Online learning is understood better and improvised with good data visualization.

By
Robert Hitler,
Srishti institute of Art, Design and Technology

Capstone project 2019

Acknowledgment

This project would have not happen without help from many people. I am fortunate to get this project in Adobe. I would like thank Mr. Ashish Garg, Mr. Ramesh Srinivasaraghavan and Mr. Gokul Krishna Paravasthu, who were my mentors from adobe. I would like to thank Mr. Pratyush Raman, who was my mentor in college. And I specially thank Mr. Tridib Roy Chowdhury and Mrs. Anuradha Roy Chowdhury, for helping me get this project in Adobe.

Contents

Abstract

The world is changing and developing rapidly which leaves us in a situation to get adapted to the changes like new technology, new culture, new job fields, etc. Online learning is one of the areas to which people are getting adapted. As information is growing and new job opportunities emerging, we leverage online learning. Institutes cannot structure and organize contents to teach the amount of knowledge and skills to students in this information age, as everyone are free to move in their direction of interest and passion. We see online learning has become an important part of sharing knowledge and skills, so it's important to structure and support the governance of online learning. Supporting by updating content, improvising it and make sure the user getting the knowledge and skills that he/she needs.

One way to support online learning is by administrating it and by creating useful and good quality content, the one more important way is by collecting the data produced by online learning and seeing how to improvise the online learning method. In this paper, I am going to show how data visualization plays an important role in supporting online learning and what makes a good visualization.

Chapter 1

Introduction

In today's world, everyone takes part in producing data in some form and also consume data or use data in some other form. In this era of the information age, data is awash in the world and one main reason for the increase in data is an increase in scientific knowledge and applications. Many companies like Netflix, Facebook, Amazon, etc. depend on this data to provide a better service to their customers. These companies collect customer's data to analyze what one likes, what are their hobbies, and their taste to provide service accordingly. From everything like a satellite to an electric bill, all the data are collected and stored. Fact is we produce 2.5 quintillion bytes of data per day. There is so much data, that we can't even imagine the amount of data we have and the amount of data produced every day.

Data is one of the important reasons why there is an increase in information, scientific knowledge and solution to many problems. This has led to many changes and solutions around us like, change in living methods, change in learning methods, change In job fields, where old jobs are eliminated and new jobs are replaced, etc. For this project I am going to work on one specific problem, that is learning, where data helps to find solutions and understanding to improve learning. How has change brought into online learning by information?

Initially, educational institutions were the places where information was structured shared or taught to students to gain knowledge and skills

but later when the internet and other technologies rise, which led data to increase and circulate rapidly. And slowly informal learning started online. Increase in knowledge and change in job opportunities, made it difficult for an educational institution to provide all the knowledge with limited time and resources, which led the technology to expand online learning and structuring it, to make it up to the traditional learning and one great advantage is collaborative learning. For example, if a new branch of skill has emerged and one from Australia wants to learn it and he/she finds none who knows about it in his surrounding but he/she can leverage collaborative learning online, where someone might have acquired the skill in America and person from Australia can collaborate to learn the topic.

Now the importance of online learning is known and it's understood the importance of supporting it. Information is extracted from data - one of the main reasons for online learning to expand and on the other side online learning is also producing data. A lot about data and how it has changed things is been spoken but it is fundamental that our brain is not capable of processing those data which is encoded with numbers and texts. Data is very abstract and unstructured which doesn’t make any sense when you just look at it. To extract information or to find a pattern or a solution or to tell a story using that data, it has to make sense. How do we structure the data, so that it makes sense? And this is where data analytics and data visualization comes in, to play a very important role and structures the data where it makes sense. “With the emergence of ‘big data’, there has been a rush by businesses, governments, research and

civil society organizations alike to make powerful use of data sources and techniques of analysis."[1]

"A key technique of database governance in education is data visualization. The turn to visualization in education is part of a wider trend where data are increasingly visualized and mobilized graphically in a 'cascade of representations."[2] I am going to explore and learn data visualization tools and create visualization or visual dashboard to show that data extracted from online learning makes a lot of sense and helps improve online learning with good data visualization.

Chapter 2

Research and insights

Data visualization was very new to me and in I knew very basics about data visualization but I know that not all the visualization created, do what it meant to do. In the process of this project, I explored and learned about data visualization, thanks to the online line courses and books which helped to acquire this knowledge.

I downloaded some online tutorials and book on data visualization and understood the importance to make a good and comprehensive visualization to get a good sense of data, which I am going show in this project. We saw why online education is important and why it important to

[1] "Digital Education Governance: Data Visualization, Predictive Analytics, and 'real-time' Policy Instruments." Taylor & Francis. Accessed April 13, 2019. https://www.tandfonline.com/doi/abs/10.1080/02680939.2015.1035758

[2] Ibid

make sense of the data produced by online learning but it won't happen without data analytics and data visualization – this help the raw data to transform into a structured data and visualize to understand it.

"Technology has enabled us to amass greater and greater amounts of data and there is an accompanying growing desire to make sense out of all of this data."[3] And, "Pretty much anyone can put some data into a graphing application (for example, Excel) and create a graph."[4] It's easy but is it an effective visualization? Or will it be able to visualize "big data" and make sense of it? The answer is no.

