Design and Implementation of a Data Visualization Course with a Real-World Project Component in an Undergraduate Information Systems Curriculum

Svetlozar Nestorov, Nenad Jukić, and Sippo Rossi

Design and Implementation of a Data Visualization Course with a Real-World Project Component in an Undergraduate Information Systems Curriculum

Svetlozar Nestorov
Nenad Jukić
Quinlan School of Business
Loyola University Chicago
Chicago, IL 60611, USA
snestorov@luc.edu, njukic@luc.edu

Sippo Rossi
School of Business
Aalto University
Espoo, Finland, 02150
sippo.rossi@aalto.fi

ABSTRACT

This paper describes a new data visualization class and its real-world project component in the information systems undergraduate program at Loyola University Chicago Quinlan School of Business. The motivation for and the evolution of the data visualization class are outlined. The fit and the position of the data visualization class in the information systems curriculum are discussed. The content of the class, including the choice of Tableau as the data visualization tool used for instruction, is discussed as well. The paper also describes the details of the project component of the class undertaken in conjunction with GE Transportation and discusses the validity and feasibility of using real-world data and scenarios. The outcomes of the project (which included the analysis of sensor data generated while testing locomotive engines) and the outcomes of the course are also discussed.

Keywords: Data visualization, Decision support system, Curriculum design & development, Data analytics

1. INTRODUCTION

As the amount of data continues to increase exponentially, so does the demand and reliance on data analyses. However, simply relying on statistical and data mining methods is not sufficient. Data visualization plays an increasingly important role in the analytical process from basic charts and graphs that elucidate trends, patterns, and outliers, to executive dashboards and guided analytics (Sharda, Delen, and Turban, 2013).

The importance of data visualization in the data analysis process is also validated by the emergence and growth of software companies and tools that enable turning data into charts. Tableau Software is a prime example. Founded in 2003, the company's main product is the eponymous software that makes data visualization as easy as point-and-click and drag-and-drop. The company went public in 2013, trades on NYSE under the ticker symbol DATA, and, as of November 2018, has a market capitalization of $8.71B (Tableau Software Market Cap 2011-2018, 2018). While Tableau is broadly used (Diamond and Mattia, 2017) and has become one of the

industry standards, there are many other tools that compete in the marketplace, including Qlikview, Microsoft's PowerBI, and Highcharts (Gartner, 2018).

Data visualization is already widely acknowledged as a topic area in the information systems curriculum (Chiang, Goes, and Stohr, 2012; Anderson et al., 2014; Gupta, Goul, and Dinter, 2015; Kang, Holden, and Yu, 2015). The importance of courses addressing the interpretation and composition of data visualization has been recognized and advocated (Nolan and Perrett, 2016), and data visualization has even been proposed and listed at some institutions in the names of majors, such as Applied Data Analytics and Visualization (Wymbs, 2016). Data visualization is being taught in areas as varied as liberal arts, computer science, and engineering (Owen et al., 2013). However, a standalone data visualization class is still a relative rarity in information systems academic programs, especially when compared to more plentiful offerings of standalone courses in other data-analysis related areas, such as data mining or data warehousing (Mills, Chudoba, and Olsen, 2016). Our own analysis of 50 peer information-systems programs (list

available at List, 2018) revealed that less than 15% have courses that include data visualization in the course titles, and even fewer (less than 10%) have standalone data-visualization dedicated classes. Since data visualization is recognized as one of the most fundamental data analytics skills, and the shortage of qualified professionals with a varied set of analytical skills is approaching hundreds of thousands in the U.S. (Tang and Sae-Lim, 2016), we felt compelled to undertake a serious effort in creating a data visualization class at our institution to complement already existing courses covering database management, data mining, and data warehousing.

2. EVOLUTION OF A DATA VISUALIZATION COURSE

Motivated by the development outlined in the introductory section above we decided to develop and offer, on a trial basis, a standalone undergraduate data visualization course as a special topics class. The decision was based on applying a quick entrepreneurial approach to the genesis of a course, as opposed to going through a formal curriculum review and student demand surveys (Jones and Liu, 2017).

Our initial approach to teaching the course was based on an informal survey of many similar courses taught at various departments at different universities. Such courses can be found in Computer Science, Data Science, Information Systems, and several other departments. The courses are taught at various levels of technical involvement ranging from web development with D3.js to using only Excel. While there is no consensus on the software and tools used, one common theme was the use, at least in some part, of the work by Edward R. Tufte (Tufte, 2001). His books and articles explore the theory and practice of data visualization through numerous examples, past and present, absurd and exquisite. Furthermore, several courses in business schools used the work of Stephen Few who has written extensively about the data visualization practice in the business world (Few, 2009, 2012, 2013, 2015).

We also considered the importance of selecting the appropriate prerequisites for the data visualization course. In particular, we identified deep knowledge of relational databases and data warehousing and proficiency in Excel, including pivot tables, as two important topics that students should know well before enrolling in the data visualization course. In our Information Systems (IS) program, these topics are covered in a two-class sequence. The first class serves as an introduction to information systems and includes a significant Excel component that teaches students data analysis skills through in-class tutorials and weekly homework assignments. The course also includes an introduction to data modeling and relational databases. This theme is explored further in the second course that focuses exclusively on data modeling, relational databases including SQL, and data warehousing. The knowledge of data warehousing, and particularly of dimensional modeling and star schemas, is critical to achieving a deep understanding of data visualization because very often the data that is being visualized is presented to the users modeled as a star schema (Jukic, Vrbsky, and Nestorov, 2016; Benghiat, 2017). If students are only exposed to relational databases used in operational environments where the schemas are normalized in order to avoid duplication and achieve data consistency, they will often be confused by the structure of the data used for analytical purposes. Analytical data may exhibit redundancy and other

properties not consistent within the context of an operational database. For example, a product or a service that has a singular base price in an operational database may have several different base price values in a data warehouse in order to account for historical values. Such scenarios are accommodated by the concept of slowly changing dimensions (Kimball et al., 2007) where the same entity instance may be represented by multiple records in a data warehouse, which is an example of the fundamental difference between operational and analytical data. Without knowing these and other particulars of analytical data design, students may experience difficulties in deciding how to map data to appropriate graphic attributes and selecting inappropriate visual mappings, which can impede analysis and lead to misleading conclusions (Primich, 2010).

