

Engaging Students in Multivariable Thinking with Social Justice Data Investigations

Project Team: Josephine Louie (PI), Education Development Center (EDC); Beth Chance & Soma Roy (Co-PIs), California Polytechnic State University; Emily Fagan & Jennifer Stiles, EDC; Bill Finzer, Concord Consortium

Overview

The Strengthening Data Literacy across the Curriculum (SDLC) project (go.edc.org/ussdata) developed and has been studying curriculum materials for non-AP high school statistics to promote interest and skills in statistical thinking and data analysis among diverse high school populations. A major goal of the project has been to support students' **multivariable thinking**. We developed data investigations in which students use large-scale socioeconomic data from the U.S. Census Bureau and the Common Online Data Analysis Platform (CODAP) to address questions with social justice themes.

Definition of Multivariable Thinking

Multivariable thinking involves "an appreciation that the relationship between two variables may depend on other variables" (Carver et al., 2016, p. 10). People who engage in multivariable thinking do not draw conclusions about bivariate associations based on first appearances; they consider (and when possible, investigate) whether and in what ways these associations may change after accounting for the influence of other variables.

Curriculum Content and Sequence

The project developed two teacher-facilitated curriculum modules that each take approximately 15 hour-long class periods to complete. Students analyze random samples of person-level data from the American Community Survey (ACS) and the decennial census, which they extract from a Microdata Portal within CODAP, to address questions about society. They build multivariable thinking by conducting data investigations of increasing complexity using dynamic graphs in CODAP.

In one module, *Investigating Income Inequality in the U.S.*, lessons unfold as follows:

- **Initial ideas and data sources** (Lessons 1-2): Students discuss their conceptions of income inequality, how it may be measured, and strengths and limitations of income variables in the ACS and decennial census.
- **Baseline concepts and trends** (Lessons 3-5): Students address: "What are typical incomes in the U.S.?" and "How have higher- and lower-income earners in the U.S. been doing over time?" by examining distributions and statistical measures of annual wages across multiple decades.
- **Explicit focus on multivariable thinking** (Lessons 6-7): Students investigate "How much income inequality exists between males and females in the U.S.?" by examining distributions and statistical measures of annual wages subset by male and female earners. Then they address "Does education explain the wage gap between males and females?" by exploring 1) wages by education; 2) education by sex, and 3) wages by sex by education (Figures 1-4).
- **Independent application** (Lesson 8): In a final investigation, students choose their own third variable and investigate whether this factor explains the wage gap between male and female earners (Figures 5-6).

Multivariable Thinking when Investigating Income Inequality in the U.S.

Students investigate whether education explains the wage gap between male and female earners by making and examining a sequence of graphs using 2017 person-level microdata from the ACS in CODAP.

