

URBAN DATA SCIENCE EDUCATION: A KEY ACTOR TOWARDS IMPROVING DATA-DRIVEN POLICY-MAKING FOR SOLVING URBAN PROBLEMS

1

Abstract

In this paper, we explained the benefits of Urban Data Science Education in improving data driving policy-making and highlight the impact of vital use of available open data by using a practical demonstration from the Summer-Winter School (SWS) at CEPT University, Ahmedabad, India. Good Urban data policies can be an important driver for smart city research and the implementation of good data management practices. Many authors are also becoming interested in the data revolution that is enhancing the way we study and understand cities. It now becomes appropriate to respond to this need in an evidence-informed manner by working with stakeholders that will encourage a more resilient approach via education for the younger enthusiast, leading to a more transparent, improved, sustainable and safer urban cities. Although the major challenges we foresee is the inability to utilize the large amount of urban data generated daily, with the drive to mobilize more education towards urban data science and analytics, generated data can be effectively utilized to make informed and evidence-based decisions. This paper points to an effective methodology for reproducibility, the key steps taken, as it views the inclusion of an Urban Data Science curriculum as priceless, and a decision that will bring into notice the importance of engaging the younger generation in scholastic learning on data principles and analytics. Data along with spatial enhancement can be the driving factor towards improving data-driven policy-making for solving urban problems in smart cities.

KEYWORDS: Urban Data Science, Data Analysis, Policy-Making, Open Data, Smart Cities

1. Introduction

Data is the lifeblood of decision-making and the raw material for accountability (UN, 2019). There has never been a more exciting time to be involved in the utilization of data, as our world today is transformed by data every day. We have in our disposal large unexplored data with such a capacity and potential to transform our living status in many different ways (Lee & Kang, 2015; MacLaughlin, 2017). And with the progress in innovation and technology, especially in the recent developments in big data analytics, a large amount of these data is collected daily and analyzed; using both the traditional data analysis method and a more sophisticated form of data science analytics (Jin, Wah, Cheng & Wang, 2015; Larson & Chang, 2016; Kambies, Roma, Mittal & Sharma, 2017). Since the end of the Millennium Development Goals (MDGs) in 2000, the world has seen an increased uptake in the use of data and technology to improve people's lives (Archenaa & Anita, 2015; Wang, Kung & Byrd, 2018).

The availability of a large amount of open data in the 21st century has made Data analytics a highly sought after skill (Ahmed, Capretz, Bouktif & Campbell, 2015; Börner, Maltese, Balliet & Heimlich, 2016; Jang, 2016; Stevens & Norman, 2016). For example, Google holds about 10-15 exabytes of data which translates to 30 Million PCs of 500GB combined (Heshmore, 2017). While data analytic skills are sort after at an exponential rate by many data-driven organizations, not many have learned to utilize its potential (Wang, Kung & Byrd, 2018; Petty, 2019). Even in many governmental organizations, the promise of a data-driven and evidence-based decision making has failed to materialize because most public institutions lack the capacity to effectively utilize these data, and so the progress in turning data into actionable insights is halted (Esty & Rushing, 2007; Bradford, 2019; Hastings et al. 2019). Also, in many developing economies, policy-making is still been driven by personal sentiment, experience, opinion or belief with little empirical inputs; and only very small proportion of government organizations are scratching the surfaces in the use of data and it's technologies in innovative ways (Almeida & Báscolo, 2006; Harris, 2015; Head, 2016).

A data-driven approach involves the methodical and orderly way of collecting, collating, managing, analyzing, interpreting, and applying data in a way to provide insight that could help a social or urban problem (Kitchin, R. (2017). These data can be open source or collectively from a government or private organization database. And if these data can be effectively and efficiently utilized, grooming enough talent to fill the demand for Data utilization through data science education becomes critical.

Every data-driven solution if effectively maximized can solve varying urban problems. A good example is observed in the United States where a Pred Pol software leverages on historical data sets using Data Analytics and Machine learning to pre-emptively prevent crimes from occurring, thereby allowing the local police departments across the cities to

more effectively deploy its resources (PredPol., 2015; Capotosto, 2017). A study conducted by Pokhriyal & Jacques (2017), shows where several sources of data are combined for better diagnosis and poverty mapping for improved poverty prediction and mapping. A study conducted in Guatemala and Honduras by Adelman et al (2018) shows how Administrative data is used to predict school drop out.

With the foreseen benefits in Urban Data Science education, the present course we proposed is to complement this drive of data explosion and its benefits in the contemporary world in creating the possibility of bringing data into reality through sound empirical research and analysis, thus minimizing uncertainty towards the best course of action in a particular topic in policy design. The paper, therefore, aims to show that *Urban data science education can both help to solve various urban problems by enhancing participants' data skills in other to nourish the culture of data science.*

2. The Urban Data Revolution for Cities Research

In the last decade, there has been an explosion of city and urban life data (Townsend, 2015; Kharrazi, Qin & Zhang, 2016). According to the United Nations, “By some estimates, 90% of the data in the world has been created in the last two years, and it is projected to increase by 40% annually; simply because data is being gathered by inexpensive and numerous information-sensing, mobile devices, which encompasses the open data movement, the rise of crowdsourcing, new ICTs for data collection, and the explosion in the availability of big data, together with the emergence of Artificial Intelligence (AI) and the Internet of Things (IOT)” (UN, 2019). In addition, many authors are becoming interested in the data revolution that is enhancing the way we study and understand cities (Batty, 2016; Kitchin, 2016; Barkham, Bokhari & Saiz, 2018; Glaeser et al. 2018), often referred to as “Urbanism” (Kitchin, 2014; Williamson, 2015; Leszczynski, 2016). These new forms of generated data can range from social media tweets via mobile phones, monitor digital camera feeds (e.g: Video vehicle detection), GPS, transponders, transportation (location/movement), and mapping surveys that monitor city operating systems. These operations are enabling us to tag locations and coordinate movements using location-based services (LBSs) (Wilson, 2012; Zhou et al 2019), thereby changing how we plan and govern our cities. For example, location and movement data generated by travel via the intelligent transportation system (ITS) can be used to create and improve models and simulations to guide future urban development – thereby helping to decongest some areas while others are improved to stimulate land values (Brennand et al. 2015; Nellore & Hancke, 2016). This can be a very effective way of solving traffic congestion for most cities (Singh & Gupta, 2015; de Souza et al. 2016). This daily transformation is enabling data-driving planning for the cities. It now becomes appropriate to respond to this need in an evidence-informed manner by working with stakeholders that will encourage a more resilient approach via education for the younger enthusiast, leading to a more transparent, improved, sustainable and safer urban cities. The major challenges we foresee are the

inability to utilize the large amount of urban data generated daily; and therefore the need to mobilize more education towards urban data science and analytics, in other to link these data together to make informed and evidence-based decisions under smart city movement.