The relationship between the two prerequisite classes and the data visualization class is depicted in Figure 1.
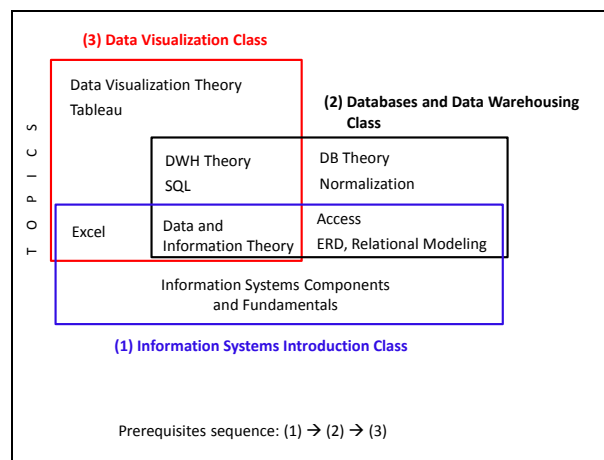

Figure 1. Relationship between the Data Visualization Class and its Prerequisites

While it is possible to create a standalone course in data visualization that has no prerequisite of knowledge of databases/data warehousing/Excel, such as data visualization courses outlined by Jafar, Babb, and Abdullat (2016) and Primich (2010), having prerequisites that cover these topics in depth provides a better foundation to deliver positive learning outcomes that not only teach new concepts, techniques, and tools, but also reinforce and augment the knowledge acquired in the perquisite courses. Having prerequisites, as outlined in Figure 1, allowed us to put maximum focus on visualization topics and avoid a predicament of using valuable class time on remedial reviews, such as in a class described in Primich (2010) where Excel review occupied a substantial amount of course time. The continuity between the courses in our prerequisite sequence is one of the key features in our program aiming to achieve better learning outcomes.

3. COURSE STRUCTURE

The class, which lasts 15 weeks, is structured as follows. The first five weeks of the class are devoted to studying the principles and theory of data visualization, dashboard design, and effective data visualization techniques, while relying on the above-mentioned work by Tufte (2001) and Few (2009, 2012, 2013, 2015). The next five weeks are focused on getting

students familiar with Tableau Desktop and translating the theory and principle learned into practice. The last five weeks of the course are used to examine advanced topics within Tableau such as table calculations as well as executive and exploratory dashboards.

We chose Tableau Desktop as the primary data visualization tool for the class for several reasons. First, it is one of the most widely used data visualization tools in the industry (Hamilton, 2018), and many job postings list it in the required or preferred qualifications. Second, Tableau has a number of features that map very well to how data is typically organized in the corporate data warehouse (Kimball et al., 2007). Third, Tableau can be used at different levels of technical sophistication, ranging from a point-and-click and drag-and-drop interface, to writing complex SQL queries, to creating complex calculated fields using advanced features such as table calculations and level of detail.

The course also included weekly homework assignments, many of them requiring the use of Tableau to create visualizations, such as interactive top-K charts, time animations, stories, and executive and exploratory dashboards.

4. PROJECT COMPONENT

After its first iteration in 2016, the second iteration of the course was made available for enrollment in Spring 2017. At that time, the decision was also made to add a semester long group-project component to the course, with the idea of fostering student engagement by providing a designed activity that promotes a motivational feedback loop for the learning process (Kennedy, 2004; Saundage et al., 2016). This iteration of the course still featured quizzes, tests, and individual homework assignments, but it was obvious that validation of learned concepts via experiencing a longer project on a larger set of data would add a meaningful and tangible experience to the learning process and also give an opportunity for students to better market themselves for internships and job opportunities. This decision was also in line with service-learning approaches that provide opportunities for students to apply their knowledge in real-world settings (Lee, 2012). The structure of the project component consisted of three parts. During the first part, students form teams ranging from two to six members. While we initially recommended teams of three to five students, the logistics of a large class required student individual preferences to be accommodated and thus the range of the team limits was extended. Teams select a dataset and formulate a list of questions that the team will explore and create visualizations that explain their findings.

We recommended that students choose publically available datasets from the Chicago Data Portal (Chicago Data Portal, 2018). The site contains hundreds of datasets ranging from crimes to food inspections as well as city officials and their salaries. Most of the teams select datasets from the Chicago Data Portal but we have had groups using other datasets, such as crowdfunding project data, public health data, and others. After the teams submit their project proposals, the instructor reviews them and provides feedback in terms of dataset suitability and question relevance. A rule of thumb is to require a dataset of at least 100,000 rows (data points) and at least 10 columns (variables).

In the second part of the project, students apply the visualizations techniques learned in class in the analysis of the datasets of their choice guided by their initial list of questions and the feedback provided by the instructor. The deliverables for the second part for each group are a set of visualizations and analyses that utilize most of the concepts and techniques learned in class, such as interactive top-K charts, data segmentation, histograms, etc. The focus is on the application of the classroom material to the real-world dataset and uncovering interesting patterns and trends.

