Master's Theses and Doctoral Dissertations

Master's Theses, and Doctoral Dissertations, and
Graduate Capstone Projects

2018

Students' interpretations of categorical data using dynamic graphical representations

Adam Eide

Students' Interpretations of Categorical Data Using Dynamic Graphical Representations

by

Adam Eide

Thesis

Submitted to the College of Arts and Sciences

Eastern Michigan University

in partial fulfillment of the requirements

for the degree

MASTER OF ARTS

in

Mathematics

Thesis Committee:

Stephanie Casey, Ph.D., Chair

Andrew Ross, Ph.D.

Carla Tayeh, Ph.D.

November 27, 2018

Ypsilanti, Michigan

Abstract

Statistical association is an important concept in statistics. An exploratory study examined how students reason about statistical association utilizing graphical representations constructed with CODAP, a dynamic statistical graphing software. Task-based interviews were conducted with three 6th grade students prior to formal instruction. Students' conceptions of a statistical relationship, proportional reasoning skill level, ability to interpret bivariate categorical graphs (particularly segmented bar graphs and two-way binned plots), and ability to identify association of two categorical variables were all investigated through interview tasks and responses to inquiry. Students were found to have developing proportional reasoning skills and struggled to correctly define and identify association. These results were compared to a previous study which asked students to analyze pre-constructed graphs. Students were more successful interpreting graphs that they constructed than pre-constructed graphs. These results have curricular and future research implications.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The ability to accurately understand and interpret statistics has increasingly become important for all people, not just experienced researchers (Franklin et al., 2015). Technology has recently made the collection of data more accurate, accessible, and available. The wealth of data that is gathered from nearly every aspect of our lives, at nearly all times, has led many to refer to current times as the "Big Data Era" (Manyika et al., 2011). With this increase in the availability and amount of data comes an amplified push for people with the advanced skills needed to understand, interpret, and communicate the results of data analysis (Franklin et al., 2015). Big data and open data make the need for statistical thinking even more important and provide contexts for discussing core statistical ideas (Ridgway, 2016). This new era of big data has inspired changes to the K-12 statistics curriculum to prepare students for success in college and in civic life.

A common desire when analyzing data is to investigate whether two, or more, variables have a relationship. In statistics, this is known as the concept of statistical association. Statistical association is one of the fundamental ideas in statistics (Baterno, Burrill, & Reading, 2011). Correlation and association are often used interchangeably, but association refers to any relationship between variables, while correlation often refers to a linear relationship between quantitative variables.

Data that is gathered can be categorized into two distinct types: categorical and quantitative. Categorical variables are defined as "a variable measured in terms of possession of qualities and not in term of quantities. Categorical variables contain a minimum of two different categories (or values) and the categories have no underlying order of quantity" (Cramer & Howitt, 2004). Quantitative variables, or measurements, are defined as follows: "Individual

observations are given a quantitative or numerical value as part of the measurement process. That is the numerical value indicates the amount of something that an observation possesses" (Cramer & Howitt, 2004). Both categorical and quantitative variables are often analyzed to determine if the variables are associated, but their analysis requires different approaches from students (Tran & Tarr, 2018). The association of two quantitative variables is usually analyzed by creating scatterplots, noticing relationships between variables, and fitting a function to the data and/or assessing the goodness of fit referred to as regression analysis. The association of two categorical variables is often analyzed by investigating two-way frequency tables and/or using a chi-square test (Tran & Tarr, 2018). Categorical data analysis skills are an important key to students developing well-rounded statistical literacy and helping prepare students for college, careers, and life (Franklin et al., 2015).

Rationale for Study

The learning of statistical association has been identified as important for today's students as can be seen by its inclusion in the Guidelines for Assessment and Instruction in Statistics Education Report (Franklin, et al., 2007) and Common Core State Standards (8.SP.4 and HSS-ID.B.5) for Mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2018). The association of two categorical variables is often addressed in school curricula by investigating two-way frequency tables with probabilities and/or using a chi-square test to measure the strength of the relationship (Tran & Tarr, 2018). Current standards call for the learning of it in middle school with eighth graders expected to "understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use

relative frequencies calculated for rows or columns to describe possible association between the two variables." (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2018).

It is vital for the statistics education community to have appropriate methods to assess the level and quality of comprehension of students as it pertains to proportional reasoning and bivariate categorical association. The increasing diversity of data usage requires the ability to analyze complex data sets and "educators should place less emphasis on small samples and linear models and more emphasis on large samples, multivariate description and data visualization" (Ridgway, 2016). This investigation hopes to encourage the development of statistical thinking in all students and explore new methods of teaching and learning statistical association.

This study seeks to explore the conceptual understanding of association and proportional reasoning skills of some local sixth graders. There is a great need in the education community to focus on increasing students' abilities to determine categorical association. This study hopes to add to the developing repository of research surrounding this topic. This study could help inform curriculum design for middle school mathematics classes. The study has the potential to contribute to the understanding of student learning methods of bivariate data to foster different approaches to teaching and learning.

To work towards this goal, the following research questions were developed for this investigation:

1) What are students' conceptions of a statistical relationship?

2) At what level are student's proportional reasoning skills?

3) How do students interpret graphs (particularly segmented bar graphs and two-way binned plots) for association?

4) How do students' conceptions of a statistical relationship and their proportional reasoning skills inform their interpretation of graphs (segmented bar graphs and two-way binned plots)?

5) Do dynamic graphical representations in CODAP improve students' understanding of statistical graphs (particularly segmented bar graphs and two-way binned plots), with and without association, when compared to pre-constructed graphical representations?

Chapter 2: Review of Literature

The goal of this research is to study the abilities, strategies, and comprehension levels of middle school students as they analyze statistical graphs for evidence of association. This section summarizes studies related to proportional reasoning and students understanding of statistical association, in particular statistical association as it relates to the analysis of association between two categorical variables. Specific studies are described in the following paragraphs.

2.1 Statistical Association

Researchers have recognized that students struggle when analyzing bivariate categorical data for association (Batanero et al., 1996; Watson & Callingham, 2014). Research has shown that students generally struggle with correctly identifying association, or lack thereof, when comparing two categorical variables. These errors are caused by a variety of different misunderstandings and misconceptions that students make when observing statistical graphs. The research of Batanero et al. (1996) and Watson and Callingham (2014) concluded that the most predominant misconceptions held by students could be categorized as lack of proportional reasoning, localist, deterministic, and ignoring the data. These will be explained in turn in the following paragraphs.

Research studies targeting middle school aged students' understanding of, and skills with, quantitative association and covariation have been fruitful (Cobb, McClain, & Gravemeijer, 2003). It has also been identified that categorical association often requires very different cognitive approaches when compared with quantitative variable association. Research illustrates curriculum differences when observing the teaching, and learning of, quantitative and categorical association. Quantitative association is frequently represented by a functional relationship with two numerical variables and teaching categorical association often involves calculating

5

conditional probabilities while analyzing a possible association between two categorical variables (Tran & Tarr, 2018). These studies motivated our decisions to select a population of middle school students and analysis of students' categorical association identification skills.

This investigation will utilize the CODAP (Common Online Data Analysis Program) software program to explore real data. Research has explored how technology-assisted explorations affect students' abilities to assess statistical association (Konold, 2002). This research concludes that students are more successful interpreting graphs that are built up in stages using statistical software such as TinkerPlots (Konold, 2002). Researchers explain that student learning benefits in a variety of ways through the use of dynamic statistics software during instruction (Konold, 2002; Konold C. , 2007). This study seeks to mirror these successful studies' approaches to using technology to investigating students' understandings of statistics and apply it to analyzing statistical association.

2.2 Proportional Reasoning

Proportional reasoning is a type of mathematical reasoning involving a sense of co-variation and the capacity to compare situations in relative (multiplicative) rather than absolute (additive) terms (Fielding-Wells, Dole, & Makar, 2013; Hiebert, Behr, Lesh, & Post, 1988). The ability to differentiate between multiplicative and additive situations is a major component of proportional reasoning (Confrey & Smith, 1995; Van Dooren, De Bock, Hessels, Janssens, & Verschaffel, 2005), yet research has consistently highlighted students' struggles with differentiating between multiplicative and additive proportional reasoning concepts (Hilton, Hilton, Dole, & Goos, 2013; Garfield & Ahlgren, 1988; Staples & Truxaw, 2012). Research suggests that struggles with proportional reasoning is one of the most frequent causes of students' difficulties in understanding and applying statistical concepts (Garfield & Ahlgren,

1988). Proportional reasoning skills include understanding ratios and part-to-whole relationships, the similarities between numbers and ratios, using proportional understanding to find equivalent ratios, and understanding how to make accurate comparisons between base proportions and equivalent ratios. Students that struggle with proportional analysis skills often mistakenly compare counts(frequencies) rather than percentages in their analysis, or claim that one cannot compare two groups of unequal sizes (Watson & Callingham, 2014; Batanero et al., 1996). A further detailed description of some of these common error tendencies will provide additional background and details for our exploration.

A localist conception describes students that examine only a portion of the displayed data and not the data as a whole. Researchers give an example of this thinking as they describe one student that "compared the relative frequencies in only one conditional distribution" (Batanero et al., 1996). Students, who tend to have localist conceptions will, in many cases, only analyze the cell with the highest frequency or only one conditional distribution to attempt to determine if an association exists.

The deterministic conception describes students that perceive association as an all-or-nothing concept. Researchers demonstrate an example of this thinking as a student considered two cells that do not agree with the association "in the [2x2] table ought to be null in order to assume association" (Batanero et al., 1996). These students often feel that all cases must show an association with no exceptions in order for an association to exist. Students also believe that the cells in the two-way table that are counter to the association should have zero frequency.

A student that has been identified as "ignoring the data" is a student that ignored the data at hand when making a conclusion about the displayed graph. Students often instead use their pre-existing knowledge about the given variables under investigation to determine if an

association exists. We wish to keep these common misconceptions in mind and explore if

students that have not received any formal instruction in determining association display any of

these common incorrect analysis techniques.

Chapter 3: Methodology

3.1 Introduction

This study attempted to answer the following research questions:

1) What are students' conceptions of a statistical relationship?

2) At what level are student's proportional reasoning skills?

3) How do students interpret graphs (particularly segmented bar graphs and two-way binned plots) for association?

4) How do students' conceptions of a statistical relationship and their proportional reasoning skills inform their interpretation of graphs (segmented bar graphs and two-way binned plots)?

5) Do dynamic graphical representations in CODAP improve students' understanding of statistical graphs (particularly segmented bar graphs and two-way binned plots), with and without association, when compared to pre-constructed graphical representations?

This chapter begins with an overview of the study and a description of its participants. The development of interview questions, assessments, and accompanying materials are then described. Implementation of the study is then discussed. Finally, the development of a coding scheme to aid in qualitative analysis along with overall study data analysis methods are described.

3.2 Overview of Study

To answer the above stated research questions, a study was conducted with sixth grade students. This study took place during the spring of 2018 near the end of the academic school year. Students participated in a 30-to-45-minute semi-structured, task-based interview. The

interview was conducted in two distinct parts. The first part was designed to introduce students

to the software program CODAP (https://codap.concord.org), inquire about a student's

conception of a statistical relationship, and how a student may determine if variables have a

relationship. It also informally assessed their proportional reasoning skills. Students were shown

how to graph and compare categorical variables from a data set about granola bars. Students

were shown a variety of ways in which the data can be displayed and analyzed. In the second

part of the interview students were directed to make graphs to analyze specific variables and

identify if a relationship exists between the variables. All interview questions and materials were

reviewed by Dr. Stephanie Casey, Eastern Michigan University, advisor on this project and

approved by the Eastern Michigan University Human Subjects Review Committee (see

Appendix A). The study was implemented by the author, who conducted all interviews,

recordings, and data gathering.

3.3 Study Participants

The students that participated in this study during the spring of 2018 were selected from

local sixth grade public school classrooms in Michigan. Students in this study were from a Math

6 course at a local middle school. Classrooms were selected that had not received formal

instruction on how to analyze whether two categorical variables are associated. Local sixth grade

students are between ages 11 and 12. Three weeks prior to the interviews, the researcher visited

local classrooms and gave a 5-minute presentation to each class interested in participating to

briefly explain the purpose of this study and what would take place during the interview. The

students were asked to participate in a 30-to-45-minute semi-structured interview. The researcher

explained that the interviews would be video recorded and screen capturing software would be

used to record computer activity. Parental permission forms were distributed to interested

students for their parental guardians to complete (see Appendix B). Students were also asked for their permission to use their responses with identifying markers removed. Students received an assent form explaining these details (see Appendix C) and were allowed to opt out of participating in this research study. A total of three students (2 females, 1 male) voluntarily participated in this study.

3.4 Interview Questions and Assessment Development

The interview protocol for this investigation was patterned after one used by Casey, Hudson, and Ridley (2018) with the hopes of making comparisons to its results. Casey et al. found that most students experienced difficulty when trying to correctly analyze pre-constructed graphs for association. One subsequent goal of this study is to analyze the effect on students' abilities to correctly identify association through assisting students in creating their own bivariate categorical graphs using technology.

The interview protocol was developed with three major goals in mind. One goal was to analyze students' current abilities to accurately reason through proportions and percentages. A second goal was to provide students with data tables and graphs that were easy to understand and relatable for their age group yet would also challenge the students with data that may go against their preconceived notions. The intent of this goal is not to try to confuse students, but instead give researchers needed information to ascertain if students are using outside knowledge or the provided displayed data to make their decisions. A final goal was to analyze students' conceptions of statistical association. The study examined how students interpret bivariate statistical graphs, students' conceptions of a statistical relationship, and determine if there was a link between what students think constitutes a statistical relationship and how they interpret

statistical graphs. The research team next sought to select variables from the Granola Bar data set that, when graphed, would allow these types of questions to be asked.

3.4.1 Interview development. The first part of the interview introduced students to a dataset detailing various aspects of 33 different granola bars (see

https://codap.concord.org/releases/latest/static/dg/en/cert/index.html?url=https://concord-consortium.github.io/codap-data/SampleDocs/Social_Science/Granola_Bars/Granola_Bars.codap) (see also Appendix D). Students utilized a software program CODAP to analyze the Granola Bars dataset. Students were shown different graphs, depending on their choice of variable to explore, and asked to analyze various displays of percentages (row percentages, column percentages, cell percentages, etc.) for the displayed graphs. A few examples of these graphs are shown below in Figure 1 and Figure 2:

Table 1

A small table for aligning figures. It has no borders.
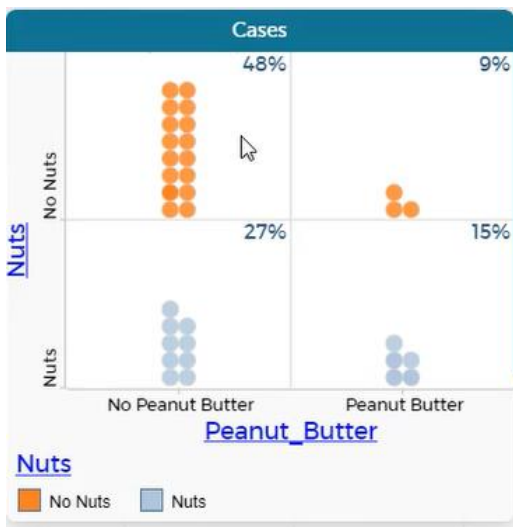
Figure 1. Example two-way binned plot graph.
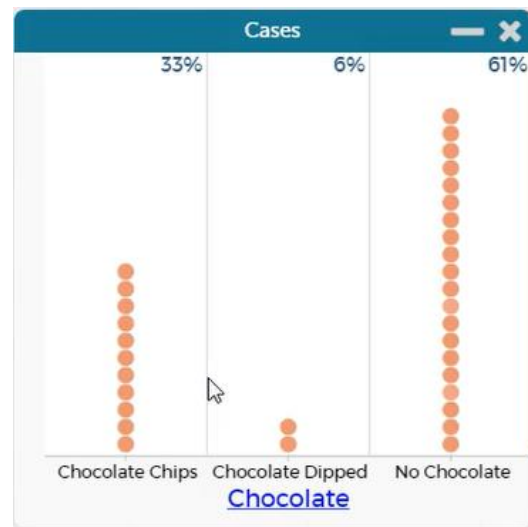
Figure 2. Example of univariate graph.

It was decided that the first portion of the interview would be a student-guided observation with freedom given to students to choose any variables that they would like from the dataset to explore using CODAP. The interviewer would then guide the students into creating two-way binned plots that would allow various proportional percentages to be displayed for the data. This portion of the interview would be used to analyze students proportional reasoning skills.

The first interviewed student was allowed to choose two categorical attributes of granola bars to explore further. The students were shown how to create two way tables, two-way binned plots, and segmented bar graphs in CODAP. Students were asked to use these analysis tools to determine if there is a relationship between the variables. The first participant chose to explore two variables that were intended to be introduced in the second part of the interview (fruit versus nut content), thus creating the possibility of exposure bias. The author decided at this point to guide the exploration portion of the following interviews to ensure that participants would not choose variables that are intended to be explored in the second portion of the interview.