3. Methodology and Curricula

The growing urban cities are increasing demands for evidence and the use of data, making Urban Data Science education a more prevalent approach. And while most governments and institutional leaders are recognizing the need to provide education in data, interpretation towards sound training on data-driven decision making (DDDM) is lacking (Marsh & Farrell, 2015; Datnow & Hubbard, 2016). With this need in focus, we organized a Summer School on Urban Data Science to emphasize the need, potential, and importance of an institutionally based program in establishing the knowledge and skills needed to enhance data science practices in governance and policymaking. The Summer Winter School (SWS) at CEPT University, Ahmedabad, India was the primary host of the program.

Two cycles of students have undertaken the course, with the first cycle in 2018 with 22 students and the second cycle in 2019 with 26 students, making a total of 48 students, who were from different disciplines like planning, architecture, civil engineering, and geomatics. Though from different disciplines, they all have one thing in common: learning about how data analytics skills can help in formulating effective policies for decision making in the different urban sectors.

and also with the goal that exposing them to the availability of new sources of data and the emergence of new data science technologies will propel an inner drive for making use of such data towards contributing and impacting significantly on public institutions. Student's enrollments in the course were both from the bachelor's and master's levels involving technical and non-technical fields. The summer school provided first-hand practical experience to students to get started.

The school adopted a simple methodology that guided the students in formulating their case studies. The process began with the students learning how to search for a research topic of interest, formulating the topic into a research title, recognizing the relevance of the topic by improving the research objectives, searching for a relevant open-source data in line with their study objectives chosen, understanding the structure, format, and dimensions of the open-source data, and if there are any flaws in the data which will require cleaning the data sets, making graphs and charts for the exploratory phase of the data analysis and then analyzing the data with the aid of the Statistical software package "R", which was the primary statistical and visualization tool used for the entire duration of the course (<https://www.r-project.org/>) as shown in **Figure 1**.

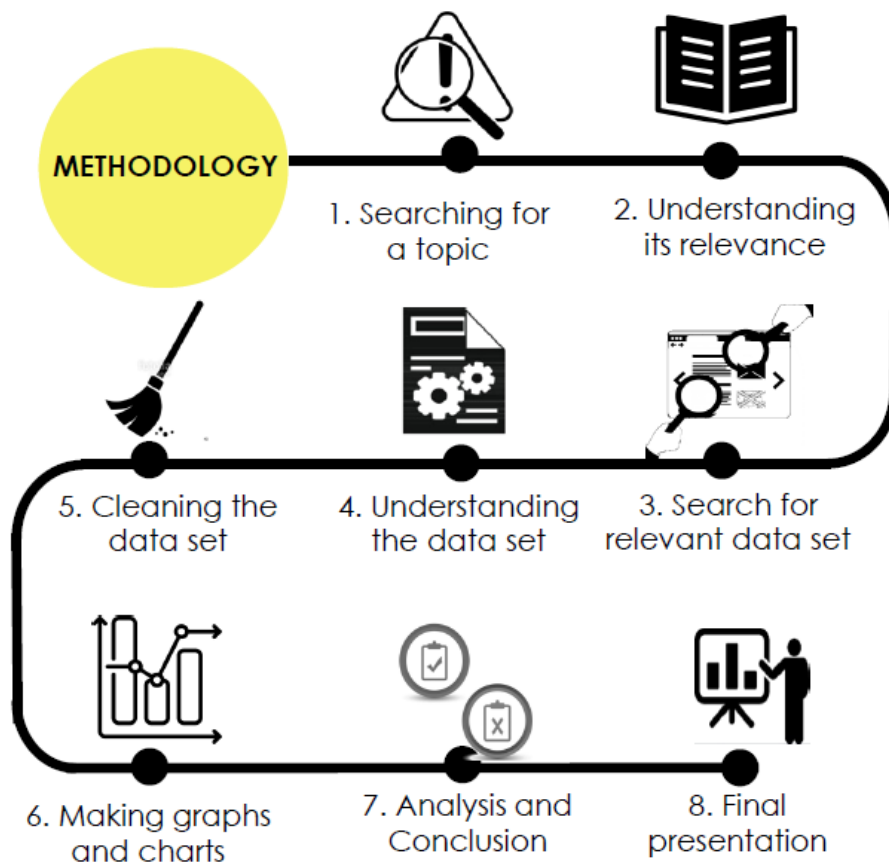


Figure 1: Methodology adopted by students for their case studies

The geographic information system (GIS) was also incorporated into the study modules in order to help students formulate a system designed to capture, store, manipulate, analyze, manage, and present spatial and geographic data (<http://www.arcgis.com/index.html>). Students were also exposed to various open-source software for data cleaning like OpenRefine. An introduction to Git and GitHub was given to students with a proper understanding of the importance of data sharing and various licenses.

The course was taught keeping in mind the heterogeneity of the students. A perfect blend of theory and practical's were planned with relevance to each topic. The course was intensive and to keep every student up to speed at the same pace in the classwork or exercises, students' feedback was taken daily which helped in the enhancement of the class activity and decisions taken by the tutors for teaching the next concept. The course was modified daily with more group activities and practicals based on students' feedbacks received as shown in **Figure 2**.

