A Case for Introducing Visual Data Storytelling in CS/CIS Curriculums

Ed Lindoo, Ph.D.
Regis University
3333 Regis Blvd.
Denver, CO 80221-1099
303-964-6358
elindoo@regis.edu

Abstract

It is safe to say that most colleges and universities do a good job of graduating programmers that continue in the field as their life profession. Often these professionals are relied upon for specific functions, for example, writing interfaces to extract data for analysis. Sometimes it is as simple as that, extract the data and let an analyst analyze it and present it. However, too many times this data can be complicated, especially in the world of big data. The programmer, having extracted the data might understand it whereas the analyst might not, rendering these professionals to collaborate on these projects. Raw data is just that, basically meaningless so it makes sense for the programmer to be able to tell a story with the data, a story that can be further enhanced by an analyst. That said, a trend we see today is that these same programmers are being tasked to fill the shoes of both the programmer and analyst. Thus, we make our case for introducing visual data storytelling in CS and CIS curriculums.

Introduction

The third industrial revolution began in the late 1900's with the advent of ARPANET in 1969, and the development of Ethernet in 1973, contributing to the popularity of the programming profession from the 1970's to present day [14]. This era experienced a rise in electronics not previously seen, from computers to new technologies of automation, advancement in telecommunications and widespread globalization. As the demand for programmers exploded, colleges around the world stepped up to teach the various languages and sought-after skills. Initially, programmers typically learned a specific language and went on to support applications that made use of these languages. While colleges did a good job, they often were not agile enough to timely add to their CS curriculum and failed to adopt quickly to changing industry needs. For example, Linux, which essentially became popular overnight. Instead, we often saw things as CICS remaining much longer than it probably should have.

Today, pundits as well as Dugan [5] state "we are now in the fourth industrial revolution". With this, we see that enterprises are taking advantage of the data that is available to them. Companies are finding that they need to do business in a new way, such as in their hiring process. New skill sets are needed such as data scientists with the responsibility of overseeing the collection, storage, and interpretation of data for businesses. Data scientist duties include sifting through data points to create organized categories, comparing data points to current company processes and writing reports outlining business predictions or proposals. These folks typically went through a college programming curriculum where they learn not only how to program, but how to use these programs to extract data out of various systems. As Yazdanian, West and

Dillenbourg [15] pointed out, The Fourth Industrial Revolution has "considerably sped up the pace of skill changes in many professional domains, with scores of new skills emerging and many old skills moving towards obsolescence."

Big Data

Companies now find themselves inundated with "Big Data" and the need to do something with it. Due to "Big Data" arriving at faster and faster speeds, a new set of efficient data analysis techniques are required. As Ahmed and Pathan [1] explained, "the term data science has gained a lot of attention from both the academic research community and the industry." Data science is a big field when we consider that it covers financial trading, Internet of Things (IoT), smart cities, cyber systems, healthcare and more. With these diverse application domains, new research challenges are arising. Stackpole [13] for example, found that countless organizations are trying to use data analytics with their enterprise data, but the endless charts, dashboards and visualizations are falling flat with their intended audience.

Ahmed et al. [1] found that there are several reasons for these failures such as overwhelming recipients with too much data, or presenting the wrong data, or not really understanding how to create an effective narrative that recipients will understand and relate to. Because of this, Ahmed stated the need for Visual Data Storytelling skills calling this "a skill set handcrafted for the era of big data." We can hypothesize that today's students enrolled in computer programming or computer information systems curriculum need to be introduced to Visual Data Storytelling.

Visual Data Storytelling

Most experts agree that data storytelling is the ability to convey data not just numbers as or charts, but as a narrative that recipients can understand. Ahmed et al. [1] explained that a good story needs to be presented without bias, and needs to have a beginning, a middle and end. In this way, business users can absorb and leverage the insights for better decision-making. Programmers or data scientists do not often make decisions, but they need to ensure that they get the right data in the right format to those who do. Burke and Kafai [4] expanded on this with the theory that when writing papers for example, simply learning parts of speech and sentence structure without any designated purpose beyond grammar, will fail to produce good writers. Learning through design ties back to project-based learning in which students simultaneously learn new information. Therefore, we conclude that introducing Visual Data Storytelling offers a new lens for accentuating the connection between coding and writing, both of which require an exact input to produce a particular output.