For the second portion of the interview, the researchers initially selected four graphs, two with association and two without, for the students to interpret. It was later discovered that the statistical analysis program used for this study treated one of the study variables as a quantitative variable instead of a categorical variable. For this reason, the fourth graph was removed from the study and only the three graphs analyzing bivariate categorical data will be discussed.

The second portion of the interview engaged the students in a focused analysis of three pairs of categorical variables using the tools learned in part one of the interview. Students created a series of three graphs, with two displaying variables that show some association and one displaying variables with very little association. Each student was asked to determine if there

was a relationship between the displayed variables and explain why. Table 2 displays the graphs the students were directed to make and briefly describes if there is association and if the data follows common intuition or challenges it.

Table 2

Graphs Used in Part 2 of Interview
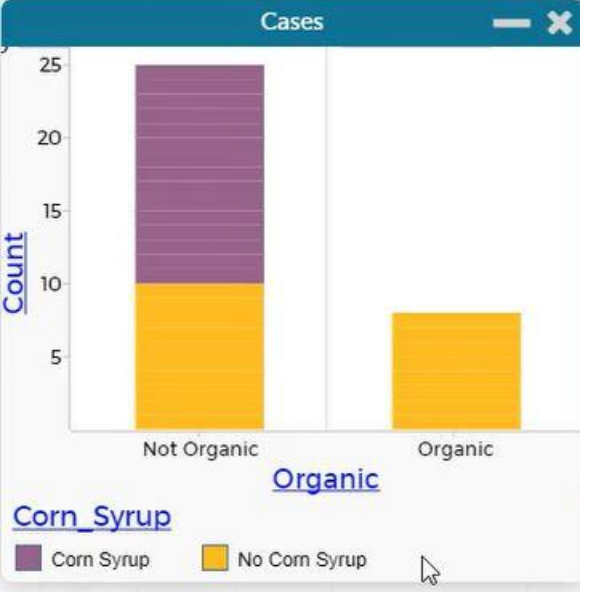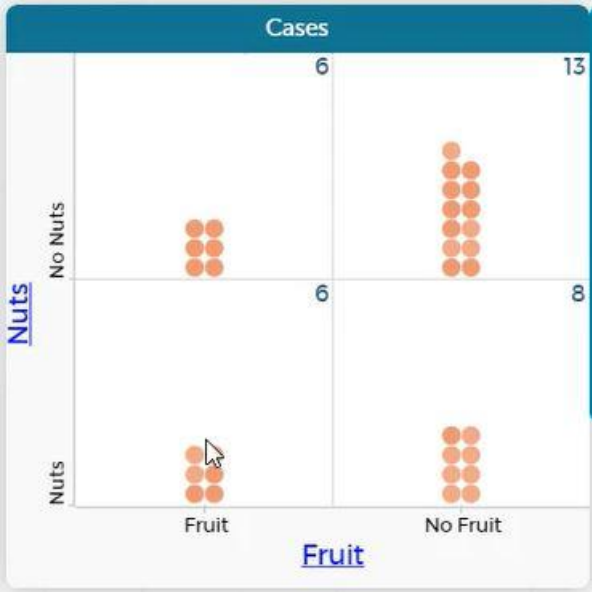
	Type	Association	Intuition
	Conditional Segmented Bar Graph	No	Graph goes against intuition as most expect nuts to be in more crunchy bars and in less chewy bars.
	Segmented Bar Graph	Yes	Graph follows intuition as granola bars containing corn syrup are not typically labeled as organic.

Table 2 *continued*

	Type	Association	Intuition
	Two-Way Binned Plot	Yes $p = .0463$	Graph goes against intuition as fruit and nuts typically have no reason to be related yet the data displayed in this graph shows that a relationship does exist.

These graphs were selected to align as closely as possible with pre-constructed graphs that were used in a previous study by Casey et al (2014). Graphs were selected to provide students with a variety of combinations of graphs with and without association, as well as data that both followed and went against intuition. The current capabilities of CODAP do not allow for the graphing of side-by-side bar graphs so I chose to create a two-way binned plot with the same variables in an attempt to closely simulate a side-by-side bar graph.

The order of the activities in the study design was carefully considered in an attempt to reduce exposure bias. It was determined that each student would need to have an idea of how the software operated and displayed graphs. This prevented randomization of the order in which the first and second parts of the interview took place. The first part of the interview would always need to be performed first as to introduce students to the software and its functionality. During

the second part of the interview, the students were asked to observe and analyze three separate statistical graphs. The order in which these graphs were created and displayed was randomized for each participant.

3.5 Study Implementation

Interviews took place near the end of the academic school year in 2018. Interviews were conducted in a conference room to help protect student privacy. I documented the interviews by video recording each session, and screencasts of the student's computer work were captured using screen casting software.

3.6 Data Analysis

After the implementation of the study, data from the students' interviews and screencasts were analyzed. All participants in this study returned a completed parental consent form as well as signed an assent form to participate. Of particular interest in this study was examining areas in which students appear to display correct understanding of whether bivariate categorical variables are associated and areas in which they still had misconceptions or misunderstandings after participating in the activities within this study. First, in order to de-identify the data, a unique pseudonym was assigned to each student. The interviews were then transcribed for qualitative research using pseudonyms for confidentiality.

To answer the research question "What are students' conceptions of a statistical relationship?" the student-created graphs from Part 1 of the interview were analyzed along with students' responses to the question "Would you say there is a relationship between these two variables?" when viewing the created graphs. One of the graphs that the participant Amy chose to analyze, in Part 1 of her interview, displayed one quantitative variable and one categorical variable. Amy's original graph was excluded from the analysis as invalid. Instead, I examined a

response that she provided in Part 2 of the interview to further analyze her concept of a statistical relationship. Students responses were analyzed qualitatively for correctness of claim of association, or not, and justification of claim. Two researchers discussed the analysis and reconciled the analysis through identifying, discussing, and classifying the responses.

To answer the research question "At what level are student's proportional reasoning skills?" students' responses to interview protocol (see Appendix E) Questions 1,9,10,11, and 15 during Part 1 of the interview were analyzed. These questions asked students to analyze a variety of different displayed percentages on univariate and bivariate graphs. For bivariate two-way binned plots, percentages were displayed by row, column, and cell, keeping the two analyzed variables the same throughout. Students responses were analyzed qualitatively for correctness of graph interpretation and justification for interpretation. Two researchers discussed the analysis and reconciled the analysis through identifying, discussing, and classifying the responses.

To answer the research question "How do students interpret graphs (particularly segmented bar graphs and two-way binned plots) for association?" students' responses to interview protocol (see Appendix E) Questions 18, 21 and 24 as well as the question "Would you say there is a relationship between these two variables?" during Part 2 of the interview were analyzed. Students responses were analyzed for correctness of determining association, correctness of interpretation of the graph, and justifications for both. Two analysts independently interpreted the students' responses and reconciled the analysis through identifying, discussing, and classifying the responses. Throughout the study, students' responses would often be inconsistent so responses may be coded, Correct/Incorrect/Correct for example, demonstrating this inconsistency.

To answer the research question "How do students' conceptions of a statistical relationship and their proportional reasoning skills inform their interpretation of graphs (segmented bar graphs and two-way binned plots)?" students' responses and analysis from Research Questions 1 and 2 were analyzed in comparison to students' responses to Research Question 3. Justifications for responses were qualitatively analyzed for details and descriptors that could be grouped together in the search for common themes or patterns. Two researchers discussed the analysis and reconciled the analysis through identifying, discussing, and classifying the responses.

To answer the research question "Do dynamic graphical representations in CODAP improve students' understanding of statistical graphs (particularly segmented bar graphs and two-way binned plots), with and without association, when compared to pre-constructed graphical representations?" students' responses to interview protocol (see Appendix E) Questions 18, 21 and 24 as well as the question "Would you say there is a relationship between these two variables?" during Part 2 of the interview were compared with results from a previous study. Side-by-side comparisons of the graphs from each study were done, noting any differences that could affect students' interpretations or analysis. Population demographics and sample size differences were also discussed and considered. Three graphs from the previous study were selected and their analysis responses compared with this current study. A direct comparison of the proportion of students that correctly interpreted the graphs, as well as the proportion of students that correctly identified association, was done between the two studies.

Chapter 4: Results

4.1 Introduction

To examine students' observations of categorical variables and understanding of bivariate categorical association, a semi-structured study interview was implemented with explanatory and assessment questions. This chapter first describes the results from the first portion of the interview as the researchers sought to determine the students' preexisting proportional reasoning skills and what they think a relationship between variables is. The chapter then examines individual subject interviews and their personal assessment of the statistical graphs when asked to determine if a relationship exists between the two given variables. Then results are compared to other subjects within this study. Finally, the results are compared with a previous study that used pre-created statistical graphs.

4.2 Analysis of Students Conceptions of a Statistical Relationship

To analyze students' conceptions of a statistical relationship, the students in this study constructed a variety of statistical graphs, from variables that they had selected, in Part 1 of the interview and asked "Would you say there is a relationship between these two variables?" Students were then asked to justify their claim with additional reasoning and details. One of the graphs that Amy chose to analyze, in Part 1 of her interview, graphed one quantitative variable and one categorical variable. Amy's original graph was excluded from the analysis as invalid. Instead, I examined a response that she provided in Part 2 of the interview to analyze her concept of a statistical relationship. All student names displayed are pseudonyms to protect participant identity.

4.2.1 Amy's conceptions of a statistical relationship. In Part 1 of the interview Amy created the graph in Figure 3, with cell percentages displayed, and was asked to identify if there was a relationship between the two variables:
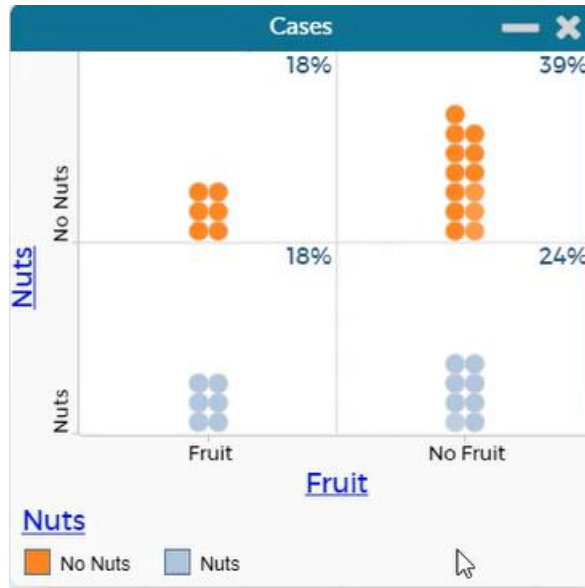
Figure 3. Amy's two-way binned plot graph of fruit versus nuts.

Amy responds, "Yeah… Because usually if there is no fruit then there's no nuts. So that's a relationship."

Later, in Part 2 of the interview, when viewing the graph in Figure 4, Amy is asked to determine if there is a relationship between the two variables:
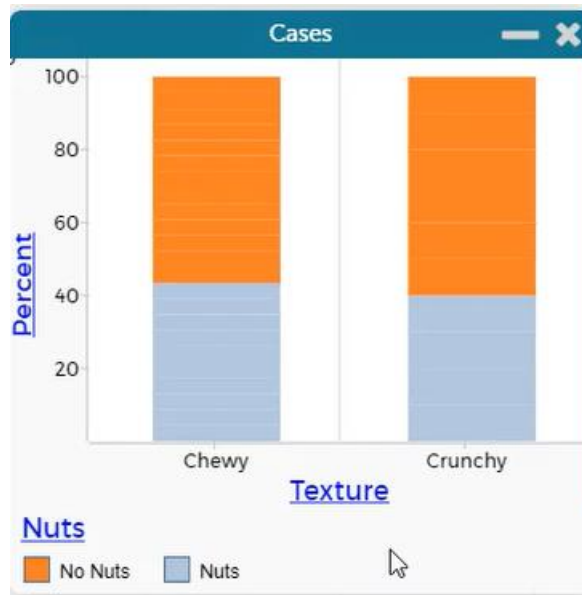
Figure 4. Amy's conditional segmented bar graph of texture versus nuts.

Amy interprets, "It's still almost to 50% so if it's chewy it's either um… if it's chewy it's sometimes there is nuts in them but more of the times there is no nuts. Then if it's crunchy most of the time there's no nuts in them." When asked to identify if this constitutes a relationship between these variables, Amy responds, "Yeah… Because if it's crunchy there's nuts in…most of the time there is no nuts in them. Then if its chewy there's like a sometimes there's nuts in them and more of the times there's no nuts in them."

Amy gives an incorrect interpretation of what constitutes a relationship between two variables. The idea that if something happens more than 50% of the time conditioned on one value of the other variable there is causation, is incorrect; the conditional relative frequency of that occurrence for other values of the other variable must also be examined and compared. It is possible for more than 50% of one variable, or both variables, to have a tendency with the given categorical variable. Varying levels of association can potentially occur, depending on sample size and relative frequency, in these scenarios. Amy at times attempted to view the dataset as a

whole and would refer to the other variable but tended to struggle linking the two together to justify a claim of overall association or not. From the above responses we were able to determine that Amy has incorrect conceptions of what constitutes a statistical relationship between two categorical variables.

       4.2.2 Brian's conceptions of a statistical relationship. Brian created the graph in Figure 5 with percentages by row displayed:
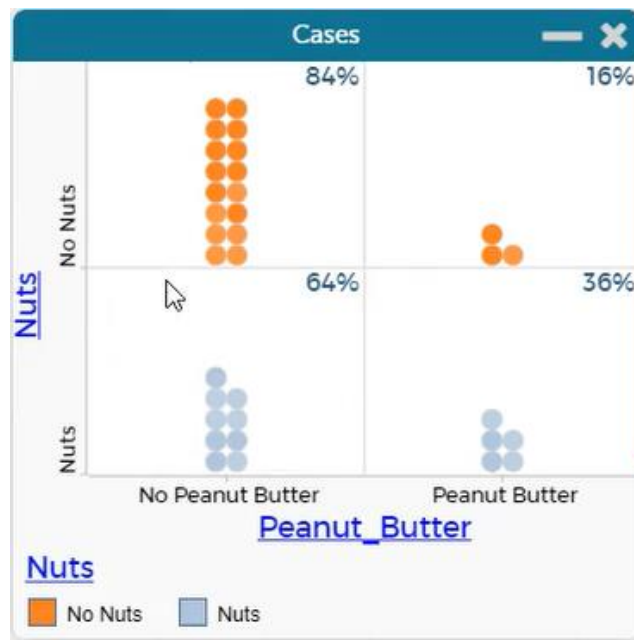
Figure 5. Brian's two-way binned plot graph of peanut butter versus nuts.

When asked to identify a variable that may have a relationship with the variable peanut butter Brian responds, "Nuts… Because sometimes nuts are like peanuts. So if they don't have nuts in them then they don't have peanut butter." This is a correct interpretation and what many people would conclude should be the result when comparing nut content with peanut butter content. When asked what one would expect to see in the data if a relationship existed between two variables Brian responds with, "There's like I guess a difference maybe." Brian is unable to

clarify this statement further beyond saying that if there is no relationship then there is no difference and the viewed proportions would be about the same. In other portions of the study Brian claims, while viewing a two-way table, that a relationship can exist granularly for two of the cells but not for the other two cells. Brian also claims that if there is a "pattern" to the graph, then a relationship exists but is not able to clarify further.

Brian gives multiple responses when asked what constitutes a relationship between the two variables and struggles to explain his claims. Brian often struggled to view the dataset as a whole and tended to have a granular view of the data in this study. From the above responses, we were able to determine that Brian has inconsistent incorrect conceptions of what constitutes a statistical relationship between two categorical variables.

4.2.3 Claire's conceptions of a statistical relationship. During part 1 of the interview, Claire created the graph in Figure 6 with non-conditional cell percentages displayed:
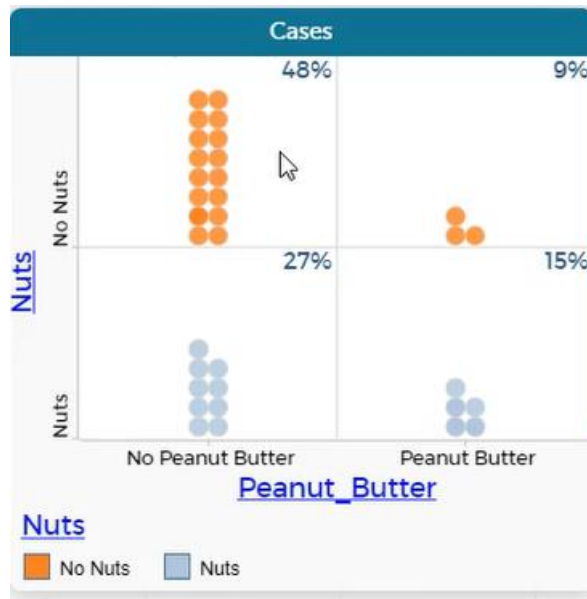
Figure 6. Claire's two-way binned plot of peanut butter versus nuts.