Data visualization is a way of telling story and humans are better in understanding stories than understanding complex data. To do the first one has to understand the context of the data, only then the data can be structured in a format that can be used to find a solution or lead to some deeper understanding or findings. By context, I mean where the data came from, who is the audience whom you want to solve a problem for or give a new understanding? Then the suitable visual chart has to be selected only when the story can be communicated in a way it has to be communicated. Now comes the important part, a neat and aesthetic visual presentation has to be created. As I told before the context is important, so data visualization has to be created according to the set of the target audience and the tone you want them to understand.

And by following the above procedures I am going to create a visual dashboard to understand the online learning data and see what can I find or solve effectively or efficiently using those data. "A dashboard is a

---

[3] Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: Wiley, 2015.

[4] Ibid

visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance."[5]

An online learning platform is using many technologies like, LMS (Learning management system), database, etc. to be more stable and effective. LMS helps an educator or trainer to create a course, publish it, monitor it and provide support to student or trainee to finish the course. Basically it administrates online learning and it integrated with a standard called SCORM (Shareable Courseware Object Reference Model) which tracks student activities like, enrollment, completion, of course, attendance and overall score but that data was not enough to visualize and improve education. And that led to the invention of a new standard called xAPI (Experience API), also known as Tin Can API, which tracks data in much more details than SCORM, for example how many time a student, watched a particular video tutorial while enrolled in the course or how many exercises the student skipped, etc.

As xAPI tracks or captures every single activity there is enormous data and more the data, more the possibilities to solve a problem. We are going to see how the data from xAPI and good visualization of that will help us to improve the online learning method.

[5] Stephen Few, *Information Dashboard Design* (Beijing: OReilly, 2006).

Chapter 3

Process and methods

From Adobe I got mentors who guided me through the project. I am going to create a dashboard for a set of data collected by online learning and this will be proof of concept for adobe to visualize their “big data” from adobe’s LMS called Adobe Captivate Prime. But I couldn’t work with Adobe’s LMS data due to company privacy or formalities. The xAPI data is anonymized and stored in a database called LRS (Learning Record Store), which stores the data, which is available for anyone to download and study it. So I downloaded an anonymized dataset produced by an online learning platform in a site called academic torrent, I will share the link below in the appendix.

And for data visualization, there are tools like Tableau, Google charts and Power BI (business intelligence) and I chose power bi to work on as per my mentor’s suggestion. Now I have needed data and tool to continue my project.

How to get a good story out the data using data visualization tools and methods? “Nobody sets out to make a bad graph. But it happens. Again and again. At every company throughout all industries and by all types of people. It happens in the media. It happens in places where you would expect people to know better. Why is that?”[6]
Because we are not taught to tell a story out of data and it is a skill important for any data visualization person to learn. We will split this part into two parts one is ‘the story we want to tell our audience’ through data

[6] .Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: Wiley, 2015.

and another is ‘the visualization we want to show’. To do so, first, we need to follow the procedures I mentioned in chapter 2, understanding context and choosing suitable visual charts and tone to tell the story.
The data which I downloaded was an online educational dataset where 115 students had enrolled for the course. The dataset a process folder where it contained every detailed activity of each student in six different sessions, then a final_grades.xlsx spreadsheet having each student marks or scores for each exercise which was totaled for 100 marks and a log.txt file having a record of students attendance for each session.

Context

Data is from the online learning platform. My target audience here are educators, trainers, people who want to do a similar project and people like Adobe who want to understand their customer to provide a better LMS service. I used both exploratory and explanatory analysis to understand the data better. I will explain the term exploratory and explanatory analysis in chapter 4.

Suitable chart (graph or table)

I believe no one chooses their data according to the visualizations it is the other way, where visualization is created or selected based on the type of data and context in data. “There are many different graphs and other types of visual displays of information, but a handful will work for the majority of your needs.” The charts many uses are scatterplot chart, vertical bar chart, horizontal bar chart, stacked vertical bar chart, stacked

horizontal bar chart, area chart, heat map, waterfall chart, box chart, etc. In chapter 4, I will show what kind of visuals I used to develop the dashboard but before going to it will explore more about how to create a good dashboard. Choosing a suitable chart is just one part but how to present it is another important part of data visualization.

Clutter and Pre-attentive attributes

Clutter refers to much noise or too much information which makes difficult for humans process or understand and which creates a sense of discomfort. “When we ask a computer to do work, we are relying on the computer’s processing power. When we ask our audience to do work, we are leveraging their mental processing power.”[7] This means we have to design our visualization in such a way that it is easily communicating with the audience and reduce the cognitive burden to human. And to reduce clutter in the visualization, Gestalt principle of visual perception helps understand how human’s brain is programmed to distinguish things and categories from one another. Gestalt principles are proximity, similarity, enclosure, closure, continuity, and connectivity, to learn read, Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic.

It is important to maintain the aesthetics of the visualization to give the audience a visual satisfaction to spend time to decipher your visualization. And leverage pre-attentive attributes (colors, size, text, and position), to design nice looking visualization and create information hierarchy. Information hierarchy is a method to use pre-attentive attributes

[7] Ibid

to organize your information for the audience to know what they have to look first. If there is too much data it should be organized in such way that only important data is shown or displayed in visualization and information hierarchy helps to organize the selected data in a way that audience sees the information in an order which it has to be visualized. I will show an example of pre-attentive attributes in chapter 4.