Before books, movies, radios or even computers, storytellers told stories for such purposes as entertaining, informing, and instilling moral and social values [11]. From a computing perspective, supporting and interpreting dashboards and visualizations by the people intended to use them exposes critical design challenges as well as data understanding. Let us step back in time and remember the concept of dashboards, which started with the automobile. These interfaces connected the driver to the car by showing critical data about speed, fuel level, lights and even alarms if something was wrong with the engine. Other less critical data such as the

time, or outside temperatures also are found. Basically, as Echeverria [6] pointed out, car dashboards are typically uncluttered interfaces that provide just the information an end user would need. In the computer dashboard world, researchers and designers can easily overlook the audience that these visualizations have been created for. Echeverria [6] also discussed, the design of visualizations and how they can often become too complex and thus hard to interpret, especially when designers and researchers try to communicate multiple insights about the data.

Özüdoğru [11] elaborated that storytelling is an effective means of simplifying complex issues while teaching, not merely through printed media or oral methods, but also by using digital media tools. Özüdoğru [11] also mentioned ten significant elements of storytelling:

1. overall purpose of the study
2. the narrator's point of view
3. dramatic question/questions
4. the choice of content
5. clarity of voice
6. pacing of the narrative
7. use of a meaningful audio soundtrack
8. quality of the images
9. economy of the story detail
10. good grammar and language use

These elements of good story telling must be taken into consideration while using technology to tell stories through software, including multimedia tools such as visuals, audio and music. Bouchrika, [3] found that societies have used storytelling to teach key principles throughout millennia. In addition, he determined that storytelling has been used as an information medium in education of all types, including dentistry, general medicine, law, and business. With all the great tools available to us in the 21st century, storytelling has been made richer and more effective using digital media such as images, videos, maps and audio files.

In its simplest form, digital storytelling can be defined as the practice of using computer-based tools to present ideas or tell stories. One of the inherent side effects of digital storytelling is that it creates space for meaningful listening. That is, digital stories provide students with the opportunity to digest information in meaningful ways. Bouchrika [3] believes that this is important in today's environment as people are bombarded with stories and information. He went on to say that digital stories can teach students the value of emotional rhetoric which allows them to explore new ways of thinking differently. In addition, these stories can elicit emotional responses and encourage students to pursue topics they are passionate about. As a result, students not only showcase their learning but also improve technical skills and improve their research and writing skills.

Humphrey [8] elaborated, stating, "Visual data stories have been shown to significantly aid the comprehension and short-term memorability of statistical facts and value messages." As a result, Humphrey believes that these visual data stories are great for communicating complex

information to target audiences. Creating these stories often requires a number of tools to execute the story creation process. Humphrey found that current information visualization tools either focus on exploration or lack sequencing models for visualization presentation. Probably the most important piece of comprehensible visual data stores is the sequencing of visualization pieces which make up the visual data story. Humphrey went on to explain that an optimal sequence conveys the data in a clear and logical manner that reduces the effort required to understand the story. This is not an easy task; it demands a good deal of time and effort often requiring a variety of tools to execute the different phases of the visual data story creation. The specifics of these tools are discussed later in this paper.

Demand for programmers and storytellers

Stackpole [13] found that Glassdoor ranked data scientist as the third most desired job in the U. S. with more than 20,000 openings. A quick check on Indeed.com in May 2022, found over 32,100 job opportunities for data scientists. Obviously, the demand for this position is rapidly growing, and as Stackpole pointed out, technical people that are fluent in languages like Python or R, or experts in statistics and math are just a part of what is required to be successful with data analytics. That is, it is one thing to be good at analytics, but one must be able to effectively communicate their analysis to the audience at hand. Stackpole further explained, data analysts and data scientists do not often have range across skill sets of analytics and storytelling. In fact, data scientists typically have "point-and-shoot skills" where they cannot explain why they are doing what they are doing. "They have a hard time working backwards from questions into practical business solutions. That's really the missing skill set" [13]. The skill of data storytelling is in removing the noise from the presentation and focusing people's attention on key insights.