When asked if there is a relationship between these two variables, Claire states,

"Yes…Well like it's not really a relationship between them because like if there was …most of

them that have no nuts also have no peanut butter but... Well that's really not like a relationship

because there's like none." Claire goes back and forth between saying there is a relationship and

there isn't one but isn't able to justify either claim with sound reasoning.

Claire later observed the graph in Figure 7 to attempt to identify association:
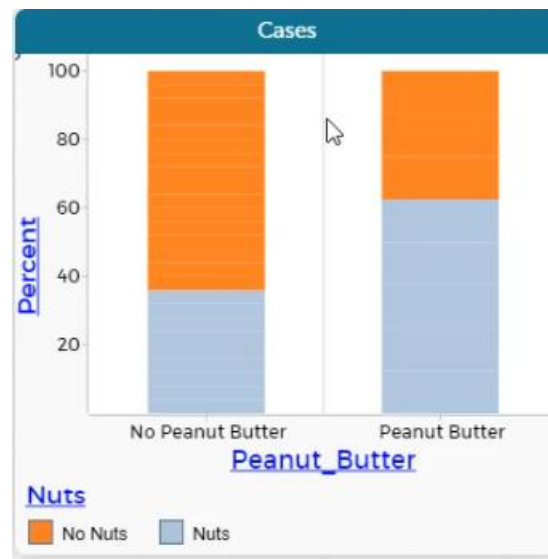
Figure 7. Claire's conditional segmented bar graph of peanut butter versus nuts.

When asked if there is a relationship between these two variables (the same two variables

that were displayed in Figure 6 but using a different type of graph), Claire states, "Yes… If it

tends to have peanut butter, then it probably has nuts too… So like all of this [student hovers

over peanut butter column with no nuts color] is like 62.5% that have peanut butter also have

nuts." Interestingly, Claire changes her reasons for determining a relationship to include the idea

if something "tends" to be with another variable then a relationship exists.

Claire gives two different incorrect responses when asked if a relationship exists between

these two variables. Claire at first demonstrates an incorrect deterministic definition of

association with here statements including "because there's like none." In the second graph interpretation, Claire states another incorrect definition of association by if something "tends" to be with the other variable then a relationship exists. This conception is similar to Amy's conception of what constitutes a relationship. This idea that if it is happening more than 50% of the time is incorrect without also observing and analyzing the other categorical variables. It is possible for more than 50% of both variables to have a tendency with the given categorical variable, in which case no association would exist. This shows that Claire struggled to view the dataset as a whole and tended to have a granular view of the data in this study and has inconsistent incorrect conceptions of what constitutes a statistical relationship between two categorical variables.

    4.2.4 Students' conceptions of a statistical relationship conclusion. From our detailed analysis of our three subjects' responses, we were able to identify if the students had correct conceptions of what identifiable markers in a graph determine if a relationship exists between two categorical variables.

    When asked to identify what constitutes a relationship existing between two categorical variables, none of the students were able to give a complete accurate definition. Students did display some correct ideas of what a statistical relationship means, reasoning that it's more likely a granola bar has one attribute if the related other attribute is present. All students struggled to justify their given definition with results from the data. All students struggled with looking at the data as a whole and tended to have a granular view of the data. Amy and Clare both have the conceptual idea of a relationship being if it's happening more than 50% of the time, then it is association. While Brian has multiple views of association but tends to most often use the

concept of a relationship existing if there is a difference when variables are compared and no difference constituting no relationship.

4.3 Analysis of Students' Initial Proportional Reasoning Abilities

The students in this study were shown a number of different types of univariate and bivariate graphs with percentages and counts displayed. Percentages displayed would be by row, column, or cell. The students were then asked to interpret the meaning of one of the displayed percentages. The students' responses are detailed below and informed the researchers' analysis of each student's proportional reasoning ability.

4.3.1 Amy's proportional reasoning analysis. During the interview, Amy created the graph in Figure 8 with row percentages displayed.
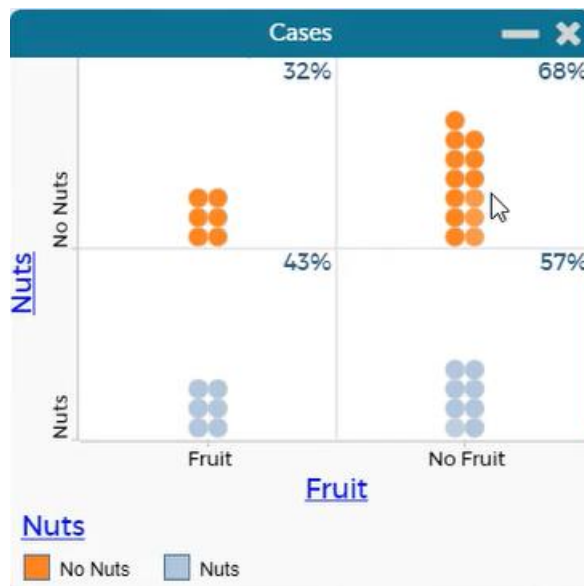
Figure 8. Amy's two-way binned plot graph of fruit versus nuts with row percentages.

When asked to interpret one of the cells in Figure 8, Amy responded with, "Um…57% of granola bars that have no fruit have 68% of them that have no nuts." The interviewer shows Amy that the displayed percentages are row percentages and asks Amy to clarify further. Amy responds with "I don't know." The interviewer asks if Amy know which way rows work to

which Amy responds, "Uh…. Up and down? I don't know." The interviewer clarifies with the student that columns go vertically and rows go horizontally. The student is asked to interpret the percentages in one of the rows now that they understand rows versus columns. Amy responds, "One of them is less than 50 the other one is more than 50." The student isn't able to interpret further, and the interviewer recommends that she attempt to add together the percentages. Now that the student sees that the top row represents 100% of something and the bottom row represents 100% of something the student is asked again to pick one of the cells and interpret the graph. Amy's responds, "60 out of 100 of them have no nuts…I mean… I don't know."

These responses show that Amy struggles to properly consider the displayed numbers in proper relative terms. Amy often struggles to properly identifying what is the representative "whole" in which the proportion is describing. The interviewer attempts to guide Amy towards the proper identification but is unsuccessful.

Amy later observed the graph in Figure 9, with column percentages displayed, and attempted to interpret the displayed percentages:
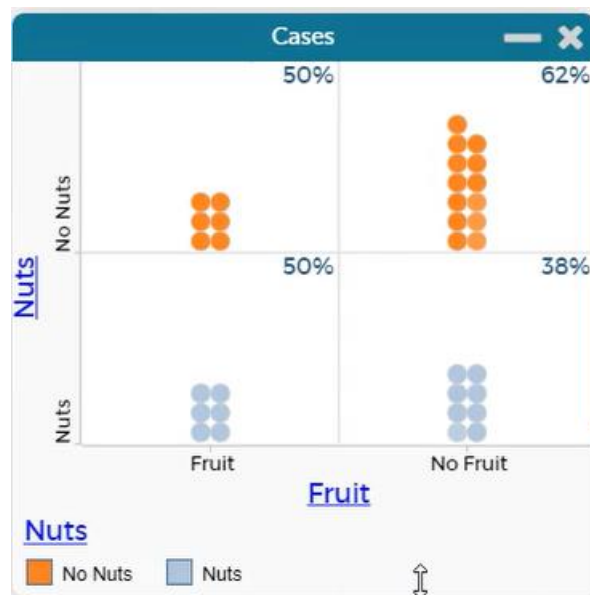
Figure 9. Amy's two-way binned plot graph of fruit versus nuts with column percentages.

Amy interprets, "50% of the granola bars have fruit in them and 50% of them also have no nuts in them and then 62% have no nuts and no fruit…I don't know." Amy incorrectly interprets that the 50% is related to all of the granola bars in this study and her statements lean towards what would be interpreted if non-conditional cell percentages were displayed. Amy is not able to properly parse out that the 50% only relates to granola bars that also contain fruit. Amy seems to understand that her interpretation is incorrect when she talks about 50%, then 50%, then 62% of somethings but isn't able to reconcile these errors.

Amy later observed the graph in Figure 10, with non-conditional cell percentages displayed, and attempted to interpret the graph:
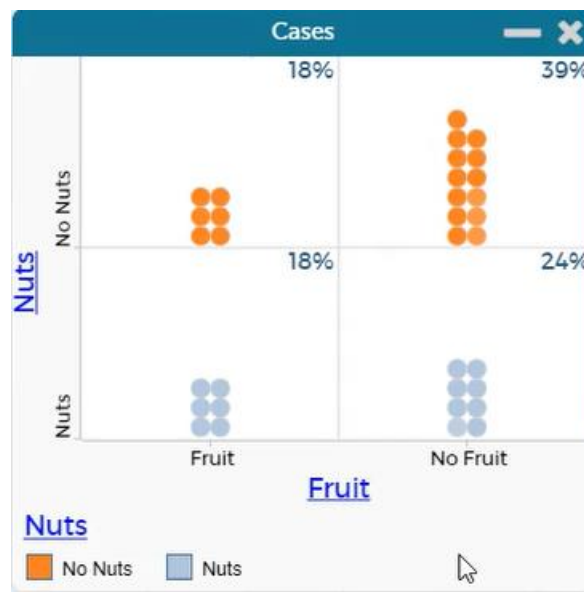
Figure 10. Amy's two-way binned plot graph of fruit versus nuts with cell percentages.

The interviewer walks Amy through the percentages displayed in the graphs, shown in Figure 8 and Figure 9, demonstrating that if you add the percentages up horizontally in the percentages by row graph you get 100% and vertically in the percentages by column graph you get 100%. The interviewer then asks Amy if she could form any hypothesis about how the

displayed non-conditional cell percentages are related in Figure 10. Amy responds "I don't know." The interviewer recommends adding them all up together and provides a calculator. "That equals 99." The interviewer explains that it is almost 100% and asks Amy to interpret one of the cells with this information.

Amy interprets that, "18% of that amount of fruit have nuts in them." With non-conditional cell percentages now displayed and the extra information about displayed percentages explained, it could be thought that Amy would return to her previous conception of each percentage representing something related to all the granola bars in this study. Amy instead says that the 18% is related to the amount of data points displayed in the lower left cell representing granola bars containing both fruit and nuts.

Amy struggled with proportional reasoning during this portion of the interview. Amy was not able to relate the proportional percentage shown in any of the graphs to the proper whole. The above responses show that Amy's proportional reasoning skills are unsound.

4.3.2 Brian's proportional reasoning analysis. During the interview Brian created the graph in Figure 11 and was asked to interpret what this graph is displaying:
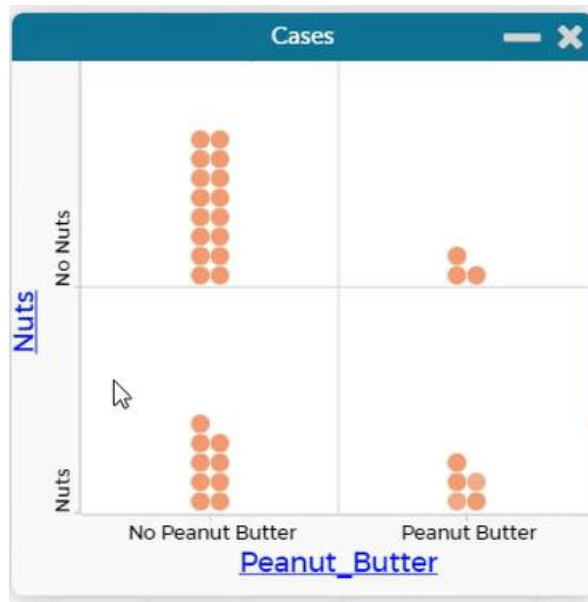
Figure 11. Brian's graph of peanut butter versus nuts with no counts/percentages displayed.

Brian responds, "There are four boxes and most of the no peanut butter have no nuts but some have nuts. Like three." Brian's response here is incorrect as a correct observation would have referred to the 9 in the lower-left cell in reference to the no peanut butter bars containing nuts.

During the interview, Brian created the graph in Figure 12 with non-conditional cell percentages displayed:
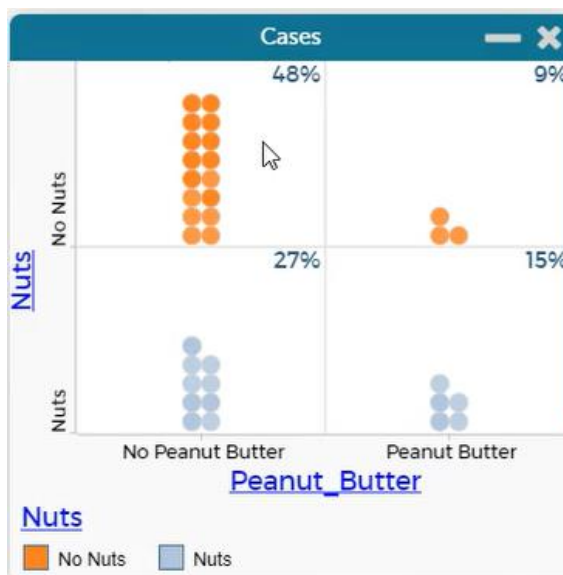
Figure 12. Brian's two-way binned plot of peanut butter versus nuts with cell percentages.

Brian's initial response when asked to interpret one of the cells was, "It [points to No Peanut Butter and No Nuts cell] has 48% of all of the rest of the boxes added together." To try to get Brian to clarify his analysis further the interviewer asks what type of things are in that box and then asks Brian to restate his claim. To which Brian responds, "The top left box [No Peanut Butter and No Nuts] has 48% of the whole thing…the whole granola bar thing which have no nuts and have no peanut butter." The student initially states that the 48% has something to do with the sum of all of the other cells but after some direction from the interviewer is able to review the graph again and correct their interpretation to relate the 48% to all granola bars. This shows that the student is correctly interpreting the non-conditional cell percentage by relating that 48% of all of the granola bars have no peanut butter and no nuts.

Brian later observed the graph in Figure 13, with percentages displayed on the vertical column, and attempted to interpret the graph:
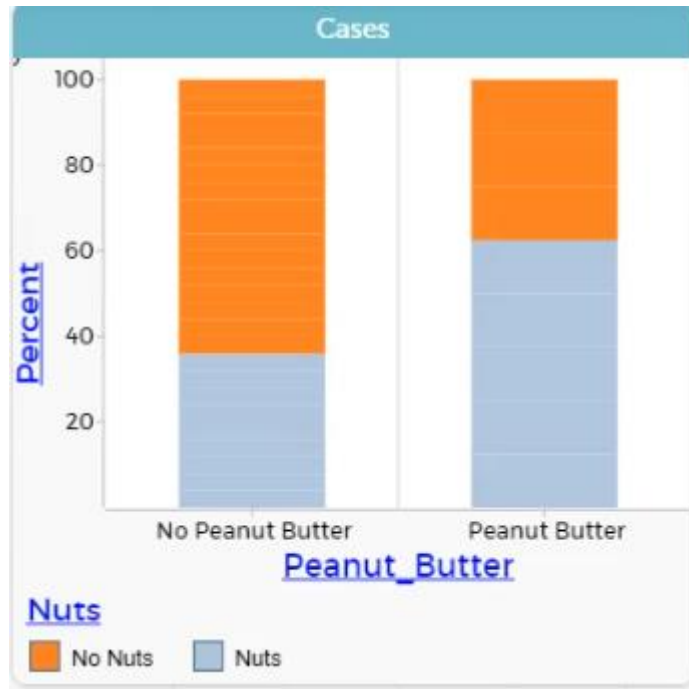
Figure 13. Brian's conditional segmented bar graph for peanut butter versus nuts.

Brian interprets, "This one [points to column labeled Peanut Butter] tells us that there are about 60% with nuts and peanut butter and 20% …I guess… with no nuts and peanut butter still." When asked to interpret the No Peanut Butter portion of the graph, Brian responds with, "It says that there are around 35 with no nuts…with nuts but no peanut butter and the rest has no nuts and has no peanut butter." Brian makes two errors in these statements with the claim that removing 60% from the whole would leave 20% remaining and claiming that around 35 with no nuts rather than 35%. As Brian saw more graphs, he seemed to be able to correct these missteps.

Brian later observed the graph in Figure 14, with row percentages displayed, and attempted to interpret the graph:
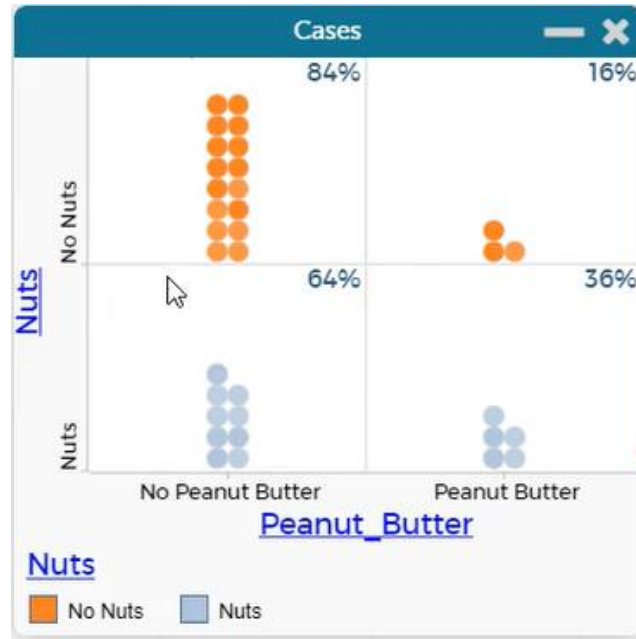
Figure 14. Brian's two-way binned plot of peanut butter versus nuts with row percentages.