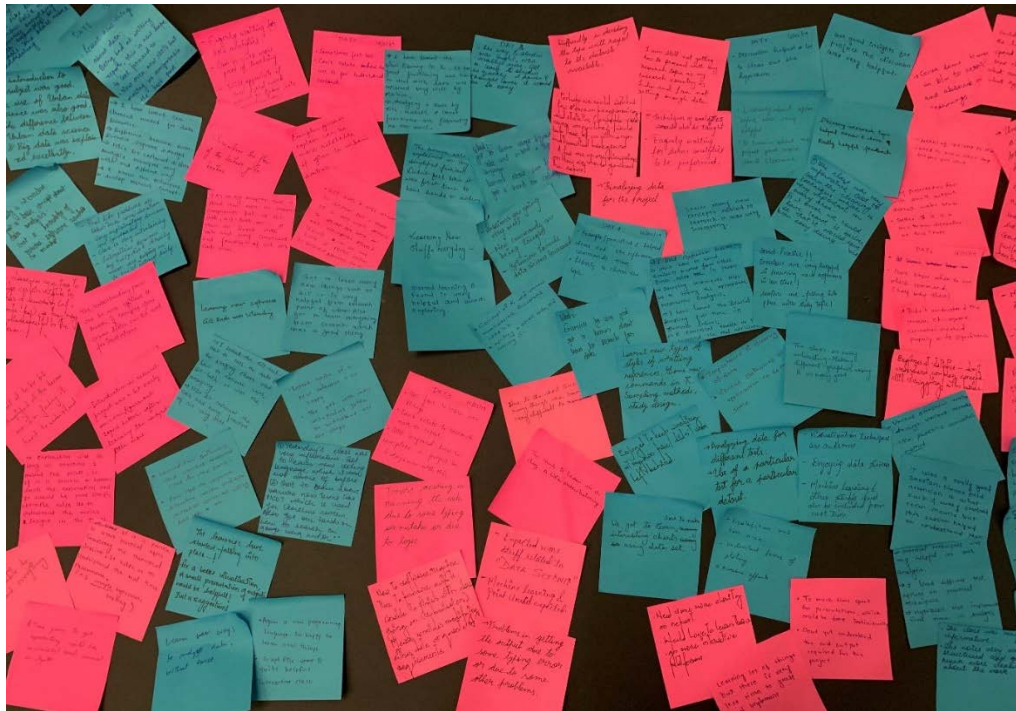


Figure 2: Student Feedback in Pink and Blue Post-it notes

This process taking regular feedback from the students and understanding the capability of the students was well appreciated. Two faculties were present in class which helped in giving individual attention to the students.

General information of the students selected responses to a post feedback survey and also students' major strengths of the Summer School were captured in an online survey during the course duration in other to improve the course methodology from their feedbacks as shown in **Tables 1-3**.

Table 1: General Information of Participants

Characteristics	Percentage (%)
Participants level of Education	
Post Graduate	70.83
Under Graduate	29.17
Gender	
Male	53.85
Female	46.15
Operating System used by participants	
Windows	96.15
Apple/Mac OS/UNIX	3.85
How often do participants use a Statistical Software	
Never	84.62
Less than once per year	11.54
Monthly	3.85
How often do participants use a Programming Software	
Never	65.99
Weekly/Daily	14.78
Monthly	11.54
Less than once per year	7.69
Participants satisfaction rating of the level of current data management and analysis skills before summer school	
Unsatisfied	30.77
Neutral	46.15
Satisfied	7.69
Not sure	15.53
Reason for Participating in the Summer School	
To learn new skills that I can apply to my current work	40.88
To learn new skills that will help me get a job or a promotion	26.92
To learn skills that I can apply to my work in the future	24.5
As a requirement for my program	3.85
To refresh or review my skills	3.85

As shown in Table 1, the distribution of the class was fairly equal among male and female students (53.85% vs. 46.15%). The class had more post-graduate than undergraduate students (70.83% vs. 29.19%). The majority of the students (96.15%) used Windows operating system for the class exercises. Around 84.62% of the participants had never used statistical software while 65.99% of the students had never used programming software. This gave the instructors a good idea about the level of the students. About 30.77% of the students were unsatisfied with the level of their current data management and analytics skills before the summer school while 46.15% were neutral. Only a minor proportion (7.69%) said they were satisfied with the level of their current data management and analytics skills, and 15.53% unsure. The most selected reason for participating in the

summer school was to learn new skills that can be applied to current work (40.88%) and to help get a job or promotion (26.92%).

Table 2: Selected responses to the post feedback survey

Characteristics	Percentage (%)
I felt comfortable learning in this summer school environment	
Strongly agree	41.67
Agree	50.0
Neutral	8.33
I can immediately apply what I learned at this summer school	
Strongly agree	29.17
Agree	50.0
Neutral	16.67
Disagree	4.17
I was able to get clear answers to my questions from the instructors	
Strongly agree	41.67
Agree	50.00
Neutral	4.17
Disagree	4.17
The instructors were enthusiastic about the summer school	
Strongly agree	83.33
Agree	8.33
Neutral	8.33
I felt comfortable interacting with the instructors	
Strongly agree	83.33
Agree	12.50
Neutral	4.17
The instructors were knowledgeable about the material being taught	
Strongly agree	83.33
Agree	16.67

As shown in Table 2, it was very satisfactory to see the results of the post-summer school feedback which observed that 91.67% of the students were comfortable learning in the summer school and 79.17% agreed that they could immediately apply the knowledge gained in this summer school to their respective study. The majority of the students (91.67%) said they were able to get clear answers to their questions from the instructors, and 95.83% said they felt comfortable interacting with the instructors. All of the participants (100.0%) agreed that the instructors were knowledgeable about the material being taught.

Table 3: Students feedback on the major strength of the Summer School translated verbatim

Responses	Percentage (%)
The Big field of data analysis and statistics is explained at the cellular level making it easy to grasp	20.04
The syllabus was broad and the course structure and study materials were well written with access to all the important concepts which can be used further on my own	19.04
Now I can manage, create and analyze larger data-sets in Excel and R	14.28
Impressive depth of knowledge of wide topics covered by the instructors	9.2
Interaction with the instructors by the students and the attention paid by the instructors to the students was great	9.2
The analysis and the capacity of handling the data by the software is good using a step by step instruction process by the instructors	9.2
Learnt programming and statistics that I can implement on my projects as well as job criterion	4.76
Misconceptions about Big Data, Machine Learning and excels were cleared	4.76
Smooth and to the point learning. A lot was taught and I didn't get confused.	4.76
The summer school was designed in a very well planned manner. I am amazed to see how easy it has become for me to work on R Studio. The course gave me so much information about how a paper should be written, how can data be cleaned easily for the first time	4.76

Table 3 reflected on the students' feedback on the Major strength of the Summer School. The most were that the Big field of data analysis and statistics is explained at the cellular level making it easy to grasp (20.04%) followed by the syllabus was broad and the course structure and study materials were well written with access to all the important concepts which can be used further on their own (19.04%).

From the students' feedback, it was clear that students understood open data and data sharing policy, which was reflected in the variety of projects chosen. Also, from the students' feedback, it was concluded that they learned new terminology and it added to a better understanding of the importance of Urban Data Science.