Figure 1. Income by sex

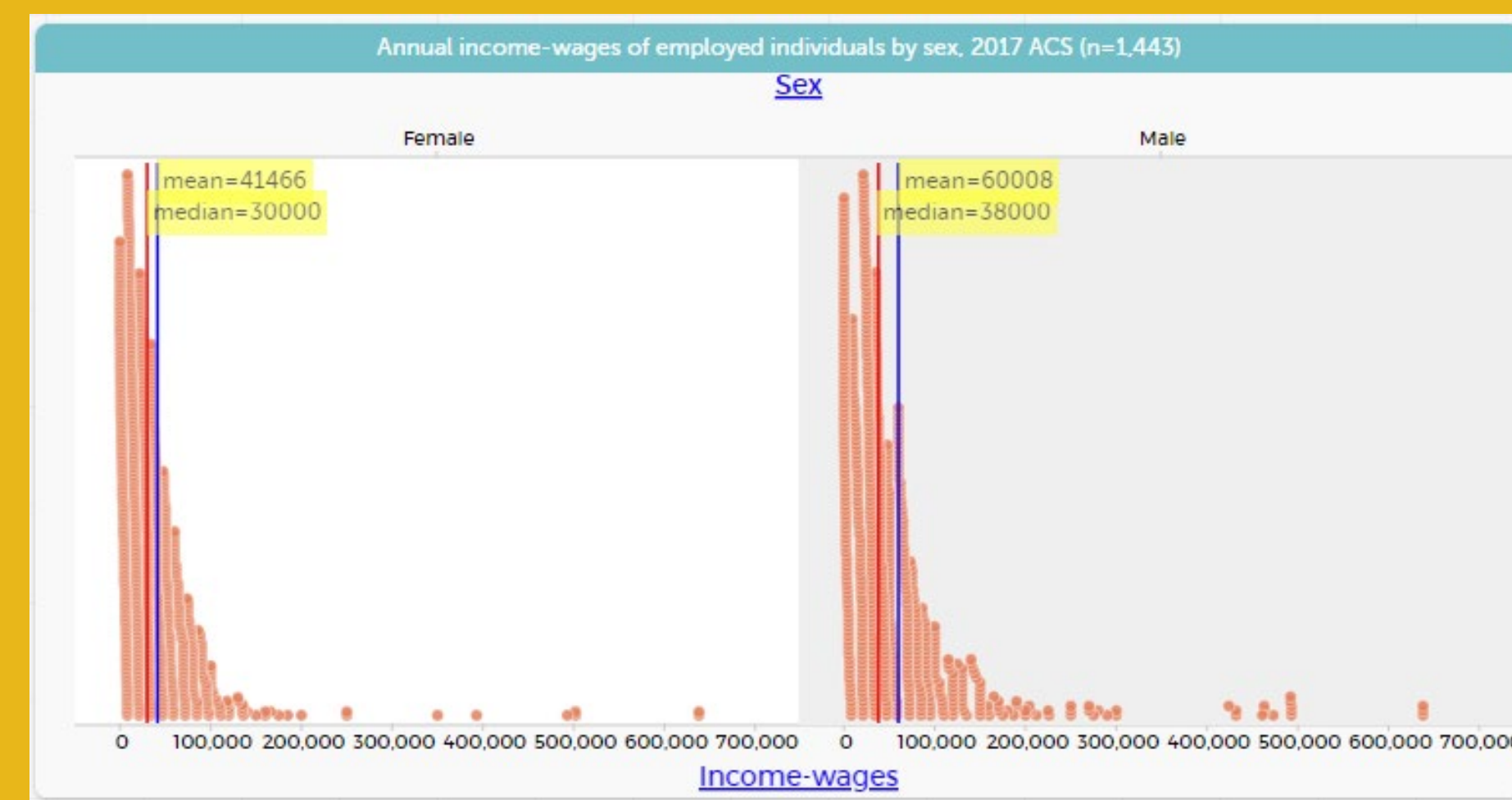


Figure 3. Education by sex



Figure 2. Income by education



Figure 4. Income by sex by education



In a Final Investigation, students choose another third variable to try to explain the male-female wage gap.

