business leaders are choosing cognitive approaches to best meet their business goals. In order to prepare professionals with skills relevant to industries we have completely changed the content of the course in Big Data Analytics including content on cognitive computing applications using IBM Watson.

Keywords: Watson Big Data, Cognitive computing, Curricula redesign

# Designing, Developing, and Delivering a Graduate Data Science Curriculum

Manjeet Rege and Bhabani Misra
Graduate Programs in Software
University of St. Thomas
St. Paul, MN 55105, USA
{rege, bsmisra}@stthomas.edu

**Abstract**

Founded in 1985, the Graduate Programs in Software (GPS) at the University of St. Thomas, is one of the largest and longest running programs in the United States. With over 3400 alumni from 39 states and 14 countries, it is also one of the most diverse graduate only software and information technology programs. To address a growing need for data professionals in Twin Cities, MN, GPS introduced a Master's of Science in Data Science degree in the Fall of 2014. Within less than three years, this degree has witnessed a tremendous success with a current enrollment of more than 200 graduate students. We will present our 3D process of *Design, Development,* and *Delivery* that we followed for successfully launching the degree.

***Design.*** One of the keys to its success and consistent growth in recent times is that the department has always been very responsive to advances in technology. To this end, GPS relies on an advisory board consisting of senior executives from the industry. Through this association, the department directly receives valuable inputs from some of the leading industry experts about business challenges as well as data science skills expected from the student workforce graduating from our programs. The Data Science degree was developed in close consultation with these experts.

***Development.*** The second step in our process was to take preparatory measures such as academic alliances and faculty training. GPS became an academic partner with *Cloudera*, a leading provider of Big Data software, support services, and training. We also partnered with *SAS, Tableau*, and *Amazon*. Students in our courses have greatly benefitted by getting free access to state of the art software, training materials, and preparation for some of the certifications. Full-time faculty participating in this degree completed a series of courses offered by these companies as well. In addition, faculty also attended training at several other data science conferences and workshops. The department purchased a high-end computing cluster that was specifically put together to store, process, and analyze large amounts of data. The cluster has been regularly used by students in our Big Data courses.

***Delivery.*** The instruction in the degree is provided by a blend of full-time as well as adjunct faculty. The full-time faculty who underwent the professional development took the initiative in developing new courses to support the degree. Some of the sections of these courses are taught by adjunct faculty who work full time in the industry. While full-time faculty lead in curriculum design and development, adjunct faculty bring in a fresh industry perspective into the classroom that our students stand to benefit from. The degree consists of 12 courses and follows a classroom face-to-face delivery model. By interviewing several students from our program, we learned that students preferred the in-person mode of instruction in the classroom as opposed to an online learning model.

# Big Data Beyond The Classroom In An Education Environment

Peter Tong
Concordia International School Shanghai, China
general@petertong.com

Patrick Love
Concordia International School Shanghai, China
patrick.love@concordiashanghai.org

Christopher Carter
Concordia International School Shanghai, China
chris.carter@concordiashanghai.org

**Abstract**

Big Data is not just a course at Concordia International School Shanghai. Since its inception 3 years ago, the school has embraced Big Data from the classroom, to counseling department and on to administrations. What began as a pilot course for students gaining exposures to Big Data Analytics three years ago has its interests and applications beyond the high school classroom settings. While the course is gaining international exposures through students presenting at international big data conferences and international collaboration with other institutions, Big Data is being utilized in the different departments within the school. The student admission department used enrollment data from surrounding schools to predict future student enrolment. Administrators used data on courses that student enrolled in to find relationships between AP courses taken and college acceptance. Curriculum department used student data to better understand student learning. Counselors are using Big Data to better understand how students experience school, and work with student data to optimize each student's school experience. Based on current research on the subjects of student success and life satisfaction, as well as risk factors that prevent success in school and life, we track student stress, happiness, sleep, workload, motivation and relationship satisfaction. Our goal is to remove barriers that prevent students from reaching their potential, while at the same time connecting students to leaning and growth opportunities within our school and beyond. While Big Data functions are being introduced in several high school courses, In AP Computer Science Principles (APCSP) a whole unit is dedicated to Big Data. Operating in teams of four, APCSP students conduct research into topics of their choice, apply Big Data concepts and tools, and present their findings formally. As this program matures, these presentations will increasingly be conducted to authentic audiences of scholars and business leaders within the community. Presenting skills are seen as fundamental to Big Data research at Concordia because communicating complex concepts in a concise, clear, and understandable manner is necessary if the findings from Big Data is going to be applied in a meaning way.