Chapter 4
Problem statements and Result

In this project, I am experimenting to create a good visualization and how good visualization can solve problems in online education methods. In the previous chapter, I mentioned that process folder had activity data of students which had 13 variables like, student unique ID, session ID, exercise number, duration, and variables related to mouse activities. I analyzed this data by mix and match of variables, slicing it to specific requirements and structuring it so that it answers some questions or leads to some findings, this is a type of exploratory analysis. While analyzing this data I noted down some common questions which data might answer. Those are given below:

1. Did all the students attend all the sessions in the course?
2. Did all the students who attended the session also attend exams?
3. How many students attended every session?
4. How many students got good grades and how many got poor grades?

5. How many students attend the exam twice and did they make some progress compared to the first one?

And I got interested in visualizing the mouse activities and keystroke of the students, so I created a relationship between mouse activities, keystroke activity with grades to see whether if I can find some pattern? My expectation was to see any behavioral change like the students who got high score might have spent more time on course and which lead to more mouse click, mouse scroll and keystroke activities or else the other way where the student was not able to understand the topic and felt difficult to understand which lead to more mouse and keystroke activities.

Let us find out, will the visualization be able to answer the above question and will it clear my doubt by showing behavioral changes?

The first table figure 1.1 answers the first and second question. Educators or trainers who created the course might want to know how many students attended every session and who missed exams because missing a session might affect student grades. And if more students didn't attend the sessions and if more students didn't attend the exam then there might some mistake in the course that needs to be updated.

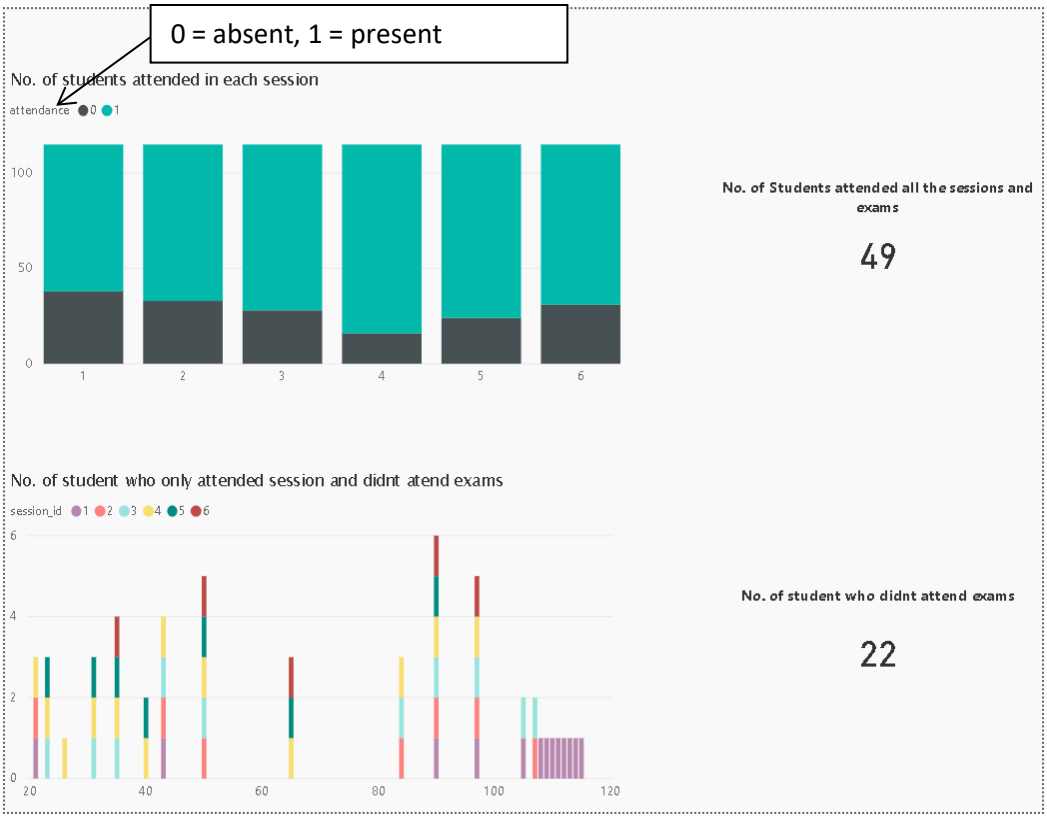

Figure 1.1 *showing how many students attend each session and number of student who failed to attend exam*

The first visualization in figure 1.1 shows or answers the majority of the students attended the sessions and very few people who did not attend. If the cursor is placed on the bar which is each session it shows the exact number who attended and not attend. And the text box next the visualization answers the third question by showing the number of students who attended every session without absences.

The second visualization shows how many students enrolled for the course and attended a few sessions but didn't attend any exam even though the exam was conducted twice. Each color in the bar indicates

each session and it is noticeable that few among them attend more than 4 sessions but didn't attend exam why? Did they feel the course difficult?