The third and last part of the project is to tell a coherent story about their chosen dataset and prepare and present their discoveries using a combination of slides, charts, and dashboards. The presentations are around 10 minutes long and involve a short Q&A portion. Groups also submit their presentation materials electronically.

The class offered in Spring 2017 had 50 students enrolled. Right before the beginning of the class, somewhat coincidentally and fortuitously, General Electric Transportation (GET) approached us with the idea that a student group engage in a data visualization and analytics project that could bring new insights and understanding to their locomotive engine test data. At that time, we had a four-year long relationship with GET that revolved around them recruiting our students for their rotational leadership internships. We decided to accept this opportunity and offer it to the students enrolled in the data visualization class. This was aligned with Wang (2015) that identifies "using real-world projects that allow students to work with industry professionals and learn how to define problems and collect, organize, analyze, and visualize data" as one of the guidelines for a "dream business analytics program."

We gave a choice to all enrolled students to either undertake a regular project that will be supervised by the class instructors and whose only deliverable is to the instructor or to volunteer for the GET project which will (in addition to the same deliverables as the regular project) also have to be delivered and presented to the GET leadership. Ten students volunteered to do the GET project as their semester-long group project. The remainder of the class worked on the regular projects supervised by the class instructors.

5. GE TRANSPORTATION PROJECT

The target of the project was locomotive engine test data collected in GET repair and testing facilities. The idea of expanding the use of test data by performing ongoing analyses in order to gain new insights formed the core of this collaborative initiative. For GET, one of the goals of the project was to provide awareness into what such expanded analysis could bring to the company. The stated goals for the undergraduate team were to determine early indicators of success and failure within the engine testing data and to create a new data visualization dashboard that provides actionable insights for the end users.

This initial setup matched closely the first phase of the group project in our undergraduate course on data visualization. The dataset was supplied by GET and its size and scope matched the recommended guidelines for the rest of the teams. The general area of interest in the dataset also was specified by GET but our students had to come up with their own specific questions that would help their understanding of the data and lead them to achieving the main goals of the project.

We divided our ten volunteer data visualization students into two groups, and we gave them faculty support on initial

contacts with GET. However, from the outset, we insisted that students independently apply the skills that they were acquiring in class and seek assistance from the faculty only if and when needed. Here we will present the results and the process of both groups combined, since their work and communication with GET overlapped significantly.

From GET's original use case, our student groups were able to narrow the focus on two specific problems: presenting earlier detection of failure in a clearly visible format and shortening the testing process. Solving these two problems would help GET improve their testing process and add value to their business model by reducing both costs and operating time. Additionally, as our students were the first students to participate in GET's student-led data analytics project, GET was hoping to use the findings to advocate for continuing collaborations with future student groups. Therefore, another one of our motivations to engage in this project was to help them explore what students could do and create a model for future student interaction and engagement.

5.1 The Approach
The project began by students speaking with GET to understand their motivation for analyzing the locomotive engine test data. Throughout this exploration, which spanned the time period from January to April 2017, the students maintained constant communication with the stakeholders to understand the data and testing processes. After identifying the use case, students were given multiple data sets as starting points for their exploration. Analysis of these data sets provided by GET allowed the students to narrow their focus to the dataset that they believed would allow them to most successfully provide solutions. Student groups, collaborating with each other, then used mind mapping to determine what key indicators of engine performance would be useful to display on a dashboard. After getting feedback from the engineers, they found they needed to be able to identify both the overall engine performance as well as what sensors led to that success or failure. This prompted them to find a way to derive a meaningful KPI from the 14 most important sensors and their corresponding specification limits. Additionally, they explored several alternatives in order to shorten the test duration based on the behavior demonstrated by successful engines. Using the knowledge of how the test result is determined, they aggregated values from those indicators to optimize the testing process.

5.2 Dataset
The initial dataset for the project consisted of time-series sensor data from the testing process of five passing engines and one failing engine. The sensors record the value of around 400 variables, such as gross horsepower and manifold air temperature, at every minute throughout the test. The dataset was extracted from a GET data lake (Miller, 2016) and was semi-structured. It contained three types of attributes: sensor name, timestamp, and value. The sensor name corresponds to what the sensor is measuring; the timestamp shows the year, month, day, hour, and minute the measure is taken; and the value column represents the measurement registered by the sensor. The total number of records in this format was 14 million. The size and width of this dataset, when transformed for analytical purposes, far exceeded the class requirements of 10 variables and 10,000 records. GET also provided a smaller dataset of specification limits for the 14 most important sensors

which helped identify and extract the relevant data from the larger files.

While the GET team provided guidance and answered questions about the organization of the data and its metadata, all data transformations were performed by the student teams. The structure of the raw data reflected the operational nature of the dataset. The sensor readings were collected while multiple locomotive engines were being tested, so capturing the data in real time, interfacing with some legacy systems, and future extensibility necessitated the generic schema of the operational database. While this project is only a single instance of this phenomenon, the differences between operational and analytical data are well studied and understood in both academia and industry (Jukic, Vrbsky and Nestorov, 2016). Understanding the fundamentally different characteristics of these two types of data and the need for manipulating, filtering, and restructuring the operational data for analysis is of paramount importance for anyone who analyzes enterprise data. As discussed earlier, a course on databases and data warehousing is a prerequisite for our data visualization course. This requirement helped prepare the student teams for the data transformation part of the project as described below.