Brian interprets, "64% of the things that have nuts… 64% of the granola bars that have nuts in it… Have nuts and no peanut butter."

While the previous errors show that Brian is still developing his proportional reasoning skills, it was determined that he accurately interpreting the displayed percentages as they relate to proper proportions of the given dataset. Brian often struggled to view the dataset as a whole and tended to have a localist view of the data. These struggles properly identifying what is the representative "whole" in which the proportion is describing lead to initial incorrect descriptions of the displayed numbers in correct relative terms. From the above responses, we were able to determine that Brian has good proportional reasoning skills that are still developing.

4.3.3 Claire's proportional reasoning analysis. During the interview, Claire created the univariate graph in Figure 15 with percentages displayed:
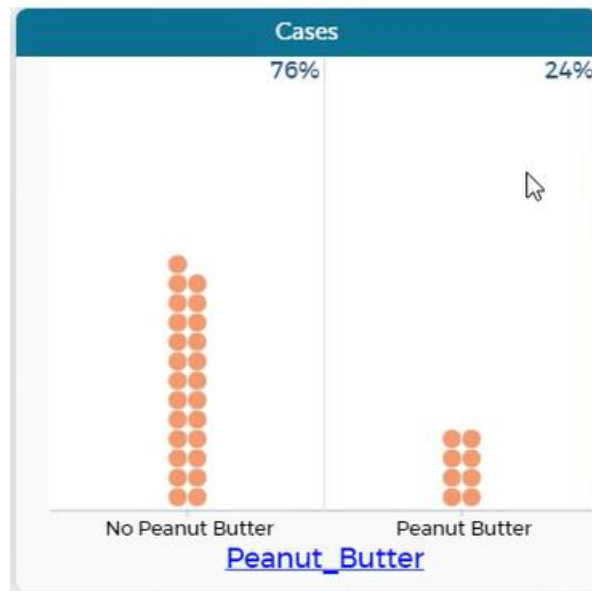
Figure 15. Claire's univariate graph of peanut butter with percentages displayed.

When asked to interpret the graph Claire responded, "76% of the granola bars have no peanut butter and 24% do." Claire was accurately able to relate the proportions displayed to all granola bars in the data set.

During the interview, Claire created the graph in Figure 16 with row percentages displayed:
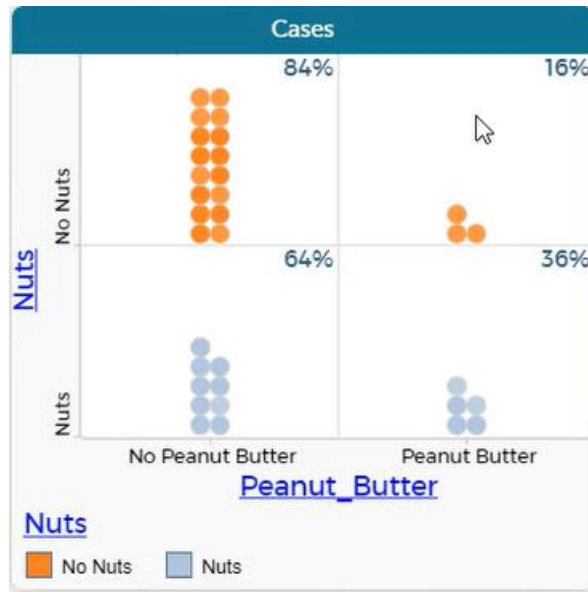
Figure 16. Claire's two-way binned plot of peanut butter versus nuts with row percentages.

When asked to interpret one of the cells Claire responded, "Ok so this one [points to Peanut Butter with No Nuts cell] says that only 16% of the granola bars that have peanut butter in them and no nuts." Claire incorrectly interprets the percentage as a column percentage instead of a row percentage.

The interviewer next has Claire change to display the graph in Figure 17 with column percentages as below:
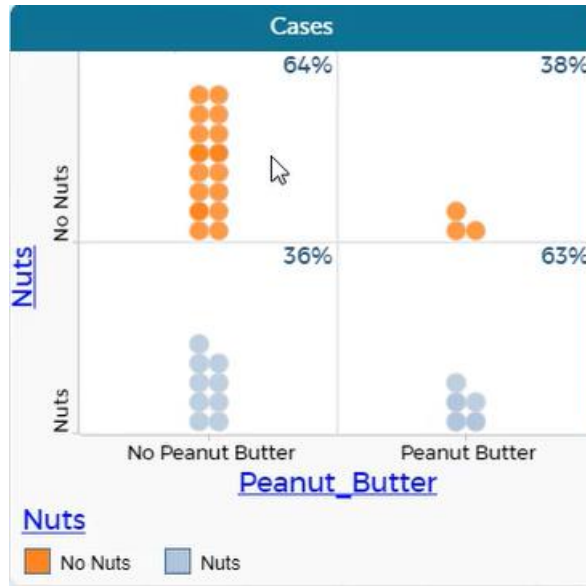
Figure 17. Claire's two-way binned plot of peanut butter versus nuts with column percentages.

Claire is asked to interpret one of the cells and responds, "Ok so this one [points to No Peanut Butter with Nuts cell] says that 36% of the bars that have no peanut butter have nuts." Claire is now able to accurately relate the displayed number to the proper whole after some guidance about how row and column percentages are displayed.

During the interview, Claire observed the graph in Figure 18 with non-conditional cell percentages displayed:
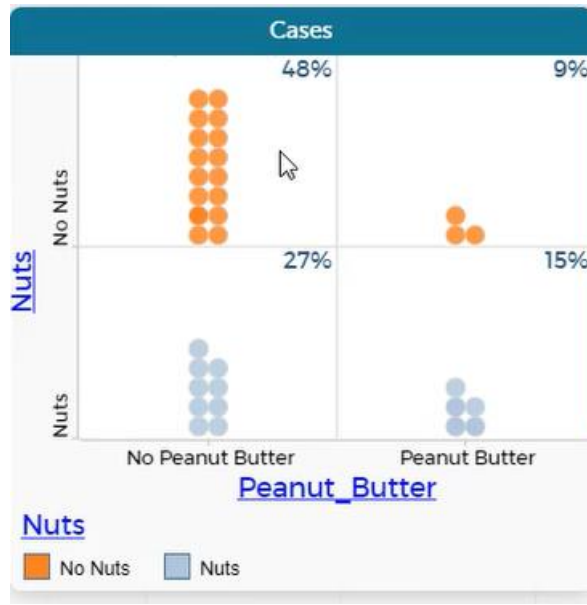
Figure 18. Claire's two-way binned plot of peanut butter versus nuts with cell percentages.

Claire's initial response when asked to interpret one of the cells was "So this one [No Nuts and No Peanut Butter] 48% of all no nuts have no peanut butter in them" but after thinking about it for a short time Claire corrects here interpretation by stating, "48% of the granola bars have no peanut butter and no nuts." The student initially states that the 48% is for granola bars with no nuts but after thinking corrected the interpretation as to all granola bars. This shows that the student is correctly interpreting the non-conditional cell percentage by relating that 48% of all of the granola bars have no peanut butter and no nuts.

Claire later observed the graph in Figure 19, with percentages displayed on the vertical column, and attempted to interpret the graph:
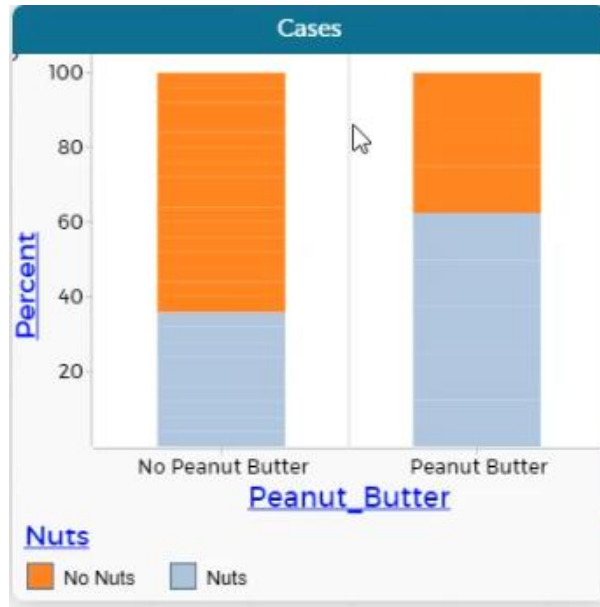
Figure 19. Claire's conditional segmented bar graph for peanut butter versus nuts.

Claire interprets, "About 40% of the granola bars that have no peanut butter have nuts and 60% of them that have no peanut butter have no nuts." And "Like almost 60% that have peanut butter have nuts and 40% that have peanut butter don't have nuts." This shows that Claire is accurately interpreting the displayed percentages as they relate to proper proportions of the given dataset.

While Claire made some initial errors in her proportional reasoning she was able to correct them with some thought and a little guidance in each case. This shows that Claire is still developing her proportional reasoning skills but is able to draw correct conclusions afterwards. From the above responses we were able to determine that Claire has good proportional reasoning skills that are still developing.

4.3.4 Proportional Reasoning Conclusion. From our analysis of our three participants, we were able to determine an initial proportional reasoning skill level for each student. We determined that Brian had sufficient but still developing proportional reasoning skills. Claire's

proportional reasoning skills are still developing, and Amy struggled throughout the study with unsound proportional reasoning skills.

From this analysis of initial skills, we may later examine if the more advanced proportional reasoning skills of students tends to lead to more accurate determinations of association.

4.4 Students' Interpretation of Graphs for Association

All three students were told to create three graphs (see Table 2) during the second part of the interview. One graph was a segmented bar graph, by percent with no association, examining the relationship between the variable texture (chewy or crunchy) and nut content. Another graph is a segmented bar graph, by count with association, which analyzes an organic labeling and the relationship with corn syrup content. The final graph is a two-way binned plot, with association, comparing fruit content with nut content. Each of the three students interpreted these graphs they constructed and identified if there was an association between the two variables. The students' responses are detailed below and informed the researchers' analysis of each student's ability to identify association.

4.4.1 Texture versus nuts (conditional segmented bar graph). Students observed the graph in Figure 20 with percentages displayed on the vertical axis:

Figure 20. Conditional segmented bar graph for texture versus nuts from Interview Part 2.

Our first student Amy correctly interprets the graph stating, "It's still almost to 50%… if it's chewy it's sometimes there is nuts in them but more of the times there is no nuts. Then if it's crunchy most of the time there's no nuts in them." This statement shows that the student is observing that the splits between the cases are close to even but that there are more cases of No Nuts when compared to Nuts.

When asked to identify if there is a relationship between the two variables Amy incorrectly concludes that there is a relationship. When asked to explain, Amy justifies here conclusion by stating, "Because if it's crunchy there's nuts in…most of the time there is no nuts in them." [Student hovers over grey portion of Crunchy column to reveal 4 of 10 (40%) Crunchy are Nuts...student then hovers over orange portion of column to reveal 6 of 10 (40%) Crunchy are No Nuts.] The student continues to explain further "Then if its chewy [Student alternates

hovering over both colors in the Chewy category to reveal 10 of 23 (43.5%) of Chewy are Nuts and 13 of 23 (56.5%) of Chewy are No Nuts] there's like a sometimes there's nuts in them and more of the times there's no nuts in them." Amy concludes that there is a relationship and justifies this by determining that a 60/40 split and 56.5/43.5 split are differences, when compared to a 50/50 split, and evidence that there is a relationship present between the two variables. While there may be some very small relationship, the displayed association is not large enough to be considered statistically significant.

Our second student Brian correctly interprets the graph in Figure 20 saying, "It tells me that that a lot of… well they're about the same the amount of nuts and no nuts that are chewy and the amount of crunchy and no nuts and nuts are basically the same." The student interprets that there is an approximate 50/50 split between No Nuts and Nuts for both Chewy and Crunchy granola bars. The student notes that there is a tiny difference between the proportional split of the two graphs.

When asked to identify if there is a relationship between the two variables, Brian initially, incorrectly, concludes that there is a relationship by analyzing the data in the raw data table. After being redirected to interpret the existence of a relationship using the graphical representations, the student changes his mind and correctly concludes that there is not enough evidence of a significant relationship.

Upon first inquiry, Brian states that there is a relationship: "Like when I look at here [student points towards raw data table on left] it shows that a lot of things that are chewy have no nuts…I meant have nuts in it. And a lot of things that have no nuts have… [trails off]." Once redirected to the graph Brian takes a few moments to look at the graphs before changing his mind and says, "There is no relationship because like if its chewy there's like…it's the same

with no nuts and nuts. And if it's crunchy there's a tiny difference but no nuts are more crunchy then nuts. So there's basically no change." Brian changes his mind to the no relationship conclusion because there is "no change" or an approximate even split between the variables.

Our final student Claire correctly interprets the graph in Figure 20 saying, "So a lot of the ones that are chewy have no nuts. Like this one [student points to crunchy column] I would have guessed that there's nuts in a lot of crunchy ones but there's not." Additionally, "Like about half of the chewy ones have nuts and about half don't have nuts. And probably the same about half of the crunchy don't have them." Claire correctly uses the counts on each bin to interpret the graph and even concedes that the data contradicts her initial assumption that most bars that are crunchy have nuts in them. This shows that Claire is interpreting the data from the graph and not using outside knowledge in her interpretation.

When asked to identify if there is a relationship between the two variables, Claire said that she doesn't know. When asked to think further and make her best deduction, Claire incorrectly concludes that there is a relationship. Claire states, "I think I'd say yes but this stuff [points to crunchy with no nuts] doesn't really prove anything." Claire adds, "It makes me think that it just like depends I guess on like how they make it or what else is in it." Claire struggles to make any conclusions from the graph and isn't able to justify her deduction with the given data. When faced with an unclear result, Claire falls back on her outside knowledge about granola bars to try to come up with a solution.

4.4.2 Organic versus corn syrup (stacked bar graph) Each of the three students in this study observed the graph in Figure 21:
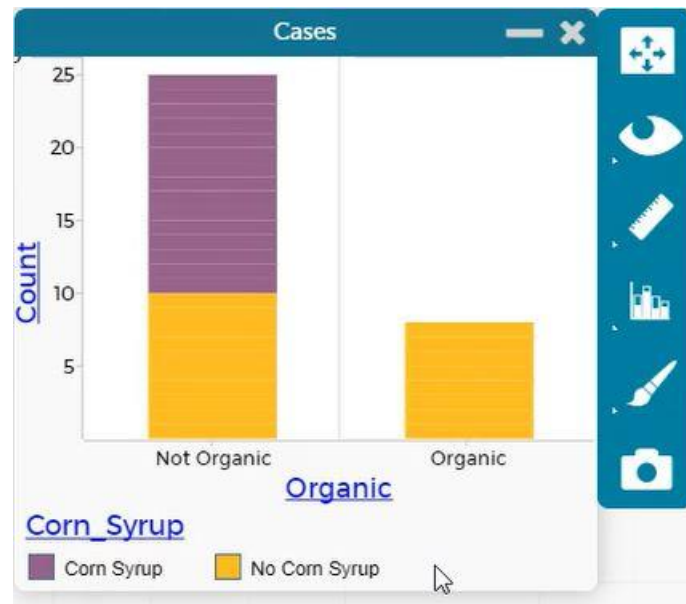
Figure 21. Segmented bar graph for organic versus corn syrup from Interview Part 2.

Our first student Amy correctly interprets the graph stating, "The organic ones have no corn syrup in them and the non-organic ones have usually corn syrup in them." Amy also observes, "For organic ones the yellow is supposed to have no corn syrup and then this one the purple is corn syrup and the yellow is no corn syrup." These statements show that Amy is correctly identifying the characteristics and proportions of each of the given variables.

When asked to identify if there is a relationship between the two variables, Amy correctly concludes that there is a relationship. Amy states, "If the granola bars are organic, then there's no corn syrup in them." Amy correctly interprets that none of the organic bars have corn syrup, and thus, there must exist a relationship between the two variables.

Our second student Brian correctly interprets the graph in Figure 21 stating, "A lot of…or all of the organic ones…the organic granola bars do not have corn syrup in it but more than half

of the non-organic ones has corn syrup. Which tells me that organic ones don't use corn syrup." We see that Brian correctly interprets the graph showing that there are no organic bars with corn syrup. The student also uses the displayed count scale on the vertical axis to estimate that more than half of the non-organic granola bars do contain corn syrup.