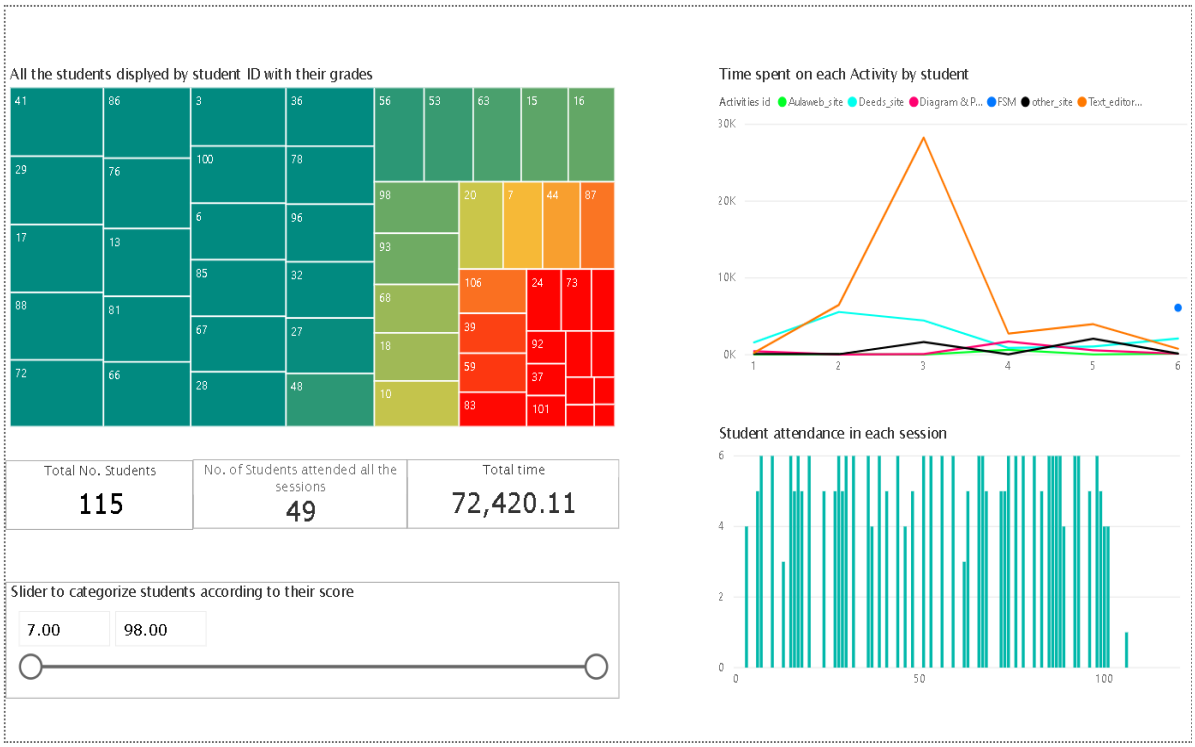

Figure 1.2 *This shows the information related to first exam.*

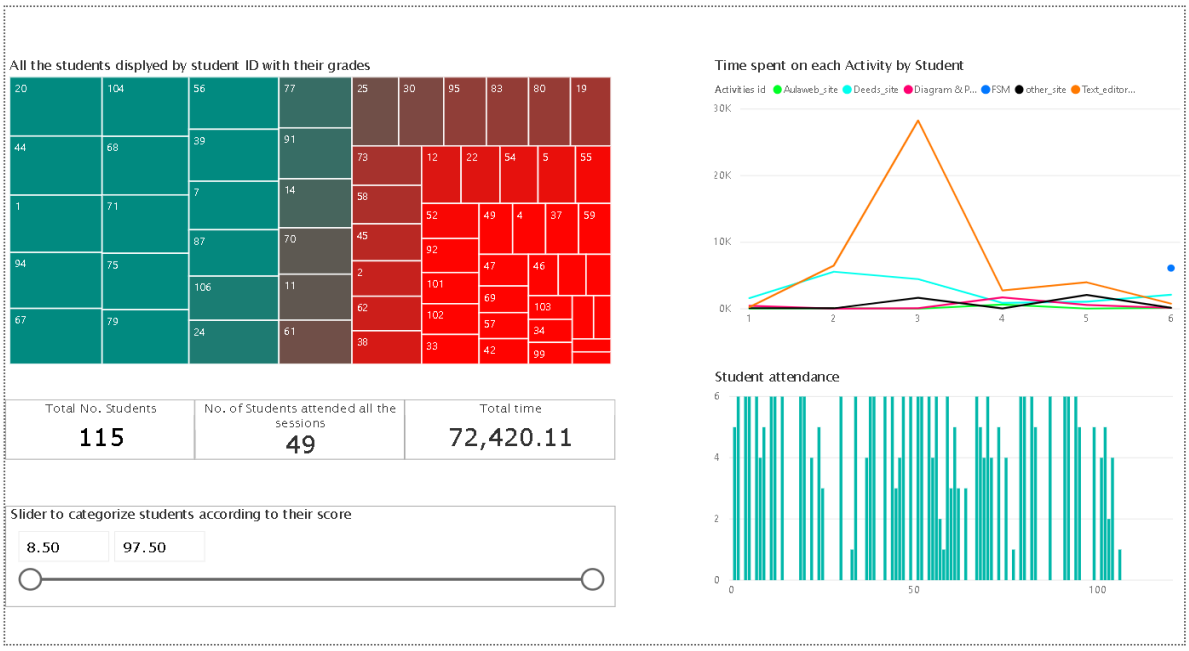

Figure 1.3. *This shows the information related to second exam.*

Figure 1.2 and figure 1.3 answers the fourth question, how many got good grades and how many got poor grades. I will use these figures to give an example to pre-attentive attributes example. First the color attribute in the first visualization, which shows three categories that is green green-red mixed color and red. So the information here is to tell that green represents students who took more than 75 for 100 and red represent students who got less than 35 for 100 and red-mixed represent students who scored between 35 and 75. Second the size of the boxes in the first visualization. So it's simple, higher the score of a student bigger the box and vice versa.

