

## **Section 3: Representing Categorical Data**

The following maps the videos in this section to the Texas Essential Knowledge and Skills for Mathematics TAC §111.47(c).

### **3.01 Frequency Tables**

- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(F)

### **3.02 Joint, Marginal, and Conditional Frequencies**

- Statistics (1)(C)
- Statistics (1)(F)
- Statistics (4)(F)
- Statistics (5)(A)

### **3.03 Pie Charts**

- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)

### **3.04 Bar Charts**

- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)

### **3.05 Comparing Distributions**

- Statistics (1)(A)
- Statistics (1)(B)
- Statistics (1)(D)
- Statistics (1)(E)
- Statistics (4)(B)
- Statistics (4)(D)

Note: Unless stated otherwise, any sample data is fictitious and used solely for the purpose of instruction.

### 3.01

## Frequency Tables

What is the difference between categorical and quantitative data?

Why are there different ways to summarize and display both types of data?

Once categorical or quantitative data has been collected, it must be summarized. The \_\_\_\_\_ of a data set shows the values of the variable and how often they occur.

A **frequency table** is constructed by arranging a set of collected data values with their corresponding frequencies.

1. Suppose 50 local high school students were randomly selected and each student was asked to rate his or her study habits. Shown below are the frequency table and **relative frequency table** summarizing the results. Interpret the tables.

| <u>Frequency Table</u> |           | <u>Relative Frequency Table</u> |                    |
|------------------------|-----------|---------------------------------|--------------------|
| Rating                 | Frequency | Rating                          | Relative Frequency |
| Excellent              | 7         | Excellent                       | $7/50 = 0.14$      |
| Very Good              | 14        | Very Good                       | $14/50 = 0.28$     |
| Good                   | 16        | Good                            | $16/50 = 0.32$     |
| Poor                   | 9         | Poor                            | $9/50 = 0.18$      |
| Very Poor              | 4         | Very Poor                       | $4/50 = 0.08$      |
| Total                  | 50        | Total                           | $50/50 = 1$        |

2. Suppose 15 random adults in a local community were randomly selected, and each person was asked to rate how concerned he or she was about a new restaurant that's being built. The responses are displayed below. Complete the frequency table and relative frequency table summarizing the results.

|   |                      |    |                      |    |                      |
|---|----------------------|----|----------------------|----|----------------------|
| 1 | Not at All Concerned | 6  | Moderately Concerned | 11 | Extremely Concerned  |
| 2 | Extremely Concerned  | 7  | Extremely Concerned  | 12 | Moderately Concerned |
| 3 | Somewhat Concerned   | 8  | Extremely Concerned  | 13 | Somewhat Concerned   |
| 4 | Somewhat Concerned   | 9  | Moderately Concerned | 14 | Not at All Concerned |
| 5 | Slightly Concerned   | 10 | Extremely Concerned  | 15 | Moderately Concerned |

**Frequency Table**

| Concerned  | Frequency |
|------------|-----------|
| Extremely  |           |
| Moderately |           |
| Somewhat   |           |
| Slightly   |           |
| Not at All |           |
| Total      |           |

**Relative Frequency Table**

| Concerned  | Relative Frequency |
|------------|--------------------|
| Extremely  |                    |
| Moderately |                    |
| Somewhat   |                    |
| Slightly   |                    |
| Not at All |                    |
| Total      |                    |

3. Complete the following frequency table and relative frequency table summarizing the results from a hypothetical study that evaluated the priority local voters placed on education in comparison to other policies in the upcoming local elections.

| <u>Frequency Table</u> |           | <u>Relative Frequency Table</u> |                    |
|------------------------|-----------|---------------------------------|--------------------|
| Priority               | Frequency | Priority                        | Relative Frequency |
| Essential              | 314       | Essential                       |                    |
| High Priority          |           | High Priority                   | 0.25               |
| Medium Priority        |           | Medium Priority                 | 0.28               |
| Low Priority           | 86        | Low Priority                    |                    |
| Not a Priority         |           | Not a Priority                  | 0.07               |
| Total                  | 1000      | Total                           |                    |

### 3.02

## Joint, Marginal, and Conditional Frequencies

Suppose Joanna surveys 20 of her friends to determine the most popular geometry teacher in the school. She asks her friends, 12 boys and eight girls, which geometry teacher they would prefer: Mr. Lewinsky, Ms. Ramirez, Ms. Gundy, or Mr. Garnet. Joanna uses a two-way table to analyze her data.