Keyword: Big Data, K-12 education, Student wellness, High school, AP Computer Science Principles

# Social Analytics for Business

Michael Bliemel
Dalhousie University, Canada
m.bliemel@dal.ca

Ali Kazerani
Dalhousie University, Canada
ali.kazerani@Dal.Ca

Sarah Lahanky
Dalhousie University, Canada
Sarah.Lahanky@Dal.Ca

**Abstract**

Organizations today can leverage the analytics of social media data to generate insight and business value using cloud based tools such as IBM Watson Analytics for Social Media (WASM). We engaged in a pilot to develop curriculum that teaches business students the benefits and challenges of social analytics, with an emphasis on critical thinking and interpretation of the results. We implemented the curriculum in a module of our graduate and undergraduate business analytics courses, which take place in our Executive Analytics classroom which was designed for hands-on team based learning with technology. The format of the activity follows a flipped classroom approach, where the readings were assigned as homework and the class time has the instructor demonstrating the social analytics process, and the class subsequently repeating the lesson on their computers. Students were then given the opportunity to build new analyses related to their term projects or any topic which interested them to reinforce the lessons learned. Based on the learning reflection submissions from the class, the majority of students were able to identify the strengths and limitations of analyzing data generated from social media.

One goal of the curriculum was to demonstrate how sentiment valence and frequency relates to other data. We chose to examine the topic of movies that were released on Valentine's day 2016, in North America. Students can easily relate to this topic, which makes it interesting to them, creating strong engagement with the exercise. We chose this topic as it allowed us to demonstrate a direct comparison between 3 movies using multiple approaches by looking at relative scores and quantity of critic and fan reviews on Rotten Tomatoes, box office sales reported by the-numbers.com, and data acquired using WASM from the weeks before and after the movies' opening. We limited the temporal period of analysis to limit the number of documents, which in our case was just over 90,000 with the parameters chosen. This dataset was created by using WASM using the topics "Zoolander 2", "Deadpool" and "How to be Single" while applying various synonyms, context items and exclude terms to improve accuracy. To go beyond simple sentiment analysis, we added themes based on Semantic Differentials for Hedonic Systems developed by Van der Heijden (2004). Data was then collected from Facebook, News, Twitter, Blogs, Videos, forums and reviews, in a process which took WASM approximately 15 minutes to complete. In the class, this dataset was saved in a shared project so students could simply dive right into analysis of the results. The social analysis had multiple facets including the examination of temporal, geographic, source, thematic, demographic and influencer angles which revealed a very rich story about what was happening during this time. The sentiment valence and frequency ratios of the three movies

in the WASM analysis also were found to be consistent with the Rotten Tomatoes and box office sales ratios. We discuss several recommendations to faculty interested in teaching social analytics for business using WASM based on our experiences.

Keywords: Watson Analytics for Social Media, Business Curriculum, Sentiment Analysis

---

# Designing and Delivering Big Data & Analytics Course Content – Challenges and Lessons Learned

Murthy Rallapalli
Colorado State University & IBM, USA
murthy.rallapalli@ibm.com

**Abstract**

Introducing Big Data & Analytics based solutions is a topic that is gaining high traction across the corporations in the United States. With the advent of globalization and real-time data multiplication, the complexity is only getting bigger by the day. One of the major challenges for the companies is to find the right skills, as universities around the world are gearing up to the challenge of introducing Big data courses as part of curriculum to build next generation of students equipped with Big data skills (both theoretical as well as hands-on).

Having designed and delivered several Big data courses for academic institutions as well as corporations, this paper will address the lessons learned and challenges along the way with a particular focus on the hands-on lab activities. The paper will address areas such as - how to provide appropriate hands-on activities on Big data platforms like Hadoop, IBM Big Insights and IBM Watson Platform. Paper will also cover defining the pre-requisites, equipping and obtaining required licenses with limited budgets, and finally evaluating the effectiveness of these courses at the work place. The goal of this paper is to provide all the lessons learned in the Big Data journey so it can help other faculty members and universities trying to create similar Big data based courses and labs in their environments. In addition, it will touch upon the details of how IBM BlueMix environment was utilized for Big data hands-on activities.