And now if you see the first visualization in figure1.2, its easily understandable that green is more so the majority got good grades. But it's the other way in figure 1.3 that is the second exam where red is more than green so there is a higher rate of low graders in the second exam. Why? The educator or trainer who created this course should give a thought about it.

The fourth question is answered but we saw a difference in grades between first visualization and second visualization, wherein first exam more green appeared and in the second exam red appeared more and the educator might want to find out why? So the other visualization shown in figure 1.2 and 1.3 serves the purpose to dig in deeper like looking into a student's attendance and activities and time spent on each activity. You can read the title and know what it meant to show.

For example, think the educator wants to see every activity of student who got poor grades he/she can just slide the slider to select below 35 and select individual student id or box to his activity as shown in figure 1.4.

Figure 1.4. *This show the least grader version of second exam.*

In above figure 1.4, the slider is used to select the only student whose grade is below 35 and you can see that in the figure where the slider is selected below 35 and the visualization show everything in red. And I have selected one student by clicking on the box to see his/her activities and attendance. In attendance visual chart, the individual student's attendance is highlighted and shows he/she attended all the session but didn't manage to get good marks and activities show he/she spent more in the text editor in the course and didn't spend time on other sites (outside the course). And the total time spent is more than 900 hours. Let's do the vice versa in figure 1.5.

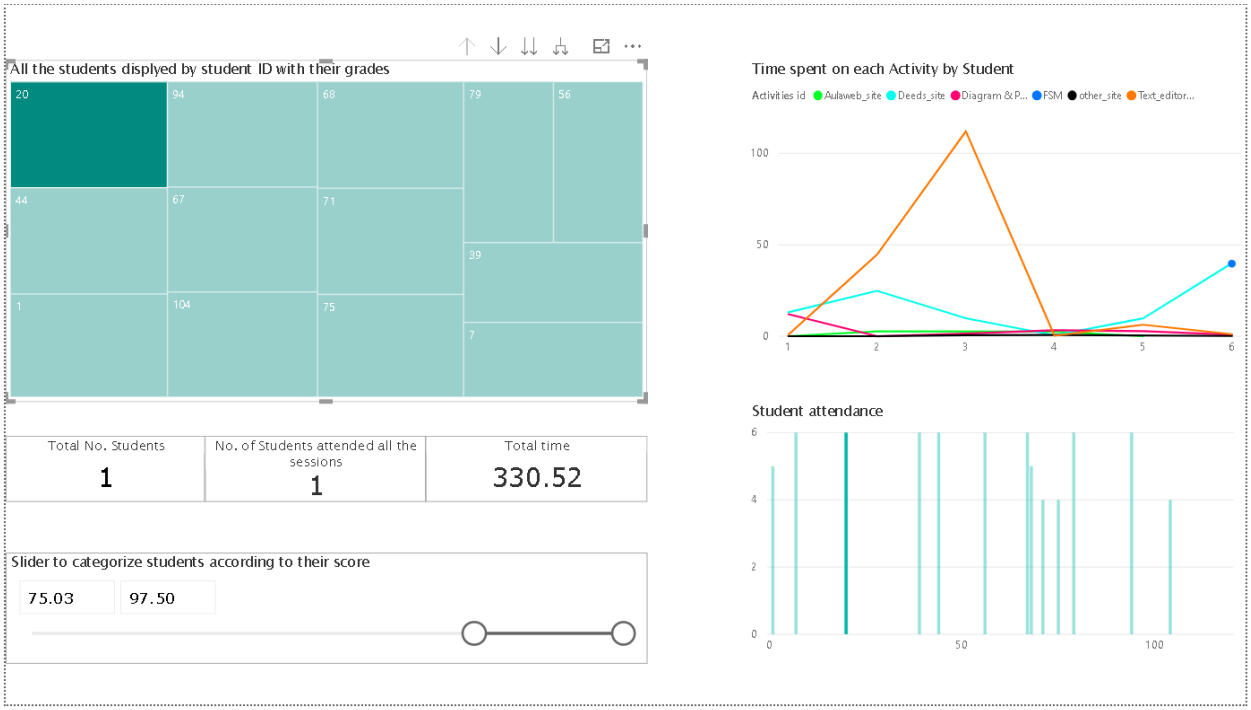

Figure 1.5. *This shows the highest graders version of second exam.*

In figure 1.5, it shows the slider has been selected above 75 and one individual is selected. I notice there is no difference between figure 1.4 and figure 1.5 in activities and attendance but the total time, it shows a big difference.

The student selected in figure 1.4 got less than 35 marks and has spent around 900 hours to finish the course but the student selected in figure 1.5 has got the highest grade and haven't spent even the half of the other students time.

This is just to show how detailed data can show multiple possibilities which needed to be explored more. Let's see figure 1.6 which answers the fifth question.