5.3 Tools and Analytics
The first step in the analysis of the data was data exploration using several tools, including Excel and Tableau. The outcome of the data exploration was a better understanding of the data structure aided by the communications with the GET team. The next step was the transformation of the operational data into structures that matched the analysis goals. The transformation process was performed using a combination of SQL and Excel (which utilized the knowledge from the prerequisite classes) as well as some basic shell scripting and R. Note that most other class projects do not involve such complex transformation processes. However, in many cases, students do manipulate and restructure their chosen datasets, albeit to a much smaller degree.

The basis of the analytics was rooted in transforming sensor data timestamps into an adjusted timeline that allowed the comparison of engines based on the elapsed time of the test. The new structure also allowed the overlay of engine tests for easy visual comparison. To determine the overall engine performance, students aggregated each sensor performance based on specification limits and reduced this to one overall indicator. Furthermore, they worked with stakeholders to identify the most important sensors, as some have a larger impact on the test outcome. Based on this, they created weighted averages to take into consideration their differing impacts on performance. To optimize the testing time, they aggregated the averaged values of each individual sensor to create a moving average calculation over the timeline. This was followed by the analysis to determine the earliest time when definitive test results could be taken.

5.4 Data Exploration
Based on the data exploration and the weekly communications with the GET team, the students developed a good understanding of the testing process, how data was being generated, and how the data is interpreted. The GET process

Moving Average Optimization Table								
	Spec Limits			% Difference SFC & Spec Limit		5 Min MAvg	10 Min MAvg	15 Min MAvg
Tagname	Min Value	Max Value	SFC Result	Min %	Max %	Consistent at:	Consistent At:	Consistent At:
TE009.Compressor discharge left temp	0	509.09	466.91	-100.00%	9.03%	131	134	138
TE009.Compressor discharge right temp	0	509.09	485.09	-100.00%	4.95%	128	133	137
TE009.ECU - Lube Oil Pressure	112.73	157.58	122.42	-7.92%	28.72%	128	132	114
TE009.ECU - Manifold Air Temp	140.61	150.3	145	-3.03%	3.66%	N/A	133	146
TE009.ECU - Pre-Turbo Left Temp	0	1393.94	1241.09	-100.00%	12.32%	122	125	130
TE009.ECU - Pre-Turbo Right Temp	0	1393.94	1250.55	-100.00%	11.47%	131	130	131
TE009.Fuel header psig	100.61	118.79	115.03	-12.54%	3.27%	126	131	136
TE009.Lube oil flow gpm	498.18	612.12	595.15	-16.29%	2.85%	126	131	136
TE009.Lube oil in psig	112.73	152.73	138.06	-18.35%	10.63%	127	131	137
TE009.Manifold air pressure	78.79	109.09	89.82	-12.28%	21.45%	130	131	139
Min Value				-100.00%	2.85%	122	125	114
Max Value				-3.03%	28.72%	131	134	146
		Value Used to Show Invidual Values within 2% in Calculated Field						
		Optimized Test Time						

Table 1. Moving Average Optimization Table

involves increasing the engine revolutions per minute (RPM) until they reach what GET refers to as "Notch 8."

During Notch 8, the sensor measurements are expected to fall into the corresponding specification limits. The engines operate in Notch 8 for about two hours and the result (named "SFC" Result for "Standard Fuel Consumption") is taken for each sensor by calculating the average of the measurement values for the last 10 minutes of Notch 8. When learning about this process, students inquired with GET whether exploring possible variations in this particular setup, such as SFC measurement at the end of the 2-hour Notch 8 process and the 10-minute threshold, would be of use. The GET stakeholders liked the idea and encouraged students to proceed with analyzing alternative scenarios. This type of analysis, examining and questioning specific numbers built into the data and data generation process, is encouraged for all project teams in the data visualization class.

5.5 Results
Through their analysis, students developed two primary results. The first finding centered around GET's existing testing process. The method and justification for the results are described below.

Table 1 summarizes the calculations for the percent difference, and it also depicts the logic behind the students' choice of 2% as the boundary for deviation from the SFC Result. When looking at the percent differences with the specification limits, the tightest measure required was 2.85%. Students used 2% to ensure that the moving average of all sensors is within the specification limits. The goal was to examine how soon the ten-minute average converges to within 2% of the SFC measurement taken at the end of the process.

