

Innovative Pedagogy for Teaching and Learning Data Visualization

Dr. Vetricia L. Byrd PhD, Purdue University at West Lafayette (COE)

Dr. Vetricia Byrd is an assistant professor in the Department of Computer Graphics Technology in the Polytechnic Institute at Purdue University in West Lafayette, Indiana. Dr. Byrd is the founder and organizer of BPViz: Broadening Participation in Visualization (BPViz) Workshops. Dr. Byrd has given numerous invited talks on visualization, and given numerous workshops nationally and internationally on visualization. Dr. Byrd received her graduate and undergraduate degrees at the University of Alabama at Birmingham, in Birmingham, Alabama which includes: Ph.D. in Computer and Information Sciences, Master's degrees in Computer Science and Biomedical Engineering and a Bachelor's degree in Computer Science. Dr. Byrd's research interests include: data visualization, data visualization capacity building, high performance visualization, big data, and collaborative visualization.

Innovative pedagogy for teaching and learning data visualization

Abstract

This paper describes a process for teaching and learning data visualization with the use of a series of worksheets for each stage of the data visualization process. The class meets twice a week, for 16-weeks with lecture on the first day of the week and in-person labs on the second day. The course consists of an online-lecture, and use of activity worksheets for in-class and homework assignments to foster student participation and engagement, and in-person labs.

Data visualization is a complex, intricate process, that fosters higher-order thinking and critical thinking in each stage of the process. Ordinarily, the class is in-person which allows for interactive engagement and organic dialog between faculty and students, which makes it easier to see when students are struggling with a concept. In the absence of these key elements, worksheets are used to introduce each stage of the data visualization process, student self-assessments are used to gauge self-efficacy and quizzes are used to assess competency. What is innovative about the approach is the automation of the process for students and faculty. Students complete a worksheet online and receive a copy of their responses by email with the option to generate a PDF version of their responses. Subsequently, the student submits the PDF version to Brightspace for review by the instructor. After submitting the worksheet, students complete a self-assessment survey to assess student's self-efficacy with content covered in class and reinforced in the worksheet. Worksheets coupled with self-assessments provide insight on student's data visualization capacity levels.

The goals of the worksheets are to enable students to understand essential elements of data visualization while fostering critical thinking throughout the process. Skills identified and showcased in each worksheet align with capabilities characteristic of higher-order thinking skills across the knowledge and cognitive dimensions of Bloom's Taxonomy hierarchy of learning. Worksheets are introduced in a linear manner for the novice; however, each worksheet is designed to stand alone to facilitate the iterative, dynamic nature of the data visualization process. Overall response to the worksheet method was positive. With 97% of Likert responses being agree total, 65% were strongly agree in response to capturing students' self-assessment of what was learned from the data visualization course. Three percent of students disagreed. The activity worksheets are used to inform pedagogy of data visualization. In this work, we describe the online-process, the worksheets, assignments and the ways in which faculty and students navigate the pandemic inspired teaching and learning environment to support higher-order thinking, and critical thinking skills, crucial to computing and engineering curriculums, among undergraduates.

Introduction

The purpose of this research is to introduce data visualization in a manner that encourages participation and engagement in a hybrid learning environment, for students with little to no background or experience visualizing data. The approach described consists of three key elements: 1) using a constructive learning environment to introduce data visualization, 2) introducing the multi-stage data visualization process incrementally to support learning and 3)

students' self-assessment of the process. The main research questions addressed in this work are:

RQ1: Do students, in an introductory data visualization course, understand basic ideas introduced for visualizing data?

RQ2: Do students, in an introductory data visualization course, understand core ideas associated with the data visualization process?

The goal of this research is to introduce students to essential concepts about visualizing data. After completing the course, students will: (1) use their understanding of the process to acquire, process and visualize data, and (2) students will understand their own individual visualization capacity and ways in which they can improve.

Six building blocks for the course are identified to facilitate understanding of the data visualization process, within the context of this work:

1. Visualizing data is a multi-stage process.
2. A basic level of understanding of the role of each stage in the process is required.
3. Output from one stage serves as input to another stage.
4. The data visualization process is an iterative process.
5. Data must be preprocessed before visualizing.
6. Creating insightful visualizations requires an understanding of the data.

These building blocks enable knowledge and skills needed to visualize different types of data as data increase in size and complexity. After completing the course, student will be able to demonstrate their factual and conceptual knowledge about the data visualization process:

1. The basic stages for visualizing data.
2. What happens in each stage of the visualization process.
3. What stages are likely to initiate the iterative nature of the process.
4. Different techniques used to better understand data.

After completing the course, students will be able to demonstrate the following procedural knowledge:

1. Demonstrate actions to acquire data.
2. Demonstrate the ability to change raw data into a useful format for further processing.
3. Implement procedure(s) to extract data of interest from a larger dataset.
4. Choose the appropriate visualization chart for the task/data.
5. Implement methods and techniques to improve the visualization.