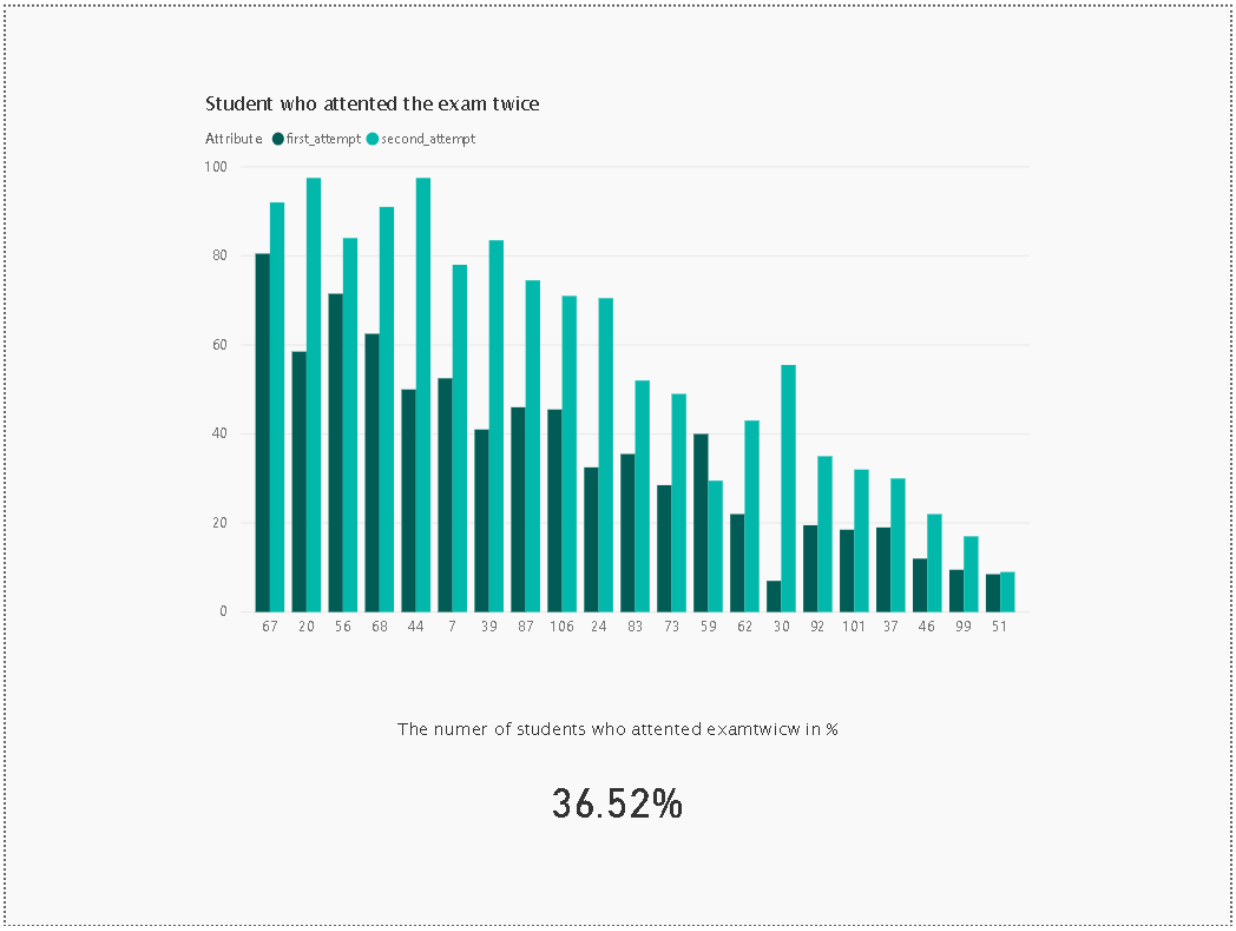

Figure 1.6. *This show the exam grades of people who attended both the exams*

Figure 1.6 clearly shows that 36.5% of students attended the exam twice and has made progress compared to the first attempt.

Now moving to the last question that identifying any behavioral pattern when mouse activities, keystroke, time spent on activities put together and related with grades the students got. Let us see it in figure 1.7.and 1.8.

