A System and Model of Visual Data Analytics Related to Junior High School Students

Dang Van Pham[1,2](✉) and Phuoc Vinh Tran[3]

[1] Faculty of Information Technology, Nguyen Tat Thanh University, Hochiminh City, Vietnam
pvdang@ntt.edu.vn, pvdang.tps@gmail.com
[2] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[3] Ho Chi Minh City Open University, Hochiminh City, Vietnam
phuoc.tvinh@ou.edu.vn, phuoc.gis@gmail.com

Abstract. The assessment of students' learning ability for career guidance in the future is a huge challenge. The development stage of students' learning ability is considered from the sixth grade to the ninth grade. Student's transcripts from grade 6 to grade 9 are used to assess students' learning abilities. A transcript comparison of grades 6 through 9 is essential for each parent and analyst from there they can guide their children to comprehensive development of knowledge. The objective of this paper is to visually analyze student data using visual analysis approach, proposes a visual analysis system for data discovery with many variables (VAS), a visual data analysis model, visual data analysis criteria, visual data variables, multidimensional cube representing student data, and some visual data analysis questions based on visual graphs related to Junior High School students (JHSSs). Visual analysis of student data helps parents or analysts observe and extract useful information that they interact visual on visual graphs by asking themselves or answering the visual data analysis questions themselves when observing visual graphs by the retina to guide their children to choose the right knowledge chain and future jobs. Visual graphs represent the correlation between subjects and especially the comparison of a subject in the academic years together to help parents and analysts see clearly the trend of the development of students' learning abilities by visual data analysis model.

Keywords: Visual graphs · Students' learning ability · Visual analysis system (VAS) · Visual data analysis model · Visual data analysis criteria · Visual data analysis questions

1 Introduction

The direction of knowledge chain and future careers for each student is a very important job. In order to properly orient knowledge chain and careers for students, studying at all levels such as Education of Primary (EP), Junior High School (JHS), and High School (HS) is always the primary concern of parents, especially the level of JHS, students have qualities that develop in many directions, students' learning ability manifests in many ways and always change over time. In order to assess the inner

P. C. Vinh and A. Rakib (Eds.): ICCASA 2019/ICTCC 2019, LNICST 298, pp. 105–126, 2019.
https://doi.org/10.1007/978-3-030-34365-1_9

learning abilities of each student, people use the academic results table in the classes at the same level to assess. The multi-column transcript shows data on student scores in the learning process for each academic year. Therefore, the transcript is a data table for each academic year that is shown in a desultory way at the same level of study. When parents or analysts look at transcripts, not everyone also understands the future development direction of their child or knows how their development trends are. From there, giving directions to help their children develop some subjects according to their ability and aptitude and have a solid knowledge of their learning and career guidance.

Assessing the students' learning ability of JHS is a necessary job for parents and analysts. Assessing the learning capability of JHSSs, we should evaluate all 4 grade years of JHS. The transcript for a school year only evaluates that student is average, fairly good, good, poor, or weak at a subject. Relying on transcripts for a school year to assess how the student learns does not yet speak of the student's true learning abilities. JHSSs are in the stages of comprehensive development both physically and intellectually, qualities develop in many directions, at this time the learning capability of students newly exposes in many ways and always changes over time. Selecting the results of studying classes at the same level to assess students' learning ability chains is always the right choosing and doing.

The objective of this paper is to visually analyze students' data by using visual analysis approach, proposes a VAS and visual data analysis model related to JHSSs. The visual analysis of the JHSSs' data set will help parents or analysts to know developmental trends of students in the future. The visual analysis of student data helps parents to orient their children to develop comprehensive learning capabilities in the future. This paper focuses on building visual data variables, proposes visual data analysis criteria, multidimensional cube representing student data, provides some visual data analysis questions, and visual graphs that can help answering some visual data analysis questions from parents or analysts that understand the meaning of students' learning outcomes when observation of visual graphs by visual viewing – thinking method.

The remainder of this paper is organized as follows. Section 2 presents an overview of related research works, comments, and suggestions. Section 3 visually presents student data using visual analysis method, proposes a visual analysis system (VAS), visual data analysis model, and visual analysis criteria of student data of JHS. Section 4 performs the results of visual analysis of student data of JHS using some visual data analysis questions based on visual graphs. Section 5 presents the conclusion of the paper and suggests the direction of development.

2 Overview

2.1 The Concepts

2.1.1 Types of Data

Data are values such as numeric, string, image, date, audio, light, etc. They are collected from various places such as schools, businesses, individuals, etc. In this study, we focused on collecting data at a JHS and performing data classification into data types such as nominal data, ordinal data, and ratio data. Nominal data are the type of data based on attributes whose data are not unequal but only differences or equalities about rank [1–3]. In this paper, nominal data are used to indicate the differences in subjects and are applied to the subject variables. Ordinal data the type of data based on the attributes of these attributes whose data are unequal, different, and equal [1–3]. In this paper, ordinal data are used to measure attitudes, opinions, preferences, perceptions, and views, ordinal data are used to represent the preceding order of time and are applied for a temporal variable. Ratio data are the type of data based on the properties that data of these properties can perform arithmetic operations such as addition, subtraction, multiplication, and division, comparison operations differences, and equal data between properties [1–3]. More specifically, ratio data allows comparing the difference in the ratio of data values, order ranking, and distance comparison. In this paper, ratio data are used for calculations such as addition, subtraction, multiplication, and division, comparison scores with each other and are applied to score variables.

2.1.2 Visual Data

Visualization is the conversion of data and information into images. Visualization is an essential tool to help us understand data. The purpose of visualization is to describe and develop previously unknown ideas to help people perceive useful information, information and knowledge that are hidden within the data through transmitted retina perception into the human brain. Vision helps people to receive information from visual models, and the brain processes to help people get useful information when looking at visual graphs.

Visual data use visual techniques to convert data into visual objects such as points, lines, or graphic object bars that are displayed on 2-dimensional (2D) computer screens. Visual data are conveying information and knowledge clearly and semantically by means of graphics. People can perceive information and knowledge by graphics better than text, numbers, scripts, etc. Users observe graphic images to extract information and seek knowledge to serve specific purposes and serve timely decision making.

2.1.3 Visual Features

Data are visualized in the form of graphic images or visual graphs that help users, analysts, etc. to perform information discovery and seek hidden knowledge within the data. A visual graph must ensure visual features. We conduct surveys and have found some visual features as follows [4–6].

- Ordinal: For user-specific purposes, users can observe visual graphs to explore information and knowledge. Therefore, a visual graph is called data visualization when it meets user requirements. A visual graph must be ordered so that users can observe before, after, over, below, left, right, large, and small.
- Selective: For user-specific purposes, users can observe visual graphs to explore information and knowledge. Therefore, a visual graph is called visual data when it satisfies user requirements. In there, a visual graph must have selective so that users can select as a certain element or detect the location of an object on the visual

graphs, can identify differences in the objects on the visual graphs and retrieve some characteristics of the object on the visual graphs.