Data Visualization Process

For this research, seven stages of visualization are identified and aligned with course outcomes described in previous work ([1]-[2]). The multi-stage data visualization workflow, adopted from [3], was chosen for its simplicity. There are many data visualization workflows to choose from; however, representing the process in seven stages, with self-explanatory names for each stage

provides a clear learning path for students to follow. Students who are new to data visualization are able to see the big picture with a clear view of deliverables. A second reason for choosing Fry’s data visualization workflow is the ease of alignment with basic ideas (BI) and core ideas (CI) identified for this research as seen in Table 1. The the first column in Table 1 (Basic Ideas) provides a high-level view of factual and conceptual knowledge students should know after completing the course. For instructors, the list of basic ideas serves as an outline of topics to cover. Basic ideas build the foundation for core ideas. The second column in Table 1 (Core Ideas) details what students should be able to do, as evident by the action verbs (identify, discuss, demonstrate, explain, examine, determine) leading each idea statement.

Table 1 Basic and core ideas for an introductory data visualization course.

Basic Ideas (BI): What students should know	Core Ideas (CI): What students should be able to do
BI1: Basic stages for visualizing data.	CI1: Identify the stages of the visualization process CI2: Discuss how output from one stage impacts other stages in the process.
BI2: What happens in each stage of the visualization process.	CI3: Demonstrate the ability to apply each stage of the visualization process.
BI3: Iterative nature of the data visualization process.	CI4: Explain the interactions between stages.
BI4: Different techniques used to better understand data.	CI5: Examine data in its most basic form.
BI5: Chart types impact how the visualization will be received by the viewer.	CI6: Determine the value of the visualization process.

Constructive Learning Environments

The intricate and iterative nature of the data visualization process lends itself to careful instruction with a focus on understanding the role of each stage in the process. A constructivist learning environment is utilized as the pedagogical underpinnings of this work. Constructivist learning theory is based on the belief that learning occurs when learners are actively involved in a process of knowledge construction [4]. Jonassen [5] describes four types of constructivist learning environments (CLEs):

- (1) question-based learning (starting with a question) [4]
- (2) case-based learning (acquire knowledge and requisite thinking skills) [6],
- (3) project-based learning (integrating complexity of multiple cases) and
- (4) problem-based learning (requiring self-directed learning). [6]

The case-based learning approach is appropriate for this work. Students are introduced to the data visualization process using a series of worksheets. The goals of the worksheets are to enable students to understand essential elements of data visualization while fostering critical thinking throughout the process. Skills identified and showcased in each worksheet align with capabilities characteristic of higher-order thinking skills across the knowledge and cognitive dimensions of Bloom's Taxonomy hierarchy of learning. Worksheets are introduced in a linear manner for the novice; however, each worksheet is designed to stand alone to facilitate the iterative, dynamic nature of the data visualization process. This paper describes a process and outcomes from for teaching data visualization with the use of a series of worksheets for each stage of the data visualization process.

Design Activity Worksheets for Visualizing Data

Content and layout for the activity worksheets were initially created using Microsoft Word as a template. To facilitate online access, each worksheet was re-created in Qualtrics. Qualtrics is an online survey tool that saves responses for later analysis.

Each stage of the visualization process is introduced incrementally to allow for in-depth examination and discussion of important concepts. This approach allows instructors to identify and teach *factual*, *conceptual* and *procedural knowledge* students should know about the data visualization process. *Factual knowledge* are the basic ideas students must know to be acquainted with data visualization in order to solve data visualization challenges [7]. *Conceptual knowledge* is the understanding of interrelationships that exist between the stages within the larger structure of the process that enable the stages to function together [7]. This approach allows students to recognize the *procedural knowledge* they should be able to demonstrate in each stage of the process. *Factual*, *conceptual* and *procedural knowledge* are critical to understanding and implementing the data visualization process.

In the context of this research, *factual knowledge* refers to the list of basic ideas shown the first column in Table 1. *Conceptual knowledge* refers to the list of core ideas shown in the second column in Table 1. Procedural knowledge is gained by the application of the visualization process. Although introduced incrementally as a linear approach, the data visualization process is non-linear. In fact, a visualizer can be in any stage of the process at any given time. This element of the process is introduced to students after learning the process.

Innovative Pedagogy

The innovative pedagogy of this work is represented in the semi-automated design activities for students and faculty. Students complete a worksheet online and receive a copy of their responses by email with the option to generate a PDF version of their responses. Subsequently, the student submits the PDF version to Brightspace for review by the instructor. Brightspace is an online management system for online teaching and learning.

Methodology

This research was implemented in an undergraduate data visualization course. The class meets twice a week, for 16-weeks with lecture on the first day of the week and in-person labs on the

second day. The course consists of an online-lecture, and use of activity worksheets for in-class and homework assignments to foster student participation and engagement, and in-person labs. A total of 37 students participated in this research.

The first step was to identify a dataset student can explore easily that is interesting enough to keep students engaged in the process. Nutritional data on 80 cereal products from the Kaggle.com website was selected as the dataset used for in-class activities for introducing data visualization. The dataset consists of one file containing 16 columns of data, with variety of different types of data (integers, strings, etc.,).