Keywords: Big Data, BlueMix, Big Insights, Analytics

# Earthquake activities in Japan with Watson Analytics

Ryo Adachi
Concordia International School Hanoi, Viet Nam
ryo.adachi@concordiahanoi.org

Peter Tong
Concordia International School Shanghai, China
general@petertong.com

**Abstract**

Earthquakes in Japan are a big issue, since they have caused several huge disasters and deaths. The Great East Japan Earthquake in 2011 and the nuclear disaster in Fukushima are very well-known disasters in the world and are still talked about by many people. This big data project undertakes the data analysis of earthquake data to seek possibilities of predicting earthquake activities in Japan using Watson Analytics. Over 85,000 data points were collected for analysis on earthquake activities that occurred in Japan for the past 40 years, from the earthquake database of the Japan Meteorological Agency. Data information on time, location, magnitude, seismic intensity and depth were collected for this project. Several insights were made known by datamining through these data sets. A relationship between the number of earthquakes and the seismic intensity was found, which can be represented using a mathematical expression. This was attained by using Watson Analytics to graph the relationship between the number of earthquakes and their seismic intensity. A trend on the number of earthquakes that occurred over a time period was also found, which is a step closer to making predictions on the future earthquakes. Another finding is the trend of earthquakes with locations, where the locations of earthquakes were mapped according to their frequencies of occurrences and their seismic intensities. This was done using Google spreadsheets and Google maps, importing the dataset from Google spreadsheets to Google maps for each seismic intensity. The outcome of this research allows predictions on how likely certain locations in Japan that are more likely to experience some greater as well as smaller earthquake activities. Overall, this project marked some important progress on prediction of future earthquakes in Japan.

Keyword: Watson Analytics, Earthquake, Big Data, Data Mining

# Ethics, Trust and Governance in Higher Education Curricula for Data Science and Analytics

Richard Self
University of Derby, United Kingdom
r.j.self@derby.ac.uk

**Abstract**

Data Science, Big Data Analytics and Machine Learning have become highly fashionable terms in business. The field figures strongly on the Gartner Hype curve presentations each year with many new terms entering the Hype Curve each year. There are a wide range of business and academic conferences each year extoling the benefits of all the latest techniques and products. There are regular reports documenting the need for ever more graduates to support all sectors.

The Higher Education sector has responded to the rapidly rising demand for graduates in Data Science and Analytics with a proliferation of courses. A significant majority of the undergraduate and post graduate courses are highly technical. However, many organisations are demonstrating through their behaviour and achievements that technical skills are not necessarily all that is needed to gain the best value from Data Science and Big Data Analytics. It is clear that undergraduates and postgraduate students need to be exposed to a range of soft skills and an understanding of the Ethics, Trust and Governance issues that result from the use and analysis of Big Data. Unfortunately, reviewing the course content of many UK based undergraduate and postgraduate courses demonstrates that this critical aspect is not covered. A small number have small elements of ethics included. The University of Derby is almost unique in ensuring that Ethics, Trust and Governance form a significant aspect of the curriculum.

This presentation will provide the justification for the need for Ethics, Trust and Governance to be firmly embedded in the curriculum. It will outline the approach and learning outcomes and assessment approaches that are employed. As a result, the graduates of University of Derby Data Science and Big Data Analytics programmes have a sound grasp of the importance of and the frameworks necessary to ensure effective governance of Big Data projects. Key aspects driving the need for this curriculum relate to the following factors:-

1) Over 60% of all Big Data Analytics projects fail to deliver the intended business value
2) There are many projects where the analytics that is carried out is of highly questionable ethics
3) The data veracity problem, identified by J Easton, leads to important governance questions about the reliability of the analytics and the use to which they are put. It is particularly important to identify the sensitivity of decision making to inaccurate data that cannot be reliably identified as correct or incorrect.
4) The new EU GRPR regulation requires “algorithmic transparency” for analytics that affect EU Citizens’ identifiable and sensitive data. This has very important consequences on the use of many forms of AI.

Examples of each of the above drivers will be discussed as an introduction to the curricula which the University of Derby have developed.