4. The Value

The outputs of the case studies were spread in varying domains of Crime, the Economy, Education, Environment, Public Health, Road Accident, Sports, Governance and Planning, etc. The Word Cloud shown below depicts the various topics covered by the students from the 2018 and 2019 batch; addressing vast domain of urban problems.



Figure 4: Topics selected by students for Case Study

One of the interesting studies presented by the student of the summer school was on the Prevalence of Obesity among Socially Vulnerable Groups in the United States (see Figure 5). Heart disease, which has been linked to obesity is one of the leading causes of death in the United States (Rao et al., 2015; Rosenthal et al., 2017; Gonzalez-Campoy, 2019). And because it's a consideration as to whether it is a disease or not is debatable, effective intervention is slow (Allison et al., 2008; Stoner, & Cornwall, 2014; Kyle, Dhurandhar & Allison, 2016). The opensource dataset used for the study was the National Health and Nutrition Examination Survey (NHANES) from the year 2013-14 (kaggle.com). The result from Pearson's correlational coefficient (r) showed a weak but significant negative correlation between annual household income and BMI (see Figure 6). It can be assumed that as annual household income increases, the BMI decreases. This may be due to being able to afford quality healthy food and access to healthcare.

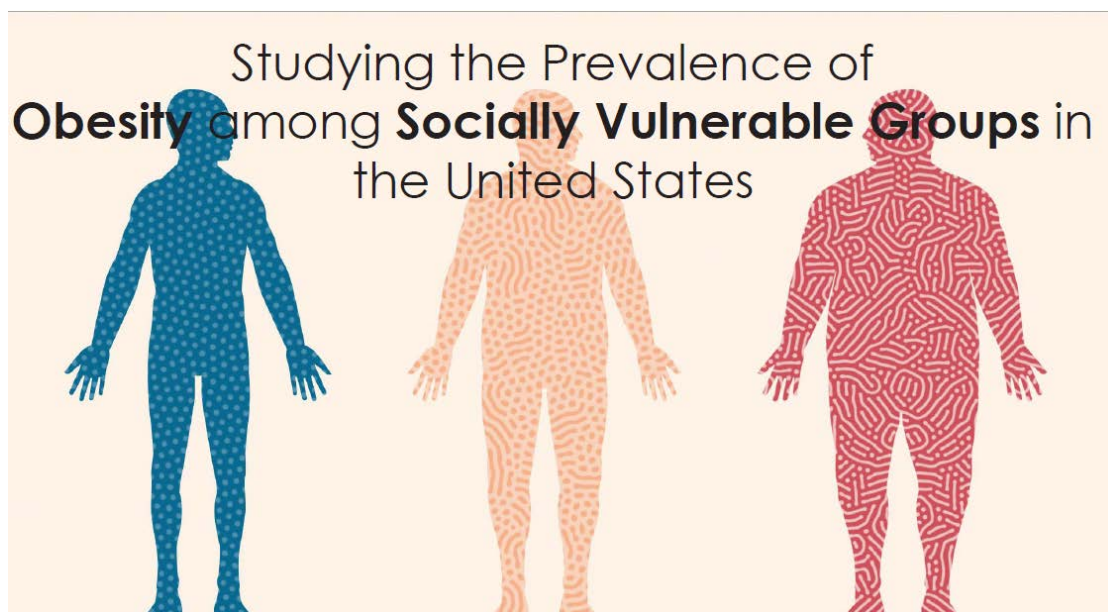


Figure 5: Project cover page for Prevalence of Obesity among Socially Vulnerable Groups in the United States

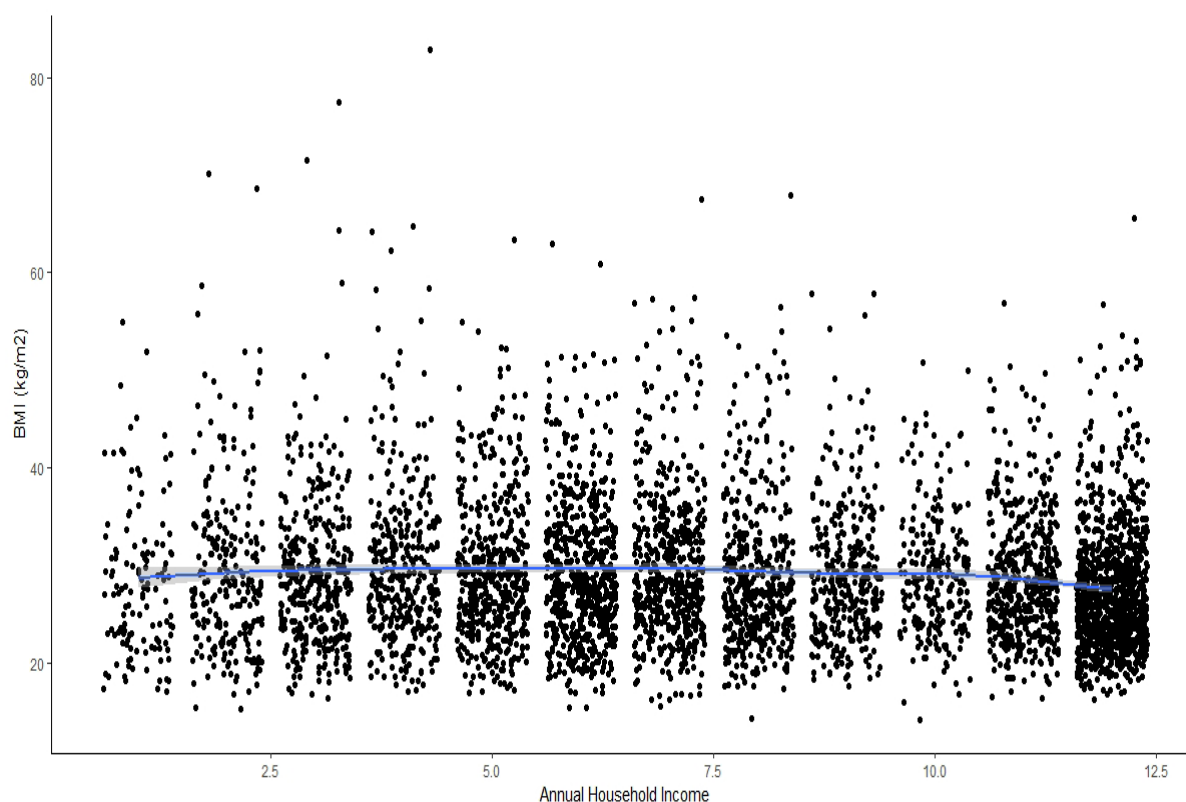


Figure 6: Scatterplot showing a weak positive correlation between annual household income and BMI

Cross-tabulations using a statistical test like Chi-Square, Pearsons Correlation, Kendall's rank correlation, ANOVA (Analysis of Variance) were performed on Race, Age, Income,

Education with its relation to Obesity. Interesting results were found proving that Obesity is prevalent among Ethnic/Racial minorities, and that socioeconomic, racial factors influence obesity in children and the elderly. The findings of the study supported that people from Low-income households having lower educational levels showed a tendency of being more vulnerable to obesity due to poor dietary options.