When asked to identify if there is a relationship between the two variables, Brian initially says there is no relationship, then changes to yes a relationship exists, before changing his mind again to incorrectly conclude that there is no relationship. When asked if there is a relationship between the two variables, Brian answers, "Nope because there are no…actually yes because ones that are organic granola bars there are no corn syrup." When asked to show where on the graph Brian can see this, he adds, "See right there [Student hovers over the organic bar which shows a pop up that 8 of 8 Organics (100%) are No Corn Syrup] there's like. It says that it's 100% are no corn syrup which means like this whole thing doesn't have corn syrup. So there's like no relationship." These statements show that Brian correctly interprets the data but is unsure whether the results indicate a relationship or not. Brian is unsure about his deduction and seems to struggle with a solidified understanding of what constitutes a mathematical relationship or not.

Our final student, Claire, correctly interprets the graph in Figure 21 by observing the graph and concluding, "If it has corn syrup it's definitely not organic." Claire correctly identifies that the organic category only contains data from the No Corn Syrup bars, while the not organic category contains data from both Corn Syrup and No Corn Syrup. Claire demarcates these details from the graph using the colors and scale of the bars.

When asked to identify if there is a relationship between the two variables, Claire correctly concludes that a relationship exists between the variables. In her analysis explanation, Claire justifies her conclusion by stating, "Because corn syrup is like not organic." Claire

correctly identifies that there is an association, yet her justification for this association appears to be based on outside knowledge that corn syrup is not organic. Claire doesn't use the graph or the data on the graph to justify her reasoning.

4.4.3 Fruit versus nuts (two-way binned plot) In this portion of the study, Brian and Claire observed the graph in Figure 22:
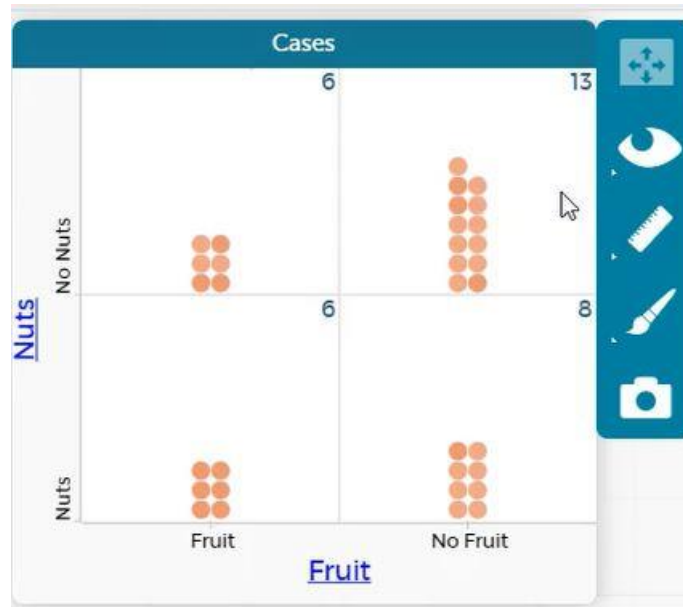
Figure 22. Two-way binned plot graph for fruit versus nuts from Interview Part 2.

However, Amy observed the graph in Figure 23 due to a deviation from the interview protocol by the interviewer:
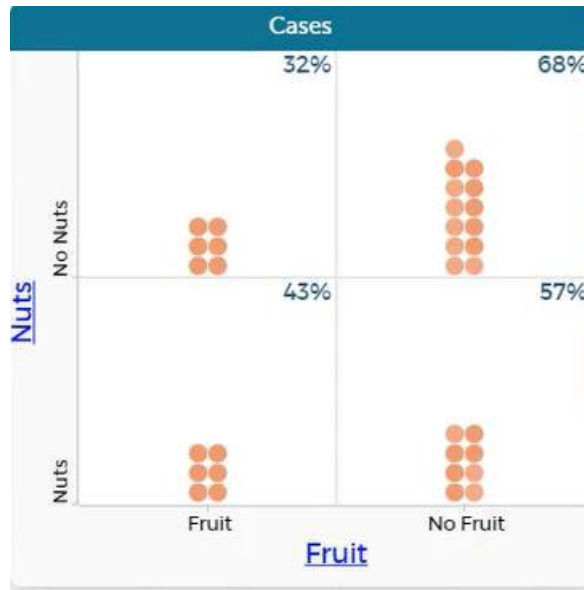
Figure 23. Amy's two-way binned plot graph for fruit versus nuts from Interview Part 2.

It is important to note that Amy's graph displayed row percentages, while Brian and Claire's graph displayed cell count. It is also important to note that Amy observed the graph displaying cell counts in an earlier portion of the interview. Amy's responses in both Part 1 and Part 2 of the interview were similar, and her conclusions were the same for all graphical representations of comparing the variable Nuts with the variable Fruit.

Our first student Amy incorrectly interprets the graph in Figure 23 stating, "There's 60…I don't know… um…uh… 57% of the ones with no fruit have nuts in them. And 68% of the ones with no fruit have no nuts in them. Then 32% of the ones with fruits have no nuts in them. Then 43% of the ones with fruit have nuts in them." Amy interprets the displayed percentages as column percentages when they are displayed as row percentages. Amy could have noticed that adding the percentages together horizontally results in 100% whereas summing them up vertically does not.

Amy correctly concludes that there is a relationship between the variables. Amy struggles to justify her claim stating, "Because a lot of them with the no fruit have no nuts in them. And

46

then…I don't know." Amy correctly states that there exists a relationship but when asked for justification cannot accurately justify their answer using the given graphs or data. Amy seems to struggle with the concept of what it means for there to be a mathematical relationship between two things.

Our second student, Brian, correctly interprets the graph in Figure 22 stating, "It tells us that most of the no fruits ones have no nuts and only some of them have not fruit with nuts. Well there's the same amount of ones with fruits that have no nuts and nuts." Brian correctly uses the displayed count for each cell to determine the number of bars in each category. He then makes the observation that there are the same number of bars with nuts as bars with no nuts in the category of granola bars containing fruit.

When asked to identify if there is a relationship between the two variables, Brian struggles to view the data as a whole and concludes that for some granola bars there is a relationship and for others there is not. When directed to make a conclusion about the data set as a whole Brian correctly states that there exists a relationship between Fruit and Nuts. Brian starts his analysis by answering the relationship question with, "For these two [points to the left two bins Fruit with nuts and Fruit with no nuts] there wouldn't be but I think for these two [points to the right two bins No Fruit with nuts and No Fruit without nuts] I think that there is." When asked to justify this hypothesis, Brian adds, "Because like here for no fruits there are no nuts but if we go down here there is like a tiny bit less like about 5 granola bars less that have nuts in it but have no fruits. Which tells me that that there's not a lot that have nuts in it but have no fruits." When asked to analyze the graphs on his own, Brian struggles to interpret the data as a whole and tries to parse the data into categories making the conclusion that for granola bars with no fruit there is a relationship with nuts/no nuts but for granola bars with fruit there is no

relationship. With some direction from the researcher, Brian comes to a correct conclusion about the existence of a relationship overall for the two variables.

The third student, Claire, correctly interprets the graph in Figure 22 stating, "There's a lot that have no fruit or nuts." Claire correctly uses the displayed count for each cell to determine the number of bars in each category. It is important to note that Claire only interprets one cell of the graph (a localist view), the no fruit or nuts cell, which is the cell with the highest count.

Claire incorrectly concludes that there is no relationship between Fruit and Nuts. Claire is initially unsure and responds, "Like maybe. Like they either put them both in or none of them." When asked to show where on the graph this can be seen, Claire states, "So like here [points to no fruit and no nuts bin] they put none and here [points to fruit and nuts bin] they put both. I know this one [points to no fruit with nuts bin] is like kinda the same." In her interpretation, Claire only analyzes the count on each bin and isn't able to draw any connections between count and association. If counts on all the bins are close to the same then the student should draw the conclusion that there is no association. Additionally, if the compared variables are proportionally close to one other when compared the students should also conclude that there is no association. If counts and/or proportions differ by a significant amount, when making all of these comparisons, students should conclude that there is an association.

4.5 Comparing Students' Strategies and Results

In an attempt to identify how students' conceptions of a statistical relationship and proportional reasoning skills inform their interpretation of graphs, I compared the results of students' graphical interpretations and claims of association within this study. Key factors examined were students' definition and justification for the existence of a relationship,

explanations and interpretations of the displayed graphs, and use of proportional reasoning in interpretations and explanations.

4.5.1 Texture versus nuts (segmented bar graph). Each of the three students in this study observed the graph in Figure 24:
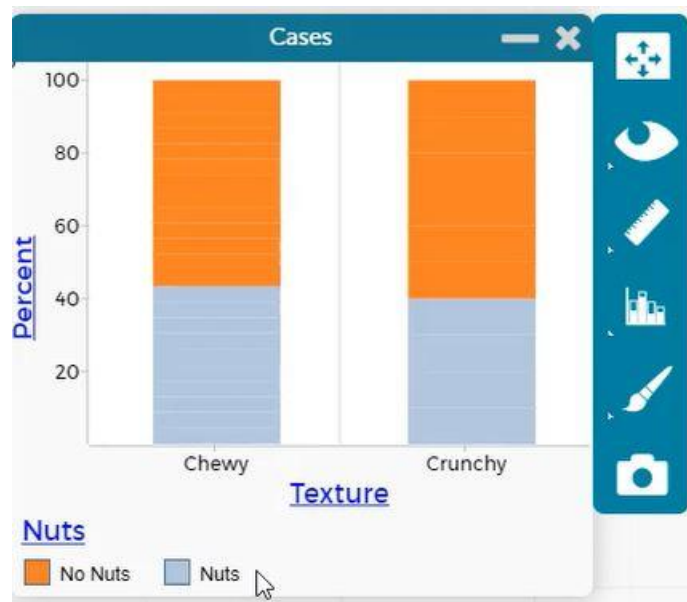
Figure 24. Conditional segmented bar graph for texture versus nuts from Interview Part 2.

All three of the students in this study were able to correctly interpret the graph with all students concluding that there is an approximate even split between the categorized variables. These data challenge most people's preconceived notions about the content of crunchy granola bars as one's intuition would generally say that crunchy granola bars are more likely to have nuts in them.

While viewing the graph in Figure 24, all three of the students in this study initially incorrectly claim that a relationship exists between these two variables. Brian's initial claim of a relationship is based off a granular view of the data table raw values rather than viewing the graphical representation of the data. After examining the graph and further consideration, Brian

concludes that his initial claim is incorrect and changes his interpretation to correctly conclude that there is no relationship. Brian justifies this claim by there is "no change" or an approximate even split between the variables which constitutes no relationship. Amy concludes that there is a relationship and justifies this by determining that a 60/40 split and 56.5/43.5 split is a difference and evidence of there being a relationship present between the two variables. Claire guesses that there is a relationship but isn't able to justify her claim using the graph. Claire states, "It makes me think that it just like depends I guess on like how they make it or what else is in it."

Brian's "no difference" conception of association seems to help him correctly interpret the association of the above variables. Amy has some similarities in her concept of association but incorrectly determines that if there is any difference at all, then it constitutes a relationship. Brian's more developed proportional reasoning skills appear to benefit his analysis when compared to Amy's. Brian is able to reason that the 60/40 split and 56.5/43.5 split are not proportionally a large enough difference to constitute a relationship. Amy seems to analyze the splits absolutely, determining that they are different rather than proportionally where they are not significantly different. Claire isn't able to find any graphical data that allows her to make a claim one way or the other and ends up relying on outside information rather than the displayed graph.

4.5.2 Organic versus corn syrup (stacked bar graph). Each of the three students in this study observed the graph in Figure 25:
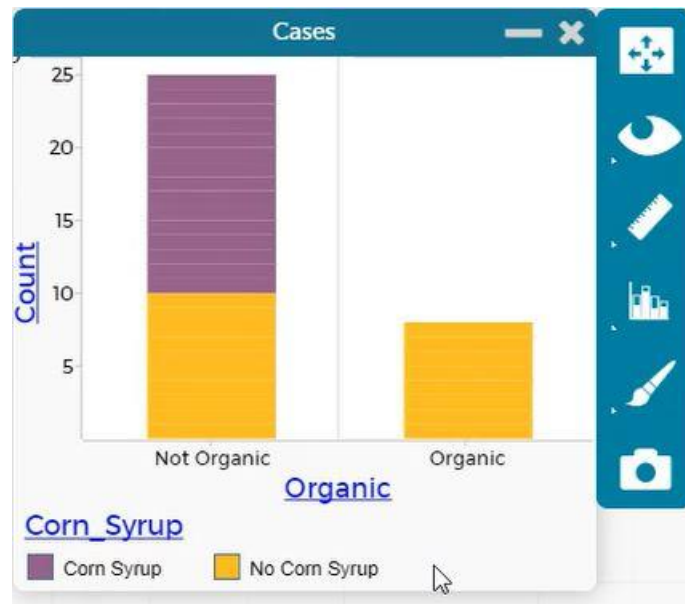
Figure 25. Segmented bar graph for organic versus corn syrup from Interview Part 2.

All three of the students in this study were able to correctly interpret the graph in Figure 25 with the students feeling comfortable that zero of the organic bars contained corn syrup. This follows most people's intuition about the content of organic granola bars and this data does not challenge these preconceived notions.

Students tended to accurately conclude that there is a relationship between the two variables with both Amy and Claire correctly coming to this conclusion. Since this graph showed 100% of organic bars did not contain corn syrup, students did not need to significantly apply proportional reasoning into their analysis. Brian struggled to determine if this data constituted a relationship or not. Brian went back and forth between claims that a relationship exists and that there is no relationship. It appears that Brian's lack of a solidified definition of what constitutes a relationship prevents him from coming to the correct conclusion. Claire concludes that there is a

relationship based on the outside knowledge that organic granola bars do not contain corn syrup. Claire is not able to justify her claim with the displayed graph and this also may be due to a lack of a solidified definition of association.

     4.5.3 Fruit versus nuts (two-way binned plot). It is important to point out two differences in Amy's interview when compared to our other two subjects. First, during the exploration portion of the interview, Amy selected to explore the relationship between fruit and nuts. Association between these two variables were explored by all participants in the second portion of the interview, and thus Amy had more exposure to data about these two variables. Amy's interpretation of the graph was the same (incorrect) in both portions of the interview. Secondly, Amy's two-way binned plot on second viewing was displayed with row percentages, while the other two participants viewed the graph with counts displayed. As a reminder, Amy viewed the graph in Figure 26:
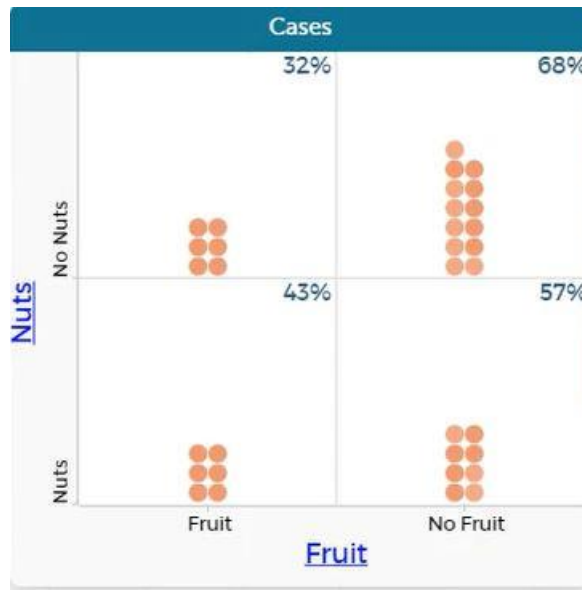
Figure 26. Amy's two-way binned plot graph for fruit versus nuts from Interview Part 2.

     The other two students in this study viewed the graph in Figure 27:
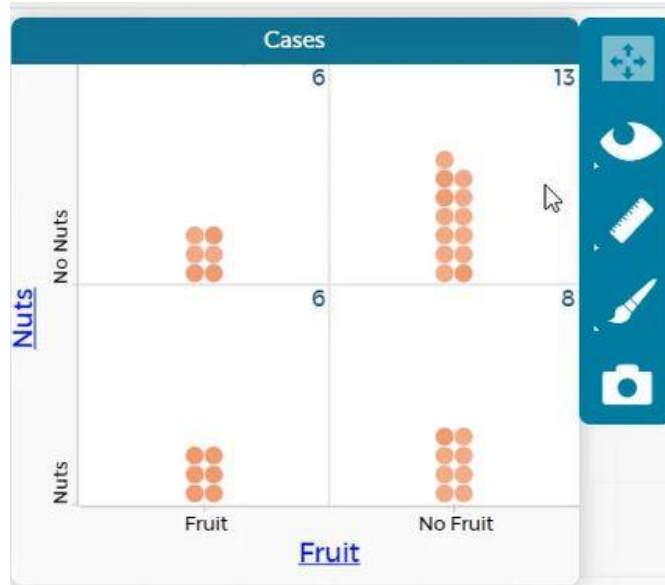
Figure 27. Two-way binned plot graph for fruit versus nuts from Interview Part 2.

Brian and Claire are correctly able to interpret the graph in Figure 27. Amy incorrectly interprets the displayed percentages in Figure 26 as column percentages when they are row percentages.