In the second step, the data visualization process was introduced to the class. The first six weeks of class was dedicated to introducing the data visualization process. Each week focused on one stage, as seen in Table 2. Activity worksheets were developed to guide students through each stage of the process using the cereal data as a training dataset. The Appendix includes examples of the Acquire, Parse and Mine Activity Worksheets using the cereal training dataset.

Table 2 Course timeline mapped to learning and instructional activities.

Timeline	Topic Description	Learning Activity	Instructional Activities
Week 1	Acquiring Data	Acquisition	Acquire Worksheet
Week 2	Parsing Data	Examining data	Parse Worksheet
Week 3	Mining Data	Identifying patterns and trends	Mine Worksheet
Week 4	Filter and Represent Data	Extracting and representing data	Filter & Represent Worksheet
Week 5	Critique	Constructive feedback	Student seek feedback from colleagues/classmates
Week 6	Refining Data	Improve data visualizations	Refine Worksheet
Week 7	Bringing it all together	Storytelling	Articulate the story supported by visualizations

In the interest of keeping the focus on the data visualization process rather than learning multiple tools, Microsoft Excel is used for data wrangling and for creating basic visualizations. Most students have some basic knowledge of Microsoft Excel. Once students become familiar with the process of visualizing data, other tools for visualization are introduced. The sections below provide examples how the cereal dataset is utilized in each activity worksheet.

Week 1: Acquire

The multi-stage data visualization process is introduced. **Learning activity:** Acquisition. Students use the Acquire Activity Worksheet as a guide for completing the acquire stage using the using the acquire activity worksheet. **Learning outcome:** students should be able to describe what happens in the acquire stage, describe how to locate and identify reputable data sources,

demonstrate actions taken to acquire data, explain how the acquire stage impacts the visualization, explain how data is acquired (the process). Students should be able to plan, generate and produce data for a given project. Figure 1 shows a snapshot of the cereal dataset downloaded in .csv format from Kaggle.com.

Detail Compact Column										10 of 16 columns
name	mfr	type	calories	protein	fat	sodium				
100% Bran	N	C	70	4	1	130				
100% Natural Bran	Q	C	120	3	5	15				
All-Bran	K	C	70	4	1	260				
All-Bran with Extra Fiber	K	C	50	4	0	140				
Almond Delight	R	C	110	2	2	200				
Apple Cinnamon Cheerios	G	C	110	2	2	180				
Apple Jacks	K	C	110	2	0	125				
Basic 4	G	C	130	3	2	210				
Bran Chex	R	C	90	2	1	200				
Bran Flakes	P	C	90	3	0	210				
Cap'n Crunch	Q	C	120	1	2	220				
Cheerios	G	C	110	6	2	290				
Cinnamon Toast	C	C	120	1	2	210				

Figure 1 Snapshot of the Cereal Dataset (.csv file) "Acquired" from <http://kaggle.com>

Week 2: Parse

Learning activity: Examining data. Students use the Parse Activity Worksheet as a guide for breaking data into its basic parts for further processing. **Learning outcome:** students should be able to describe what happens in the acquire stage, describe how to locate and identify reputable data sources, demonstrate actions taken to acquire data, explain how the acquire stage impacts the visualization, explain how data is acquired (the process). Students should be able to plan, generate and produce data for a given project.

Figure 2 shows the parsing for the cereal dataset. The first column lists each of the 16 variables. The second column indicates the variable type (string, character, integer, etc.). The third column show the range of values for each variable. The parsing stage allows for a closer examination of the data as well as an opportunity to determine valid values for each variable. For example, in Figure 2 shows a number of variables with negative numbers; however, it is unlikely negative numbers are valid values for the cereal dataset. The Parse Activity Worksheet asks students to list any assumptions they may have about the data based on what is known about the data so far.

Variable	Type	Values
Cereal	String (38)	a-z, A-Z, alpha numeric
Manufacturer	Character (1)	Q, P, G, K, R, N
Type	Character (1)	C-cold, H-Hot
Calories	Integer	50 – 160
Protein	Integer	1-6
Fat	Integer	0-5
Sodium	Integer	0 – 320
Fiber	Integer	0 – 5
Carbohydrates	Integer	-1 – 23
Sugars	Integer	-1 – 15
Potassium	Integer	-1 – 330
Vitamins	Integer	0 – 100
Weight	Integer	-1 – 1.5
Shelf	Integer	1 – 3
Cups	Floating point	-1 – 1.5

Figure 2 Sample parsing of cereal dataset.

The most frequent assumptions include: the value of -1 for carbohydrates, sugars and potassium is a data entry error, Cups is serving size and measurements are in grams. The list and number of assumptions will vary for each person.