Recommendations were made from this study based on the findings which include: Food stamps, raw food materials could be provided for economically weaker households; The current school lunch program in the United States is not focused on providing a menu of healthy foods, and could be corrected; Unhealthy fast foods and sugary sodas should be taxed higher to discourage their popularity; Better medical care, especially the public medical care and insurance system in the United States needs to be strengthened and strong socialist reforms are needed in this sector to provide free healthcare for all.

A second student analyzed the performance of Indian states and union territories in terms of Sustainable Development Goals (SDG) for the year 2018. NITI Aayog has been entrusted with the role to co-ordinate 'Transforming our world: the 2030 Agenda for Sustainable Development' at the national level. From the study, evaluation is carried out at the country level as well as for each state and Union territory and a score is given between 0 to 100; data was collected from <https://niti.gov.in/>. The spatial distribution of the Scores is shown in Figure 7 which rightly indicates that none of the states have been able to achieve 100 scores to date. By analyzing each goal individually it helps in identifying the direction in which the government should take action. "Gender Equality" is still one of the alarming situations in India for all the states.

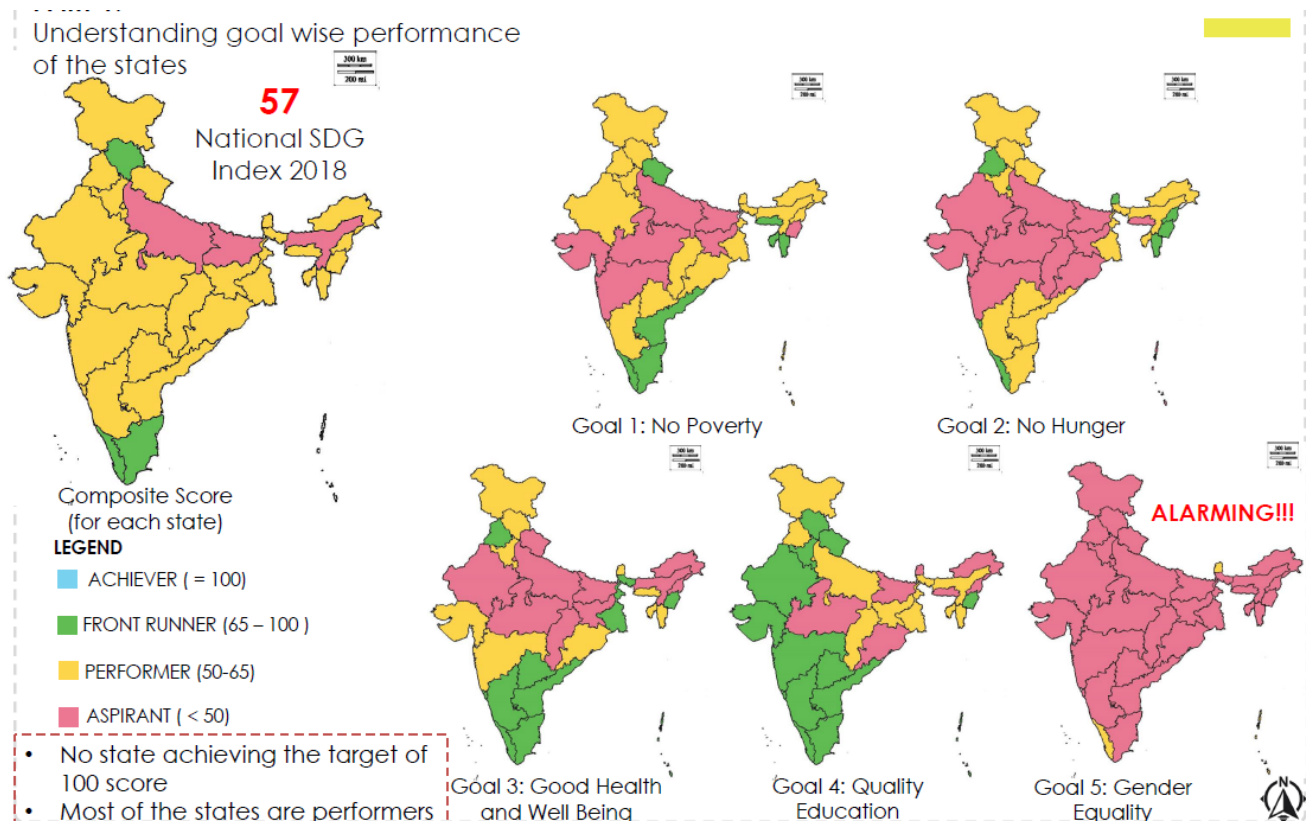


Figure 7: Understanding the Status of SDG goals 1-5 for the states of India

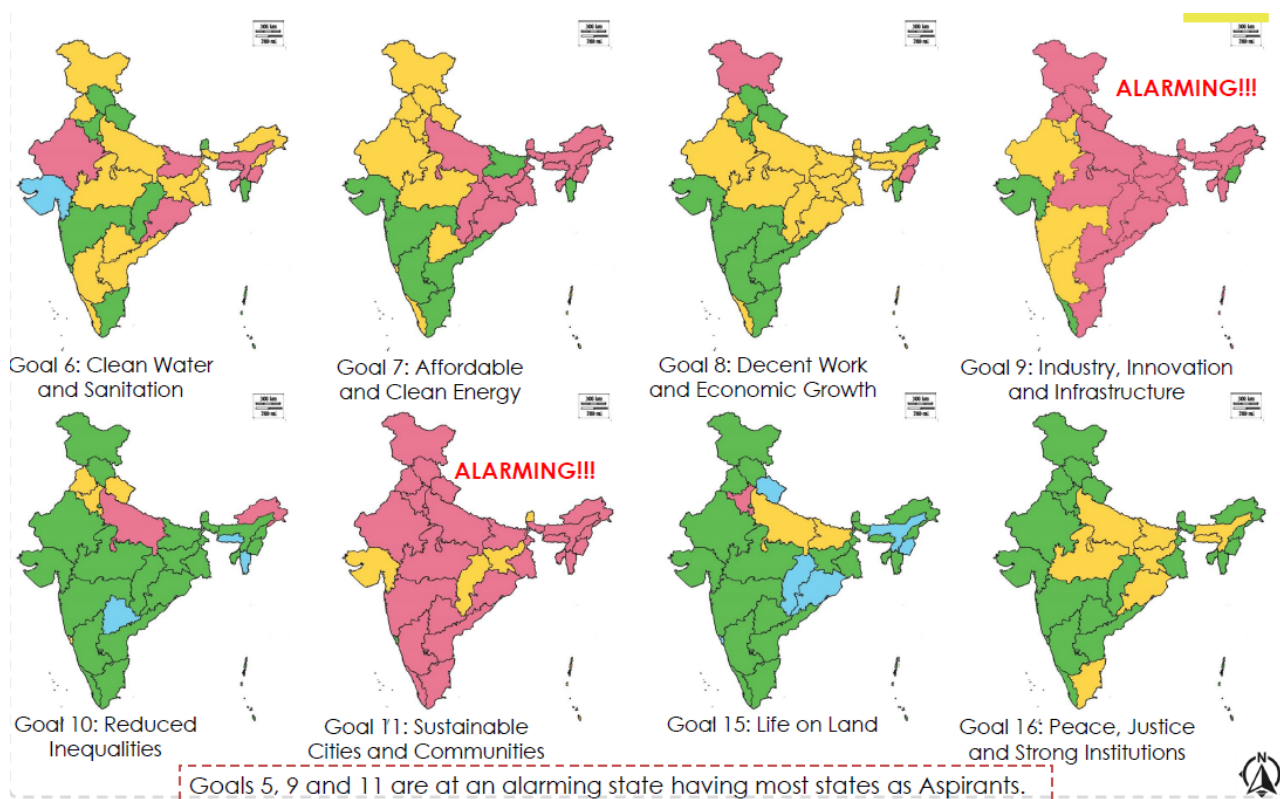


Figure 8: Understanding the Status of SDG goals 6-16 for the states of India