- Associative: For user-specific purposes, users can observe visual graphs to explore information and knowledge. Therefore, a visual graph is called visual data when it satisfies user requirements. In there, a visual graph must be associated so that users can group objects that are related to each other on visual graphs to extract information and knowledge.
- Quantitative: For user-specific purposes, users can observe visual graphs to explore information and knowledge. Therefore, a visual graph is called visual data when it satisfies user requirements. In there, a visual graph must be quantitative so that users can identify the ratio of two values when they are visualized representation.
- Lengthy: For user-specific purposes, users can observe visual graphs to explore information and knowledge. Therefore, a visual graph is called visual data when it satisfies user requirements. In there, a visual graph must have lengthy so that the user can identify the length relative to the number of elements of a variable represented on a coordinate axis, then the user can perceive each element on the visual graphs.

2.1.4 Visual Analytics

Visual analytics is a fundamental foundation in the field of analytical psychological science. This area has facilitated people to discover information and seek knowledge from data by visual interaction techniques between people and computers through a graphical interface [7]. People use visual techniques and visual analysis tools to discover information and knowledge from data, simultaneously, have an insight into the big data sets that are related to each other that these data sets aren't heterogeneous, vague, rich, diverse and always changing characteristics by the time. The field of visual analysis is an area that goes through professions and assists people to gain a broad sense of large data sets through assessment, strategic planning, and decision making timely. In visual analysis, there are visual interactions and visual representation techniques, these techniques are used to take advantage of the human's retinal to acquire information and knowledge when people observe visual graphs by methods of viewing - thinking intuitiveness. Therefore, the main goal of visual analysis is to support the maximum of the reasoning process and analyze human data conjunction with computers through the creation of argument methods, techniques for visual data analysis and visual representation of data to help maximize the ability to perceive, understand, and reasoning of humans, in which humans play a major role in the process of amplifying knowledge from data.

2.2 Related Works

Making decisions and making decisions in a timely manner as well as carrying out human daily tasks effectively depends on who receives information or knowledge in which context or background and how reliable is it? Therefore, human always desire to find the best methodologies and visual data analysis tools to support users to extract more information as well as perceive more hidden knowledge inside data. Until now, the scientific achievements of knowledge discovery from data have been studied very much by scientists. In it, there are fields such as data visualization, visual analytics, exploratory data analysis, data mining, knowledge discovery in databases, etc. In this paper, we perform systematization some of the research works related to these areas to demonstrate that data scientists are very interested in these areas, especially when human increasingly aim toward awakening the mind (also or effecting the mind or touching the mind) every human being to support policy making, decision doing, and decision making.

Data visualization is a process of creating images that represent diverse data sources, qualitative or quantitative data, where users can extract information and seek hidden knowledge within the data. Data visualization having input parameter is a variety of data sources and having the output parameter is a visual graph, the visual graph must be as simple as possible, good suggestions for users to see the entire range of visual data, do not miss important information, not to misunderstand the meaning does not exist in the data, must take advantage of the retinal viewing ability of human beings and humans can apply the visual viewing - thinking method to view visual graphs. To date, many models of data visualization systems (also or data visualization processes) and data visualization processes have been proposed by many authors groups.

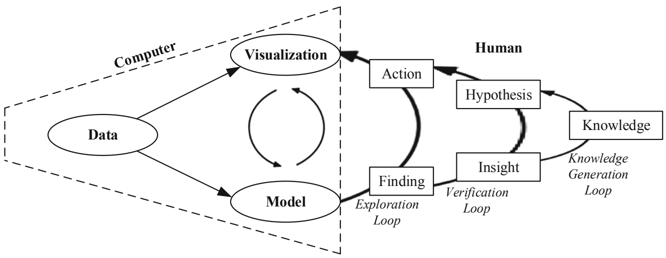

Fig. 1. Model generates knowledge from data [8]

The model arising from knowledge in the field of visual analysis is proposed by Sacha et al. [8] (see Fig. 1). The main purpose of this model is to propose the role of human participating in the process of knowledge development and the rest is a visual analysis system due to the computer performed automatically. The Shacha team divided the model into two main phases, the first stage illustrating a system of visual analysis, the remaining stage illustrating the human join in the process of knowledge development. Human implement the process of generating knowledge that includes sub-processes: the mining process, the testing process, and the process of generating knowledge. The two stages in the model that generate knowledge from data are active interactions between human and computers, in which human play a central role in controlling the whole process.

The process of visual analysis (also known as the visual analysis system model) was proposed by Keim et al. [9], this process combines the model method and visualization method with the main purpose is to help human participate in the process of knowledge mining from data. This process is divided into two main branches, the first branch, human using the visualization method to discover knowledge from data, the second branch, human using the model method to discover knowledge from data. Between these two branches, model methods and visualization methods are mutually supportive in the process of discovering knowledge from data. With the model method, the data is analyzed automatically when the user requires, while the visualization methods, human play a major role in the process of knowledge mining from data. The strength of the visualization method is thanks to the knowledge of existing data analysis expertise of experts in many fields, thanks to the knowledge and experience available in many professional areas that experts have can extract a lot of information and find knowledge in many interdisciplinary branches.

The process of visual analysis proposed by Keim authors [10] is converted from the visualization model proposed by Wijk [11], this process clearly demonstrates the three stages of the knowledge discovery process from data. Stage 1, the process of receiving raw data sources, the original data must be classified and analyzed by data analysis techniques. Stage 2, visualization method will be applied to convert the analyzed data into visual image to help users easily feel. Stage 3, users can apply viewing - thinking methods to perceive or deeply recognize visual images to crystallize into knowledge, from existing knowledge and newly acquired knowledge, human can simultaneously apply to generate new knowledge based on existing assumptions and continue to perform deeper analysis and discovery knowledge from data by the technical characteristics applied to the visualization method of stage 2. In the second stage, the application of visual data analysis technical characteristics is very important, normally human interact on visual representation images will be easier to understand the nature of data.

The reference visualization model is still considered as the classic model proposed by Card et al. [12]. The main purpose of the reference visualization model is to model the steps needed to turn data into an interactive visual form. In particular, the model is divided into two main stages, stage 1 is data formed in order from large data sources converted into data table structures, stage 2 is visual form formed in order from visual abstraction to transform into views. In the reference visualization model, the order is converted from stage 1 to stage 2 by visual mappings controlled by the user. In stage 1, the user controls with data transformations, stage 2, the users control with view transformations. So, in this model, the role of human is very clear, human can decide when to perform transformations, data format, choose which visual structure to use, and choose the way which displays.

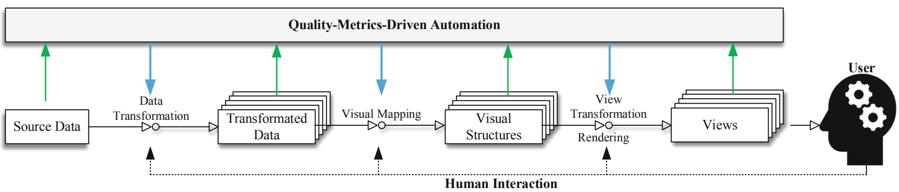

Fig. 2. Visualization process automatically adjusts to quality [13]

The automatic visualization process of quality adjustment [13] (see Fig. 2) is extended by Bertini authors group from the [12] reference visualization model, this group suggests adding at the top of the one-layer reference visualization model automatically adjusting to quality. This layer obtains information from rectangular-shaped boxes of each stage in the reference visualization process [12] and influences the processes of the reference visualization process through the data that the visualization process automatically adjusts according to the calculated quality. Users are always in control of the entire process and the end result is that users get visual structures and different display ways.