Week 3: Mine

Learning activity: Identifying patterns and trends. Students use the Mine Activity Worksheet as a guide for learning how to identify basic trends and patterns in the data. **Learning outcome:** students should be able to describe what happens in the acquire stage, describe how to locate and identify reputable data sources, demonstrate actions taken to acquire data, explain how the acquire stage impacts the visualization, explain how data is acquired (the process). Students should be able to plan, generate and produce data for a given project.

As seen in Figure 3, in the mine stage basic statistical functions are identified for further analysis of each variable. In the class, Microsoft Excel is used to examine the data at this level. Excel functions are listed in the fourth column are specific to the variable data type. For example, the MAX function can be used to find the maximum length of the variable Cereal which is string. An initial visualization might consist of the number of cereals produced by each manufacture. In this case, the COUNTIF function could be used to determine the number of cereals for each manufacturer and if the cereal is hot or cold. The remaining variables can be further examined using the basic statistical functions to calculate the maximum, minimum and average values.

Week 4: Filter & Represent

Filter and represent are two separate stages in the data visualization process. The two stages are combined into one worksheet for the Design Activity Worksheet method.

Learning activity: Extracting and representing data. Students use the Filter & Represent Activity Worksheet as a guide for learning how to identify data pertinent to the project and extract only the data needed to address the visualization challenge. Once the data is extracted, students demonstrate their ability to graphically represent the data. **Filtering Learning**

outcomes: students should be able to describe what happens in the filter stage, describe and extract data of interest from a larger dataset, implement procedure(s) to extract and normalize data if needed, examine if only the data of interest was extracted, examine filtered data and

Variable	Type	Values	Functions
Cereal	String (38)	a-z, A-Z, alpha numeric	Max (Len) to find maximum string length
Manufacturer	Character (1)	Q, P, G, K, R, N	COUNTIF to determine number of cereals for each manufacturer
Type	Character (1)	C-cold, H-Hot	COUNT IF to determine number of cereal types (hot vs cold)
Calories	Integer	50 – 160	MAX, MIN, AVERAGE
Protein	Integer	1-6	MAX, MIN, AVERAGE
Fat	Integer	0-5	MAX, MIN, AVERAGE
Sodium	Integer	0 – 320	MAX, MIN, AVERAGE
Fiber	Integer	0 – 5	MAX, MIN, AVERAGE
Carbohydrates	Integer	-1 – 23	MAX, MIN, AVERAGE
Sugars	Integer	-1 – 15	MAX, MIN, AVERAGE
Potassium	Integer	-1 – 330	MAX, MIN, AVERAGE
Vitamins	Integer	0 – 100	MAX, MIN, AVERAGE

Figure 3 Basic descriptive statistical functions identified for further examination of the dataset.

determine the value of filtered data (i.e., if it is useful for the project). **Represent Learning**

Outcomes: students should be able to describe what happens in the represent stage, describe what stages are impacted by the represent stage, use the appropriate tools to effectively communicate data, choose the appropriate chart (layout, etc.,) determine if sufficient data is available to visually represent data in a way that informs viewers, determine the best way to represent data, determine if additional data or more specific data is needed (the need to acquire more data), determine if the dataset should be smaller/larger/broader (return to the filter stage) before moving on to the refine stage. Lastly, students should be able to generate visualizations based on data received from previous stages using the appropriate model (chart, layout, etc.,) for the data.

As seen in the previous stages, the represent stage is the first time in the process where data are visually displayed in a graphical way. In Figure 4, four of the 16 variables in the cereal dataset are visually represented by bar charts.

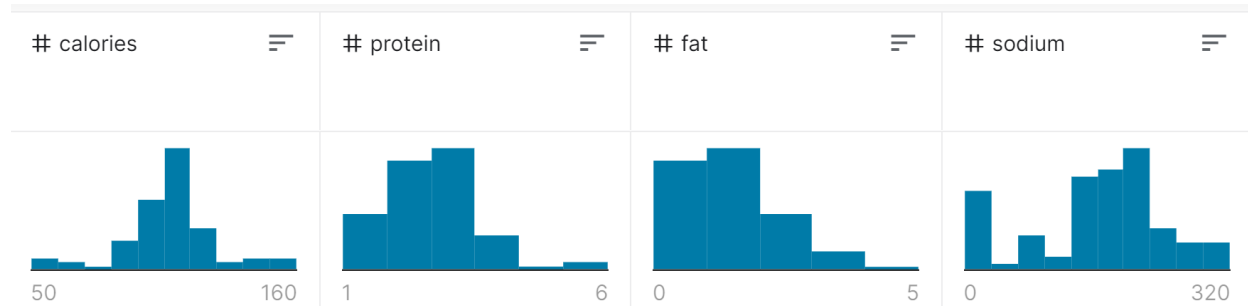


Figure 4 Visual representation of calories, protein, fat and sodium across all cereals (Kaggle.com)

Week 5: Critique

Learning activity: Constructive feedback. **Learning Outcome:** Students learn how others view and interpret their work. This step helps students if what they intended to convey translates to viewers. Output from the critique stage is qualitative and is used to inform the next stage, Refine. An example critique could be a recommendation to look more closely at the number of cereal produced by each manufacturer.