Being literate with data and able to explain it though stories is now considered a core skill which cuts across all divisions and roles within most companies. Much like communications, some roles will require a better understanding than others, but everyone who's job it is to inform via data will not escape the need to understand and explain that data to others. In fact, as reported by Stackpole [13] Althea Davis, enterprise data governance manager at Etihad Aviation Group, said that data storytelling is a much-needed enterprise skill. This statement is backed up by a quick search on Indeed.com (5/20/22) searching for "data storytelling" which came up with 15,152 jobs across the United States. Even just looking for programmers with analytic skills on Indeed revealed over 11,000 job openings in May 2022.

Tools

There are many tools available for visual data storytelling, starting with probably the most basic visualization tool, Microsoft Excel. This is fine for simple graphs and charts but it is not very powerful as compared to tools such as Microsoft Power BI, SAP Analytics Cloud, or Tableau analytics. In our university program we have used all of these but have settled in on using Tableau, which has free student licenses that are good for one year. Tableau supports the task of producing visualizations from raw data, either as a part of the exploration process or the first phase of the visual data story creation process. This is often done by simply connecting

Tableau to an Excel sheet loaded with data or connecting to an SQL database. The visualizations can then be recorded (i.e., exported or saved) for presentation purposes outside the system. Obviously, there are plenty of tools, but more importantly is how these tools are used. Ryan [12] and Knaflic [9] identified a set of "golden principles" that should be applied when creating stories:

1. Because data storytelling is goal oriented, the visualization needs to be aligned with a purpose or intention. In doing this, it provides designers and researchers with clearer boundaries about what needs to be communicated and what does not.
2. Visual and narrative elements should be used to drive the audience's focus of attention and create meaning in the visualization. This can be accomplished by using specific elements such as lines, weight, shapes, size, colors, and contrast to emphasize key aspects of the visualization.
3. Data storytelling relies on using an appropriate visual and various techniques for certain purposes. As an example, line charts can effectively show changes over time [12]. By contrast, Knaflic [9] dedicates an entire chapter to justifying why pie charts should not be used.
4. One must realize that clutter in data visualizations adds complexity to the graph and makes it harder to understand [9, 12]. Decluttering can be accomplished by removing such elements as unnecessary headers, chart features, borders, and grids that do not add value to the graph. The use of color, shape, and texture are design decisions that can have a substantial impact on decluttering.

Another tool, we call it a subliminal tool is that of computational thinking (CT). While computers can be used to help solve problems before a problem can be tackled the problem itself needs be understood. This is where CT comes in, allowing us to take a complex problem, understand it and develop possible solutions. These solutions can then be presented in such a way that a computer, human, or both can understand [2]. By default, analysts and computer programmers start out with some level of CT and continue to build on that as they progress through their college training.

Parsazadeh et al., [10] proved this in their controlled study using 52 students. Two groups were established for their study, control (CT which used computational thinking) and experimental. In the pre-test, both groups were asked to write a story related to their daily life. In the post-test, students wrote a story to introduce their family. The results found a significant difference between the control and experimental groups. The outcomes revealed that students who used the CT strategy had higher storytelling scores than students who utilized a traditional method of storytelling. Another finding that came out of this research is the fact that the storytelling process can help develop students' language skills in reading, writing, listening, and speaking. As Parsazadeh et al., [10] explained, learning activities in storytelling fosters students' motivation. Specifically, through the synthesis of imagery and verbal representations, integrating technology with CT provides in-depth learning. Because computer programmers and data analysts are typically submerged in CT, it is the position of this author that these folks should be introduced to visual storytelling within their chosen curriculum.

While data itself can be hard to understand, within it there is a story that can be brought to life. Zdanovic [16] explained that this approach can be described as information compressing, where we take complicated information and turn it into manageable pieces. Take for example a large amount of data that shows by state where there were more deaths than births last year.