A **two-way table** is used to examine relationships among categorical data. The entries in the cells of a two-way table can be \_\_\_\_\_ or \_\_\_\_\_. This type of table is also known as a \_\_\_\_\_ table.

Joanna's two-way table is shown below:

|       | Mr.<br>Lewinsky | Ms.<br>Ramirez | Ms.<br>Gundy | Mr.<br>Garnet | Total |
|-------|-----------------|----------------|--------------|---------------|-------|
| Boys  | 4               | 2              | 3            | 3             | 12    |
| Girls | 1               | 1              | 2            | 4             | 8     |
| Total | 5               | 3              | 5            | 7             | 20    |

Entries in the "Total" column and "Total" row are called \_\_\_\_\_ **frequencies**.

Entries in the body of the table are called \_\_\_\_\_ **frequencies**.

Marginal \_\_\_\_\_ are found by dividing the row and/or column totals by the total sample size.

1. Complete the following tables:

| <u>Marginal Distribution</u> |             |
|------------------------------|-------------|
| Gender                       | Probability |
| Boys                         | $12/20 =$   |
| Girls                        | $8/20 =$    |
| Total                        | $20/20 = 1$ |

| <u>Marginal Distribution</u> |             |
|------------------------------|-------------|
| Teacher                      | Probability |
| Mr. Lewinsky                 | $5/20 =$    |
| Ms. Ramirez                  | $3/20 =$    |
| Ms. Gundy                    | $5/20 =$    |
| Mr. Garnet                   | $7/20 =$    |
| Total                        | $20/20 = 1$ |

Also, \_\_\_\_\_ frequencies are found by dividing the cell counts by their row and/or column totals.

2. Complete the following tables:

| <u>Distribution for Boys</u> |             | <u>Distribution for Girls</u> |             |
|------------------------------|-------------|-------------------------------|-------------|
| Teacher                      | Probability | Teacher                       | Probability |
| Mr. Lewinsky                 | $4/12 =$    | Mr. Lewinsky                  | $1/8 =$     |
| Ms. Ramirez                  | $2/12 =$    | Ms. Ramirez                   | $1/8 =$     |
| Ms. Gundy                    | $3/12 =$    | Ms. Gundy                     | $2/8 =$     |
| Mr. Garnet                   | $3/12 =$    | Mr. Garnet                    | $4/8 =$     |
| Total                        | $12/12 = 1$ | Total                         | $8/8 = 1$   |

3. Shown below is a contingency table classifying responses for the variables *Social Life Rating* and *Gender*.

|        | Excellent | Very Good | Good | Poor | Very Poor | Total |
|--------|-----------|-----------|------|------|-----------|-------|
| Male   | 2         | 5         | 10   | 6    | 3         | 26    |
| Female | 5         | 9         | 6    | 3    | 1         | 24    |
| Total  | 7         | 14        | 16   | 9    | 4         | 50    |

- i. Determine the marginal relative probabilities for Gender and Social Life Rating by completing the tables below.

| <u>Marginal Relative Probabilities</u> |             | <u>Marginal Relative Probabilities</u> |             |
|----------------------------------------|-------------|----------------------------------------|-------------|
| Gender                                 | Probability | Social Life                            | Probability |
| Male                                   |             | Excellent                              |             |
| Female                                 |             | Very Good                              |             |
| Total                                  | 50/50 = 1   | Good                                   |             |
|                                        |             | Poor                                   |             |
|                                        |             | Very Poor                              |             |
|                                        |             | Total                                  | 50/50 = 1   |

- ii. What is the marginal relative frequency of people with a “Good” social life rating?
- iii. What is the joint frequency of males with a “Very Good” social life rating?
- iv. What is the conditional relative frequency of females with an “Excellent” social life rating?