Week 6: Refine

Learning activity: Improve data visualizations. Using the Refine Activity Worksheet in parallel with feedback from the critique stage and determine what aspects of the visualization(s), created in the represent stage, to improve/refine/change that would provide more insight into the data. use the Refine Activity Worksheet as a guide for braking data into its basic parts for further processing. **Learning outcome:** Students should be able to describe what happens in the refine stage, implement some method or technique to improve the visualization, make recommendations on what features should be refined, explain the impact of decisions made to refine the visualization on the other stages of the visualization process, devise steps to improve the visualization. Lastly, students should be able to generate/produce new and/or improved data visualization(s).

Following the qualitative feedback from the previous section, Figure 5 is an example of a refined visualization showing the percentage of cereals produced by each manufacturer.

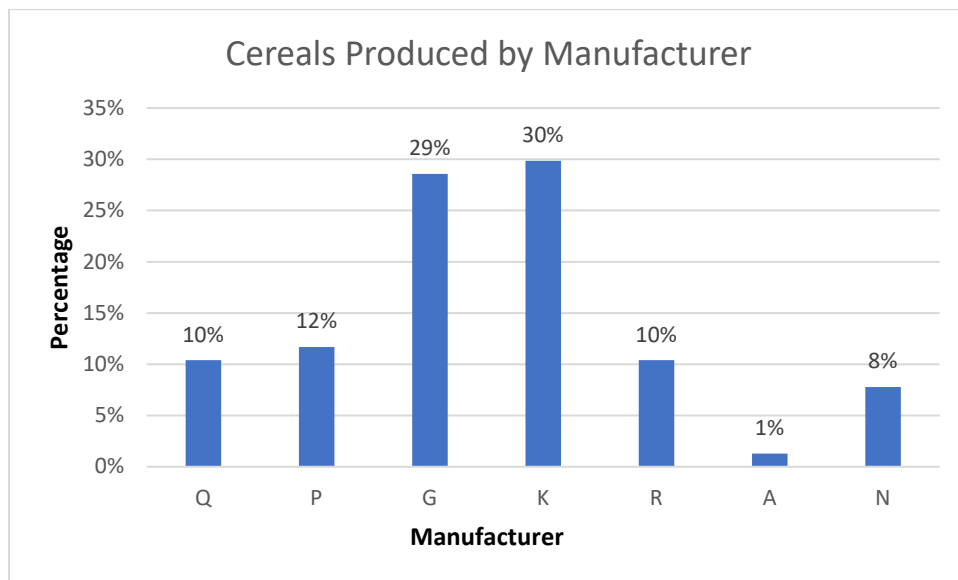


Figure 5 Refined visualization showing the percentage of cereals produced by each manufacturer.

Week 7: Bringing it all together

Online Activity Worksheets were tested for this research project. Students complete each worksheet online and receive a copy of their responses by email with the option to generate a PDF version of their responses. Subsequently, the student submits the PDF version to Brightspace for review by the instructor. Submitting the PDF version to Brightspace serves two purposes: (1) students receive an electronic copy of their responses, which replaces paper copies, and (2) submitting the PDF version to the learning management system serves as a backup for students and serves as evidence of their participation in the hybrid learning environment. Data visualization is storytelling. In week 7 students are asked to use their responses to the worksheets to help tell a story about the data. The PDF versions of their responses are used to show the visualization process they implemented to visualize the data. In subsequent assignments, all PDF documents are submitted as part of the final product for the visualization assignment.

After completing the data visualization process using the Activity Worksheets, students were asked to provide feedback to assess their self-efficacy about the process. Students were given a feedback survey that consisted of seven Likert-scale response questions and two written response questions. Self-assessment feedback is designed to help students to answer ‘what do I know?’ and ‘what does it mean questions’ [5] (p.227). Table 3 shows the types and list of questions. For this manuscript only, Likert-scale responses are analyzed and reported. Table 2 also shows how the questions align with basic and core ideas outlined for the course.

Table 3 Students’ self-assessment questions. LR: Likert Response, WR: Written Response.

Question	Basic Idea/Core Idea	Statement/Question
Q1 (LR)	BI-1, CI-1	I can list the stages of the data visualization process.
Q2 (WR)	BI-1, CI-1	List the stages of the data visualization process
Q3 (LR)	BI-2, CI-2	I can explain the interactions between each of the stages.
Q4 (LR)	BI-2, BI-3, CI-3	I can demonstrate my ability to apply each stage of the visualization process to a given dataset.
Q5 (LR)	BI-2, CI-1, CI-2, CI-4	I can discuss how output from one stage impacts other stages in the data visualization process.
Q6 (LR)	CI-6	Based on my understanding of the data visualization process, I can critique the process implemented to visualize data.
Q7 (LR)	BI-5, CI-6	Based on my experience with the data visualization process, I can critique the output that results from applying the data visualization process.
Q8 (WR)	BI-5, CI-6	What is the purpose of data visualization? (Use complete sentences)
Q9 (LR)	CI-5, CI-6	Based on my understanding and experience with the data visualization process, I can determine the value of the visualization process.