Each of the three students struggled with identifying a relationship with these variables. Amy correctly claims that there is a relationship but is not able to justify her claim with the displayed graph. Brian fails to interpret the data as a whole and tries to parse the data into categories, making the conclusion that for granola bars with no fruit there is a relationship with nut content but for granola bars with fruit there is no relationship. Claire also attempts to analyze the data and claims that maybe there is a relationship sometimes and maybe not. Claire only analyzes the count on each bin and isn't able to draw any connection between count and association.

Overall, students were most successful in analyzing the stacked bar graph with organic versus corn syrup and struggled to correctly interpret and analyze the other two graphs. From our analysis of students initial proportional reasoning skills, we found that all students in this study

struggled to view and interpret the data as a whole and displayed localist tendencies. This factor

appears to be the key concept that leads to each student's struggle to correctly identify

association and justify their claim using the displayed graph.

4.6 Comparing Results to Previous Study

In a previously conducted study (Casey, Hudson, & Ridley, 2018), 13 middle school

students were surveyed and asked to interpret pre-constructed graphs analyzing categorical

variable from the same granola bar dataset. Of the 13 students, four were finishing sixth grade

and nine were finishing seventh grade. This study found that "…the number of students making

correct claims about the variables' association…was quite small" (Casey et al., 2018). We look

to compare results from this study to the findings of the previous study, noting that sample size,

graph appearances, and study population differ from this study.

4.6.1 Texture versus nuts (segmented bar graph). In the previous study, students were

shown the graph in Figure 28:

	NO NUTS	NUTS
CHEWY	13	10
CRUNCHY	6	4
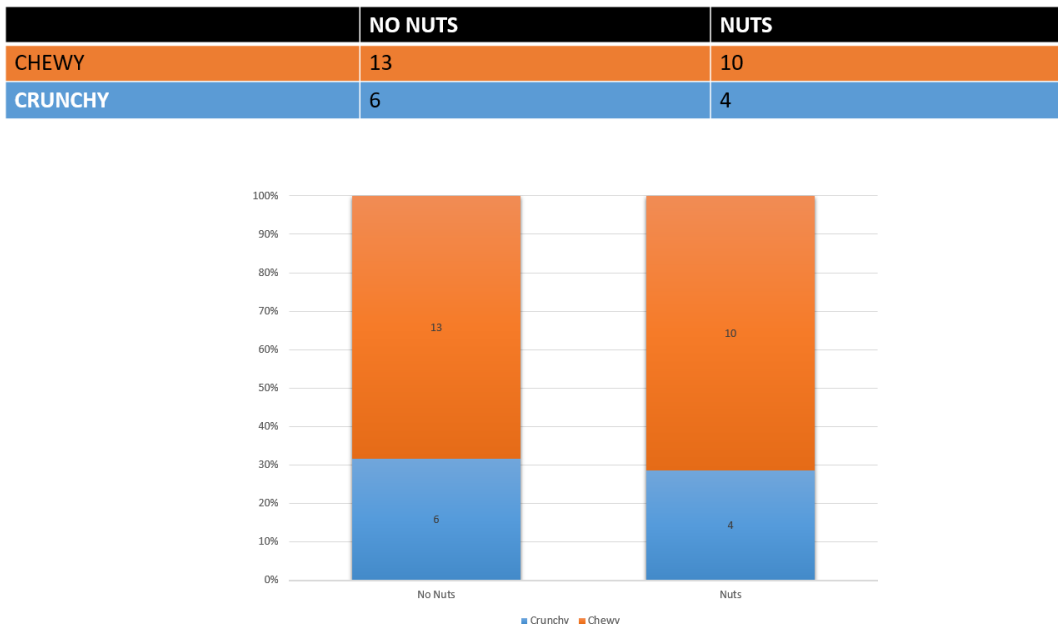
Figure 28. Texture versus nuts graph. Adapted from "Students' Reasoning about Association of
Categorical Variables," by Casey et al., 2018, Proceedings of Tenth International Conference on

Teaching Statistics (ICOTS10, July, 2018)). Copyright 2018 by the International Statistical Institute. Adapted with permission.

However the students in the present study constructed the graph shown in Figure 29:
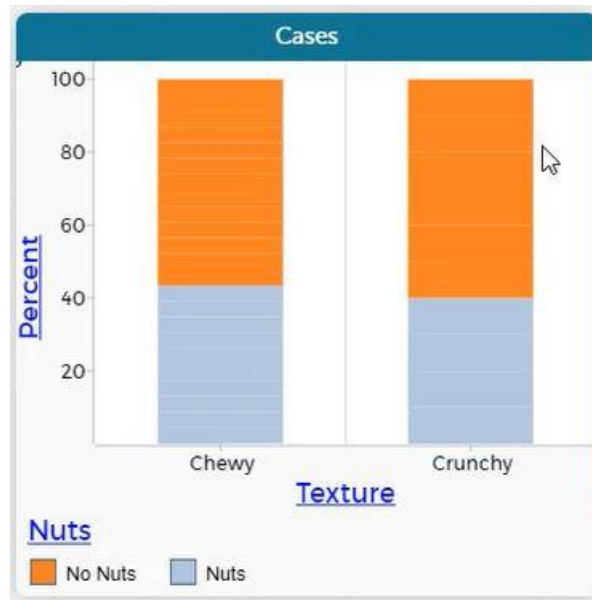
Figure 29. Conditional segmented bar graph for texture versus nuts from Interview Part 2

We note that there are differences in the above graphs as the previous study graph in Figure 28 displays the count, both on the bar graphs and in the data table above, and relative percentage for the variables, while our study graph only displays percentages. The graph from this study in Figure 29 also has columns listed as Chewy and Crunchy with colors delineating No Nuts versus Nuts. The previous study inverts this listing columns as No Nuts and Nuts with colors delineating Chewy versus Crunchy.

The previous study found that 62% of surveyed students correctly interpreted the graph (Casey et al., 2018). This study concluded that 100% (all three) of the students correctly interpreted the graph in Figure 29. The previous study also found that 23% of interviewed

students correctly identified association (Casey et al., 2018). This study similarly found that 33% (only one) of the interviewed students correctly identified association.

4.6.2 Organic versus corn syrup (stacked bar graph). In the previous study, students were shown the graph in Figure 30:

	ORGANIC	NOT ORGANIC
CORN SYRUP	0	15
NO CORN SYRUP	8	10
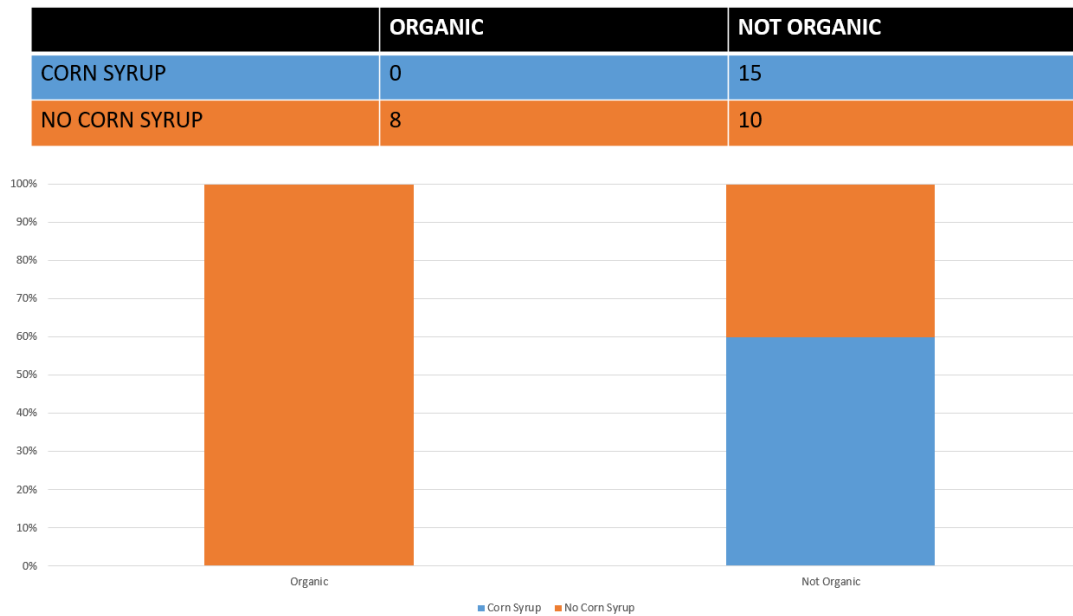
Figure 30. Organic versus corn syrup graph. Adapted from "Students' Reasoning about Association of Categorical Variables," by Casey et al., 2018, Proceedings of Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018)). Copyright 2018 by the International Statistical Institute. Adapted with permission.

However, the students in the present study constructed the graph in Figure 31:
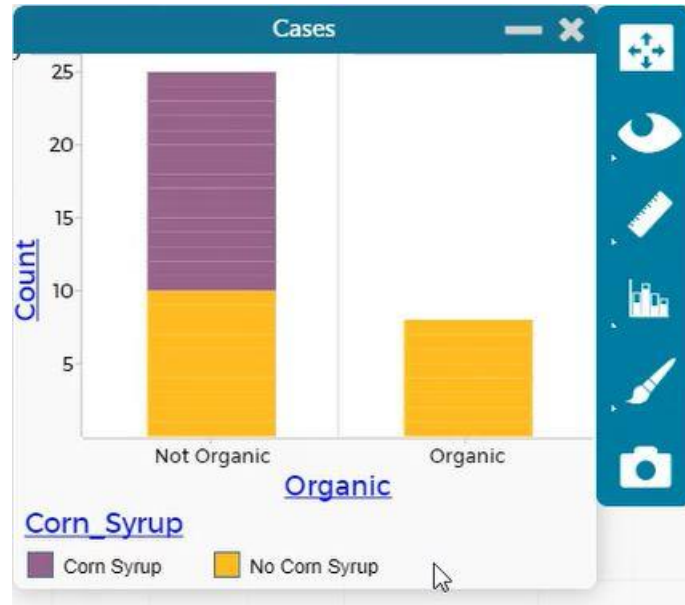
Figure 31. Segmented bar graph for organic versus corn syrup from Interview Part 2

There are important differences to note in the above graphs which may affect comparing given results. The previous study graph, Figure 30, displays percentages on the vertical axis, while this study displayed count. The previous study graph also displays the counts in numbers in the data table at the top, while this study's graph does not display this information.

The previous study found that 92% of interviewed students correctly interpreted this graph (Casey et al., 2018). Our study similarly concluded that 100% (all three) correctly interpreted the graph. The previous study also found that 62% of students correctly interpreted association (Casey et al., 2018). In this study, we similarly found that 66% of interviewed students correctly identified association.

4.6.3 Fruit versus nuts (two-way binned plot). In the previous study, students were

shown the graph in Figure 32:

	FRUIT	NO FRUIT
NUTS	6	8
NO NUTS	6	13
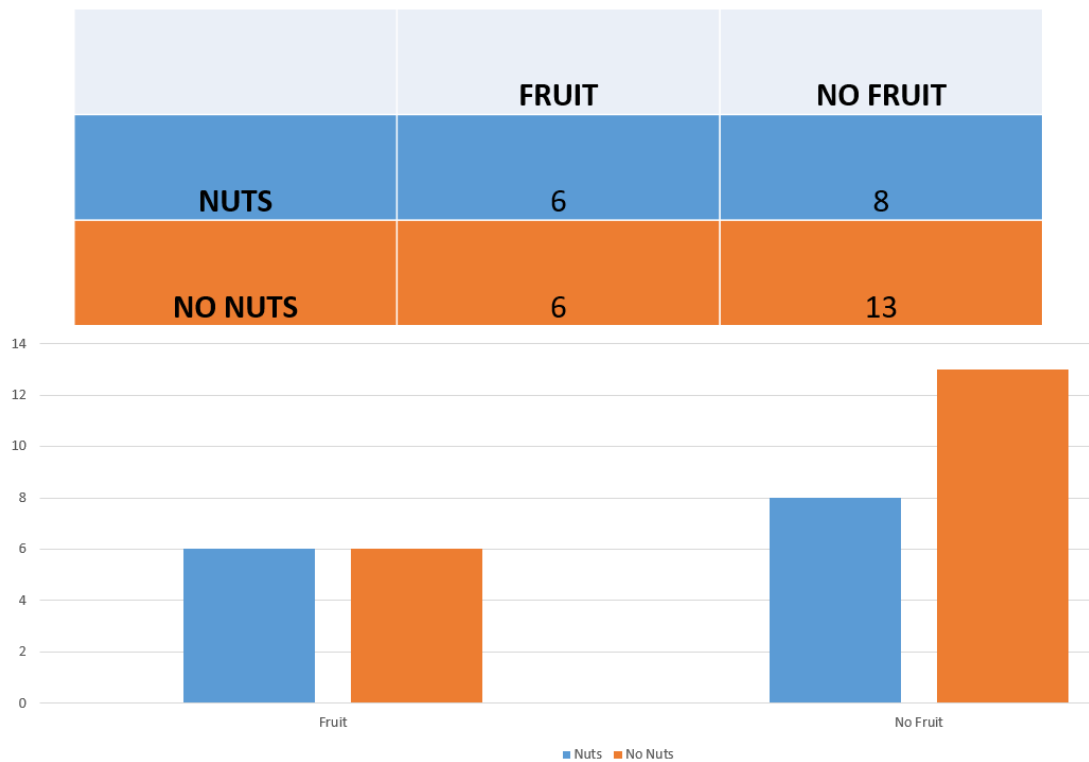
Figure 32. Fruit versus nuts graph. Adapted from "Students' Reasoning about Association of

Categorical Variables," by Casey et al., 2018, Proceedings of Tenth International Conference on

Teaching Statistics (ICOTS10, July, 2018)). Copyright 2018 by the International Statistical

Institute. Adapted with permission.

However, the students in the present study constructed the graphs in Figure 33 and Figure

34:
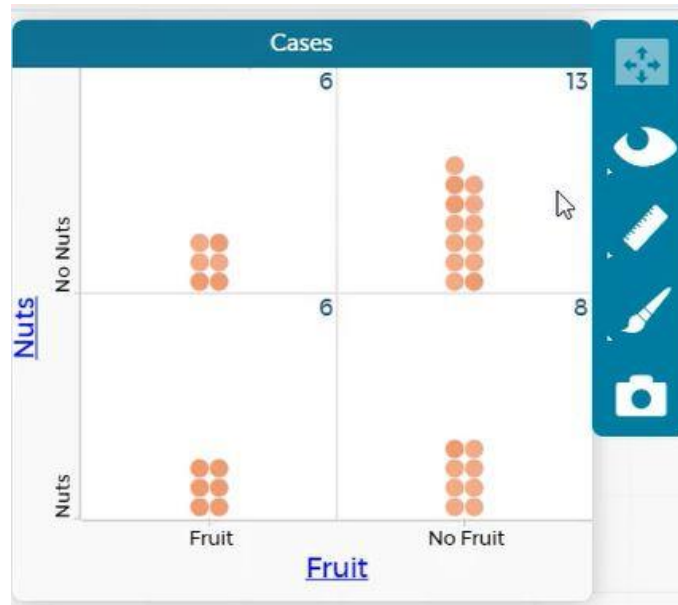
Figure 33. Two-way binned plot graph for fruit versus nuts from Interview Part 2.

and
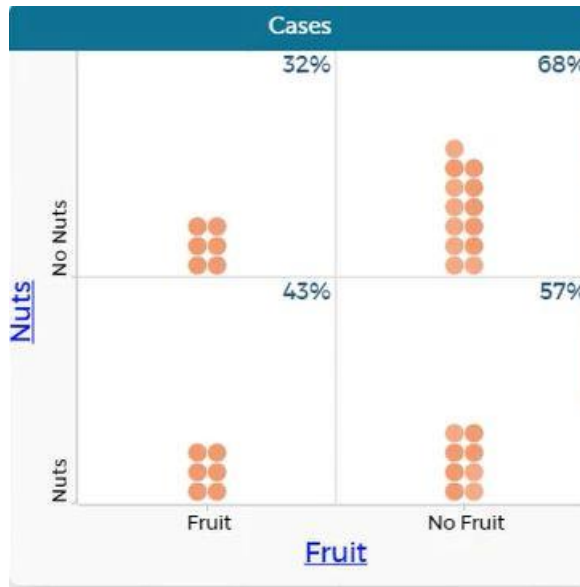
Figure 34. Amy's two-way binned plot graph for fruit versus nuts from Interview Part 2.

There are several important differences to note in the above graphs which may affect

comparing given results. The previous study graph, Figure 32, displayed side-by-side bar graphs.

Due to technological limitation this study was not able to construct side-by-side bar graphs and

instead chose to display the data as two-way binned plots. This study also chose to display,

Figure 33 and Figure 34, relative row percentages for two of the students and counts for one

student, while this previous study's graph, Figure 32, only displayed counts and not percentages.

The previous study graph also delineated the two variable columns using different colors, while

this study displayed variables using a monochromatic schema.

The previous study found that 46% of interviewed students correctly interpreted this

graph correctly (Casey et al., 2018). This study concluded that 66% of students correctly

interpreted the displayed graph. The previous study also found that 23% of students correctly

identified association with the displayed graph (Casey et al., 2018). This study similarly found

that 33% of students correctly identified association.