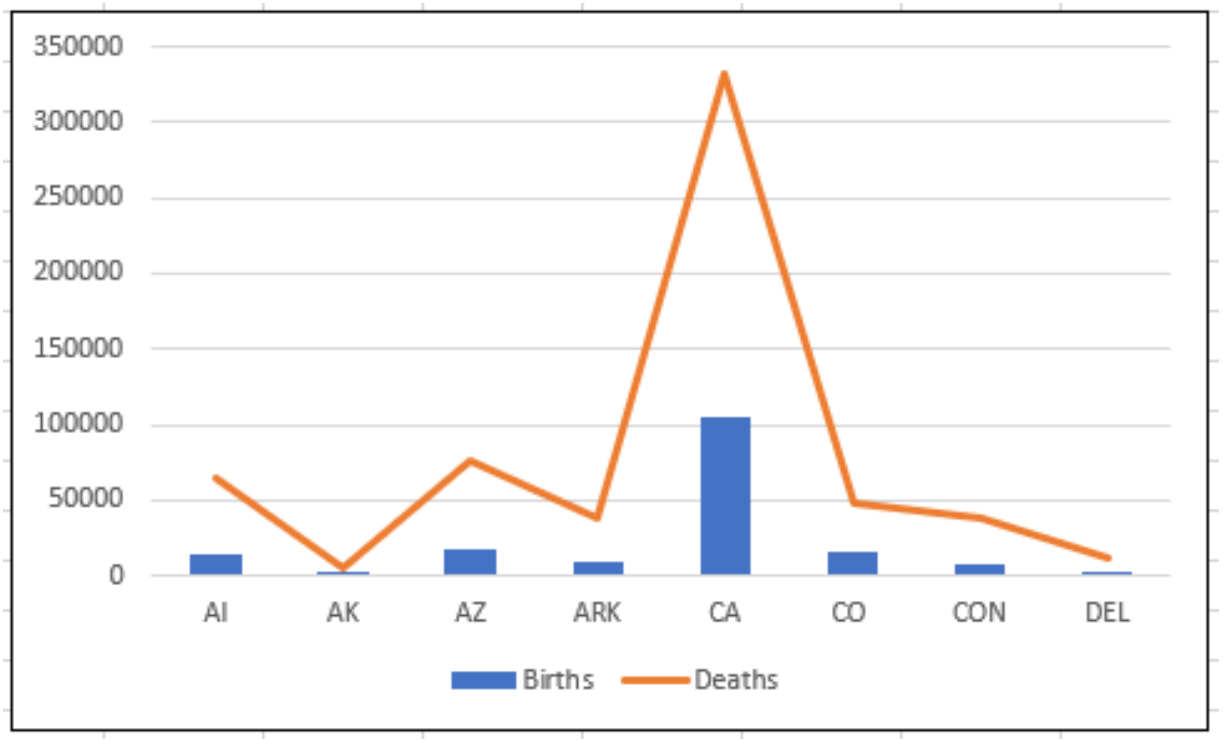

Figure 1, Births vs Deaths 2020-2021

Using a simple chart, we can see in all cases that in these eight states, there were more deaths than births, with California being exceptionally high. However, this does not tell a very good story. First, to show all 50 states in a chart like this would be difficult. To further complicate this, what if the need was to visually show births vs deaths by county, in all 50 states. Not many charts would work well, and this is where the visual storyteller is trained to use the right chart to properly convey the message. A better way to visually show this follows:

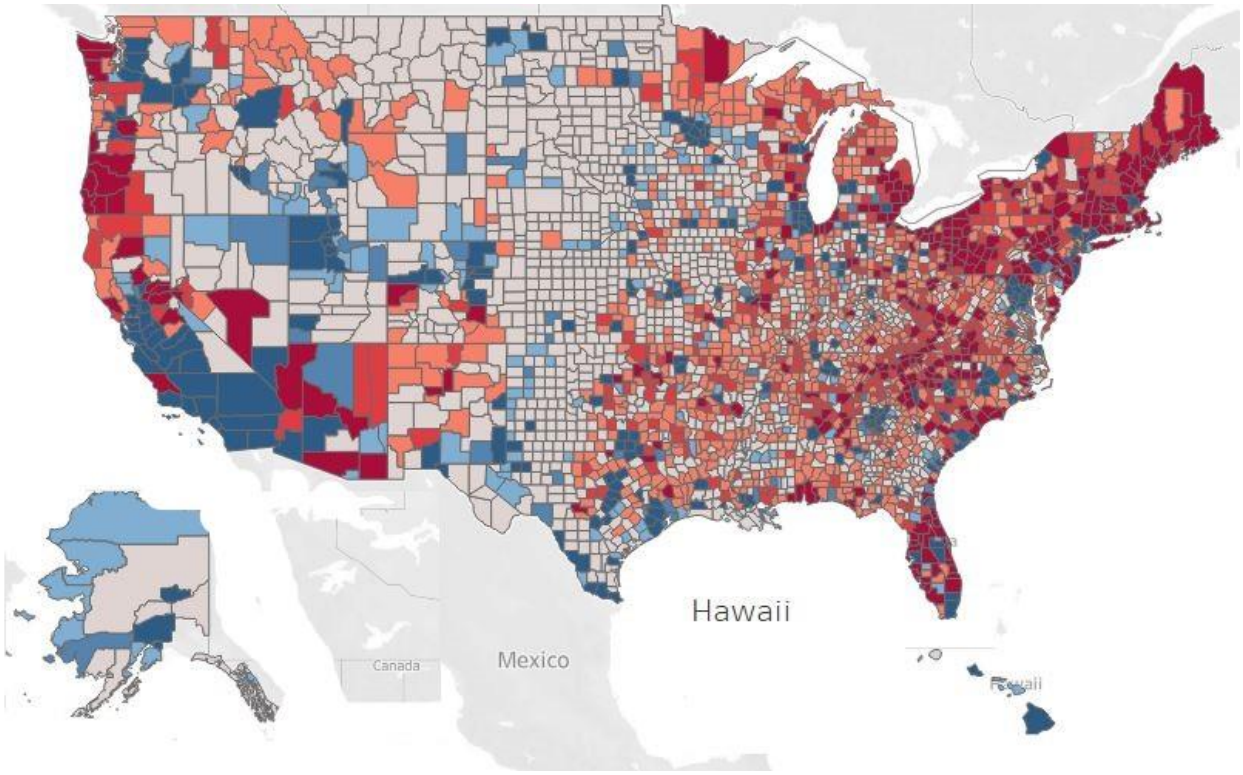

Figure 2, Natural change per 1,000 residents, 2020 and 2021 combined.

As can be seen in figure 2, there is a massive amount of data fed into this map to produce the results of 3,143 counties in the United States. Imagine trying to show this in an Excel spreadsheet, a bar chart, or a heat or tree map, it simply would not work. But here we can quickly identify positive birthrates in blue and negative birthrates in red. The darker the red or blue, the higher the rate, and everything in gray shows minimal change in births versus deaths. While it is impossible to show in a paper all the details imbedded in this map, a similar online version of can be found on the Business Insider website [7] at https://www.businessinsider.com/map-shows-births-deaths-natural-change-by-us-county-2022-4 where you can hover over any county and get detailed results.

Results

In the case of our institution, we introduced a new course, Visual Data Storytelling in the fall of 2021, and quickly had eye-opening success. One of students who works for a Fortune 500 company as a programmer/analyst, had an enormous amount of data that tracked various errors from their systems around the world. The data was overwhelming, but he tried to put graphs and charts together to show upper management using MS-Excel. Unfortunately, with all his efforts, management did not seem to get the big picture. About three-fourths of the way through this course he had an assignment to analyze data and present it in a visual storytelling methodology.