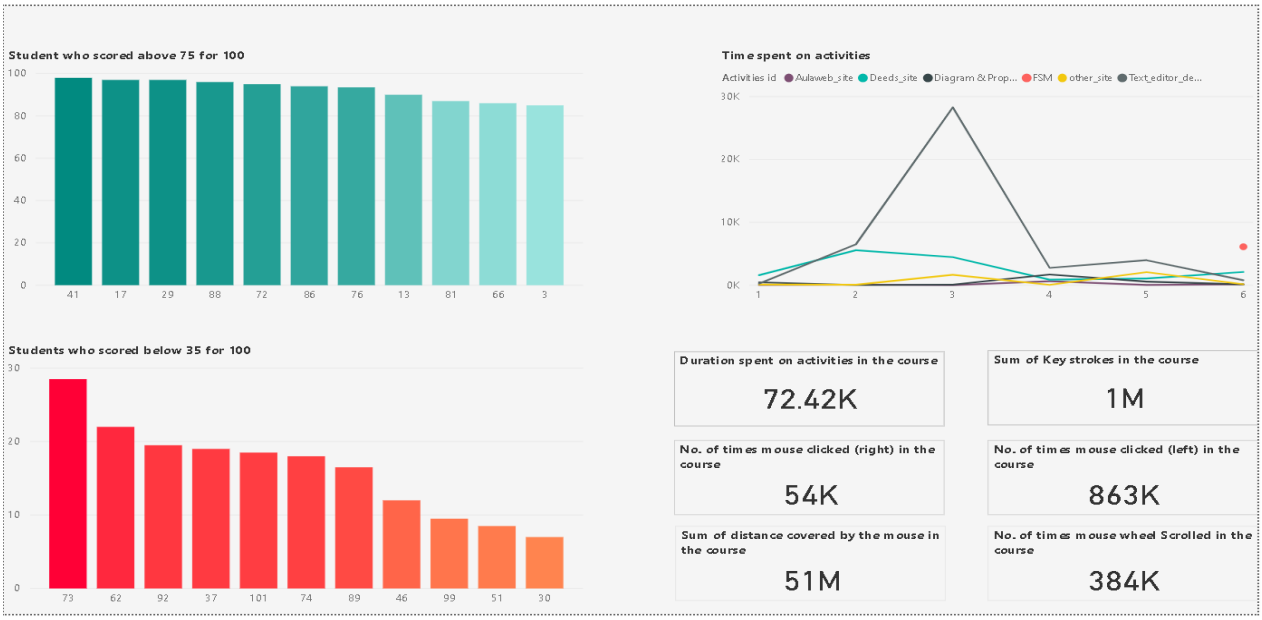

***Figure 1.7**, This shows relation between every data along with first exam*

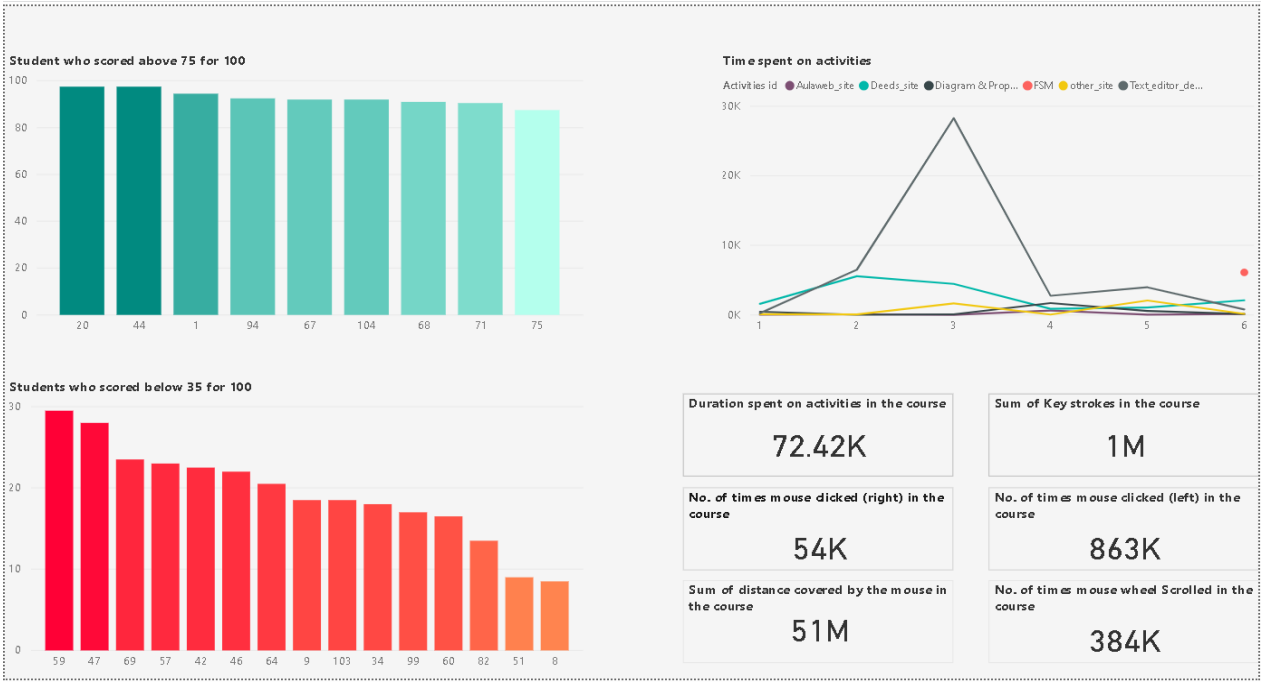

***Figure 1.8,** This shows relation between every data along with second exam*

Figure 1.7 shows data of the first exam and figure 1.8 shows data of the second exam. In these figures I categorized the data in two parts, one is students who scored above 75 for 100, which is represented with color green and another is students who scored less than 35, which is represented in color red, to compare this two category to see pattern with their mouse activities, keystroke activities and time spent on each activity to find and pattern. So each bar shows each student and the height of the bar shows the grades they scored.

Like figure 1.5 each can be selected individually to check the activities and compare with one another so that I can find any pattern but I couldn't find any pattern. If the person who scored good marks have a record of spending more time on the course activities, had more mouse-clicked, scrolled, and the number of keystrokes, the same thing is seen even of a person who scored less.

In this visualization I was able to observe that some of the students who got high score didn't attend all the session and had less number of time spent on the course activities and some of the student who got low score they have attended all the session and have spent more on the course activities but I couldn't connect mouse and find pattern it varied according to individuals. The dashboard has answered them all the questions but if I had data of more number of students I may find patterns using mouse activities and keystroke counts but with a different method and visual representation.

Chapter 5

Conclusion

We saw how important online education is and how important data visualization in online education which led the importance of creating a good visualization. A good visualization helps to tell a wonderful story with data which creates and sense of attachment or connection to the target audience. The dashboard I presented is created based on the dataset that I downloaded and analyzed, so I had data of 115 students which I was able to visualize them and look into each person's activity but this may not happen with a dataset where it has thousands of student's data. Every data produced is different and have different variables and it's the responsibility of people in the data visualization field to understand each data and visualize it. Everyone perception of seeing data varies and so let us use that ability to look at different possibilities and solve a problem, especially in online education to prepare the future generation.