*Keyword: Ethics, Trust, Governance, Higher Education, Curriculum*

# Visualizing Story Telling with Big Data

Peter Tong
Concordia International School Shanghai, China
general@petertong.com

Craig Gingerich
Concordia International School Shanghai, China
craig.gingerich@concordiashanghai.org

**Abstract**

Concordia International School Shanghai has been offering the big data analytics courses at the high school and the middle school. These courses have exposed students to the conceptual understandings of big data analytics and have given these young minds an intellectual curiosity for more applications of data analytics. To harvest insights from raw data, analytics tools have to be applied to data mine the information that is hidden within the colossal amounts of data. One challenge to applying data analytics to K-12 education is in presentation; how are the complex analytics processes presented in a simplified manner while not losing the essence of data analytics and on the output of the analyses. Additional challenges are how to get the data to tell a story? What suitable analytics tools can be employed at the K-12 level? How is data analysis presented to a general audience with limited programming skills other than to simply have the data to tell a story? The answer to these challenges are in cognitive computing and interactive visual graphs and charts! Through cognitive computing, where natural language processing is used to mimic how the brain works, analytics can be made simple. Through interactive visualization, mined data can be cleaned and visualized in entangled interactive ways to create meaning. The power of storytelling and the way to make data conveys a visual that is more than words, sentences and paragraphs.

Through the collection of data students will create effective graphs to tell their next story. Storytelling is not an inherent skill; it is our 'human currency'. When it comes to data visualization, and the many various tools at our disposal on the Internet, students can go beyond conventional tools to use data to create an engaging, informative, compelling story. Students will integrate their data and think like a designer and utilize concepts of design to create strong visualizations. Through this short course, students will learn how to leverage the power of storytelling to help tell their story so it will resonate with their audience. Two tools that provide for this are being used at Concordia International School. One such analytics tool is Watson Analytics, a data analytics tool that coupled with data visualization is made easy for any users. The school is piloting the use of Watson Analytics to 9th and 10th grader during an exploration term which consists of 10 days of 2 hours each. In this pilot course students are exposed to the use of Watson Analytics using short online modules from Big Data University. Students then apply Watson Analytics to data mining real life practical datasets to story tell with big data. Tableau is also being used, but in the lower grades. Students in the 7th and 8th grades just completed a semester long pilot elective in which Tableau, a visualization program, allowed students to interact with raw data to find correlations and meaning. Tableau allows for interactive visualization and the students do the analysis. Indeed, both of these platforms have the potential to "let the data speak" in multiple ways, at multiple levels, and ignite a true passion for all things data.

Keywords: Big Data, Data visualization, Watson Analytics, Tableau

# Analytics Education: Curriculum Development, Practical Assessment and Best Practices

William Yeoh
Department of Information Systems and Business Analytics
Faculty of Business and Law
Deakin University
70 Elgar Road, Burwood, VIC 3125 Australia
william.yeoh@deakin.edu.au

## Abstract

This abstract presentation aims to introduce the contemporary analytics curriculum development, certification programs, available teaching and learning resources, practical assessment, as well as the technical support provided by IBM for university and academics. The presentation will also share 'best-practice' teaching experiences and go through sample assignments that apply IBM Watson Analytics suite. The insights will help educators to adapt the Watson analytics suite and associated curriculum into business analytics or data science teaching programs.

Keywords: Analytics Education, Curriculum Development, Practical Assessment, Best Practices, Certification Programs

# Examining the Relationship Between Economic Activity and Environmental Impact

Ali Kazerani, Sarah Lahanky, Kayla Sabharwal
Dalhousie University, Halifax, Nova Scotia, Canada
m.bliemel@dal.ca

Presentation Link: https://www.youtube.com/watch?v=V8gXhS0FEaY&feature=youtu.be

This report examines the relationship between economic activity and environmental impact across 43 countries. Human activity, reflected in economic activity is responsible for the climate change affects that have occurred, credited mainly to the use and burning of fossil fuels. As the global economy continues to industrialize and develop, environmental impact has become an increasingly important topic and has resulted in global investment in renewable energies. The purpose of this study is to determine how much of an environmental impact higher economic activity implies and to demonstrate how the environmental impact of nations has changed over the last 20 years and in which direction they are moving towards in the future. It is hypothesized that countries with higher economic activity (GDP per capita) will have a larger negative impact on the environment. It is also believed that countries with higher usage of renewable energy will have higher GDPs per capita, due to the large investment required to transition a nation to renewable energy sources.