Figure 8 shows that for goals 6, 10, and 15 some of the states have reached Achievers level. It has been highlighted that there is an alarming situation for Goal 9 and 11 in most of the

states in India. Progress on SDG 12, 13, and 14 could not be measured because relevant state-level data could NOT be consolidated or found. SDG 17 was left out because the goal is focused on international partnerships, being less relevant for domestic level policy actions. After studying the individual states SDG scores, it was observed that Uttar Pradesh is having the lowest score i.e. 42, while Kerala and Himachal Pradesh are having the highest score (69). The range of scores is from 42-69. Hence, there is a tremendous scope of improvement to bring the National SDG Index to 100 before 2030. Various states can adopt successful models from other states to achieve every goal efficiently in less time. The recommendation from the analysis of this study is that political willingness via effective policy decision-making from the study findings could help towards enhancing Indian’s sustainable development targets.

Another case study from a student accessed the behavioural pattern of Terrorist Attacks in India. The study findings showed a high prevalence of attacks in two Indian cities: Jammu and Kashmir. The case findings can be a catalyst towards innovative strategy in the ministry of defense in India in preparation for future attacks. (see Figure 9).

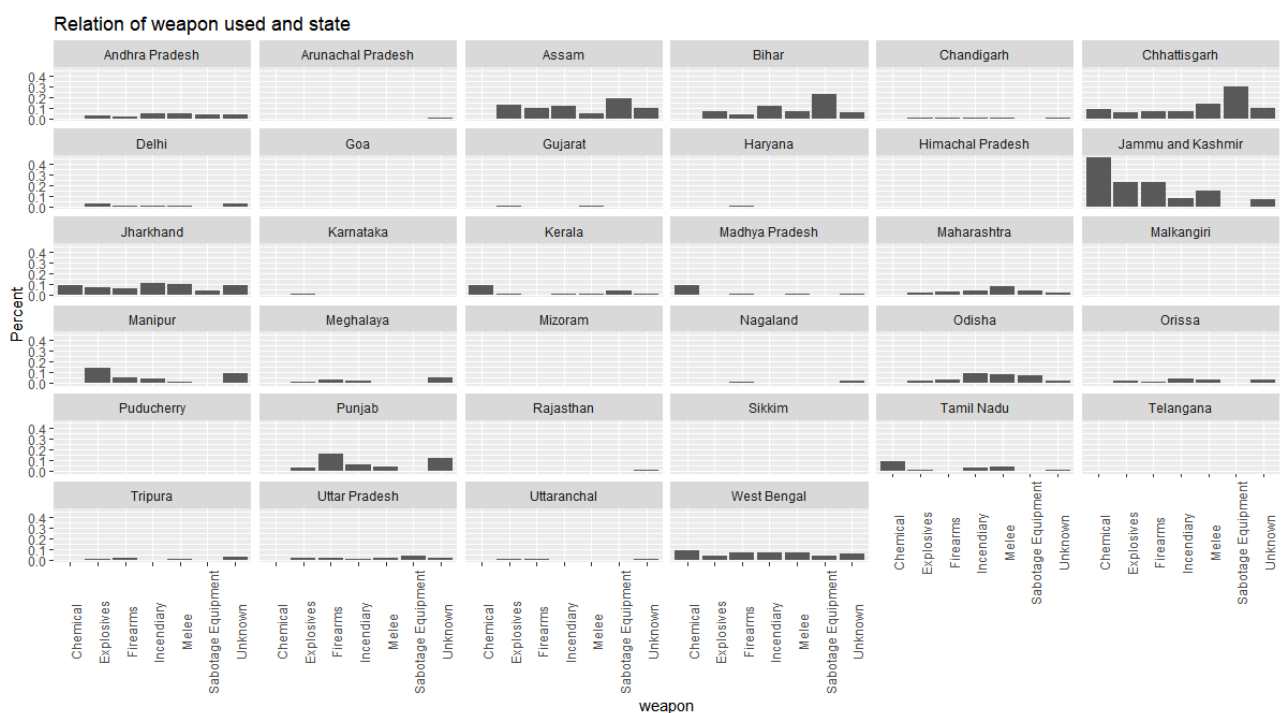


Figure 9: Relation of the weapon used in the terrorist attacks with the state

Many more interesting case studies were presented by the students, captured in Table 4.