The data analysis system developed by Thi Nguyen authors (see Fig. 3) [6] expanded from the model of knowledge generation from data proposed by Sacha et al. [8], and the model of visual analysis system proposed by Keim et al. [9]. The data visualization system is capable of transforming data into information and knowledge by model methods in data analysis to build mathematical models combined with visualization methods along with human's contributions participate in the process of information discovery and knowledge search from data. In particular, the data visualization system highlights the principle with the strategy of knowledge discovery from data, data is converted into information by model method combined with visualization methods, human and computers work together through the use of existing human knowledge and experience to participate in the process of extracting information and discovering knowledge from data by means of visualization or called visual perception method.

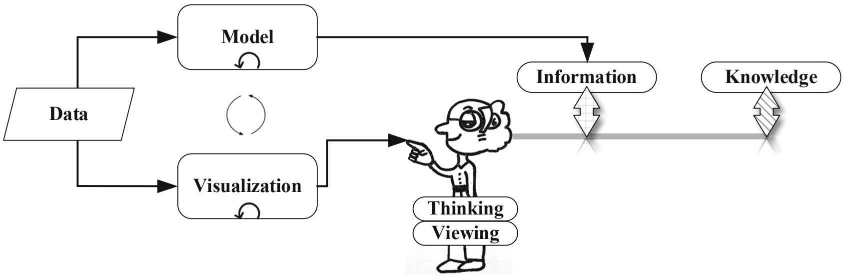

Fig. 3. The system analyzes data by model and visual method with human contribution [6].

Through the systematization of models and processes related to the field of visualization and the field of visual analysis above, we have some comments as follows. The models and processes meet some of the problems of converting data sources into information and knowledge. Models and processes have combined the interaction between computers and humans join in the process of information and knowledge discovery. The models and processes are all focused on using computer technology as the main, while the participation of people in the process of information discovery and knowledge search only speaks of basic interactions like selection, search, and control only. There is no model or process that provides clear criteria or methodology to support users in the process of information discovery and knowledge search from data.

The data analysis system proposed by Thi Nguyen authors [6] refers to the contribution of humans to the process of information and knowledge discovery from data, but human contributions are not grouped how this presents in a specific way.

Through some observations above, we find that the above models or processes are mainly focused on building processes or stages of information and knowledge discovery from data. However, in order for people to make good decisions and make timely and breakthrough decisions, people must know how to apply their existing knowledge and experience to apply to processes is a key to success. The philosophy of von Szent-Gyorgyi *"Discovery consists of seeing what everybody has seen and thinking what nobody has thought."* [14] indicates that each person needs to participate directly in the process of information discovery and knowledge search from data. Based on the above observations and based on the philosophy of Albert von Szent-Gyorgyi, this paper proposes to build a VAS of multiple variables based on visual graphs, building visual graphs, visual data analysis model, and visual data analysis criteria related data for JHSSs. These proposals will be detailed in the next sections of this paper.

3 A System and Model of Visual Data Analytics of JHSSs Data

3.1 Proposing a Visual Analytics System for Data Discovery with Many Variables Based on Visual Graphs

The discovery of information and the search for knowledge from visual graphs by viewing – thinking method are great challenges in the fields of visual analysis. In order to assist people in discovering information and seeking knowledge from data when observing visual graphs effectively, this paper proposes to build a VAS for multi-variable data discovery based on visual graphs (see Fig. 4), this system was expanded from a visual analysis model proposed by the group of authors Keim [9], a model of knowledge generation proposed by the group of authors Sacha [8], automatic quality-adjusted visualization process proposed by the group of authors Bertini [13], data analysis system by visualization method combined with a model method that namely mathematical model together with human contributions [6], and the process of knowledge discovery in the databases with data mining techniques [15]. This system is divided into two main layers, the first layer is managed by computers, and the second layer is carried out by humans. The first layer is divided into two main branches, branch 1 applies visualization method in visual data discovery, and branch 2 applies the model method in the process of knowledge discovery in the database. Branch 1, with the visualization method, the reference visualization model [12] coordinates with the model method in data analysis to render into visual graphs. Branch 2, with the model method and the process of knowledge discovery in the databases [15] cooperates with visualization methods in data analysis to crystallize information and knowledge. Branch 2, with the model method and the process of knowledge discovery in the database [15], cooperates with visualization methods in data analysis to crystallize information and knowledge based on visual graphs.

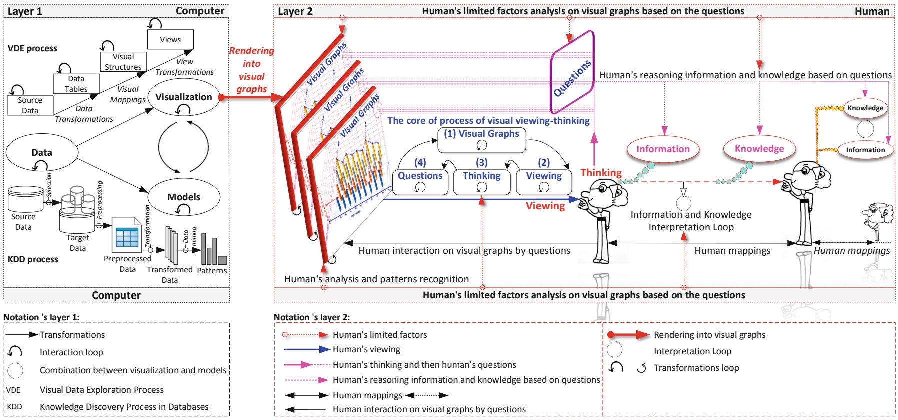

Fig. 4. Visual analytics system (VAS) for data discovery with many variables based on visual graphs

The paper is based on the philosophy of von Szent-Gyorgyi [14] to propose the construction of the second layer to attach to the existing visual data analysis system, this newly built system is called a visual analysis system for data discovery with many variables based on visual graphs, this new system is called VAS for the purpose of serving people to make visual interactions on visual graphs (see Fig. 4), by asking themselves questions and answering their own questions in themselves information discovery process and knowledge search from data. Indeed, humans are always inherently able to see-understand, capable of reasoning or interpretation, being able to remember according to the deductive mechanism or the ability to recall memories, and there are other possibilities. However, people still have limited factors that need to be analyzed need to be studied to show the limited factors of people to overcome on the second layer of this visual analysis system, both methodological and other restricted factors. The philosophy of von Szent-Gyorgyi [14] shows that people need to participate directly in the process of knowledge discovery from data. Therefore, on the second layer of this visual analysis system, the paper analyzes the restricted factors of humans when people interact on visual graphs by relying on asking or answering your own questions. It can be said that data analysis cannot be performed if no questions are asked. Questions are the first step and the basis for promoting the process of visual analysis of data conducted.