Visualization A in Figure 2 shows the moving average for one sensor, Manifold Air Temperature (MAT), as the percent difference from the SFC Result during Notch 8. The red lines show the specification limits as percent differences from the SFC Result. The blue and yellow indicate whether the measurements, not just the moving average, are within 2% of the SFC Result. Visualization A shows that for the 10-minute average, the oscillation has decreased to smaller limits and falls within the specification limits for a longer time than the 5- and 15-minute averages. This led to the conclusion that the 10-minute average is the most effective time to take the SFC Result.
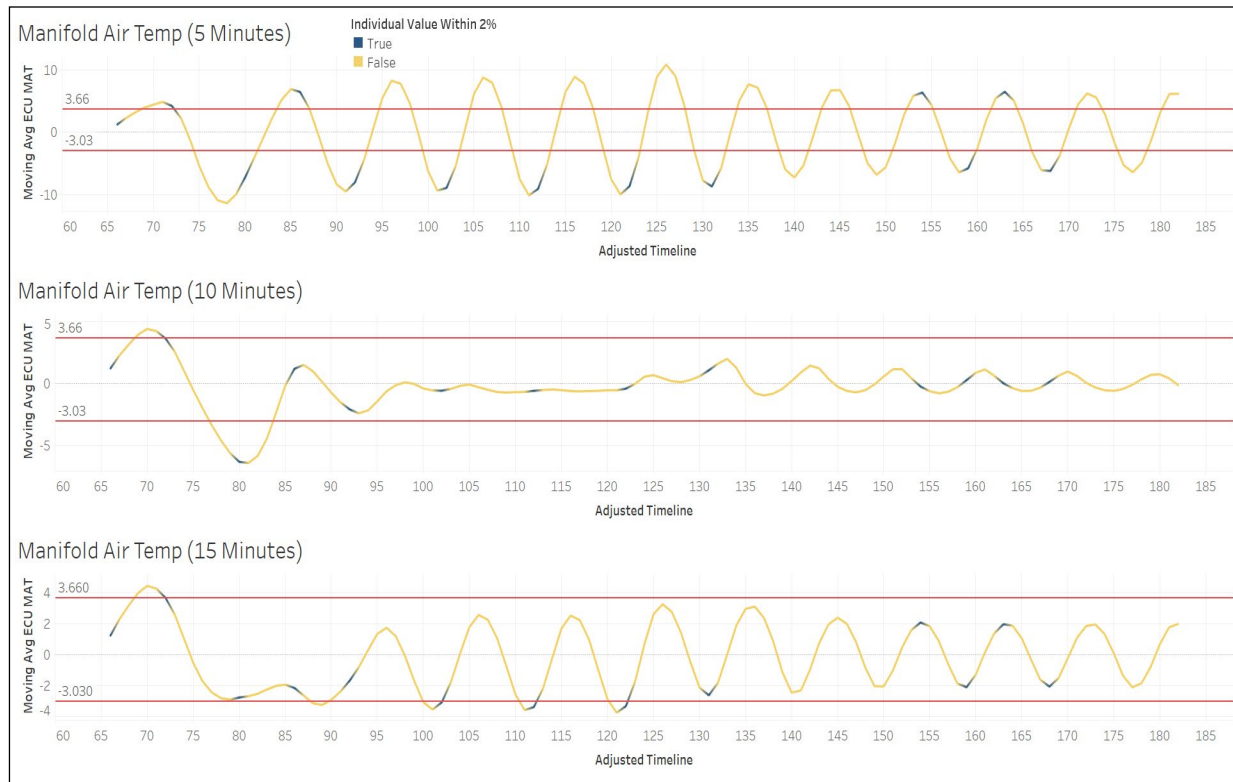
Figure 2. Visualization A

This is a great example of using data visualization to communicate concrete ideas. While the choice of a 10-minute average is likely based on the understanding of the physical and mechanical characteristics of the testing process, Visualization A (Figure 2) clearly shows that the choice is justified.

The next goal was to consider whether the testing process can be shortened. The student teams explored finding a point in the testing process, prior to when they currently take the SFC results, where the moving average is steady in order to take the SFC result sooner and cut down their cycle time. The percent difference was a determining factor in this because if the SFC was taken sooner, the value should be as close as possible to the SFC result towards the end of Notch 8. Students looked at the 10-minute moving average graphs for the sensors and found the point at which each graph started to flatten out or became as close to consistent as possible. They concluded the SFC result could be taken on average 33 minutes earlier than when GET was currently taking it.