Figure 5. Student example: Race

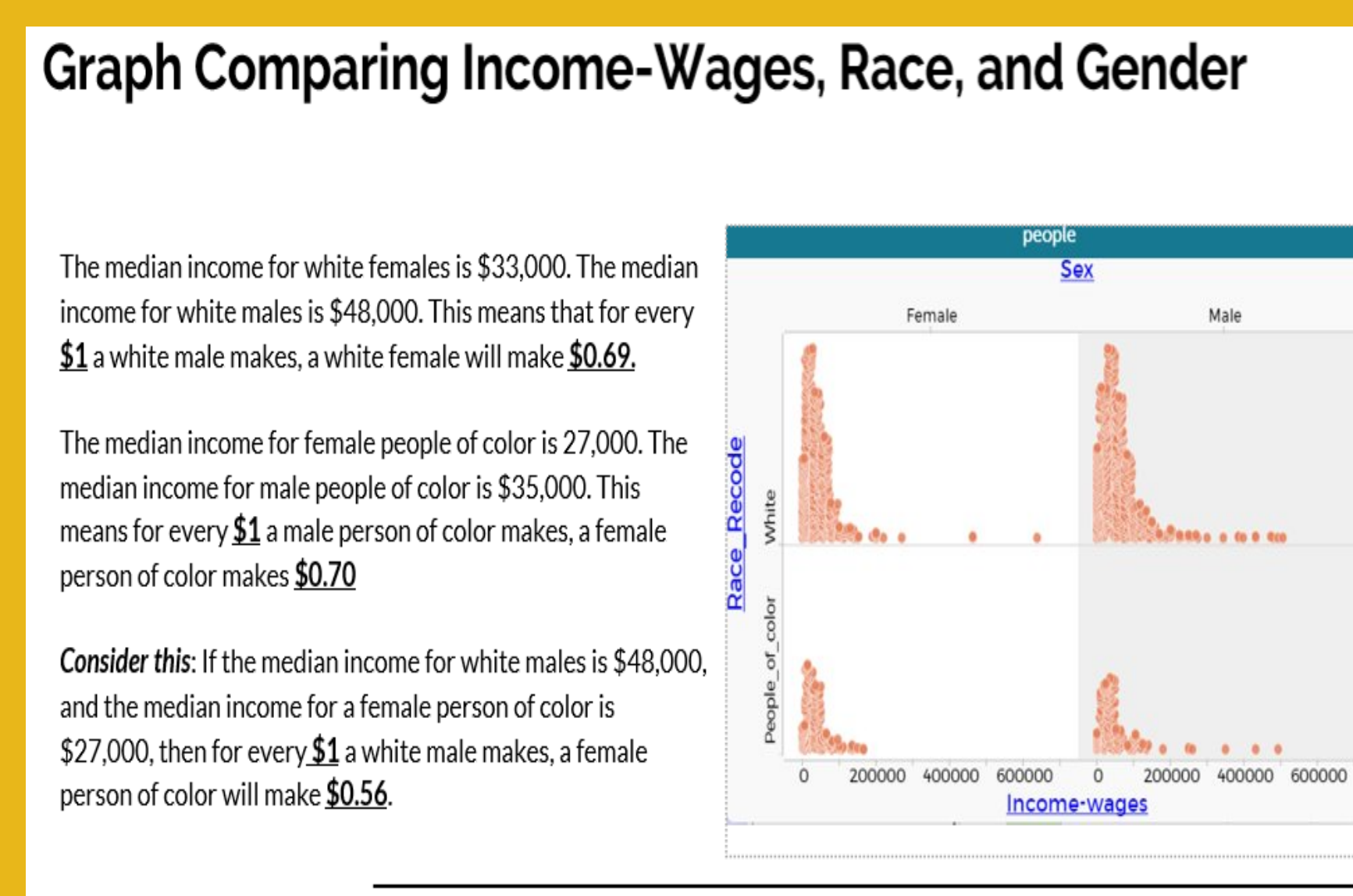
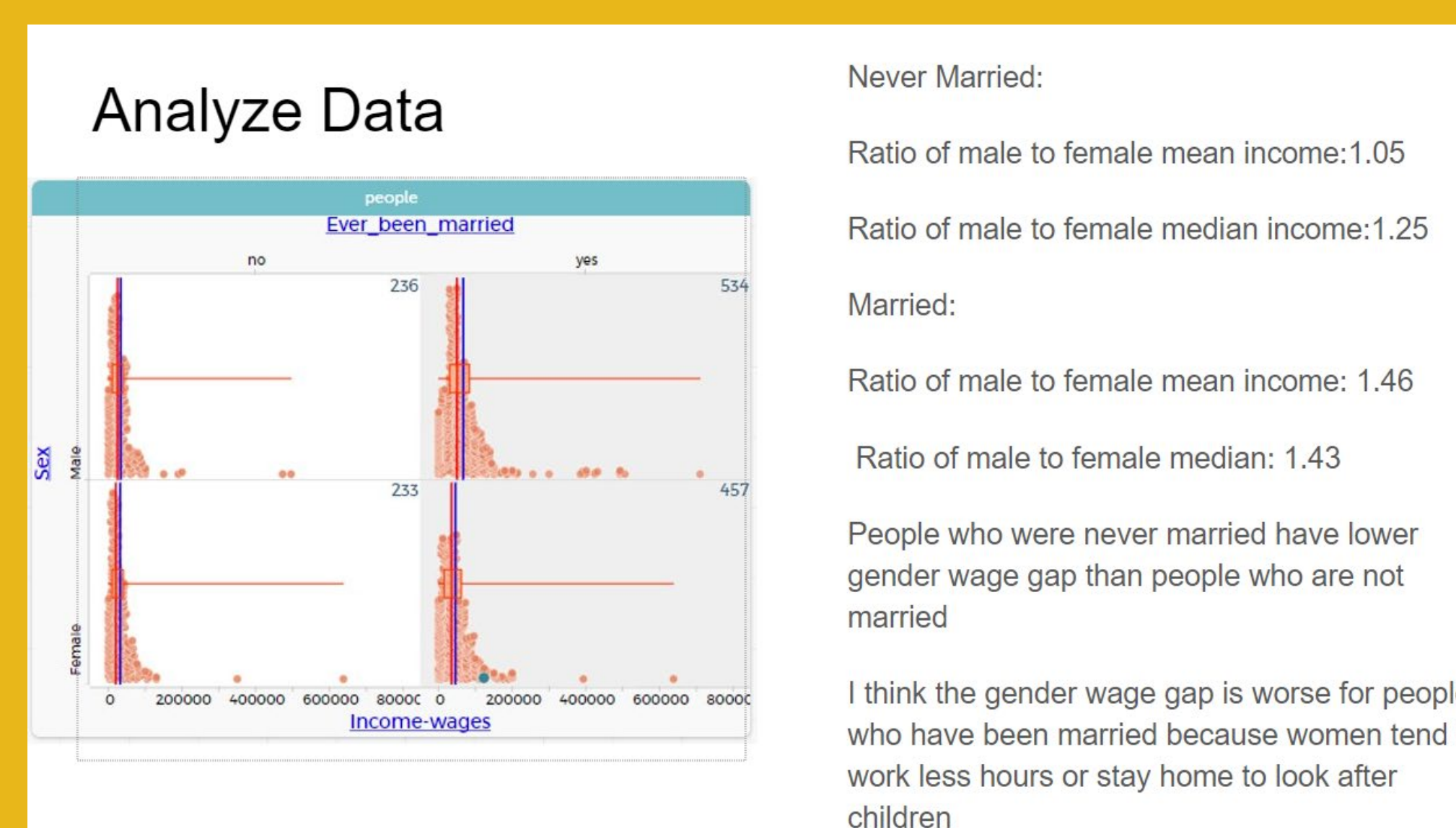