Also in the second layer of this VAS (see Fig. 4), we propose to build a core of process of visual viewing - thinking following these steps, step **(1)** with human appearance and visual graphs, step **(2)** human view visual graphs, step **(3)** human think visual graphs and render into information and knowledge, step **(4)** then people interact visually on visual graphs by asking themselves or answering the data analysis questions themselves when observing visual graphs, and continue to repeat step **(1)** if people need to explore information and seek knowledge from visual graphs for the next time. This is an infinite process of a data processing flow in a visual data discovery system with many variables based on visual graphs. In addition, on the second layer of this system, there are some limited factors for each person when participating in this system to perform information discovery and knowledge search on visual graphs based on asking yourself or answering the questions myself when observing the visual graphs. The restricted factors of humans following we are shown that need to be clearly analyzed (see Fig. 4). In particular, the limited factors 1, 2, 3, 4, and 5 were conducted in the next sections of this paper.

- **Limited factor 1:** People interact on visual graphs by posing questions themselves or answering themselves with visual data analysis questions in the process of information discovery and knowledge search.
- **Limited factor 2:** People rely on a visual data analysis model combined with visual data analysis criteria to ask themselves or answer the visual data analysis questions themselves.
- **Limited factor 3:** People use questions and answers in which questions must be classified and stratified into a visual data analysis question system as well as user answers.
- **Limited factor 4:** People use visual viewing – thinking method to ask themselves or answer the visual data analysis questions themselves.
- **Limited factor 5:** People interact on visual graphs with different viewing angles by using visual data analysis criteria to see and better understand the nature of data.
- **Limited factor 6:** People use questions that are reasoned previously or answers to perform analyzes and discoveries that add information and knowledge to existing ones.
- **Limited factor 7:** People use the loop to deduce information and knowledge that was discovered the previous time to perform analysis and discovery to add useful information and knowledge for the next time.
- **Limited factor 8:** People look at visual graphs and then analyze and recognize patterns based on visual graphs for the purpose of discovering information and searching for hidden knowledge on visual graphs.
- **Limited factor 9:** People receive visual graphs by eyes, sense of viewing, human eyes as a means of receiving more data than other means, sense of hearing, gnawing, tasting, and touching. Sensing visual visuals from the eyes performed at different levels of treatment need to be shown.

3.2 Visual Cube Representing Student Data

In this section, we introduce data variables related to students of JHS. Temporal variable (var_T), in turn, is school years of sixth, seventh, eighth and ninth grade. Subject variable (var_Sub), in turn, is Mathematics (M), Physics (P), Biology (B), Literature (L), History (H), Geography (G), Foreign languages (FL), Civic education (CE), and Technology (T). Score variable (var_Sco) is the average of the subjects. From these data variables, we construct visual variables for each variable as a coordinate (0x, 0y, 0z). Then, we built a multidimensional cube representing student data based on the data variables defined above.

3.2.1 Data Variables Related to Students

School and parents monitor the learning process of JHSSs through the academic transcripts that the school records over the school years. A transcript is a data table made up of multiple columns and rows to show students' learning outcomes by school year. The transcript represents the school years of grade 6, grade 7, grade 8 and grade 9; subjects as M, P, B, L, H, G, FL, CE, T, Informatics, Chemistry, Physical education, Music, and Art; Each subject corresponds to an average of that subject.

In order to fully appreciate the academic ability of JHSSs, we need to rely on the subjects that students are fully learning in the four school years of 6th through 9th grade. In the scope of the paper, we focus on surveying and researching a number of subjects as M, P, B, L, H, G, FL, CE, and T (see Table 1). These subjects are assessed by number scores very clearly, there are other subjects that students can only study for two years, those subjects are as chemistry and informatics. If we only assess the students' learning ability trend but only base on two years of study, the learning results do not fully reflect the students' learning ability. In addition, a number of other subjects such as gymnastics, music, and fine arts are not studied in this paper because these subjects are assessed as having achieved or not achieved in the results of the end of the school year but it does not assess students' true learning ability.

Table 1. Data table show the average score of each student in 4 years of JHS [vnedu.vn]

Y (Year)	FN (Full Name)	M (Mathematic)	P (Physics)	B (Biology)	L (Literature)	H (History)	G (Geography)	FL (Foreign Language)	CE (Civic Education)	T (Technology)
C_6	Pham Thao Nguyen	9.0	7.0	6.5	6.0	6.0	6.5	10	8.0	7.0
C_7		8.5	8.0	5.0	9.0	9.0	8.0	10	9.0	9.5
C_8		9.8	9.2	8.5	7.0	7.0	5.0	9.0	7.0	7.6
C_9		10	9.5	7.0	9.5	9.5	8.1	10	9.0	9.5
C_6	Pham Dang Khoi	5.4	5.0	6.1	5.7	5.8	6.3	3.9	6.7	6.0
C_7		6.5	7.8	4.5	5.5	5.5	6.9	4.0	5.5	5.5
C_8		8.5	6.2	8.8	5.7	6.8	6.2	7.7	5.9	5.8
C_9		9.8	7.2	9.4	7.7	7.8	6.6	9.7	6.9	5.8

Table 1 is shown as follows, the school year column contains data elements {C_6, C_7, C_8, C_9} are the school years {Class 6, 7, 8, and 9} respectively; the full name used to contain the student's first and last name; M, P, B, L, H, G, FL, CE, and T columns used to contain student's scores from 0.0 to 10 marks; on each line of Table 1 (except for the first line), each unit of time {C_6, C_7, C_8, C_9} will combine with full name {Pham Thao Nguyen, Pham Dang Khoi, etc.}, and the average score of each subject {M, P, B, L, H, G, FL, CE, T} indicates the students' learning ability in each grade 6, 7, 8, and 9.

Through the data analysis in Table 1, we found that the school year column (Y) is a unit of time considered as an independent variable; the columns of M, P, B, L, H, G, FL, CE, and T are dependent variables that change over time; the var_Sco, in turn, include the scores of the subjects. The variable is a data set whose components are the elements of the set. Therefore, in Table 1, elements of the set include subjects such as M, P, B, L, H, G, FL, CE, and T; school years like C_6, C_7, C_8, C_9; average grades of subjects such as M, P, B, L, H, G, FL, CE, and T. Therefore, the data variable is a variable data value, the data value can be a decimal form (e.g. a score), a string form (e.g. subject, time). The paper presenting the data variable used includes the following variables as temporal variable (var_T), subject variable (var_Sub), and score variable (var_Sco).

The var_T shows the school years, with the unit is class and the school years from [2014–2015] to [2017–2018] respectively is the 6th, 7th, 8th and 9th grade. The JHS program is stipulated every student that will learn in 4 years, every year a student is allowed to take a class, the class is arranged in order from grade 6, grade 7, grade 8, then finally grade 9, students are not allowed to learn beyond class. Due to the order of such classes, the var_T is an ordered data type, the var_T consists of elements C_6, C_7, C_8, C_9. Let var_T is the temporal variable and call t_1, t_2, t_3, t_4 are the set of elements of var_T. The var_Sub includes combinations of subject elements such as M, P, B, L, H, G, FL, CE, and T. The var_Sub only shows the difference between subjects or equally in the hierarchy. Let var_Sub is the subject variable and call m_1, m_2, $m_{3,\ \ldots,}$ m_9 are the set of elements of var_Sub. The var_Sco shows average scores with numerical values from 0.0 to 10.0 scores of subjects. For score variables, we can perform basic calculations such as adding, subtracting, multiplying, dividing, and performing comparisons between numerical scores. Therefore, the var_Sco is the ratio of data type. The var_Sco consists of a set of values of $\{0.0,\ 3.6,\ 3.7, \ldots 10\}$. Let var_Sco be a score variable and call $s_1, s_2, s_3, \ldots s_n$ is a set of elements of var_Sco.