The last goal was determining an indicator for failure. Visualization B in Figure 3 and Visualization C in Figure 4 show the dashboard approach to this problem. Visualization B depicts a successful engine, and Visualization C depicts a failed engine. The bottom graph on each visualization shows the sensor, Engine RPM, which is the driving force behind the other sensors as Engine RPM is what determines if the engine is in Notch 8. The top graph in each visualization shows the percentage of the sensor readings that are within their specification limits for each minute. The percentage increases as Engine RPM reaches Notch 8. In terms of the colors, red indicates when the engine is not in Notch 8, blue indicates when the engine is in Notch 8, and green indicates when the engine is in Notch 8 and 87% or more of the sensors are in their specification limits. In the graph for the failed engine there is no point in time for which 87% of the sensors are in their specification limits.
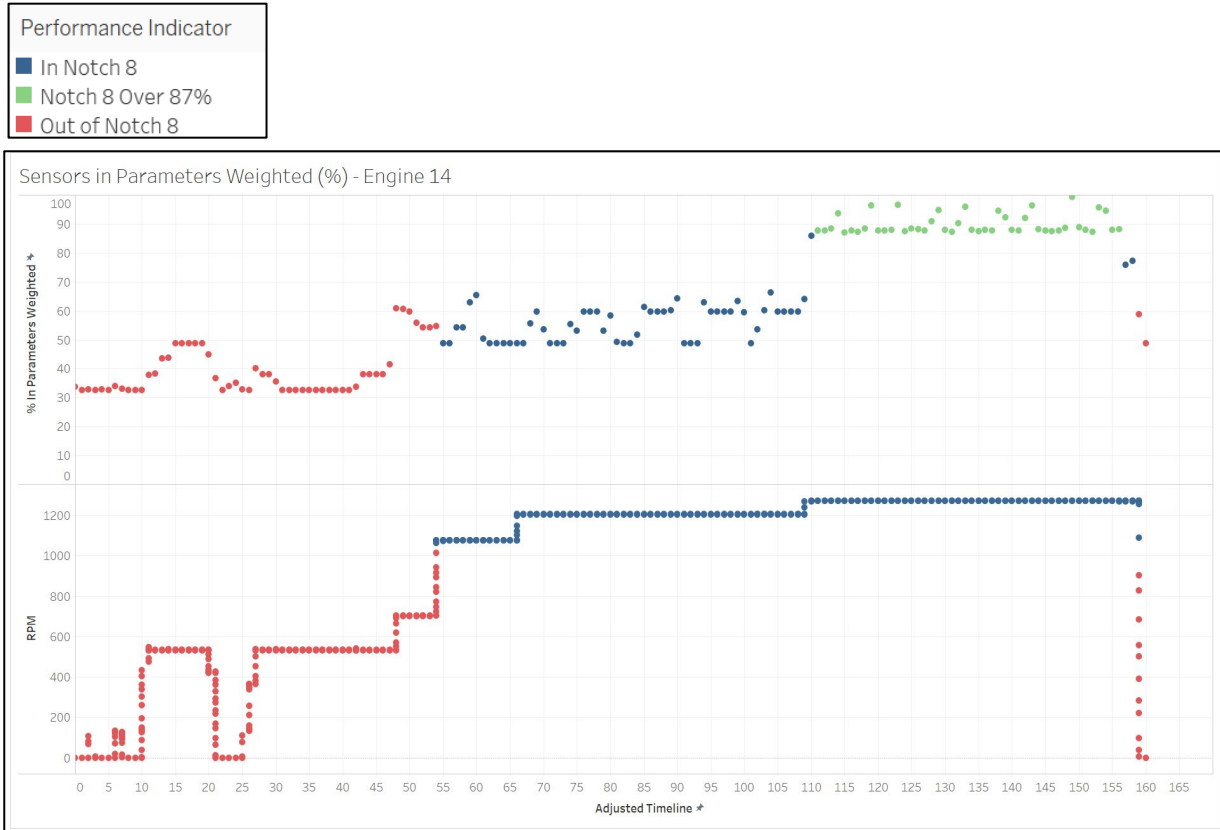
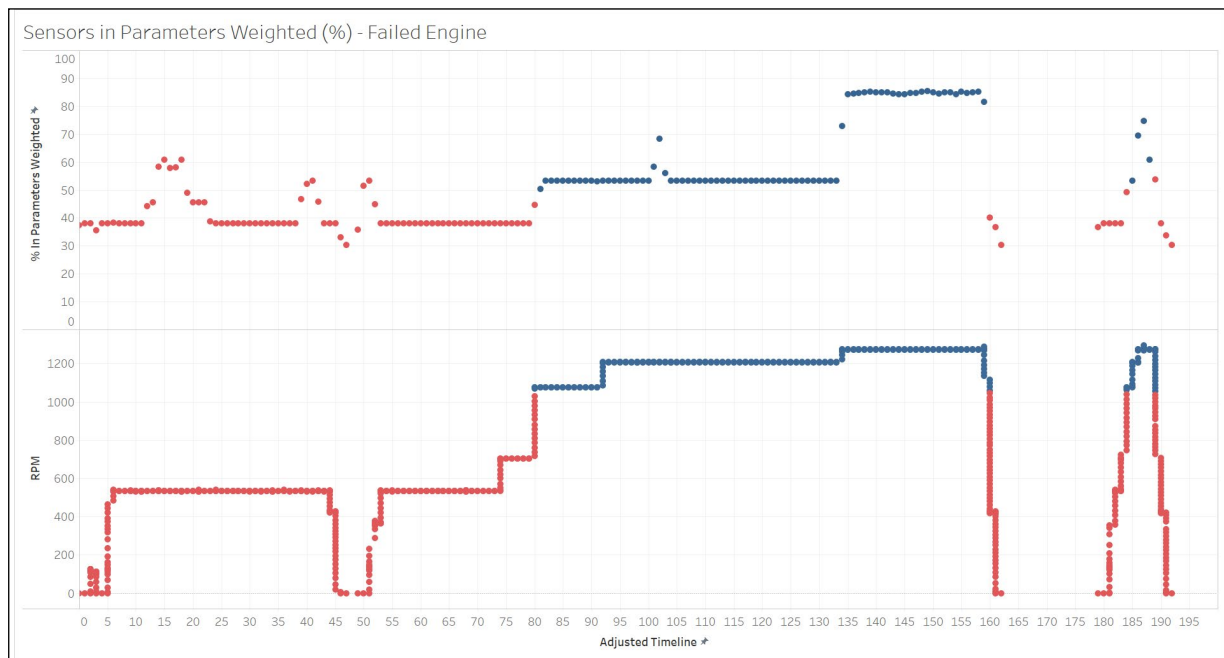**Figure 3 - Visualization B (Successful Engine)**

Figure 4 - Visualization C (Failed Engine)

Visualizations B and C represent two snapshots of a dashboard that keeps track of the progress of the engine test. The top line shows the value of the KPI metric that the teams derived. This KPI represents the percentage of sensors that are within specification limits. While looking at the raw percentage is useful, there are some sensors that are more important than others. Thus, the student teams created a weighted average to give more weight to the five most important sensors. In the dashboard for this data, the viewer is able to see three different versions: the weighted version, the un-weighted version, or the Tier 1 version that shows only the important sensors. Ultimately, GET stakeholders determined that the weighted version is of most interest to them. This is the version shown in Visualizations B and C.

The process of defining this KPI represents a great example of the iterative and collaborative nature of data visualization. While the developers (student teams in our case) possess the technical knowledge and expertise to suggest possible derivation methods, the appropriate formulas can be determined only in collaboration with the stakeholders who have the domain knowledge. For the general projects in the class, this process is simulated through the feedback that the instructor provides after the submission of the first and second parts of the projects.