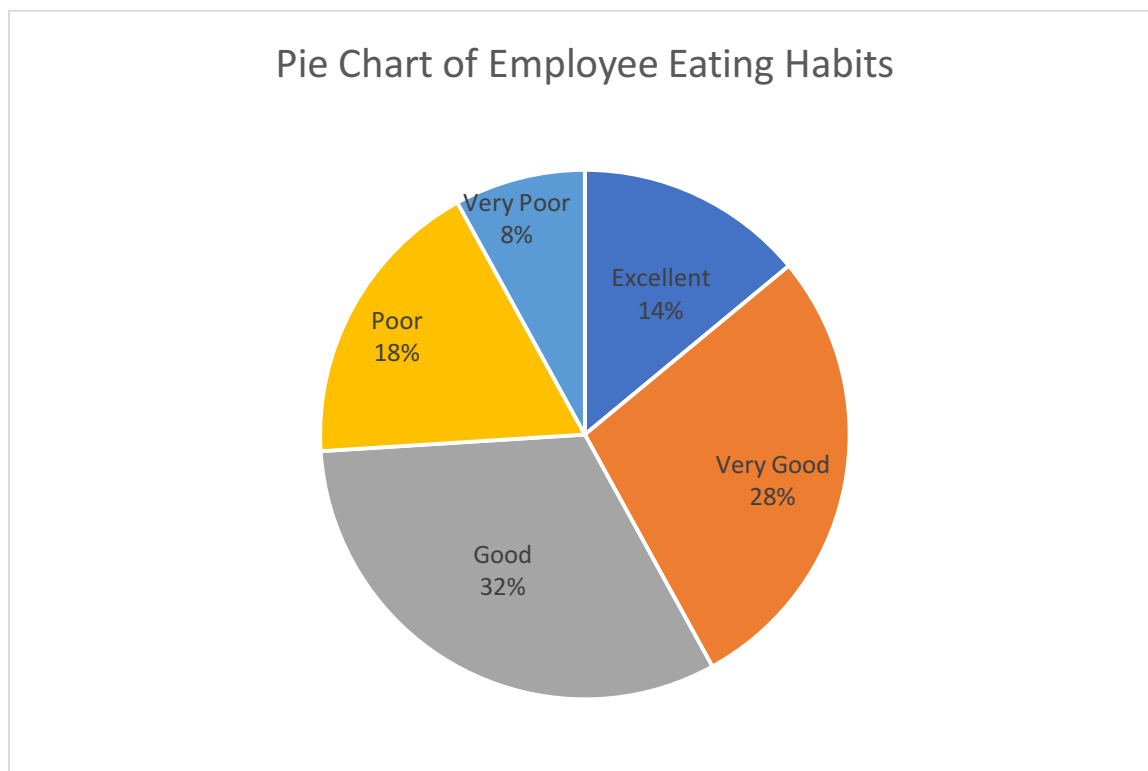
### 3.03

## Pie Charts

A **pie chart** displays data in a \_\_\_\_\_ graph.

- The pieces of the graph are \_\_\_\_\_ to the fraction of the whole in each category.
- The area of each slice is determined by each category's count or percentage.

1. Consider the following pie chart, which represents the eating habits of 50 employees at a local company.



How can we interpret this pie chart?