Figure 6. Student example: Marital status



Research Question: What strengths and challenges do students display in multivariable thinking?

Project Participants

- From 2018/19 - 2020/21: 12 high school teachers of non-AP statistics; 5 high school social studies teachers; over 500 students (gr. 12)
- Schools have high % of students who are Black, Latinx, low-income, and ELs in Boston, MA region

Data Collection and Analysis

- In fall 2019/winter 2020 (pre-Covid), Lesson 7 multivariable exit task responses collected from gr. 12 non-AP mathematics students (n= 121)
- Multi-dimensional scoring rubric developed by team, informed by Watson & Callingham (2003); consensus scores reached by two coders

Results

Students responded to the following prompt: *Do you agree or disagree with the claim that differing levels of education explain why males tend to make more money than females? Use data to support your answer.*

Scoring dimension I: Number of relevant variables cited

Score	Example response	Freq. (n=121)
Three	• "Disagree, women still tend to make less money even with the same education level"	97 (80%)
Two	• "I agree because when you compare the medians together the males always have a higher median which means that they will make more money."	22 (18%)
One/none	• "Depends on the jobs males have"	2 (2%)

Scoring dimension II: Discussion of variable relationships

Score	Three relevant variables cited (n=97)		Two relevant variables cited (n=22)	
	Example response	Freq.	Example response	Freq.
A. Condition correctly on third variable	1	"We are inclined to disagree with this statement, because males are paid more on every level of education"	66 (68%)	NA
	0	"I disagree because males make money in other ways besides education..."	20 (21%)	NA
B. Refer to more than one level of third variable	1	"I disagree... The median for males with a doctoral is \$110,000 and females are \$75,000. Males with some school have a median of \$22,000 while females have \$14,700."	22 (23%)	NA
	0	"No because they have the same level and still makes less."	50 (52%)	NA
C. Compare relevant outcomes	1	"No, males have always dominated females, i.e. even with a masters degree, males get paid an average 10,000-15,000 more"	83 (86%)	"Agree, with higher education, people tend to get higher positions/ higher paying jobs." 15 (68%)
	0	"I think males and females should make an equal amount of money if they have the same education level."	10 (10%)	"I'm agree, because the males work more than females" 6 (27%)
D. Cite accurate quantities	1	"No... Females with a doctoral Degree average \$75,000 whereas males make \$110,000."	34 (35%)	"Yes, there is a big gap in the income wage for every dollar that a male makes a female makes about \$0.62" 3 (14%)
	0	"I disagree... Females on average are more educated than males, but no matter the education level males always make more money than women do."	56 (58%)	"I disagree" 19 (86%)

Note: Partial credit responses not shown.

Scoring dimension III: Conclusion is accurate and consistent with the data

Score	Number of relevant variables cited			
	Three (n=97)		Two (n=22)	
	Example response	Freq.	Example response	Freq.
1	"No because all the levels of education the men made more than the woman."	58 (60%)	"No because, females tend to have higher education than males"	3 (14%)
0.5	"well no, i disagree, because level of education affects both male and female for making more than other levels of education, there are other factors that affect why men make more than women in the same class."	15 (16%)	"No, the data didn't explain why males make more money than females"	2 (9%)
0	"Agree, the graph shows that more males attend a higher level of schools than females, so they are given more income as a result."	24 (25%)	"It depends on the job the males have so they make more money."	17 (77%)
			"I disagree"	1 (50%)

Discussion

- About 4/5 responses cited the three relevant variables in the investigation. Of these responses, 68% conditioned correctly on education, 60% provided an accurate and consistent conclusion, and only 35% cited quantities.
- Just under 1/5 responses cited two relevant variables. Of these responses, only 14% provided an accurate and consistent conclusion, and only 14% cited quantities.
- Most students could engage in multivariable thinking, but more supports are needed to help students organize their responses and to discuss conclusions quantitatively.
- Examining a real-life question highly relevant to students' lives may have both supported multivariable thinking and distracted from addressing the target question.



This project is funded by the National Science Foundation, grant #1813956. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.