- The var_T = $\{C_6, C_7, C_8, C_9\}$, with C_6, C_7, C_8, C_9 turn in order is grade 6, grade 7, grade 8, and grade 9, in it: $(t_1, t_2, t_3, t_4) \in$ var_T, $t_1 = \{C_6\}, t_2 = \{C_7\}, t_3 = \{C_8\}, t_4 = \{C_9\}$.
- The var_Sub = {M, P, B, L, H, G, FL, CE, T}, in it: (sub_1, sub_2, sub_3, … sub_9) var_Sub, sub_1 = {M}, sub_2 = {P}, sub_3 = {B}, sub_4 = {L}, sub_5 = {H}, sub_6 = {G}, sub_7 = {FL}, sub_8 = {CE}, sub_9 = {T}.
- The var_Sco = $\{0.0, 3.6, 3.7, \ldots 10\}$, with 0.0, 3.6, 3.7,…10 is the numerical value indicating the average score, in it: $(s_1, s_2, s_3, \ldots s_n) \in$ var_Sco, s_1 = [0.0–2.0], s_2 = [2.0–4.0], s_3 = [4.0–6.0], s_4 = [6.0–8.0], s_5 = [8.0–10].

Users can perform the assessment of JHSSs' results through observing visual graphs. The visual graph is a graph showing students' learning data into images that make it easy for observers. In order for the observation to gain more information and knowledge, the user must combine with the regulations in assessing students' learning outcomes, which are clearly stated in the circular 58 of Ministry of Education and Training (MET) [16]. In order for the observation to gain more information and knowledge, the user must combine with the regulations in assessing students' learning outcomes, which are clearly stated in the circular 58 of the MET. The provisions in circular 58 combined with visual graphs observation will help the observer understand better the learning process of JHSSs. These provisions are expressed in the following terms.

- Students' learning ranking is divided into 5 categories as good, fair, average, weak, and poor is defined in item 2, article 5, chapter 3 in circular 58 of the MET.
- Evaluating in a way that scores for some subjects such as M, P, B, L, H, G, FL, CE, and T through tests are given the scores on a scale of 0 to 10 scores specified in point d, clause 1, article 6, chapter 3 in circular 58 of the MET.

- The calculation of the subject results for each semester and the whole school year for the subjects assessed by giving the scores specified in point a-b-c, clause 1, article 10, chapter 3 in circular 58 of the MET.
- The calculation of the score average of the semester or the whole school year gives in the integer or decimal number being taken to the first decimal after rounding the number specified in clause 3, article 11 in the circular 58 the MET.
- Ranking semester and a whole year of students are defined in article 13, chapter 3 in circular 58 of the MET. To classify students, it must be based on the following classification criteria.
  - Good: If the average score is 8.0 or higher and there are no subjects that the average score is below 6.5 points and the other rules are attached.
  - Fair: If the average score is 6.5 or higher and there are no subjects, the average score is below 5.0 and the other rules are attached.
  - Average: If the average score is 5.0 or higher and there are no subjects, the average score is below 3.5 and the other rules are attached.
  - Weak: If the average score is 3.5 or higher and there are no subjects, the average score is below 2.0 and the other rules are attached.
  - Poor: The other cases.

3.2.2 Proposing a Visual Data Analysis Model of JHSSs Data

Visual variables are divided into two types, plane variables and retinal variables [4]. In which, plane variables are dimensions, in a plane variable coordinate system is the coordinate system; retinal variables include variables such as magnitude, brightness, fineness, color, direction, and shape; retinal variables enhance capacity the visualization of data for people to observe, perceive the most information from data, extract useful information from data.

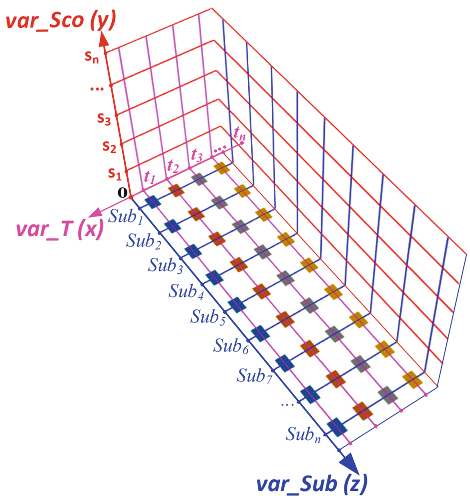

Fig. 5. Visual data analysis model of 3 student data variables based on 3D coordinate system

From the data variables described above, we propose to build a visual data analysis model of JHSSs' data through three variables including temporal variables (var_T), subject variables (var_Sub), and score variables (var_Sco). The construction of visual data analysis model of JHSSs' data by visual images will help people to observe easily and increase the ability to extract a lot of potential information inside the data. Each data variable will be represented as a plane variable and the corresponding visual variable. The visual data analysis model of JHSSs' data (see Fig. 5) helps users to visualize student data intuitively, from which users can answer some analytical questions themselves and evaluate development orient of students' learning ability of JHS. Each data variable is represented as a visual variable and a visual variable is represented on a column of the visual data analysis model.

With the visual analysis of student data based on three data variables (see Fig. 5), it will allow users to observe and extract a lot of potential information within student data. With y coordinate axis denotes var_Sco are score average of the whole school year of the subjects. With x coordinate axis denotes var_T changes of the school year. With z coordinate axis denotes var_Sub. Each of the three values of the var_Sub, var_Sco, and var_T will be represented as a column in the 3-dimensional (3D) space Oxyz (see Fig. 5). In which, the position in the Oxz plane represents the value pair of the var_Sub and var_T; y coordinate axis represents the height of the var_Sco, the value of the var_Sco will always change over time; the relationship of altitude of the same var_Sub that changes over time will help the observer see the change of the var_Sub of that student over the school years. Based on the visual data analysis model of three data variables (var_T, var_Sub, var_Sco) of students (see Fig. 5), we propose and develop some visual data analysis criteria used to observe students' learning outcomes that help users discover information and seek knowledge from student data.

Criteria 1: Users can combine a set of any three values in $\{(sub_1, \ldots sub_n), (t_1, \ldots t_n), (s_1, \ldots s_n)\} \in 3$ variables (var_T, var_Sub, var_Sco) together to discover students' learning information. For example, a combination of three values (sub_3, t_1, s_3) in $\{(sub_1, \ldots sub_n), (t_1, \ldots t_n), (s_1, \ldots s_n)\} \in 3$ variables (var_T, var_Sub, var_Sco) located in the Oxz plane, this combination of a value three (sub_3, t_1, s_3) constitutes the learning result of a subject (sub_3) of a class (t_1) with an average score (s_3) of a student, see Fig. 5.

Criteria 2: Users can combine a set of any three values of three variables (var_T, var_Sub, var_Sco), we can choose a set of three values in the Oxz plane to observe the learning results of a student. For example, we choose the set of three values of three variables (var_T, var_Sub, var_Sco) as follows {(sub_1, t_4, s_5), (sub_2, t_4, s_4), (sub_3, t_4, s_5), (sub_4, t_4, s_4), (sub_5, t_4, d_4), (sub_6, t_4, s_3), (sub_7, t_4, s_5), (sub_8, t_4, s_3), (sub_9, t_4, s_3)} in the Oxz plane. Combining these triple set at the same time (t_4) will help observers discover the learning outcomes of the subjects on the same school year (t_4), see Fig. 5.