The student decided to use the big data he had been working with using Excel and bring it into Tableau. The first result was eye opening to upper management (Figure 3):

Figure 3, “Bringing order to chaos”. 51,548 alarms

This is obviously overwhelming, and the story is meant to be just that. As mentioned, this got the attention of upper management, but they immediately wanted to know where (regionally) the problems were. This was an easy task as shown in figure 4.

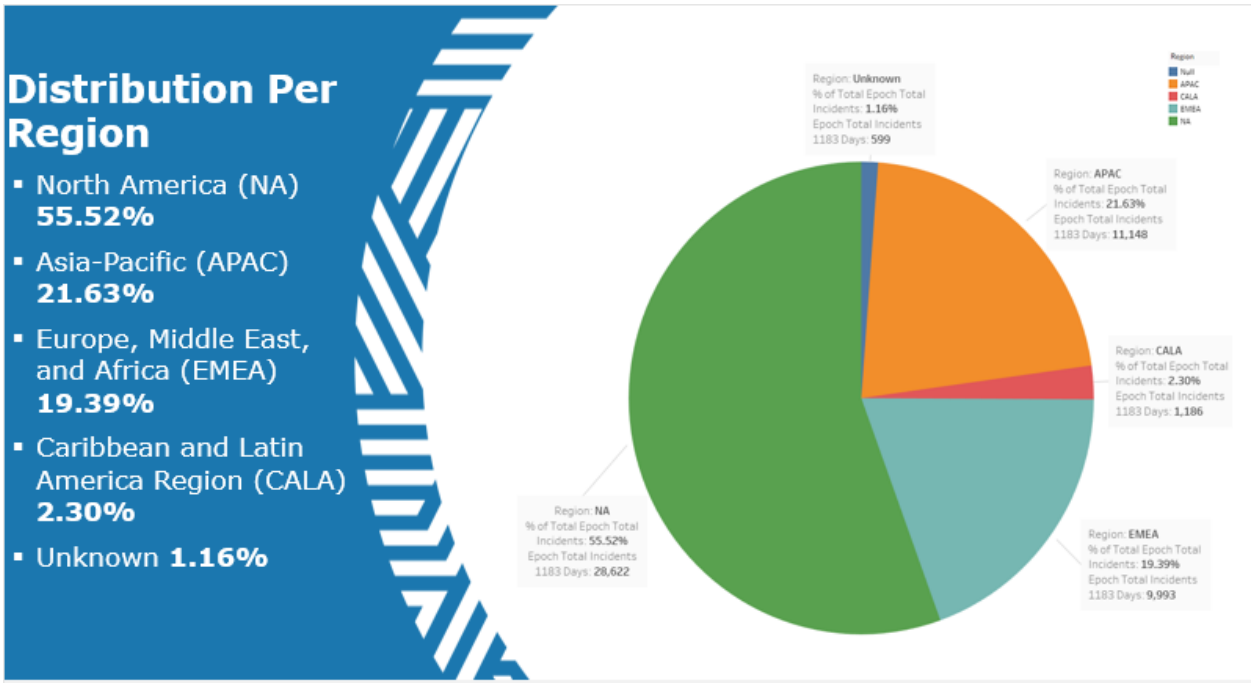

Figure 4, Alarms by region

The student continued with a total of 19 slides, storytelling the issues and how they might be solved. Upper management was so impressed they gave the greenlight to start fixing the problems that were identified. Within weeks they had reduced average monthly alarms from thousands, to about 400, saving 260 manpower hours per month. They also identified an issue in their ESXi hosts and were able to reduce the incidents from 180 per month to less than 10. As a result of this exercise, the company purchased Tableau for all their data analysts and provided them with the Visual Data Storytelling books we use in our course.

Conclusion

As demonstrated, we can no longer expect programmers to simply be programmers. Today they are often asked to extract huge amounts of data that then need to be analyzed and presented to management. Sometimes these programmers will work with an analyst who has a lot of training in storytelling, but often they will find themselves in the position of having to tell the story themselves. We also found via Indeed, a huge need for data scientists, data storytellers and programmers with analytical skills, and that this demand will only continue to grow. Therefore, we make the case that CS and CIS students should be introduced to visual data storytelling. It should be done throughout the program/curriculum using tools as Tableau, with a final, formal course in Visual Data Storytelling. This final course teaches the students how to present to management in a logical, easy to understand way.