Students indicated their agreement or disagreement to Likert-scale statements using a four-point Likert-scale: 1 – Strongly agree, 2 – Agree, 3 – Disagree, and 4 – Strongly Disagree. Written

responses were examined for trends in qualitative responses provided for questions two and eight.

Results

A total of 258 responses were analyzed. Figure 1 shows the percent responses for each Likert-response.

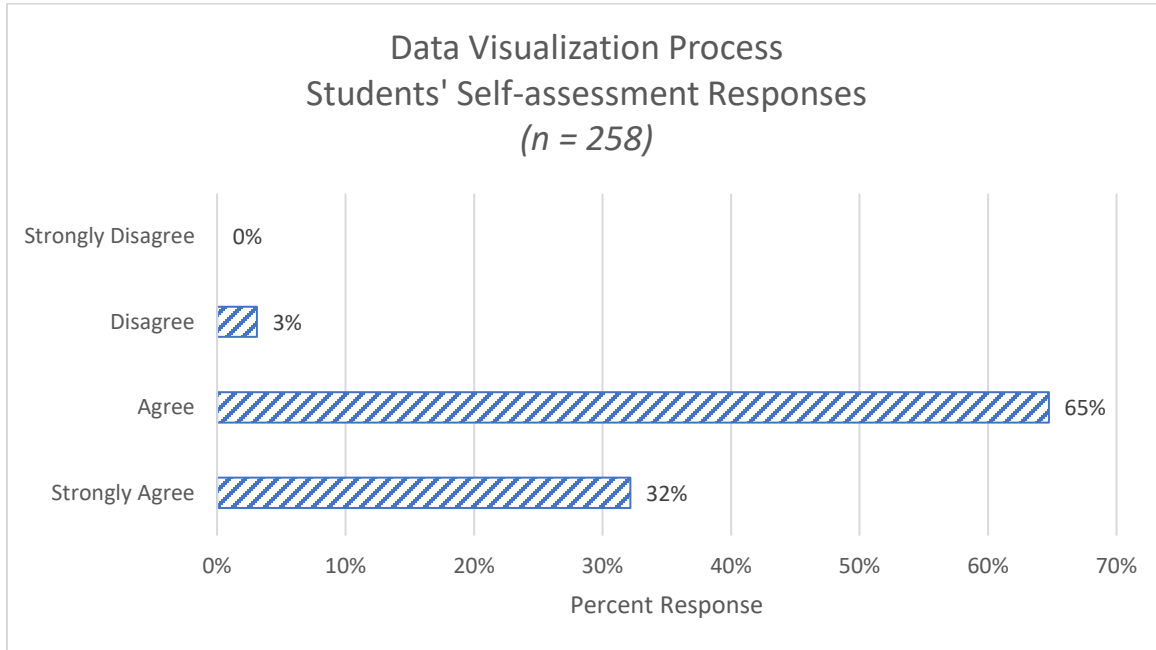


Figure 6 Likert-scale responses

Table 4 shows the percent response count and descriptive statistical calculations for Likert-scale questions (Q1, Q4 – Q7, and Q9). An invalid response was provided for question one by one student; that response was excluded from analysis as seen in the Total column in Table 4.

Table 4 Percent response count and descriptive statistical calculations for Likert-scale questions.

Question	SA	A	D	SD	Total*	Mean	Mode	Median	Std. Dev
Q1	42%	58%	0%	0%	36	1.58	2	2	0.049
Q3	32%	68%	0%	0%	37	1.67	2	2	0.468
Q4	32%	65%	3%	0%	37	1.70	2	2	0.512
Q5	30%	65%	5%	0%	37	1.75	2	2	0.540
Q6	19%	76%	5%	0%	37	1.86	2	2	0.474
Q7	32%	62%	5%	0%	37	1.72	2	2	0.552
Q9	38%	59%	3%	0%	37	1.64	2	2	0.530

* Total refers to the total number of responses analyzed for each question.

Discussion

Overall, most responses were positive. Thirty-two percent strongly agreed and 65% were agree, as seen in Figure 1. It is worth mentioning, there were some disagree responses for questions four through nine (not including the written response for question eight). Three percent of students indicated they did not feel they could demonstrate their ability to apply each stage of the visualization process to a given dataset (Q4). Interestingly, all students either agreed or strongly agreed to Q3: I can explain the interactions between each of the stages (of the visualization process). This suggest, students are more confident in factual and conceptual knowledge about the process, but lacking confidence in procedural knowledge required in each stage of the process. The worksheets were developed using Bloom's Taxonomy of hierarchical learning with the premise of factual and conceptual knowledge would improve students' data visualization capacity. Factual and conceptual knowledge maps to lower-order skills in Bloom's Taxonomy with procedural knowledge aligning with higher-order thinking skills. The relationship between higher-order skills and fact learning remains an open area for research. Agarwal [8], reported building a foundation of knowledge via fact-based retrieval practice may be less effective than engaging in higher order retrieval practices. Agarwal used a mixture of methods (fact questions, different question types) to examine the optimal type of retrieval practice for enhancing higher-order thinking. In this work, higher-order thinking is part of the data visualization process. The process is not examined, in this work, as described by [8]; however, Agarwal's findings suggest further investigation and research using the worksheet method is needed.

