Learning goal: Use a two-way table to analyze the association between two categorical variables.

Introduction:

In the previous unit, we studied the association between two quantitative variables. Now we will examine the association between two categorical variables.

1) What is the difference between categorical and quantitative data? Give an example of each.

2) Which of the following survey questions yields categorical data?

   a) *Are you male or female?*
   b) *What is your height in inches?*
   c) *What is your college GPA?*
   d) *Where do you tend to sit in class: front, middle or back?*
   e) *Do you feel that you are overweight, underweight, about right in weight?*

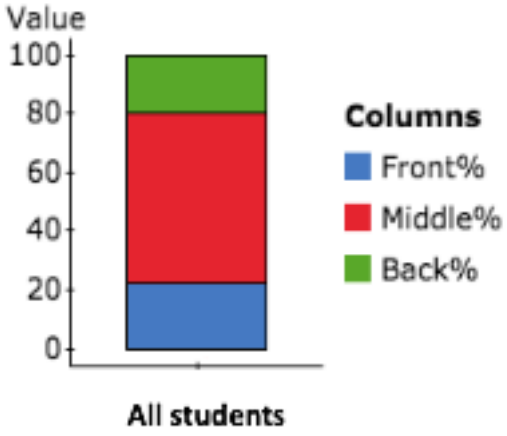
3) Which of the following research questions involve an association between two categorical variables?

   a) *Are males or females more likely to feel that they are overweight?*
   b) *Is there a relationship between college GPA and where a student tends to sit in class?*
   c) *Is there a relationship between high school GPA and college GPA?*
   d) *Are students who sit at the front of class more likely to be shorter than students who sit at the back of class?*

4) Is there an association between gender and where college students sit in class? Let's see what this data suggests.

	Front	Middle	Back	Total
Female	37	91	22	150
Male	15	46	25	86
Total	52	137	47	236

a) Describe the distribution of class seating ignoring gender by calculating appropriate percentages and labeling the stacked barchart.



Of all the students in the study, what percentage sit in the front class?

In the middle of class?

In the back of class?

b) If gender is associated with class seating, we expect the seating distribution to be (circle one: the same, different) when we take gender into account.

c) Use percentages to describe the distribution of class seating for female students. These are called *conditional* percentages.

Of all the female students in the study, what percentage sit in the front class?

In the middle of class?

In the back of class?

d) Use *conditional* percentages to described the distribution of class seating for male students:

Of all the male students in the study, what percentage sit in the front class?

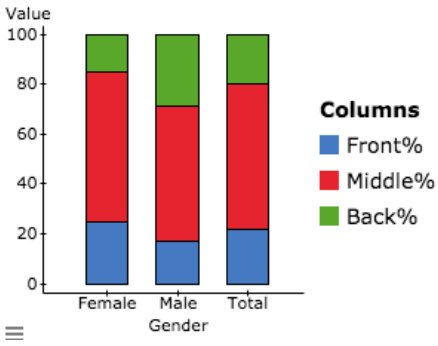
In the middle of class?

In the back of class?

e) Who is more likely to sit at the front of class, females or males?

Who is more likely to sit at the back of class, females or males?

f) Here are stacked bar graphs for all three class seating distributions: female, male, everyone.



This data suggests that gender is (circle one: not associated, weakly associated, strongly associated) with class seating for college students.

(Note: This is a judgement call. Later in the course will develop more concrete ways to determine if observed differences are significant.)

g) Why do we use percentages instead of counts to describe and compare the seating distributions?

5) Is there a relationship between a college student's major and whether they use the college's math tutoring services? Use this fictitious data to calculate relevant percentages to support your answer.

	Math tutoring NO	Math Tutoring YES	Total
STEM major	36	53	89
Non-STEM major	75	174	249
Total	230	108	338

a) What percentage of the students in this study use the college's math tutoring services?

b) Who is more likely to use math tutoring services, STEM or non-STEM majors?

c) This data suggests that major (STEM vs. non-STEM) is (circle one: not associated, weakly associated, strongly associated) with use of math tutoring services at this college.

Briefly support your choice.