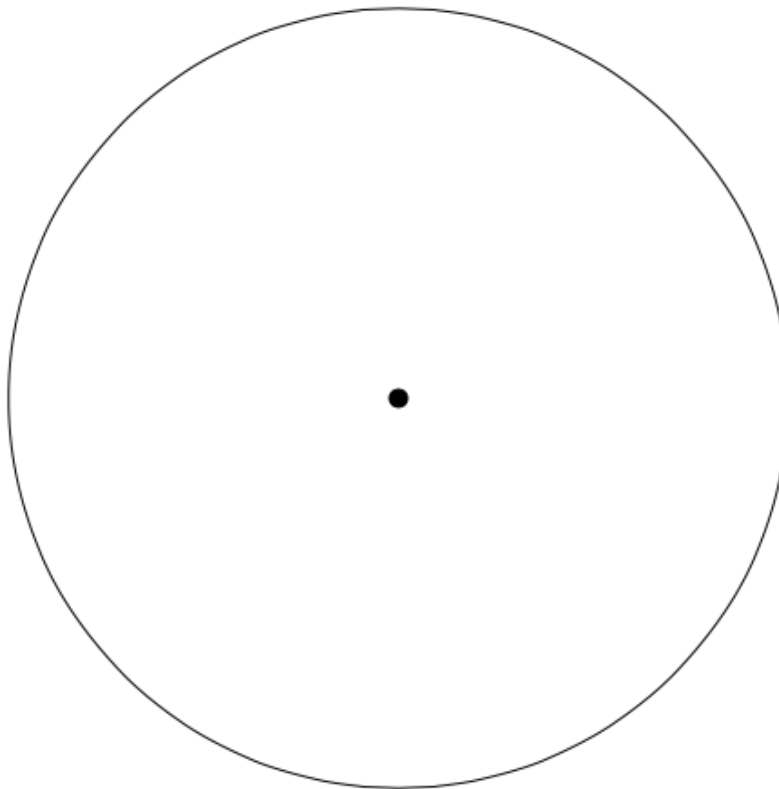
2. Assume that the following data represents the educational attainment of persons in a local county.

i. Complete the relative frequency table.

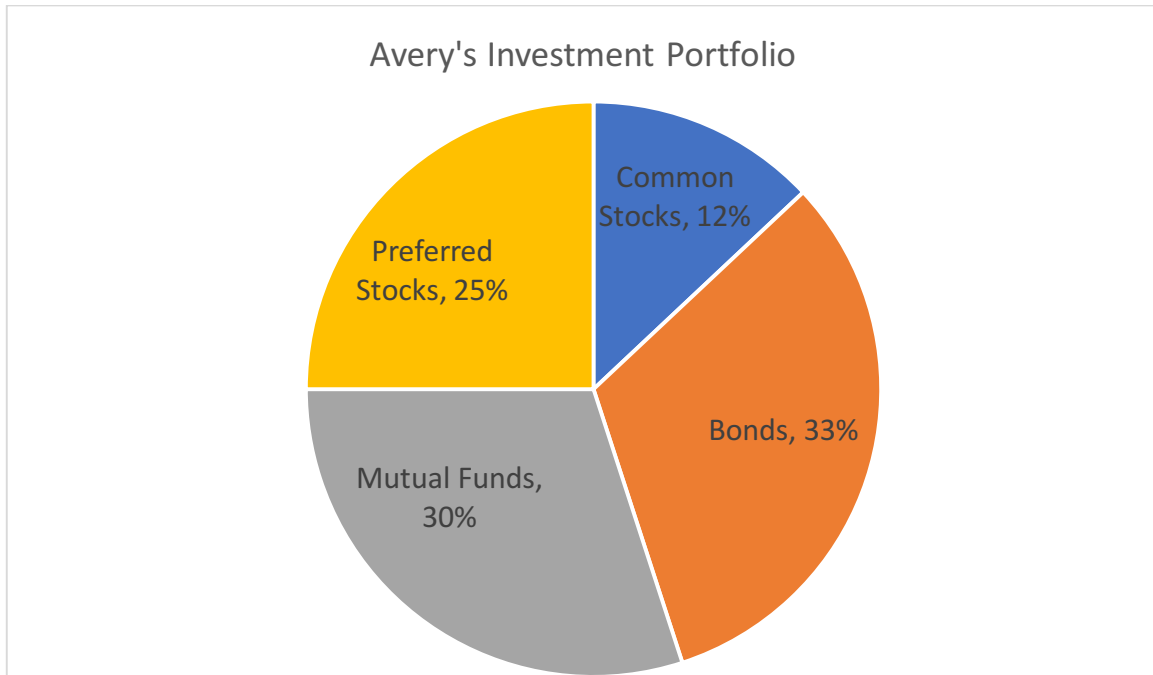
| <u>Frequency Table</u> |           |
|------------------------|-----------|
| Educational Attainment | Frequency |
| No High School         | 15,000    |
| High School            | 25,000    |
| Some College           | 15,000    |
| Associate's            | 25,000    |
| Bachelor's             | 100,000   |
| Advanced               | 20,000    |
| Total                  | 200,000   |