Five percent of students indicated: they did not feel they could discuss how output from one stage impacts other stages in the data visualization process (Q5); they did not feel they could critique the data visualization process (Q6); and they did not feel they could critique output that results from the implementing the process (Q7). Once again, responses to Q5 – Q7, are not consistent with responses given in Q3. Specially, all students either agreed or strongly agreed they can explain the interactions between each of the stages (Q3). It is unclear as to what contributed to or caused the level of disagreement indicated in questions Q4 – Q9. Closer examination of students' understanding of factual, conceptual and procedural knowledge in each stage will inform how to better scaffold the process so that students' self-efficacy in the process grows. Further analysis of data collected for this research will include analysis of written responses provided by students in response to questions two and eight in the self-assessment survey.

Conclusion

A worksheet method was introduced as a supplemental teaching tool for introducing data visualization. Although there were some noticeable disagreements among responses, outcomes of this work answered the research questions and suggest, students in an introductory data visualization course understand basic ideas introduced for visualizing data and students understand core ideas associated with the data visualization process. Outcomes from this work further suggest the Activity Worksheet approach could be serve as a supplemental tool for introducing the process. Worksheets coupled with self-assessments provide insight on student's self-efficacy and data visualization capacity levels. Further research opportunities include: pre- and post-analysis to assess students' understanding before and after introducing the Activity

Worksheets; using a control group of students who are introduced to the data visualization process without using the worksheet method and compare the Likert responses of the control group to students who use the worksheets. A next step for this research is to assess students' perception of the process incrementally, after completing each stage of the process. The results of this work suggest the worksheet method has the potential to inform pedagogy of data visualization.

References

- 1) V. Byrd, "Introducing data visualization: a hands-on approach for undergraduates," In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Association for the Advancement of Computing in Education (AACE'18), pp. 730-736, Oct. 2018.
- 2) V. Byrd, "Using Bloom's Taxonomy to Support Data Visualization Capacity Skills," In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Association for the Advancement of Computing in Education (AACE), pp. 1039-1053, Nov. 2019.
- 3) B. Fry, *Visualizing data: Exploring and explaining data with the processing environment.* " O'Reilly Media, Inc., 2008.
- 4) M. Hu, S. Cleland, and S. Burt, "Build up a Constructivist Learning Environment for Teaching First-year Students Data Flow Diagrams," In 2019 IEEE Frontiers in Education Conference (FIE) (pp. 1-8). IEEE, 2019.
- 5) D. Jonassen, "Designing constructivist learning environments", Instructional Design Theories and Models: A New Paradigm of Instructional Theory, vol. 2, pp. 215-239, 1999.
- 6) I. J. Sistermans, "Integrating competency-based education with a case-based or problem-based learning approach in online health sciences," Asia Pacific Educ. Rev. 21, 683–696 (2020). <https://doi.org/10.1007/s12564-020-09658-6>
- 7) B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, D.R. (1956), Taxonomy of Educational Objectives: The Classification of Educational Goals, David McKay Company, New York, NY.
- 8) P. K. Agarwal, "Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning?" *Journal of Educational Psychology*, 111(2), 189–209. 2019. <https://doi.org/10.1037/edu0000282>

APPENDICES

Activity Worksheets (from students' submissions)

ACQUIRE WORKSHEET

Goal: Identify appropriate data sources, analyze the data, identify data types, variables, list assumptions about the data

Objectives: Students will identify and acquire data from appropriate data sources

Outcomes: Data for the current visualization challenge

Generate

9) Identify appropriate data sources

Cereal data provided with in-class assignment.

10) Data format

The data is provided in csv delimited format and in table format. One data source, multiple formats of the same data was provided. This is considered secondary data.

11) Data types

Data types include strings, integers, characters, floating point numbers.
The csv file is a basic text file. The excel file is a spreadsheet.

Evaluate

12) Variables

List the data variables. What are the parameters? Give them names. What are the dependent variables and independent variables?