Criteria 3: Users can combine sets of three values $\{(sub_1, \ldots sub_n), (t_1, \ldots t_n), (s_1, \ldots s_n)\} \in (var\_T, var\_Sub, var\_Sco)$, and then can perform comparison operations (=, <, $\leq$, >, $\geq$, $\neq$) based on proportional values (var_Sco of s_1, s_2, s_3 … s_n) to discover information and seek knowledge from results learning of students. Compare pairs of value in turn {**(sub_1, t_1, s_3)** and **(sub_1, t_4, s_5)** based on var_Sco s_3 and s_5; **(sub_2,**

t_1, s_3) and **(sub$_2$, t_4, s_4)** based on var_Sco s_3 and s_4; **(sub$_3$, t_1, s_3)** and **(sub$_3$, t_4, s_5)** based on var_Sco s_3 and s_5; **(sub$_4$, t_1, s_3)** and **(sub$_4$, t_4, s_4)** based on var_Sco s_3 and s_4; **(sub$_5$, t_1, s_3)** and **(sub$_5$, t_4, s_4)** based on var_Sco s_3 and s_4; **(sub$_6$, t_1, s_3)** and **(sub$_6$, t_4, s_3)** based on var_Sco s_3 and s_3; **(sub$_7$, t_1, s_2)** and **(sub$_7$, t_4, s_5)** based on var_Sco s_2 and s_5; **(sub$_8$, t_1, s_3)** and **(sub$_8$, t_4, s_3)** based on var_Sco s_3 and s_3; **(sub$_9$, t_1, s_3)** and **(sub$_9$, t_4, s_3)** based on var_Sco s_3 and s_3}, see Fig. 5.

Criteria 4: Users can select sets of any three values $\{(sub_1,\ldots sub_n),(t_1,\ldots t_n),(s_1,\ldots s_n)\} \in$ three variables (var_T, var_Sub, var_Sco) according to any subject $(sub_1,\ldots sub_n)$ in each school year $(t_1,\ldots t_n)$. For example, the user selects the set of three value sets {(sub_3, t_1, s_3), (sub_3, t_2, s_2), (sub_3, t_3, s_5), (sub_3, t_4, s_5)} $\in$ 3 variables (var_T, var_Sub, var_Sco) by subject (sub_3) to explore information and seek knowledge from students' learning result, see Fig. 5.

Criteria 5: Users can select sets of any three values $\{(sub_1,\ldots sub_n),(t_1,\ldots t_n),(s_1,\ldots s_n)\} \in$ three variables (var_T, var_Sub, var_Sco) in two any subjects $(sub_1,\ldots sub_n)$ in each school year $(t_1,\ldots t_n)$, then it is possible to perform arithmetic comparisons ($=$, $<$, $\leq$, $>$, $\geq$, $\neq$) on the proportional value var_Sco to discover information and search for knowledge from the students' learning ability, see Fig. 5.

3.3 Multidimensional Cube Representing Student Data

As observed in Table 1 above, we noticed the structure of this data table is like a 2D matrix consisting of columns and lines. Each column represents an attribute that includes a set of values of the variable and each row is a data set representing a relationship between variables. Visual data is studied as a transformation of an element of the table's data variable into a visual variable as an axis or in other words each axis represents an element of the variable. Each axis displays the data of a column as an element of the data variable. The intersecting axes of var_T, var_Sub, and var_Sco create a multidimensional cube that visually represents student data.

Currently, graphs have many forms, including column graphs, line graphs, pie graphs, bar graphs, etc. In particular, one of the most common graphs used to compare numerical values is the column graph. In a column graph, the data is arranged in columns and rows on the worksheet. Column graphs often display genres across the horizontal axis and numerical values along the vertical axis. Because the column graph has advantages in data analysis, this type of graph is used to denote the values of the subject variables such as the average score of the school year of M, P, B, L, H, G, FL, CE, and T in JHS year. The columns indicate differences in height, color, and similarity in the width of the columns.

In order for a visual analysis of JHS data for good results, we performed a visual representation of JHSSs' data on 2D column graphs or 3D column graphs. For a 2D column graph (see Fig. 6), on each column of the graph shows the subject value, the different colors represent the class by each subject, in this graph the data variables are continuous representation on one axis, this should make it difficult for the user to observe and compare the relationship of the data when the data table has many data.

In order to assess the inner learning abilities of each student, it is necessary to rely on the students' learning outcomes of the JHS years, the data variables must be represented on the same visual graph by visual data analysis model. The 2D column graph has the above limitations, to solve the limitations of 2D column graphs, a 3D coordinate axis system is attached to data variables, it is called visual data analysis model of three student data variables based on 3D coordinate system (see Fig. 5). In it, the x coordinate axis represents the var_T (unit is the class), the z axis coordinate represents the var_Sub in turn, and the coordinate axis expresses the average score of the students' subjects. For 3D column graphs (see Fig. 5), on each column represents the average score of each subject by grade 6, grade 7, grade 8, and grade 9, the subjects are shown visually of each individual school year makes it easy for users to observe and compare the subject of the school year compared to the same subject in other school years, or to compare the subjects together in same school year, see Figs. 7 and 8 representing Pham Thao Nguyen's and Pham Dang Khoi's student learning outcomes in turn.

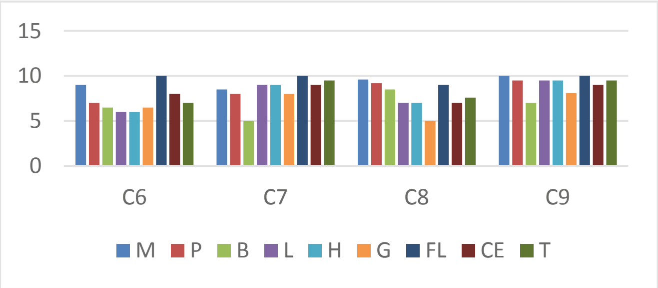

Fig. 6. 2D visual graph showing Pham Thao Nguyen's student learning results (Color figure online)

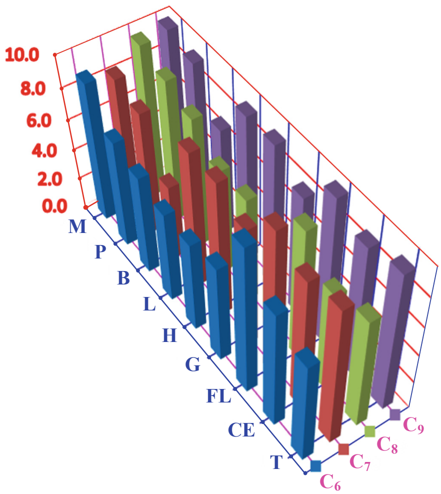

Fig. 7. 3D visual graph showing Pham Thao Nguyen's student learning results

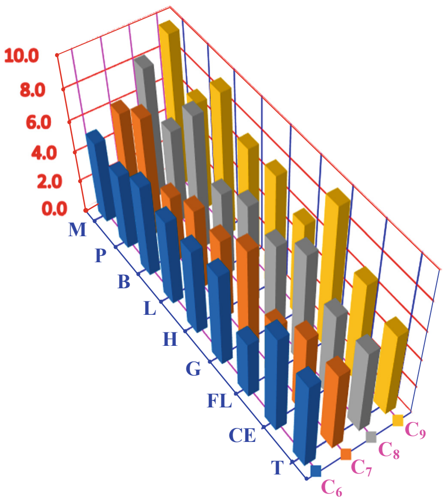

Fig. 8. 3D visual graph showing Pham Dang Khoi's student learning results

4 Analyzing Student Data

4.1 Analyzing Data by Questions

Every object that existing in this real world has a certain use meaning, people perceive it through visual recognition to view - think visually from the existing image of objects in the real world. From there, people will ask visual data analysis questions because of the need to explore objects. The visual data analysis question consists of two main parts as hypothesis and conclusion. In it, people play a major role in the process of analyzing information discovery and finding knowledge from data by applying data analysis process including the following steps [17].