4.6.4 Conclusions from cross study comparison. We summarize the results of our

study comparison in Table 3. Students responses to graph interpretation and identifying

association are listed as either correct or incorrect and the proportion of correct responses is

compared to a previous study (Casey et al., 2018).

Table 3

Study Comparison Results Summary

Graphs	Tasks	Amy	Brian	Claire	Totals	Previous	% Change
Conditional Segmented Bar Graph	Graph Interpretation	Correct	Correct	Correct	100%	62%	+ 38%
Nuts versus Texture	Identifying Association	Incorrect	Correct	Incorrect	33%	23%	+ 10%
Segmented Bar Graph	Graph Interpretation	Correct	Correct	Correct	100%	92%	+ 8%
Organic vs Corn Syrup	Identifying Association	Correct	Incorrect	Correct	66%	62%	+ 4%
Two-Way Binned Plot	Graph Interpretation	Incorrect	Correct	Correct	66%	46%	+ 20%
Fruit vs Nuts	Identifying Association	Correct	Incorrect	Incorrect	33%	23%	+ 10%

Note. Compares results of this study with study results from "Students' Reasoning about Association of Categorical Variables," by Casey et al., 2018, Proceedings of Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018)). Copyright 2018 by the International Statistical Institute. Adapted with permission.

We can see that results from the current study where students created dynamic graphs in CODAP yielded increases in students' abilities to correctly interpret what the graph was demonstrating about the data. Results for students' correctness of identifying association was either similar or slightly better in this study when compared to previous results. We recognize that the small sample size of this case study (3 subjects) prevents the generalization of these results to sixth grade students in general. However, the results are encouraging though and will hopefully motivate future researchers to explore similar studies with larger student samples.

Chapter 5: Discussion

5.1 Introduction

This final chapter will summarize the results of this study, identify study limitations, make suggestions for further research, and discuss the implications of this investigation. Students in this study had no prior experience with interpreting bivariate categorical data.

5.2 Summary of Results

The results of this study provide insight into the statistical analysis skills of sixth grade students prior to receiving instruction about how to interpret bivariate categorical data. Through this investigation we were able to determine that sixth grade students do not have a well formulated definition of what constitutes a statistical relationship. Multiple students demonstrated a conceptual idea of a relationship being defined as happening more than 50% of the time. Sixth grade students still have developing proportional reasoning skills. These results aligned with previous studies that have shown that students' do not always develop proportional reasoning skill naturally and often develop gradually over a period of time (Hilton, Hilton, Dole, & Goos, 2013). The results showed that better developed proportional reasoning skills supported students' ability to accurately identify categorical association by assisting in graph interpretation, especially with two-way binned plots, and identifying if proportional splits are relatively equivalent.

Students struggled to correctly identify association when analyzing bivariate categorical graphs. These results aligned with previous studies that have demonstrated students' similar struggles (Batanero et al., 1996; Watson & Callingham, 2014). Our study differed from these previous studies in its use of CODAP's dynamical graphing utilities but similar results were shown. Students demonstrated an increased ability to correctly interpret statistical graphs using

dynamical graphical representation under this designed study methodology when compared to a previous study where students were provided pre-constructed graphs (see Table 1). These results aligned with previous studies that explore the benefits of dynamic statistical graphing programs such as TinkerPlots (Konold, 2002). Without a control group to compare this study with, we are not able to pinpoint exactly which factors are associated with these increased results. Students' abilities to correctly identify association were similar across both compared studies in this investigation (see Table 1).

5.3 Study Limitations

There were a number of limitations that should be taken into consideration when interpreting the findings. First, the sample size observed in this study was small (3 participants). Even though this is a major limitation of the study, the conclusions that arose from our analysis were in line with other studies having larger sample sizes.

A second limitation of this study was that the research was conducted at one middle school, in one sixth grade math class, in only one location. There is no indication of how results would compare across wider demographics or alternate locations.

Another limitation of this study is that there were also issues with one of the participant's questions and graphs not matching the other participants. Through a deviation from the interview protocol and some unforeseen technical issues, one of the student's interviews did not match the other two. This restricted the consistency of data across the study and limited analysis in areas.

A further limitation of this study was the current technological capabilities of the CODAP software program. It was not possible to recreate all types of graphs that were shown to students in a previous study which limited our scope of comparison.

An additional limitation is that this study was performed solely in English. It is possible that other languages have a stronger tie between their commonly used word for "relationship" or "association" and the statistical definition of these words than English does. Similarly, the study was done in the United States at a public school. Other countries and other curricula could show differing performance levels.

5.4 Recommendations for Further Research

The study results provide indication of areas that may benefit from future research. Research studies with similar methodologies to this investigation would benefit from investigating these findings with larger sample sizes and more diverse populations.

More research is needed on students' statistical knowledge of categorical association. Research dedicated towards identifying essential, pre-requisite skills for students to understand categorical association, such as specific aspects of proportional reasoning, would be especially impactful. As more research is conducted in regard to students' knowledge of categorical association, that research can be used to improve the pedagogical knowledge of current and future teachers.

5.5 Implications

These findings show that there are opportunities in the classroom to further develop proportional reasoning skills and accurate conceptions of association through the exploration of student-created bivariate categorical graphs using dynamic statistical graphing software. These findings also confirmed that students are not likely to develop these skills naturally on their own and imply that students need carefully planned learning experiences to develop an understanding of categorical association. In doing so, this study serves as a means to inform current curriculum and teaching methods. Teachers have the opportunity to correct common misconceptions

regarding statistical association and help students develop a sound understanding and definition of association through teaching and guided student-led investigations. Student-led statistical investigations using dynamic statistical graphing software can yield fruitful results and have the potential to increase students' understanding of categorical association. This study shows if students create graphs themselves, rather than viewing pre-made graphs, it appears to improve their ability to interpret them. The study has the potential to contribute to the understanding of different approaches to teaching bivariate data to foster productive student explorations and learning.

These results may also provide insight for future researchers by providing methodologies and results which can be compared to. Such information has the potential to offer background for further research focusing on student-led investigations of bivariate categorical data, students' abilities in relation to proportional analysis and identifying association, or dynamic statistical graphing software and their suitability for the targeted students.

References

Batanero, C., Estepa, A., Godino, J., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151-169.

Baterno, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics: Challenges for teaching and teacher education : A joint ICMI/IASE study.* Dordrecht; London: Springer.

Casey, S., Hudson, R., & Ridley, L. (2018). Students' reasoning about association of categorical variables. *In M. A. Sorto, A. White, & L. Guyot (Eds.), Looking back, looking forward. Proceedings of Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018)), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1-78.

Confrey, J., & Smith, E. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education*, 26(1), 66-86.

Cramer, D., & Howitt, D. (2004). *The SAGE dictionary of statistics a practical resource for students in the social sciences.* London: SAGE.

Fielding-Wells, J., Dole, S., & Makar, K. (2013). Inquiry pedagogy to promote emerging proportional reasoning in primary students. *Mathematics Education Research Journal*, 26(1), 47-77.

Franklin, C., Bargagliotti, A. E., Case, C. A., Kader, G. D., Scheaffer, R. L., & Spangler, D. A. (2015). *The Statistical Education of Teachers (SET).* Alexandria, VA: American Statistical Association (ASA).

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A PreK-12 curriculum framework.* Alexandria, VA: American Statistical Association.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education,* 19(1), 44-63.

Hiebert, J., Behr, M., Lesh, R., & Post, T. (1988). Number concepts and operations in the middle grades. *Research agenda project conference on number concepts in the middle grades* (pp. 93-118). Reston, VA: National Council of Teachers of Mathematics.

Hilton, A., Hilton, G., Dole, S., & Goos, M. (2013). Development and application of a two-tier diagnostic instrument to assess middle-years students' proportional reasoning. *Mathematics Education Research Journal*, 25(4), 523-545.

Konold. (2002). Alternatives to scatterplots. *Proceedings of The Sixth International Conference on Teaching Statistics* (pp. 1-6). Cape Town, South Africa.

Konold, C. (2007). Designing a data tool for learners. In M. C. Lovett, & P. Shah, *Thinking with data* (pp. 267-291). New York, NY: Lawrence Erlbaum Associates.

Manyika, J., Chui, M., Brown, B., Bughin , J., Dobbs, R., Roxburgh, C., & Hung Byers, A.

  (2011, May). Big data: The next frontier for innovation, competition, and productivity.

  McKinsey Global Institute.

National Governors Association Center for Best Practices, Council of Chief State School

  Officers. (2018). *Common Core State Standards.* Washington, DC: National Governors

  Association Center for Best Practices, Council of Chief State School Officers. Retrieved

  from Officers., National Governors Association Center for Best Practices and Council of

  Chief State School: http://www.corestandards.org/Math/Content/SP/

Ridgway, J. (2016). Implications of the data revolution for statistics education. *International

  Statistical Review*, 84(3), 528-549.

Staples, M., & Truxaw, M. (2012). An initial framework for the language of higher-order

  thinking mathematics practices. *Mathematics Education Research Journal*, 24(3), 257-

  281.

Tran, D., & Tarr, J. (2018). Examination of bivariate data tasks in US high school textbooks

  through the statistical investigation and cognitive demands frameworks. *International

  Journal of Science and Mathematics Education*, 16(8), 1581-1603.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not

  everything is proportional: Effects of age and problem type on propensities for

  overgeneralisation. *Cognition and Instruction*, 23(1), 57-86.

Watson, J., & Callingham, R. (2014). Two way tables: Issues at the heart of statistics and

    probability for students and teachers. *Mathematical Thinking and Learning*, 16(4), 254-

    284.

APPENDICES

Appendix A: IRB Approval Letter

Re: Expedited Review - Initial - UHSRC-FY17-18-242 Students Interpretation of Categorical Data Using Dynamic Graphical Representations

Dear Adam Eide:

The Eastern Michigan University Human Subjects Review Committee has rendered the decision below for Students Interpretation of Categorical Data Using Dynamic Graphical Representations. You are approved to conduct your research.

Decision: Approved

Selected Category:

Findings: You must use stamped copies of your recruitment and consent forms.

To access your stamped documents, follow these steps: 1. Open up the Dashboard; 2. Scroll down to the Approved Studies box; 3. Click on your study ID link; 4. Click on "Attachments" in the bottom box next to "Key Contacts"; 5. Click on the three dots next to the attachment filename; 6. Select Download.

Renewals: This approval is valid for one year and expires on April 18, 2019. If you plan to continue your study beyond April 18, 2019, you must submit a continuing review application in Cayuse IRB at least 14 days prior to April 18, 2019 so that your approval does not lapse.

Modifications: All changes to this study must be approved prior to implementation. If you plan to make any changes, submit a modification request application in Cayuse IRB for review and approval. You may not implement your changes until you receive a modification approval letter.

Problems: All deviations from the approved protocol, unanticipated problems, adverse events,

subject complaints, or other problems that may affect risk to human subjects *or* alter their willingness to participate must be reported to the UHSRC. Complete the incident report application in Cayuse IRB.

Please contact human.subjects@emich.edu with any questions or concerns.

Sincerely,

Eastern Michigan University Human Subjects Review Committee

RESEARCH @ EMU

Parental Consent Form

The person in charge of this study is Adam Eide. Adam Eide is a student at Eastern Michigan University. His/her faculty adviser is Dr. Stephanie Casey. Throughout this form, this person will be referred to as the "investigator."

Purpose of the study

The purpose of this research study is to learn how students reason about statistical association of two-variable data when using novel, dynamic representations in a new software program called CODAP.

Eastern Michigan University is paying for this research.

What will happen if my child participates in this study?

- Your child would be interviewed one time for approximately thirty minutes during school hours.
- The Investigator will ask your child questions and have him/her complete tasks involving analysis of statistical data for association using the new software program, CODAP.
- These interviews would be done at your child's school in a private setting, such as a conference room.
- They would be done at a mutually agreeable time for your child, the interviewer, and your child's mathematics teacher.

We would like to video record your child for this study. If your child is video recorded, it will be possible to identify him or her through his or her image. However, your child's real name and school name will never be used in association with the video. If you agree to allow your child to be video recorded, sign the appropriate line at the bottom of this form.

What are the anticipated risks for participation?

There are no anticipated physical or psychological risks to participation.

The primary risk of participation in this study is a potential loss of confidentiality.

Are there any benefits to participating?

You and your child will not directly benefit from participating in this research.

Approved by the Eastern Michigan University Human Subjects Review Committee
UHSRC Protocol Number: UHSRC-FY17-18-242
Study Approval Dates: 4/19/18 – 4/18/19

Benefits to society include helping educators in all fields that teach students about association of variables (e.g., nutrition, sociology) how students reason about association of variables when using dynamic representations. Understanding how students' ideas develop and connect is at the core of high-quality teaching.

What are the alternatives to participation?

The alternative is not to participate.

How will my child's information be kept confidential?

We will keep your child's information confidential by using a pseudonym to label all data collected from your child's interview. The linking document which lists each participating student's name and assigned pseudonym will be stored separately from the data on the hard drive of the Investigator's password-protected computer. The data will be stored on the hard drives of password-protected computers used by the Investigator. The child's school will never be named either; we will use a pseudonym for it as well. We will make every effort to keep your child's information confidential, however, we cannot guarantee confidentiality. There may be instances where federal or state law requires disclosure of your child's records.

Other groups may have access to your child's research information for quality control or safety purposes. These groups include the University Human Subjects Review Committee, the Office of Research Development, the sponsor of the research, or federal and state agencies that oversee the review of research. The University Human Subjects Review Committee is responsible for the safety and protection of people who participate in research studies.

We may share the data from your child's interview with other researchers outside of Eastern Michigan University. If we do so, it will only include the pseudonym assigned to your child so that your child cannot reasonably be identified.

The results of this research may be published or used for teaching. Identifiable information will not be used for these purposes.

Storing study information for future use

We would like to store your child's information from this study for future use in research studies concerning mathematics education. Your child's information will be labeled with a pseudonym and not your child's name. Your child's information will be stored in a password-protected file. Your child's de-identified information may also be shared with researchers outside of Eastern Michigan University. Please initial below whether or not you allow us to store your child's information:

_____Yes   _____No

Are there any costs to participation?

Participation will not cost you or your child anything.

Will my child be paid for participation?

Your child will not be paid to participate in this research study.

Study contact information

If you or your child has any questions about the research, you can contact the Principal Investigator, Adam Eide, at aeide@emich.edu or by phone at 734-487-1444. You can also contact Adam Eide's adviser, Dr. Stephanie Casey, at scasey1@emich.edu or by phone at 734-487-1664.

For questions about your child's rights as a research subject, contact the Eastern Michigan University Human Subjects Review Committee at human.subjects@emich.edu or by phone at 734-487-3090.

Voluntary participation

Participation in this research study is your and your child's choice. Your child will be asked independently for his/her assent to participate in the study. You and your child may refuse to participate at any time, even after signing this form, with no penalty or loss of benefits to which you and your child are otherwise entitled. If you and your child leave the study, the information your child provided will be kept confidential. You and your child may request, in writing, that your child's identifiable information be destroyed. However, we cannot destroy any information that has already been published.

Statement of Consent

I have read this form. I have had an opportunity to ask questions and am satisfied with the answers I received. I give my consent to for my child to participate in this research study.

Signatures

_____
 Name of Child

_____
 Name of Parent

_____   _____
Signature of Parent   Date

I agree to allow my child to be video recorded for this study.

_____   _____
Signature of Child   Date

I have explained the research to the parent and answered all his/her questions. I will give a copy of the signed consent form to the parent.

_____
Name of Person Obtaining Consent

_____   _____
Signature of Person Obtaining Consent   Date

RESEARCH @ EMU

Assent Form

Introduction

- You are being asked to participate in a research study. Research studies are conducted by scientists or other researchers to answer questions and learn new things.
- The researcher conducting this study is Adam Eide. Adam Eide is a student. His supervisor is Dr. Stephanie Casey. In this form, Adam Eide will be referred to as the investigator.
- The purpose of this study is to learn how students reason about statistical association of two-variable data when using novel, dynamic representations in a new software program called CODAP.
- Please read this form carefully and ask any questions you have before deciding to participate in this study.

Study Procedures

- If you agree to participate in this study, you will be interviewed one time for approximately thirty minutes.
- The Investigator will ask you questions and have you complete tasks involving analysis of statistical data for association using the new software program CODAP.
- These interviews will be in a private setting such as a conference room.
- Your participation will last for 1 study sessions, with each session lasting 30 MINUTES.
- We would like to video record you as you complete the interview for this study. If you are video recorded, it will be possible to identify you through your image. However, your real name and school name will never be used in association with the video. If you agree to be video recorded, sign the last line at the bottom of this form.

Risks

- There is a risk that people outside of the research study might find out some of your information. The investigator will do his best to protect your information, but cannot guarantee complete confidentiality.
- You might feel uncomfortable answering some of the questions in the interview. If any questions make you feel uncomfortable, you do not have to answer them. Instead you can also talk to the investigator about this, take a break, or stop the study.

Benefits

- You will not benefit from participating in this study.