After the required data was sourced and uploaded to Watson, the Refine, Explore, Predict, Social Media, and Assemble functions were utilized to enhance each stage of the analysis. All datasets were exported from the OECD statistics website as CSV files, converted into Microsoft Excel files, and then combined into one data sheet on Excel. 5 data sources were used during this analysis, which assessed the 35 OECD countries and the 8 non-OECD member countries that are part of the top 20 CO2 emitting nations, all from 1994-2014. GDP per capita was used as the indicator for economic activity. The four indicators for environmental impact were GHG emissions, usage of renewable energy, crude oil production, and forest resources (deforestation).

The results from the Explore function on Watson indicate that GDP per capita and GHG emissions are not strongly correlated, however GDP per capita and usage of renewable energy are. Crude oil production and usage of renewable energy has been growing globally at the same rate. Watson's Predict function highlighted that GDP per capita and GHG emissions together were strong predictors of usage of renewable energy. The social media sentiment analysis performed on The Paris Agreement and Carbon Taxes revealed an overall global positive sentiment for both topics, indicating strong support for environmental protection.

Limitations of this study include missing data points which left incomplete data sets for certain indicators, as well as the risk that the four environmental indicators used do not provide a comprehensive enough picture of environmental impact. A key business insight gained from this study are that globally, there is increasing focus on renewable energies and that more developed economies are choosing to invest. This implies that the renewable energy sector will continue to grow rapidly offering new business, research, and investment opportunities. Future research to be done includes analyzing trends in the proportion or percentage of renewable energy used across countries, as well as an analysis of environmental policies.

# Climate and Landscape Change Effects on Lyme Disease in the United States

Tracy Carr, Elizabeth Handy, Abebaw Zeleke
University of Maryland University College
elena.gortcheva@umuc.edu

Presentation Link: https://youtu.be/lNyjaDJC_7g

This project will explore the relationship between key climate change indicators and reported cases of Lyme disease in the United States. Lyme disease is an illness transmitted by certain species of infected ticks in the United States that live on host animals and spread the disease through bites (EPA, 2016b). The incidence of Lyme cases in the United States has nearly doubled since 1991, from 3.74 per 100,000 people to 7.75, and it is the most common vector-borne illness in the United States (EPA, 2016b). Cases are currently still limited to specific geographic regions within the United States, primarily the Mid-Atlantic, New England, and North Central regions, however the volume and spread of those cases has grown over time.

The symptoms of Lyme disease include fever, headache, fatigue, and a specific type of skin rash, and if identified quickly, it can be treated within a few weeks with antibiotics (CDC, 2016). Left untreated, however, it can spread to joints, the heart, and the nervous system, which can lead to arthritis, facial palsy, heart palpitations, inflammation of the brain and spinal cord, and problems with short-term memory, among many other complications (CDC, 2016).

Climate change has become a factor in the transmission and incidence rate of Lyme disease because of its likely influence on the survival of the tick population in a growing land area of the United States. Studies have previously linked high and low temperature extremes and precipitation patterns to the prevalence of vector-borne diseases, or those carried by insects such as ticks and mosquitos (Beard et al., 2016). However other factors may contribute to the rise in this disease – environmental factors like changing ecosystems and landscapes, human behavior like time spent outside and a coordinated response to the disease, and changes in the ticks and bacteria causing the disease itself (Beard et al., 2016). Another study attempted to analyze the density of blacklegged tick populations, and found that in a limited studied area, tick density tracks with the rise in Lyme disease cases in that area, and found correlations with tick density and similar environmental indicators such as mild temperatures, low precipitation, low forest cover, and high urbanization (Khatchikian et al., 2012). Monitoring the population of ticks, however, is a challenging endeavor, and there are no largescale data sources tracking the population of ticks over time in the United States.

This project will search for any link between the known linked indicators of temperature and precipitation, additional climate change indicators, and variables measuring our changing landscape. Due to the challenges in documenting tick population, these indicators will hope to inform the spread of Lyme disease without ticks as a specific variable.

The team will develop a dashboard in Watson Analytics exploring the role of climate indicators and reported cases of Lyme disease, including data refinement, exploration, predictive modeling, and social media analytics. The findings will be presented in an interactive dashboard in Watson Analytics in addition to this report.

The benefits of this study would be to further shine light on the spread of this disease and to inform public health officials of its likely path going forward based on climate indicators. The Environmental Protection Agency has said that further study of this indicator, the incidence of Lyme disease, is “critical” for informing decisions affecting public health (EPA, 2016b). If indicators point to a continued increase and geographic spread, public health officials may want to resource prevention and treatment responses appropriately.

----- End-----