- Step 1: Build questions. In this step, the researcher builds visual data analysis questions to find information or knowledge from the data based on visual graphs.
- Step 2: Choose analytical methods. The analytical method consists of two main methods, which is the method of using model method and visualization method. The modeling method is to use mathematical models to explore information and knowledge from data. Visualization method is a method of combining computer and human knowledge to extract information and knowledge from data by viewing – thinking way.
- Step 3: Prepare data to apply the method. In this step, data is collected through a specific origin. Data is collected, then discarded unnecessary data types and left only the data types for thematic analysis.
- Step 4: Apply methods for data sets which were collected. After the data is available, we apply the method selected in step 2 to the analytic process of existing data.
- Step 5: Explain and evaluate the results. Researchers observe and extract information and knowledge by looking at visual graphs. Information and knowledge are extracted when observing visual graphs, and then perform the evaluation of achieved results.

In the above process, building a question is an initial step, a basis for promoting the data analysis process. It can be said that data analysis cannot be performed if no questions are asked. Through a survey of research works, we found many author groups propose ways to classify data analysis questions. Based on the relevance of the question with data variables and data variable values, Bertin divided analytical questions into three levels, elementary level, intermediate level, and the overall level [4]. In which, the question at the elementary level relates to a value of a variable, the question at the intermediate level relates to a value group of a certain variable, the question is at an overall level related to can variables. Andrienko classifies analytical questions into two levels, elementary questions, and synoptic questions [17]. In addition, approach the use of data variables and data of data variables to answer questions for information extraction or knowledge search objectives, the authors Thi Nguyen et al. classified the analysis question into three levels [6], elementary questions related a value of a data variable, variation questions related to a group or whole value of a certain variable to understand the properties or rules of variable variation, correlation questions related many data variables to find the correlation between data variables to detect new rules of relationship between attributes.

4.2 Visual Data Analysis Questions

To visually analyze the data of JHSSs, the paper proposes to build a visual data analysis model of JHSSs through data variables analyzed by visual analysis approach (see Fig. 5). To assess the true learning abilities of a student, we perform visual analysis of student data on visual graphs, thereby posing some questions to verify the visual graph is reasonable and effective. Users can ask questions to analyze student data and can answer this question based on visual graphs, from which users extract useful information by viewing - thinking visual graphs to serve specific purposes. To better understand the visual graph, let's look at a visual graph that visually demonstrates a student's learning score data and poses some questions to analyze data of JHSSs and see the student's development trend.

Users ask questions and answer questions themselves when observing the visual graphs in Figs. 7 and 8 for the purpose of exploring information and seeking knowledge from visual graphs of visual analysis JHSSs. Types of questions that users can ask include primary questions, variable questions, and correlated questions, along with five visual data analysis criteria proposed by the paper in Sect. 3.2.2. These types of questions have been detailed below and now we perform a visual analysis of student data below.

Visual data analysis questions are questions asked by users, analysts, or etc. when observing things, phenomena, etc. Visual data analysis questions used to explore information and knowledge in existing large data sets. Because of the essence of the data that users or analysts can be ask for appropriate analytical questions as they observe, the visual data analysis questions are divided into different categories and each segment question and the data area consists of two parts as the hypothetical part and the conclusion part. To highlight the students' learning ability development trend of JHSSs, the paper incorporates the criteria for visual data analysis of student data in Sect. 3.2.2 and the types of questions of the author group Thi Nguyen [6] to study data analysis of JHSSs as follows.

4.2.1 Coordinating the Type of Primary Questions and Criteria 1

Question 1: How many marks does Pham Dang Khoi get when he learns **F**oreign **L**anguage (**FL**) in grade 9? See Figs. 8 and 9.

- Supposition: Pham Dang Khoi, **FL**, grade 9
- Conclusion: Marks
- Looking at Fig. 8, parents or analysts extract Pham Dang Khoi's student learning results is shown in Fig. 9 based on coordinating between primary questions type and criteria 1 for parents results in Fig. 9.

Question 2: Pham Thao Nguyen has the results of learning in grade 7 which is the average subject and which subject is good? See Figs. 7 and 10.

- Supposition: Pham Thao Nguyen, grade 7.
- Conclusion: Which is the average subject and which subject is good?
- Looking at Fig. 7, parents or analysts extract Pham Thao Nguyen's student learning results is shown in Fig. 10 based on coordinating between type of primary questions and criteria 1 for parents results in Fig. 10.

4.2.2 Coordinating the Type of Variation Questions and Criteria 4, Criteria 5

Question 3: In the 6th, 7th, 8th, and 9th grade, how does Pham Dang Khoi tend to develop **M**athematic (M)? See Figs. 8 and 11.

- Supposition: The 6th, 7th, 8th, and 9th grade, Pham Dang Khoi, **M**.
- Conclusion: Tend to develop **M**.
- Looking at Fig. 8, parents can see that the direction of Pham Dang Khoi's development of **M** is shown in detail in Fig. 11 depend on combining between type of variation questions and criteria 4 and 5 for supporting parents and analysts results in Fig. 11.

4.2.3 Coordinating the Type of Correlation Questions and Criteria 2, Criteria 3

Question 4: Which subjects are Pham Dang Khoi good at in grade 9? What relationships do they have with each other? See Figs. 8 and 12.

- Supposition: Pham Dang Khoi, good subjects, grade 9
- Conclusion: Subject, the relationship between subjects
- Looking at Fig. 8, parents can know that which subjects Pham Dang Khoi good at in grade 9 are and what relationships they have with each other are detailed in Fig. 12 rely on coordinating between type of correlation questions and criteria 2 and 3 for serving parents and analysts results in Fig. 12.