Table 4: Student’s Research Titles for the Urban Data Science Summer School 2018 and 2019

Topic	Year	Title
Crime	2018	The crime rate in the City of Chicago
		Analyzing determinants of gun violence In the United States of America
		Understanding and interpreting crime trends in Greater London, England in 2016.
	2019	Crime Against Women in India and its impact on the economy of the Country
		Understanding the pattern of terrorist attacks in India
Disaster	2018	Vegetation fire in Cape Town
Economy	2018	Understanding the economic trend based on different sectors in the USA
		Suicides in India: A Case study
	2019	Impact of Foreign direct investment on the Human Development Index in the past 27 years
		Impact of Contraction and Expansion of Establishments towards Employment in the US
		Agriculture input versus the urban population and its effects
		Current Health expenditure per capita in the USA
		Food import value and volume changes with respect to time in the USA
		Comparison of Suicide Rate to GDP and HDI
Education	2018	Educational attainment for adults age 25 and older for the U.S., States, and counties, 1970 – 2016
	2019	Contribution of Education in the development of countries across the world
Food	2018	Homemade beer recipes
		Food Accessibility System of US Counties
Governance	2018	Accidental deaths & suicides in India
	2019	Comparing the number of reported confirmed cases to reported deaths
Governance and Planning	2019	Analyzing the performance of Indian states and union territories in terms of sustainable development goals for the year 2018
		Recognizing the disparity between ‘planned’ and ‘unplanned’ parts of the city and its effects on people’s choices; a case of Ahmedabad
Pollution	2018	Factors degrading air quality level in Gujarat.
		Analyzing Trends of Pollution: A Spatio-Temporal Study of US air pollution
		Air Quality for the Major Urban States of US: Focussed Study On Particulate Matter (PM 10, PM 2.5) from 2008 to 2017
	2019	Assessing policy need for Delhi (NCR): Comparing pollution levels in Delhi and Beijing
Public Health	2018	The relation of Demographic Parameters on mortality of Cancer incidents in the States of USA.
	2019	A study on the prevalence of obesity among Socially Vulnerable groups in the United States
		Correlation and interdependence of diseases across 500 cities in the USA
		Effect of socioeconomic parameters on suicidal deaths
		The Prevalence of HIV in the Year 2002-2011 Globally
		Social Demographic Outcomes Of Death Caused By Drugs In United States 2003-2017
Understanding the mortality rate and its relation to the health condition of children around the world		
Real Estate	2018	Predicting housing prices: a case for Seattle housing crisis, USA

	2019	Factors affecting Real Estate Prices in IOWA city, USA
Road Accident	2018	Mortality from land transport accidents
	2019	Road accident collision analysis and pedestrian safety study in Great Britain(UK)
		An assessment of the association of environmental factors and road fatalities in Britain.
Social Media	2018	YouTube videos
Sports	2018	Indian premier league matches
	2019	Comparison popularity of football player
Transportation	2018	Monitoring the trend of Cross border crossing vehicles in the USA
		The inflow of Foreign Tourists to India
		Traffic Violations in Montgomery County
		Statistical analysis of Annual average daily flow of traffic of east midlands, Britain

5. The future

While there are some concerns if the motivation for data science education is commercial rather than intellectual developments according to Donoho (2017), we cannot argue the impact it has on data-driven developments in scientific research (Clauzet, Larremore & Sinatra, 2017; Lytras, Raghavan & Damiani, 2017). In other words, there is a need for the further development of such courses in various institutions as the opportunities are rapidly growing and will continue to do so in years to come (Baumer, 2015). Urban data science is in its formative stages of development and will evolve among professionals and students desiring to shape the future of smart cities (Williamson, 2015; Acuto, Parnell & Seto, 2018). The adoption of the study of urban data science in various fields will constitute an emerging trend towards value for open data generation, thus creating a culture with dedicated data science units that are experimenting on machine learning models and cloud infrastructural solutions towards a digitalized world. Advances in this emerging field will make it possible to generate new insights from such data which will complement government-generated survey data thus adding to the well of information and knowledge on human well-being.

To conclude, Urban data science education in our view can provide the evidence required to monitor and promote effective development policies that could serve as a stepping stone for bringing awareness amongst the urban practitioners on the importance of data in decision making. It can also inform the general population on the effectiveness of education as a vital form of resources for improving data usage in improving governance via impactful decision making and policies essential for quality delivery for sustainable development. This inclusion should also stir up a wide effort towards the structured development of an Urban Data Science curriculum which has the potential towards improving data-driven policy-making for solving urban problems in Smart cities.

References

- Acuto, M., Parnell, S., & Seto, K. C. (2018). Building a global urban science. *Nature Sustainability*, 1(1), 2.
- Adelman, M., Haimovich, F., Ham, A., & Vazquez, E. (2018). Predicting school dropout with administrative data: new evidence from Guatemala and Honduras. *Education Economics*, 26(4), 356-372.
- Ahmed, F., Capretz, L. F., Bouktif, S., & Campbell, P. (2015). Soft skills and software development: A reflection from the software industry. arXiv preprint arXiv:1507.06873.
- Allison, D. B., Downey, M., Atkinson, R. L., Billington, C. J., Bray, G. A., ... & Tremblay, A. (2008). Obesity as a disease: a white paper on evidence and arguments commissioned by the Council of the Obesity Society. *Obesity*, 16(6), 1161-1177.
- Almeida, C., & Báscolo, E. (2006). Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cadernos de Saúde Pública*, 22, S7-S19.
- Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, 50, 408-413.
- Barkham, R., Bokhari, S., & Saiz, A. (2018). Urban big data: city management and real estate markets. *GovLab Digest: New York, NY, USA*. Available at: <https://mitcre.mit.edu/wp-content/uploads/2018/01/URBAN-DATA-AND-REAL-ESTATE-JAN-2018-1.pdf> (Accessed 3 September, 2019)
- Batty, M. (2016). Big data and the city. *Built Environment*, 42(3), 321-337.
- Baumer, B. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician*, 69(4), 334-342.
- Bradford, L (2019). Why All Employees Need Data Skills In 2019 (And Beyond) (online). Available at: <https://www.forbes.com/sites/laurencebradford/2018/10/11/why-all-employees-need-data-skills-in-2019-and-beyond/#174c6e27510f>. Accessed 24 June 2019
- Brennand, C. A., de Souza, A. M., Maia, G., Boukerche, A., Ramos, H., Loureiro, A. A., & Villas, L. A. (2015). An intelligent transportation system for detection and control of congested roads in urban centers. In 2015 IEEE Symposium on Computers and Communication (ISCC) (pp. 663-668). IEEE.
- Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3), 198-213.