As part of their dashboard, our teams developed options to select an individual sensor during the testing process. In the focused visualization, sensor measurements that are within the specification limits are highlighted by a separate color. This is useful for the stakeholder to determine which sensor(s) are causing the failure. For the failure shown in Figure 4, the dashboard for a specific sensor showed it rarely within its specification limits, and thus identified it as a problem sensor. This kind of visualization gives the stakeholder an easy way to understand the success/failure of any sensors that they suspected of questionable performance.

One of the unique components of this project was that the students worked directly with GET analysts and engineers to create solutions for a real-world business problem. Working with a real client allowed the groups to apply tools that they have learned in the class to an optimization problem that many companies may face, in different ways. It also helped students encounter challenges that project analysts face on a daily basis. The results of the project and visualizations may also have substantial business impacts on GET. By creating visualizations like Visualizations B and C, and using the accompanying dashboard system, GET can now monitor the sensors for failure in real time as well as conduct further research on what factors lead to failure.

6. PROJECT AND CLASS OUTCOMES

All teams in the data visualization class presented their final projects in a formal setting with a 10-minute time limit for the presentation and 5 minutes of Q&A with the instructors, other faculty, and the rest of the students in the class. The opportunity to craft and present a story about data with visualizations was rated very highly by most of the students in the class. Since several of the teams chose similar datasets, students also were able to compare their performance and receive immediate feedback.

The teams that worked on the GET project also presented their findings. In addition to the classroom audience, their presentation was attended virtually by a dozen GE executives and engineers via teleconference. The presentation focused on the two main findings of the teams: Visualization A and Table 1 that show the rationale for taking the SFC result as a 10-minute average as well as why the SFC results can be taken 33 minutes earlier in the testing process. Direct benefits of the SFC results being taken 33 minutes sooner include saving GET labor time, cycle time, and customer satisfaction that they can now implement into their organization. In the current testing lab, GET can test approximately 912 engines a year (5 engines every 2 days). It takes three hours to run an engine test and one hour to change the engine out. The 17% reduction in cycle time per test can result in saving 501 hours a year. With these 501 extra hours, GET can service an additional 132 engines a year, which would increase their capacity by 15%. If GET's demand for engine repairs stays constant, they can also reduce labor hours needed in the testing department and move employees over to other departments which need more labor power.

While we were able to deliver actual conclusions, this is a project GET is looking to continue into the future with subsequent undergraduate student teams. Our work provided insights for GET to guide future student groups towards confirming our proposed methods to optimize the testing process.

Another outcome of this class was that one of the student teams that participated in the GET project decided to take part in the Data Analytics student competition organized by the Teradata University Network (Gupta, Goul, and Dinter, 2015) by presenting their work on this project at the Teradata Annual Conference in Anaheim, California, in November 2017 attended by several thousand practitioners and academics. Their submission and presentation won them one of the top awards in this contest, "The People's Choice Award for Best Presentation" voted by the participants at the conference.

7. CONCLUSIONS AND FUTURE WORK

One of our main conclusions after the delivery of the course and projects is that, somewhat surprisingly, the amount of time and supervision needed by the GET groups was not any larger than the amount required by the other, regular groups. This was very encouraging, and it provided us with the motivation to allow for more real-world projects in future iterations of this course.

Based on very positive student evaluations, and on uniformly affirmative feedback from the employers that are recruiting from our program (who are increasingly listing data visualization skills in their internship and job postings), in 2018 the decision was made to give the class a permanent course number (INFS 360) and include the class in the information systems curriculum. In addition, a decision was made to create a permanent equivalent graduate level course (INFS 592) and offer it to the students in our graduate business programs. In 2018/2019, we are scheduled to offer six sections of a standalone data visualization class at both undergraduate and graduate levels (including an on-line section), and preliminary registration data indicates that most of the sections (and likely all of them) will be over-subscribed.

Our experience indicates that the field of data visualization is a sought-after, viable, and necessary part of an information system curriculum. In our future work, we plan on continuing to evolve our data visualization courses and expand our partnerships with corporations and organizations. We also plan

to conduct a longer-term study among our students and new alumni whose purpose would be to observe the impact of the data visualization class on their career opportunities and choices.

8. REFERENCES

Anderson, P., Bowring, J., McCauley, R., Pothering, G., & Starr, C. (2014). An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience. *ACM Special Interest Group on Computer Science Education (SIGCSE)*, Atlanta, GA.

Benghiat, G. (2017). Why Do I Need a Star Schema? Retrieved February 11, 2018, from https://medium.com/data-ops/why-do-i-need-a-star-schema-338c1b029430.

Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. *ACM Transactions on Management Information Systems*, 3(3), 12.11-12.13.

Chicago Data Portal. (2018). Retrieved November 25, 2018, from https://data.cityofchicago.org/.

Diamond, M. & Mattia, A. (2017). Data Visualization: An Exploratory Study into the Software Tools used by Businesses. *Journal of Instructional Pedagogies*, 18.

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten* (2 ed.). Analytics Press.

Few, S. (2013). *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring* (2 ed.). Analytics Press.

Few, S. (2015). *Signal: Understanding What Matters in a World of Noise*. Analytics Press.

Gartner. (2018). Magic Quadrant for Analytics and Business Intelligence Platforms. Retrieved November 25, 2018, from https://www.qlik.com/us/lp/sem/gartner-magic-quadrant-2018.

Gupta, B., Goul, M., & Dinter, B. (2015). Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs. *Communications of the Association for Information Systems*, 36(23), 449-476.

Hamilton, A. (2018). What Makes Tableau the Market Leader in Data Visualization Tools? Retrieved February 11, 2018, from https://www.quora.com/What-makes-Tableau-the-market-leader-in-data-visualization-tools.