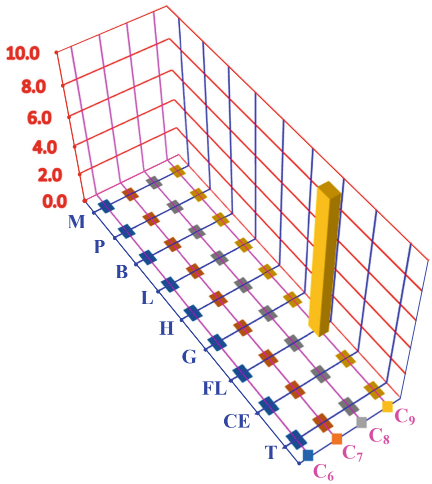

Fig. 9. Results of visual analysis Pham Dang Khoi's data

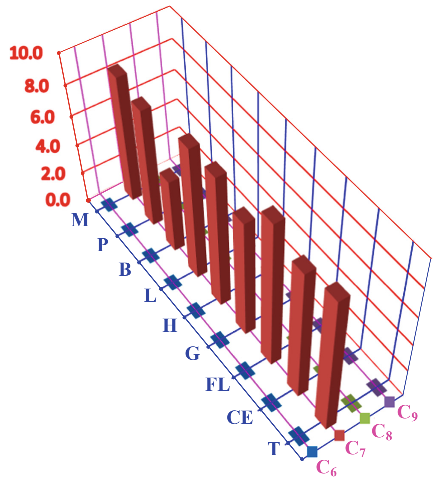

Fig. 10. Results of visual analysis Pham Thao Nguyen's data

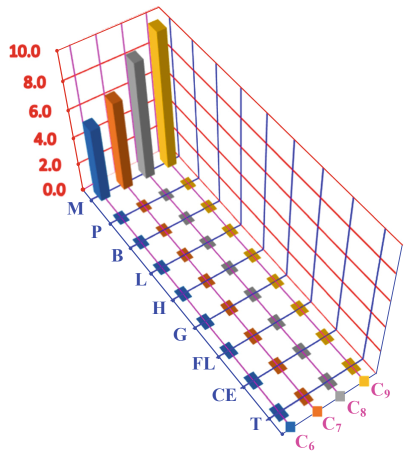

Fig. 11. Results of visual analysis Pham Dang Khoi's data

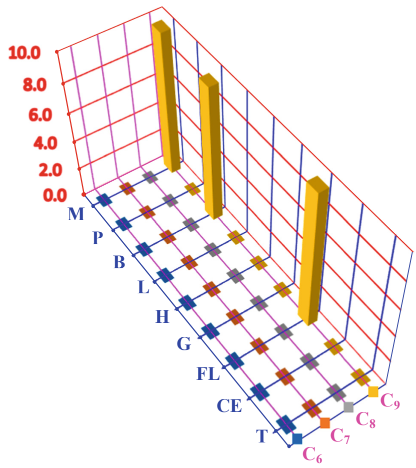

Fig. 12. Results of visual analysis Pham Dang Khoi's data

5 Conclusion

This paper has systematized the research works related to the areas of data visualization, visual analysis and processes of data visualization or visual analysis. This paper proposes to build a VAS, visual data analysis model and visual data analysis criteria for JHSSs' data whereby the paper applies to information discovery and knowledge search from JHSSs' data. Parents or data science analysts use VAS, visual data analysis model, visual data analysis criteria and visual graphs in combination with visual analysis of student data for information discovery, and seeking knowledge towards the knowledge chain as well as the future career for each student. This paper has also focused on building student-related data variables to become visual variables, then proposing visual data analysis criteria and multidimensional cube representing student data. Based on the visual graphs, we build a number of visual data analysis questions that help parents or analysts answer visual data analysis questions and understand the meaning of students' learning outcomes of JHS when observing the visual graph by visual viewing - thinking method.

With a VAS of multivariate data based on visual graphs to help parents or analysts discover information and search for knowledge from JHSSs data when observing the visual graph an effective way. With the visual data analysis model of multivariate data coordinated with the visual graph, it helped us build some visual data analysis criteria of JHSSs data, these criteria combine the provisions in circular 58 of the MET to help parents or data science analysts use it as a basis for applying to the process of exploring information and searching knowledge from data sets related to JHSSs. The visual data analysis model of JHSSs' multivariate data is shown in separate academic years to help parents or analysts perform arithmetic comparisons between subjects of this school year compared to subjects of another school year or making comparisons of subjects together in the same school year is easy.

In addition, the VAS and visual data analysis model of JHSSs' multivariate data can also be developed and expanded to compare the learning outcomes of many students in the same subject and class, or at the same level and specially developed for the High School student data set or develop and apply to the management data set in the University aimed at the knowledge chain as well as career orientation for future pupils and students.

References

1. Stuart, C.K., Jock, M., Ben, S.: Readings in Information Visualization: Using Vision to Think. Academic Press, Norwell (1999)
2. Card, S.T., Mackinlay, J.D., Scheiderman, B.: Readings in Information Visualization, Using Vision to Thinks. Graphic Press, Cheshire (1990)
3. Stevens, S.S.: On the theory of scales of measurement. Science **103**, 677–680 (1946)
4. Bertin, J.: General theory, from semiology of graphics. In: Dodge, M., Kitchin, R., Perkins, C. (eds.) The Map Reader. Theories of Mapping Practice and Cartographic Representation. Wiley, pp. 8–16 (2011)
5. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube integrated retinal variable to visually represent multivariable data. EAI Endorsed Trans. Context.-Aware Syst. Appl. **4**, 1–8 (2017). https://doi.org/10.4108/eai.6-7-2017.152757
6. Thi Nguyen, H., Thi Pham, T.M., Thi Nguyen, T.A., Thi Tran, A.V., Vinh Tran, P., Van Pham, D.: Two-stage approach to classifying multidimensional cubes for visualization of multivariate data. In: Cong Vinh, P., Alagar, V. (eds.) ICCASA/ICTCC -2018. LNICST, vol. 266, pp. 70–80. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-06152-4_7
7. Thomas, J.J., Cook, K.A.: Illuminating the path the research and development agenda for visual analytics. National Visualization and Analytics Center – NVAC (2005)
8. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Trans. Vis. Comput. Graph. **20**(12), 1604–1613 (2014). https://doi.org/10.1109/tvcg.2014.2346481
9. Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F.: Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics Association, Goslar (2010). ISBN 978-3-905673-77-7
10. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, John T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_7
11. Wijk, J.J.: The value of visualization. In: IEEE Visualization, p. 11 (2005)
12. Card, S.T., Mackinlay, J.D., Scheiderman, B.: Readings in Information Visualization, Using Vision to Think. Academic Press, Norwell (1999)
13. Bertini, E., Tatu, A., Keim, D.: Quality metrics in high-dimensional data visualization: an overview and systematization. IEEE Trans. Vis. Comput. Graph. **17**(12), 2203–2212 (2011). https://doi.org/10.1109/TVCG.2011.229
14. von Szent-Gyorgyi, A.: American (Hungarian-born) Biochemist (Nobel Laureate, 1937) who said: "Discovery consists of seeing what everybody has seen and thinking what nobody has thought" ed. by Irving Good, The Scientist Speculates (1962)

15. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**(3), 37 (1996)
16. Ministry of Education and Training (MET), Abstract: “Released circular regulations in assessing and ranking of Secondary School Students and High School Students” in menu of legal documents at page 20, Effect: in force. https://en.moet.gov.vn/document/legal-documents/Pages/detail.aspx?ItemID=744. Published date: 12/15/2011, Date to: 01/26/2012, Official number: 58/2011/TT-BGDĐT. Accessed Apr 2019
17. Andrienko, N., Andrienko, G.: Exploratory Analysis of Spatial and Temporal Data – A Systematic Approach, p. 2006. Springer, Heidelberg (2006)