Capotosto, J. (2017). Data opportunities and risks: the dynamic of public, personal, and commercial interest. *Journal of community safety and well-being*, 2(1), 18-21.

Clauset, A., Larremore, D. B., & Sinatra, R. (2017). Data-driven predictions in the science of science. *Science*, 355(6324), 477-480.

de Souza, A. M., Yokoyama, R. S., Maia, G., Loureiro, A., & Villas, L. (2016). Real-time path planning to prevent traffic jam through an intelligent transportation system. In 2016 IEEE Symposium on Computers and Communication (ISCC) (pp. 726-731). IEEE.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17(1), 7-28.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766.

Esty, D., & Rushing, R. (2007). The promise of data-driven policymaking. *Issues in Science and Technology*, 23(4), 67-72.

Gandhi, S. R., & Anyiam, F. E. (2019, May 31). CODATA Blog : Urban Data Science School from May 13 – May 23, 2019. Tratto da CODATA:
<http://codata.org/blog/2019/05/31/urban-data-science-school-from-may-13-may-23-2019>

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, 56(1), 114-137.

Gonzalez-Campoy, J.M (2019). Obesity in America: A Growing Concern (Online). Available at: <https://www.endocrineweb.com/conditions/obesity/obesity-america-growing-concern> [Accessed September 15, 2019]

Harris, R. (2015). The impact of research on development policy and practice: this much we know. In *Impact of information society research in the global south* (pp. 21-43). Springer, Singapore.

Hastings, J. S., Howison, M., Lawless, T., Ucles, J., & White, P. (2019, March 15). Unlocking Data to Improve Public Policy. <https://doi.org/10.31219/osf.io/28krq>

Head, B. W. (2016). Toward more "evidence-informed" policy making?. *Public Administration Review*, 76(3), 472-484.

Heshmore (2017). How much data does google handle?(Online). Available at: <https://www.heshmore.com/how-much-data-does-google-handle/> Accessed 20 May 2019

- Jang, H. (2016). Identifying 21st century STEM competencies using workplace data. *Journal of Science Education and Technology*, 25(2), 284-301.
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, 2(2), 59-64.
- Kambies, T., Roma, P., Mittal, N., & Sharma, S. K. (2017). Dark Analytics: Illuminating Opportunities Hidden within Unstructured Data. *Deloitte Insights Tech Trends*, 2017. Available at: <https://www2.deloitte.com/insights/us/en/focus/tech-trends/2017/dark-data-analyzing-unstructured-data.html> Accessed 24 June 2019
- Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban big data and sustainable development goals: challenges and opportunities. *Sustainability*, 8(12), 1293.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.
- Kitchin, R. (2016). The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 20160115.
- Kitchin, R. (2017). Data-driven urbanism. In *Data and the city* (pp. 44-56). Routledge.
- Kyle, T. K., Dhurandhar, E. J., & Allison, D. B. (2016). Regarding obesity as a disease: evolving policies and their implications. *Endocrinology and Metabolism Clinics*, 45(3), 511-520.
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700-710.
- Lee, J. G., & Kang, M. (2015). Geospatial big data: challenges and opportunities. *Big Data Research*, 2(2), 74-81.
- Leszczynski, A. (2016). Speculative futures: Cities, data, and governance beyond smart urbanism. *Environment and Planning A: Economy and Space*, 48(9), 1691-1708.
- Lytras, M. D., Raghavan, V., & Damiani, E. (2017). Big data and data analytics research: From metaphors to value space for collective wisdom in human decision making and smart machines. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 13(1), 1-10.
- MacLaughlin, S. (2017). The Data All Around Us (online). Available at: https://www.huffpost.com/entry/the-data-all-around-us_b_58ade156e4b0598627a55fa5 Accessed 24 June 2019

- Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, 43(2), 269-289.
- Nellore, K., & Hancke, G. (2016). A survey on urban traffic management system using wireless sensor networks. *Sensors*, 16(2), 157.
- Petty, A. (2019). 6 Data Challenges Managers and Organizations Face (online). Available at: <https://www.thebalancecareers.com/challenges-managers-and-organizations-face-with-data-4100645> Accessed 24 June 2019
- PredPol. (2015). About PredPol (online). Available at: <https://www.predpol.com/> Accessed 22 June 2019
- Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, 114(46), E9783-E9792.
- Rao, G., Powell-Wiley, T. M., Ancheta, I., Hairston, K., Kirley, K., Lear, S. A., ... & Rosal, M. C. (2015). Identification of obesity and cardiovascular risk in ethnically and racially diverse populations: a scientific statement from the American Heart Association. *Circulation*, 132(5), 457-472.
- Rosenthal, R. J., Morton, J., Brethauer, S., Mattar, S., De Maria, E., Benz, J. K., ... & Sterrett, D. (2017). Obesity in America. *Surgery for Obesity and Related Diseases*, 13(10), 1643-1650.
- Singh, B., & Gupta, A. (2015). Recent trends in intelligent transportation systems: a review. *Journal of Transport Literature*, 9(2), 30-34.
- Stevens, M., & Norman, R. (2016). Industry expectations of soft skills in IT graduates: a regional survey. In *Proceedings of the Australasian Computer Science Week Multiconference* (p. 13). ACM.
- Stoner, L., & Cornwall, J. (2014). Did the American Medical Association make the correct decision classifying obesity as a disease?. *The Australasian medical journal*, 7(11), 462.
- Townsend, A. (2015). Cities of data: Examining the new urban science. *Public Culture*, 27(2 (76)), 201-212.
- United Nations, UN (2019). Big Data for Sustainable Development (online). Available at: un.org/en/sections/issues-depth/big-data-sustainable-development/index.html [Accessed 10 September 2019]

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

Williamson, B. (2015). Educating the smart city: Schooling smart citizens through computational urbanism. *Big Data & Society*, 2(2), 2053951715617783.

Wilson, M. W. (2012). Location-based services, conspicuous mobility, and the location-aware future. *Geoforum*, 43(6), 1266-1275.

Zhou, X., Ding, Y., Wu, C., Huang, J., & Hu, C. (2019). Measuring the Spatial Allocation Rationality of Service Facilities of Residential Areas Based on Internet Map and Location-Based Service Data. *Sustainability*, 11(5), 1337.