Therefore, the example visualization which I created has a lot of corrections that have to be done to communicate more effectively even though it was able to answer questions I noted down and lead to some different questions. But the visualization can be more improved or structured by the leverage of procedures like understanding context, by using Gestalt principle, pre-attentive attributes (aesthetics and information hierarchy) and your imagination.

And the best example of using imagination to create creative visuals based on the data in the book Dear Data. Two people named Giorgia Lupi and Stefanie Posavec from different place communicated

only through data visuals via postcard. They created beautiful visual to communicate and understand each other. For example, the figure 2.1 below shows one of the visuals, which shows all of the laughter that they had for that weak, categorized them based on people who provoked them to laugh, or whom they provoked to laugh and also tracked how big or small was the laughter. These visuals are creative visualizations and are very much suitable to communicate complex data easily. “Dashboards are not for show. They’re for communication.”[8]

[8] Stephen Few, *Information Dashboard Design* (Beijing: OReilly, 2006).

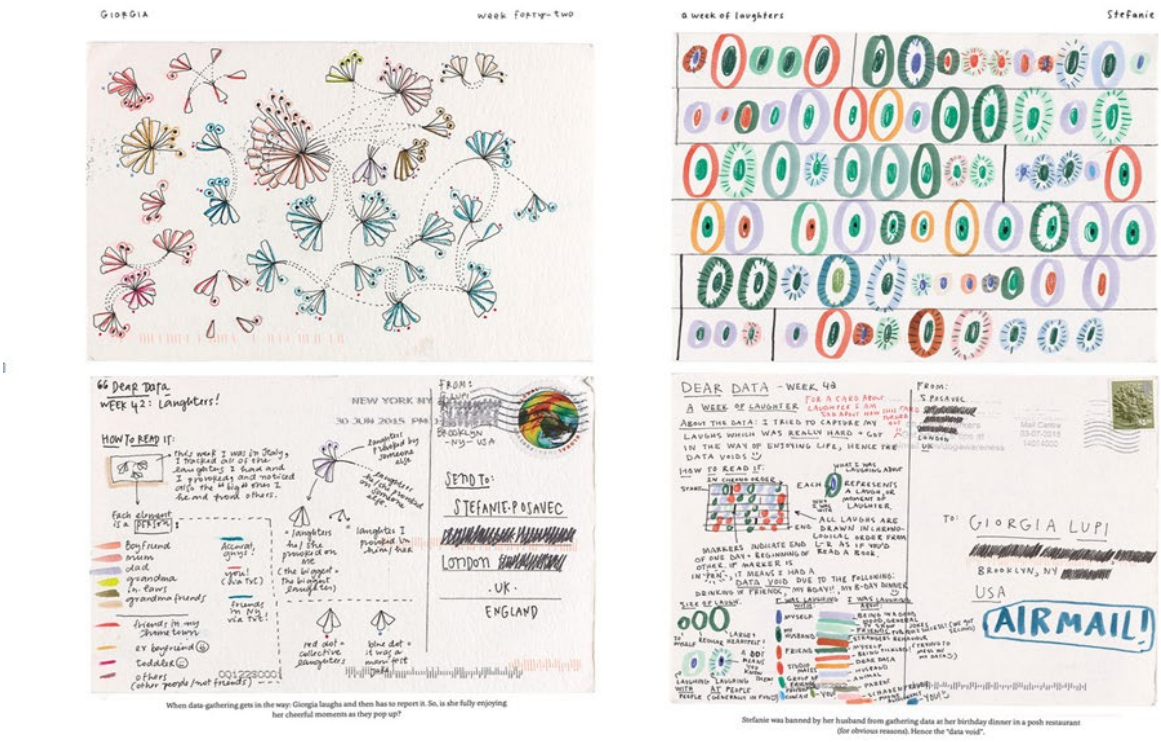

Figure 2.1, *example of dear data book.*

Therefore, it is very important to create good and simple data visualization as data visualization very necessary to help online learning.

Reference

1. AnyChart, "What Is Data Visualization? Definition, History, and Examples," Hacker Noon, December 03, 2018, accessed May 06, 2019, https://hackernoon.com/what-is-data-visualization-definition-history-and-examples-e51ded6e444a.
2. "Digital Education Governance: Data Visualization, Predictive Analytics, and 'real-time' Policy Instruments." Taylor & Francis. Accessed April 13, 2019. https://www.tandfonline.com/doi/abs/10.1080/02680939.2015.1035758.
3. Stamps, UM. YouTube. March 13, 2018. Accessed April 02, 2019. https://www.youtube.com/watch?v=6wzu0Kvmvw4.
4. Stephen Few, *Information Dashboard Design* (Beijing: OReilly, 2006).
5. Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: Wiley, 2015.
6. Vimeo. May 09, 2019. Accessed April 09, 2019. https://vimeo.com/157474716.

Appendix

1. Link that to download the data set from - http://academictorrents.com/details/e24e083cc337695bb84a2b68707695579c0ab4d8
2. Book to refer for good data visualization - *Storytelling with Data: A Data Visualization Guide for Business Professionals by*, Cole Nussbaumer Knaflic.