Variable	Parameters
Cereal	Cereal name (string)
Manufacturer	One letter code: Q-Quaker Oats, P-Post, G-General Mills, K-Kellogg's, \$-Ralston, P-Purina, N-Nabisco
Type	Hot or cold
Calories, Protein, Fat, Sodium, Fiber, Carbohydrates, Sugar, Potassium, Vitamins, Weight, Shelf, Cups	Number

13) Audience & Assumptions & Inconsistencies

Audience: could be the class, consumers

PARSE WORKSHEET

Goal: To understand the structure of the data

Objectives Students will change data into a format that tags each part of the data with its intended use

Outcomes: Every element of the data will be broken into its individual parts

Understand

1) Parse Data

Variable	Type	Values
Cereal	String (38)	a-z, A-Z, alpha numeric
Manufacturer	Character (1)	Q, P, G, K, R, N
Type	Character (1)	C-cold, H-Hot
Calories	Integer	50 – 160
Protein	Integer	1-6
Fat	Integer	0-5
Sodium	Integer	0 – 320
Fiber	Integer	0 – 5
Carbohydrates	Integer	-1 – 23
Sugars	Integer	-1 – 15
Potassium	Integer	-1 – 330
Vitamins	Integer	0 – 100
Weight	Integer	-1 – 1.5
Shelf	Integer	1 – 3
Cups	Floating point	-1 – 1.5

2) Assumptions

List any assumptions you are making about the data and/or the visualization challenge (aka the project).

- A value of -1 for carbohydrates, sugars and potassium does not make sense,
- How do we treat -1 values?
 - There are two cereals that have the same manufacturer's code: P-Post and P-Purina
- One cereal: Maypo is on the list but has an unknown manufacturer's code: A
- Not clear what the column Shelf represents, could be shelf-life (days, weeks, months, years) or it could be number of boxes currently on the shelf
- Cups is serving size
- Measurements are in grams

MINE WORKSHEET

Goal: To identify patterns, extreme and subtle features about the data

Objectives: Students will identify basic descriptors for the data, and categorize the data according to the specifications from the Parse Worksheet

Outcomes: Specific questions to be answered using the data

Analyze

1) Basic Descriptors

For each data component (aka field or variable) from the Parse Worksheet, identify basic descriptors (i.e. basic statistics). Explain.

Excel functions will be used to further define basic descriptors for each data component.

Variable	Type	Values	Functions
Cereal	String (38)	a-z, A-Z, alpha numeric	Max (Len) to find maximum string length
Manufacturer	Character (1)	Q, P, G, K, R, N	COUNTIF to determine number of cereals for each manufacturer
Type	Character (1)	C-cold, H-Hot	COUNT IF to determine number of cereal types (hot vs cold)
Calories	Integer	50 – 160	MAX, MIN, AVERAGE
Protein	Integer	1-6	MAX, MIN, AVERAGE
Fat	Integer	0-5	MAX, MIN, AVERAGE
Sodium	Integer	0 – 320	MAX, MIN, AVERAGE
Fiber	Integer	0 – 5	MAX, MIN, AVERAGE
Carbohydrates	Integer	-1 – 23	MAX, MIN, AVERAGE
Sugars	Integer	-1 – 15	MAX, MIN, AVERAGE
Potassium	Integer	-1 – 330	MAX, MIN, AVERAGE
Vitamins	Integer	0 – 100	MAX, MIN, AVERAGE
Weight	Integer	-1 – 1.5 (value of -1 does not make sense)	MAX, MIN, AVERAGE
Shelf	Integer	1 – 3	MAX, MIN, AVERAGE
Cups	Floating point	-1 – 1.5 (value of -1 does not make sense)	MAX, MIN, AVERAGE

FILTER & REPRESENT WORKSHEET (Output)

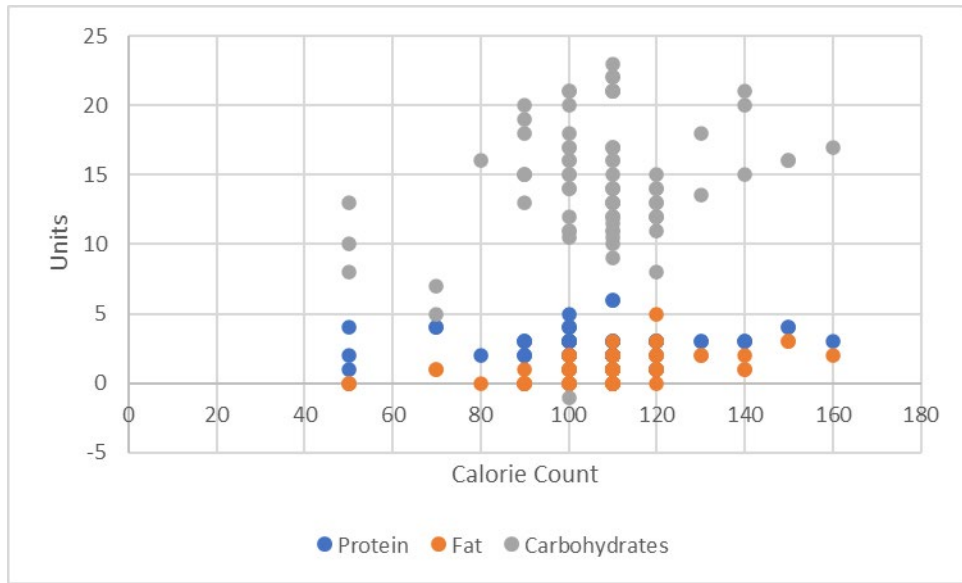


Figure 7 Student generated scatter plot of three primary energy sources for different calorie count.