Approved by the Eastern Michigan University Human Subjects Review Committee
UHSRC Protocol Number: UHSRC-FY17-18-242
Study Approval Dates: 4/19/18 – 4/18/19

- Other people might benefit from this study. For example, educators in many fields like sociology and biology can benefit from what is learned from this study to improve their teaching of statistical association.

Confidentiality
- The investigator will do everything he/she can to protect your information. However, the investigator cannot guarantee complete confidentiality.
- We will label all data collected with a pseudonym, so that your real name is never attached to your data. We will also never use the real name of the school that you attend.
- All of the collected data will be stored on the hard drives of the password-protected computers of the investigator.

Payments
- You will not be paid to participate in this study.

Voluntary Participation
- The decision to participate is up to you. You can refuse to participate in this study now or at any time. You can choose to participate and then, at any time during the study, choose to stop participating.
- Whether or not you participate in this study will have no effect on your grade in mathematics class.
- Your parents will also be asked to give permission for you to participate. Even if your parents let you participate, you can still refuse to participate.
- If you choose to participate and change your mind, you can ask the investigator to destroy all of your collected information. Please be aware that any published information cannot be destroyed.

Contact Information
- If you have questions about this study at any time, you can contact the investigator, Adam Eide at 734-487-1444 or aeide@emich.edu . You can also contact Adam Eide's advisor, Dr. Stephanie Casey, at 734-487-1664 or scasey1@emich.edu with any questions.
- If you have questions about your rights as a research participant, you can contact the Eastern Michigan University Human Subjects Review Committee (UHSRC) at 734-487-3090 or human.subjects@emich.edu. The UHSRC reviews and monitors research studies to make sure that participants' rights are respected.

Assent Statement

- By signing below, you indicate that you have read this form, that all of your questions have been answered to your satisfaction, and that you agree to participate in this research study.

Signatures

Name of Participant (print): _____

Signature of Participant: _____   Date: _____

Signature of Investigator(s): _____   Date: _____

··

I agree to be VIDEO recorded:

Name of Participant (print): _____

Signature of Participant: _____   Date: _____

Appendix D: Granola Bar Data Set

index	Brand	Type	Calories (kcal)	Fat (g)	SatFat (g)	Sodium (mg)	Carbs (g)	Fiber (g)	Sugar (g)	Protein (g)	Nuts	Chocolate	Peanut_Butter	Corn_Syrup	Texture	Fruit	Organic
1	Clif Bar and Compa.	Chocolate Chip	190	8	1.499	105	28	3	12	3	No Nuts	Chocolate Chips	No Peanut Butter	No Corn Syrup	Crunchy	No Fruit	Organic
2	Meijer Inc	Sweet & Salty	160	7	2.499	135	22	3	9	3	Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
3	Great Value	Almond	170	8	2.499	140	20	1	10	3	Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
4	Target	Sweet & Salty	160	6	1.998	125	23	1	11	3	Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
5	Family Dollar	Sweet & Salty Nut	170	8	1.998	160	20	2	6	4	Nuts	No Chocolate	Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
6	Schnuck Markets	Peanut Butter Chocolate Ch.	100	2.5	1.001	80	18	1	7	2	No Nuts	Chocolate Chips	Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
7	AHOLD	Red Fruits	130	3.5	0	105	22	3	10	5	Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Chewy	Fruit	Not Organic
8	KIND	Peanut Butter Berry	150	4.5	0.5	75	24	2.5	7	3	Nuts	No Chocolate	Peanut Butter	No Corn Syrup	Crunchy	Fruit	Organic
9	Essential Everyday	Oats & Honey	190	7	0.5	150	29	2	12	4	No Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
10	365 Everyday Value	Blueberry Vanilla	130	1	0	55	28	2	11	2	No Nuts	Chocolate Chips	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Organic
11	Nice!	Chocolate Chip	100	2	0.499	45	18	1	6	2	No Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
12	Envirokidz	Strawberry	100	2	0	65	18	2	6	2	No Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Chewy	Fruit	Not Organic
13	Atkins	Peanut Butter	210	11	3	240	19	5	1	15	Nuts	No Chocolate	Peanut Butter	No Corn Syrup	Chewy	No Fruit	Not Organic
14	Sunbelt Bakery	Chocolate Chip	140	5	3	50	21	1	9	2	Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
15	Special K	Dark Chocolate	110	3	1.5	85	17	4	7	4	No Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
16	Annie's	Peanut Butter Chocolate Ch.	100	3.5	1	105	16	1	6	2	No Nuts	No Chocolate	Peanut Butter	No Corn Syrup	Chewy	No Fruit	Organic
17	Cascadian Farm	Dark Chocolate Cranberry	140	4.5	1.001	65	24	1	10	3	Nuts	Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Not Organic
18	Bakery On Main	Peanut Butter & Jelly	130	4.5	1	110	21	4	8	4	Nuts	No Chocolate	Peanut Butter	No Corn Syrup	Chewy	No Fruit	Not Organic
19	Shurfine	Oats & Honey	190	7	0.5	150	29	2	12	4	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Crunchy	No Fruit	Not Organic
20	Big Y	Sweet & Salty	170	8	2.499	140	20	1	10	3	Nuts	No Chocolate	No Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
21	Market Pantry	Dipped Carmel Apple	130	3.5	2.499	55	23	1	16	1	No Nuts	Chocolate Dipped	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
22	Simply Balanced	Apple Berry	90	1.5	0	15	18	1	8	1	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Organic
23	Nature's Path	Peanut Choco Crunch	190	8	1	60	27	3	8	4	No Nuts	Chocolate Chips	Peanut Butter	No Corn Syrup	Crunchy	No Fruit	Organic
24	Select Choice	Oat & Raisin	100	1.5	0	40	19	1	7	2	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Not Organic
25	Food Lion	Chocolate Chip	140	5	3.999	45	22	1	14	1	No Nuts	Chocolate Dipped	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
26	Giant	Blueberry Vanilla	100	2.5	1.001	35	18	1	8	1	No Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Crunchy	No Fruit	Not Organic
27	ICA	Chocolate Chip	100	3	1.5	75	18	1	7	1	No Nuts	Chocolate Chips	No Peanut Butter	Corn Syrup	Chewy	No Fruit	Not Organic
28	Kashi	Cherry Dark Chocolate	130	2	0.5	50	24	3	8	4	Nuts	Chocolate Chips	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Not Organic
29	Nature's Promise	Raspberry	120	2	0.5	90	24	1	11	2	Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Not Organic
30	Kroger	Peanut Butter Chocolate Ch.	100	2.5	0.499	80	17	1	7	2	Nuts	Chocolate Chips	Peanut Butter	Corn Syrup	Chewy	Fruit	Not Organic
31	Cascadian Farm	Oatmeal Raisin	80	1	0	50	17	1	7	1	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Organic
32	Annie's	Gluten Free Oatmeal Cookie	110	3	1	10	21	3	7	1	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	Fruit	Not Organic
33	Nature's Path	Pumpkin-N-Spice	140	4	0.5	80	23	2	10	3	No Nuts	No Chocolate	No Peanut Butter	No Corn Syrup	Chewy	No Fruit	Organic

Cases (37 cases)

Granola Bars

WHAT DO YOU LOOK FOR IN A GRANOLA BAR?

This data set has information from the United States Department of Agriculture's Branded Food Products Database about 33 granola bars sold in the United States in 2017.

EXAMINE THE TABLE

1. Look at the column headers of the table. These are the attributes this data set has for the 33 granola bars.
2. What attributes do you care about in a granola bar?
3. What attributes would a nutritionist care about in determining whether a granola bar is 'healthy'?
4. From the table, can you identify what type of granola bars tend to have the most calories?

Next: CREATE A GRAPH

Data on comes from United States Department of Agriculture's Branded Food Products Database, June 2017

Product of ESTEEM: Enhancing Statistics Teachers Education with E-Modules

Interview Protocol

Supplies needed:

- Laptop computer running a Screencast, CODAP loaded with Granola Bar files
- Computer mouse
- Video camera positioned from side to capture student's face and gestures
- Basic calculator
- 2 pieces of blank paper for scratch work if needed
- Paper copy of interview protocol
- Paper copy of student assent form (pdf in Google Drive folder).

Research Question: Can dynamic graphical representations in CODAP improve students understanding of statistical graphs (particularly segmented bar graphs and binned dot plots), with and without association, when compared to pre-constructed graphical representations?

Introduction:

Introduce yourself.

Complete student assent

- *Talk through*
- *If student does not assent, thank student for his/her time and tell to return to class.*
- *If student assents, student and investigator sign in pen. Let student take paper copy of pages 1 and 2.*

Begin videorecording and screencast (if student assented).

Show GranolaBars CODAP file. Have you ever had a granola bar? We are going to take a look at a data set about 30 different granola bars and their attributes, like if it is crunchy or chewy or and whether it has peanut butter or fruit in it. Explain that the data comes from the United States Department of Agriculture's Branded Food Products Database, June 2017. Have student hover mouse over column headings and talk through meanings of attributes of the granola bars in this data set. Make sure to use 'attributes' vocabulary.

Interview Protocol

Note: Use probing and follow-up questions to clarify what student is doing and thinking. Never ask leading questions. Always have student use the mouse.

PART 1

1. Have students pick one categorical attribute (Nuts, Chocolate, Peanut_Butter, Corn_Syrup, Texture, Fruit, Organic) to investigate, and show them how to graph it with a binned plot in CODAP (click Graph, drag attribute to horizontal axis). Ask what they notice about this variable. Introduce count and percents. Ask what is something they learned about these granola bars from this analysis.
2. Ask if they think there's another attribute that might have a relationship with the first attribute they chose. What kind of relationship do you think they have? [If need an example to understand this prompt, state "For example, Peanut Butter comes from peanuts so do you think all granola bars with Peanut Butter in them must have nuts? What do you think?"]
3. How could you determine whether those attributes do have that kind of relationship? Try to do what the student suggests in CODAP. Can also use calculator/pen and pencil. Make sure student is making the decisions about what to do for analysis, you are just facilitating use of CODAP to do it (if possible).

Note: The following attributes can yield interesting analyses:

*Peanut_Butter vs. Nuts

*Chocolate vs. Fruit

*Fruit vs. Peanut_Butter

MAKE SURE COUNTS AND PERCENTS IS UNCHECKED BEFORE DOING THE FOLLOWING

If student needs assistance or has not come up with this approach on own, lead through following (NOTE: randomly assign half the students to do step 4 (binned plot graph with color legend). Other students skip directly to step 5 (two-way table)).

4. Create binned plot graph for one attribute. Add color for other attribute as a legend variable (drag to middle of graph).

5. Make two-way table in CODAP.
6. Pick a cell and ask: What do the dots in this cell mean?
7. Ask: Does this table help you see whether there's the relationship you thought there would be?
   a. If so, interpret the table.
   b. If not, what further information would you need?
   c. Is there evidence to support the relationship you expected?

Example

Nuts
No   Yes
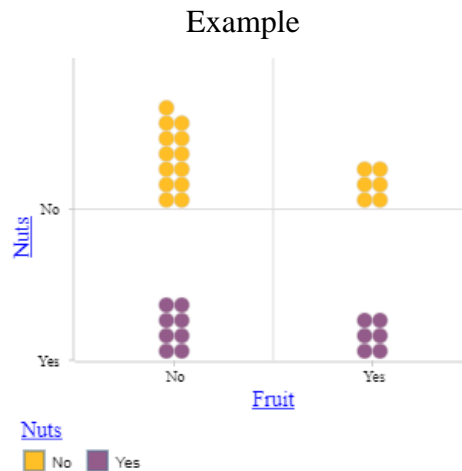
8. Ask them to determine whether there is a relationship between the attributes.

9. Add counts and percents in CODAP. Ask the students to interpret one of the percents. (Row Percent by default) (Use it in a sentence…e.g., 44% of ____ are ____.

10. Change to column percents. Ask student to interpret one of the percents.

11. Change to cell percents. Ask student to interpret one of the percents.

12. Ask: Of the row, column, and cell percents, which of these are useful in determining whether a relationship exists? How do you use [selected percents] to figure out if there is a relationship?

13. If there had been no relationship between these attributes, what would be true about the percents in the table?

14. Create a segmented bar graph with the students chosen variables. (Drag first attribute to horizontal axis, next drag second attribute to middle of graph to create segmented dots graph, finally select Fuse Dots into Bars from the Configure Display option on the side menu). By default the graph is set to count on the vertical axis. Ask student to interpret the graph with counts.

Example
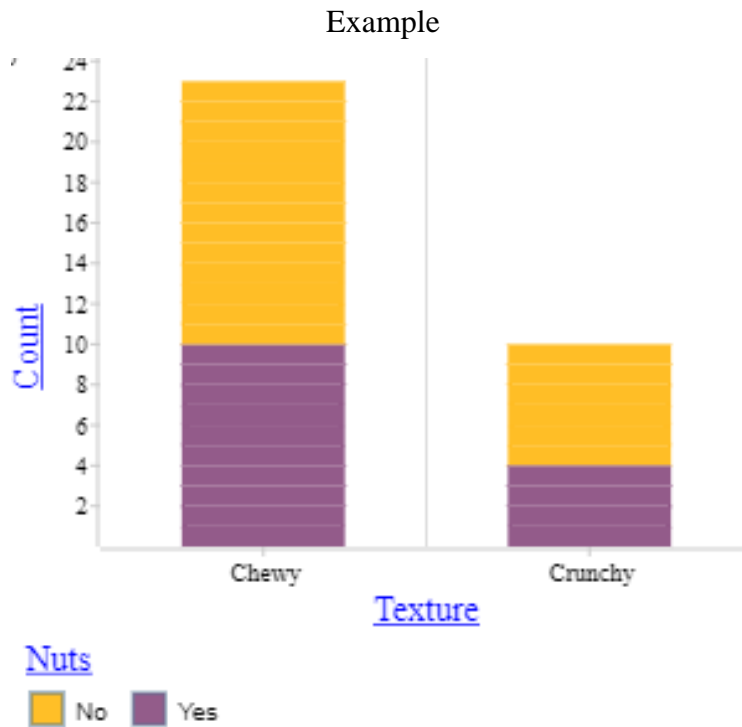
15. Change to percent (click configuration on the right side menu and select the percent button). Ask the student to interpret the graph with percent.

PART 2

NOTE: Randomize the order of creation for the below four graphs

Graph 1:

16. "Let's create a graph to look at the relationship between Texture (Chewy/Crunchy) and Nuts/No Nuts."

17. Have students drag attributes to create **segmented bar graph** for Chewy/Crunchy and Nuts/No Nut (Drag first attribute to horizontal axis, next drag second attribute to middle of graph to create segmented dots graph, finally select Fuse Dots into Bars from the Configure Display option on the side menu)

18. "What story is this graph telling you?" If student wants to use the table for analysis, prompt them to use the graph. "How can you see that in the graph? What is the graph showing you?"

Graph 2:

19. "Now let's create a graph to look at the relationship between Fruit/No Fruit and Nuts/No Nuts."

20. Have students drag attributes to create **binned dot plots** (with counts listed) for Fruit/No Fruit and Nuts/No Nut (Drag first attribute to horizontal axis, next drag second attribute to vertical axis, and finally drag second attribute to the middle of graph to create colored binned dot plots)

21. "What story is this graph telling you?" If student wants to use the table for analysis, prompt them to use the graph. "How can you see that in the graph? What is the graph showing you?"

Graph 3:

22. "Now let's create a graph to look at the relationship between Organic and Corn Syrup."

23. Have students drag attributes to create **segmented bar graph** for Organic and Corn Syrup. (Drag first attribute to horizontal axis, next drag second attribute to middle of

graph to create segmented dots graph, finally select Fuse Dots into Bars from the Configure Display option on the side menu)

24. "What story is this graph telling you?" If student wants to use the table for analysis, prompt them to use the graph. "How can you see that in the graph? What is the graph showing you?"

Graph 4:

25. "Now let's create a graph to look at the relationship between Fiber and Fruit/No Fruit."

26. Have students drag attributes to create **binned dot plots** (with counts listed) for Fruit/No Fruit/No Fruit and Fiber. (Drag first attribute to horizontal axis, next drag second attribute to vertical axis, and finally drag second attribute to the middle of graph to create colored binned dot plots)

27. "What story is this graph telling you?" If student wants to use the table for analysis, prompt them to use the graph. "How can you see that in the graph? What is the graph showing you?"

28. "What are the strengths and weaknesses of the different graphs?" [Student free to go back to look at all graphs]. "Do any graphs give you more or less information than the others? Explain."

29. "Which graph is most helpful to you in telling the story of the data? What about it makes it the most useful?" [If student doesn't address looking for relationship, ask which is most useful for determining if there is a relationship.]

We have reached the end of my questions. Do you have any additional questions or comments for me?
Thanks so much for doing this interview with me! You may return to your class now.

<u>Data to collect and upload for each student:</u>

- Screencast
- Video recording
- Completed GranolaBars CODAP file

<u>IRB documents to collect and upload for each student:</u>

- Parental consent form
- Student assent form

All student and school names will be removed from data prior to uploading, and replaced with pseudonyms. Input pseudonyms to "Pseudonym linking document".