Jafar, M., Babb, J., and Abdullat, A. (2016). Emergence of Data Analytics in the Information Systems Curriculum, *Proceedings of the EDSIG Conference*, Las Vegas, NV.

Jones, C. G. & Liu, D. (2017). Approaches to Incorporating IT Entrepreneurship into the Information Systems Curriculum. *Journal of Information Systems Education*, 28(1), 43-58.

Jukic, N., Vrbsky, S., & Nestorov, S. (2016). *Database Systems: Introduction to Databases and Data Warehouses*. Burlington, VT: Prospect Press.

Kang, J. W., Holden, E. P., & Yu, Q. (2015). Pillars of Analytics Applied in MS Degree in Information Sciences and Technologies. *Association for Computing Machinery's Special Interest Group for Information Technology Education*, Chicago, IL.

Kennedy, G. (2004). Promoting Cognition in Multimedia Interactivity Research. *Journal of Interactive Learning Research*, 15(1), 43–61.

Kimball, R., Moss, M., Thornthwaite, W., Mundy, J., & Becker, B. (2007). *The Data Warehouse Lifecycle Toolkit* (2 ed.). NJ: Wiley.

Lee, R. (2012). Experience is a Good Teacher: Integrating Service and Learning in Information Systems Education. *Journal of Information Systems Education*, 23(2), 165-176.

List. (2018). Retrieved from https://www.dropbox.com/s/z82eho4z1etx7hd/ListOfUniversitiesDataVis.xlsx?dl=0.

Miller, J. (2016). Data Lakes Offer Smooth Sailing for Egypt's Transport Sector. Retrieved November 25, 2018, from https://www.getransportation.com/data-lakes-offer-smooth-sailing-egypts-transport-sector.

Mills, R. J., Chudoba, K., & Olsen, D. H. (2016). IS Programs Responding to Industry Demands for Data Scientists: A Comparison between 2011 – 2016. *Journal of Information Systems Education*, 27(2), 131-140.

Nolan, D. & Perrett, J. (2016). Teaching and Learning Data Visualization: Ideas and Assignments. *The American Statistician*, 70(3), 260-269.

Owen, S., Domik, G., Ebert, D., Kohlhammer, J., Rushmeier, H., Santos, B., & Weiskopf, D. (2013) How Visualization Courses Have Changed over the Past 10 Years. *IEEE Computer Graphics and Applications*, 33(4), 14-19.

Primich, T. (2010). A Semester-Long Seminar in Statistical Visualization for Undergraduates as Taught by a Science and Engineering Librarian. *Science & Technology Libraries*, 29(3), 181-188.

Saundage, D., Cybulski, J., Keller, S., & Dharmasena, L. (2016). Teaching Data Analysis with Interactive Visual Narratives. *Journal of Information Systems Education*, 27(4), 233-248.

Sharda, R., Delen, D., & Turban, E. (2013). *Business Intelligence: A Managerial Perspective on Analytics* (3 ed.). Boston, MA: Pearson.

Tableau Software Market Cap 2011-2018. (2018). Retrieved November 25, 2018, from https://www.macrotrends.net/stocks/charts/DATA/tableau-software/market-cap.

Tang, R. & Sae-Lim, W. (2016). Data Science Programs in U.S. Higher Education: An Exploratory Content Analysis of Program Description, Curriculum Structure, and Course Focus. *Education for Information*, 32(3), 269–290.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (2 ed). Cheshire, CT: Graphics Press.

Wang, Y. (2015). Business Intelligence and Analytics Education: Hermeneutic Literature Review and Future Directions in IS Education. *Proceeding of Twenty-First Americas Conference on Information Systems (AMCIS)*, Puerto Rico, 3193-3202.

Wymbs, C. (2016). Managing the Innovation Process: Infusing Data Analytics into the Undergraduate Business Curriculum (Lessons Learned and Next Steps). *Journal of Information Systems Education*, 27(1), 61-74.

AUTHOR BIOGRAPHIES

Svetlozar Nestorov is an assistant professor of Information Systems at the Quinlan School of Business at Loyola University Chicago. Previously, he worked at the University of Chicago as a senior research associate at the Computation Institute, an assistant professor of Computer Science, and a leader of the data warehouse project at the Nielsen Data Center at the Kilts Center for Marketing at the Booth School of Business. He is a co-founder of Mobissimo, a venture-backed travel search engine that was chosen as one of the 50 coolest Web sites by Time magazine in 2004. His research interests include data visualization, data mining, high-performance computing, and Web technologies.

Nenad Jukić is a professor of Information Systems at the Quinlan School of Business at Loyola University Chicago. He conducts research in various information management–related areas, including data visualization, database modeling and management, data warehousing, data mining, and big data. In addition to his academic work, he provides expertise to database, data warehousing, business intelligence, and big data projects for corporations and organizations.

Sippo Rossi is a Master's student specializing in information and service management at the Aalto University School of Business in Finland. Alongside his studies he has worked both in academia as a research assistant and in management consulting as an analyst at Deloitte.

Information Systems & Computing
Academic Professionals

EDSIG
Serving Information Systems Educators
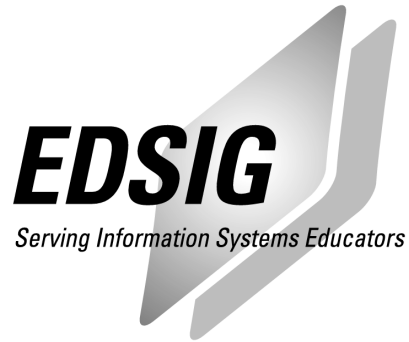
STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.