References

[1] Ahmed, M., & Pathan, A., (2018). Data analytics: concepts, techniques, and applications. *CRC Press.*

[2] BBC (2022). Introduction to computational thinking. https://www.bbc.co.uk/bitesize/guides/zp92mp3/revision/1

[3] Bouchrika, I (2021) Digital storytelling: benefits, examples, tools & tips. *Research.com*. https://research.com/education/digital-storytelling

[4] Burke, Q & Kafai, Y. (2012) The writers' workshop for youth programmers. Digital Storytelling with Scratch in middle school classrooms. *SIGCSE '12: Proceedings of the 43rd ACM technical symposium on Computer Science Education*. https://dl.acm.org/doi/10.1145/2157136.2157264

[5] Dugan, A. (2020). The fourth industrial revolution has begun: Now's the time to join. *MIT Technology Review Insights*. https://www.technologyreview.com/2020/10/15/1010365/the-fourth-industrial-revolution-has-begun-nows-the-time-to-join/?utm_medium=search&utm_source=google&utm_campaign=BL-ACQ-DYN&utm_content=categories&gclid=Cj0KCQjw3v6SBhCsARIsACyrRAk0nkIKQaC9ZD0pHaj0zeN51-viOUY-X6jSdAtnH8OUeu5EcPcRMrsaAtVrEALw_wcB

[6] Echeverria, V., Martinez-Maldonado, R., Shum, S., Chiluiza, K., Granda, R., Conati, C. (2018).Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics.* Volume 5(3), 72–97. http://dx.doi.org/10.18608/jla.2018.53.6

[7] Hoff, M. (2022). Natural change per 1,000 residents, July 1, 2020 to June 30, 2021. Business Insider, https://www.businessinsider.com/map-shows-births-deaths-natural-change-by-us-county-2022-4

[8] Humphrey O., Chua, C., Avazpour, I., Abdelrazek, M., Grundy, J., Bdnarz, T. (2020) Authoring logically sequenced visual data stories with Gravity. *Journal of Computer Languages.* https://www.sciencedirect.com/science/article/abs/pii/S2590118420300216

[9] Knaflic, C. 2015. A data visualization guide for business professionals. John Wiley & Sons

[10] Parsazadeh, N., Cheng, P., Wu, T., and Huang, Y. (2021). Integrating computational thinking concept into digital storytelling to improve learners' motivation and performance. *Journal of Educational Computing Research*. DOI: 10.1177/0735633120967315

[11] Özüdoğru, G. (2020). An investigation into the opinions of pre-service teachers toward uses of digital storytelling in literacy education. *Participatory Educational Research* (PER) Vol. 7(1), pp. 242-256, March 2020. http://www.perjournal.com http://dx.doi.org/10.17275/per.20.14.7.1

[12] Ryan, L. (2016) Visual Data Storytelling with Tableau, *Addison-Wesely Professional PTG*

[13] Stackpole, B. (2020). The next chapter in analytics: data storytelling. *MIT Sloan School of Management.* https://mitsloan.mit.edu/ideas-made-to-matter/next-chapter-analytics-data-storytelling

[14] Ward, K. (2019). Four industrial revolutions – How we went from the lightning rod to the IIoT. *Manufacturing Data Summit*. https://manufacturingdata.io/newsroom/timeline-of-revolutions/

[15] Yazdanian, R. & West, R. & Dillenbourg, P. (2021). Keeping up with the trends: Analyzing the dynamics of online learning and hiring platforms in the software programming domain. *International Journal of Artificial Intelligence in Education: Official Journal of the International AIED Society*. 31(4):896-939

[16] Zdanovic, D., Lembcke, T., and Bogers, T. (2022). The influence of data storytelling on the ability to recall information. *In Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval* (CHIIR '22), March 14–18, 2022, Regensburg, Germany. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3498366.3505755