| <u>Relative Frequency Table</u> |                    |
|---------------------------------|--------------------|
| Educational Attainment          | Relative Frequency |
| No High School                  |                    |
| High School                     |                    |
| Some College                    |                    |
| Associate's                     |                    |
| Bachelor's                      |                    |
| Advanced                        |                    |
| Total                           |                    |

ii. Construct a pie chart of the data.



3. Consider the pie chart below:



Suppose this pie chart displays the details of Avery's investment portfolio. The portfolio's value is \$300,000. There are three varieties of mutual funds: equity funds (25%), fixed-income funds (35%), and money market funds (40%).

- i. What percentage of the portfolio consists of equity funds?
- ii. How much money does Avery have invested in bonds?
- iii. How much money does Avery have invested in money market funds?

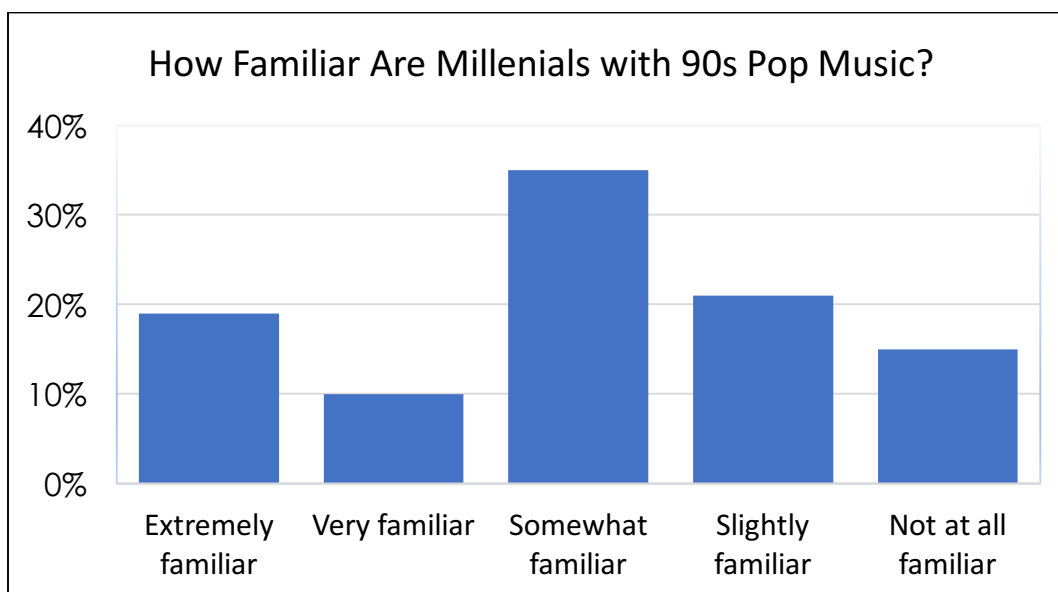
### 3.04

## Bar Charts

A **bar chart** displays the frequency or relative frequency for each category of a categorical variable using bars on a graph.

- The bars are positioned with a label that represents a \_\_\_\_\_.
- The bars can be drawn either vertically or \_\_\_\_\_.
- The height or length of the bar indicates the \_\_\_\_\_ of the group, defined by the column or row label.
- For most categorical variables, the bar order is \_\_\_\_\_.

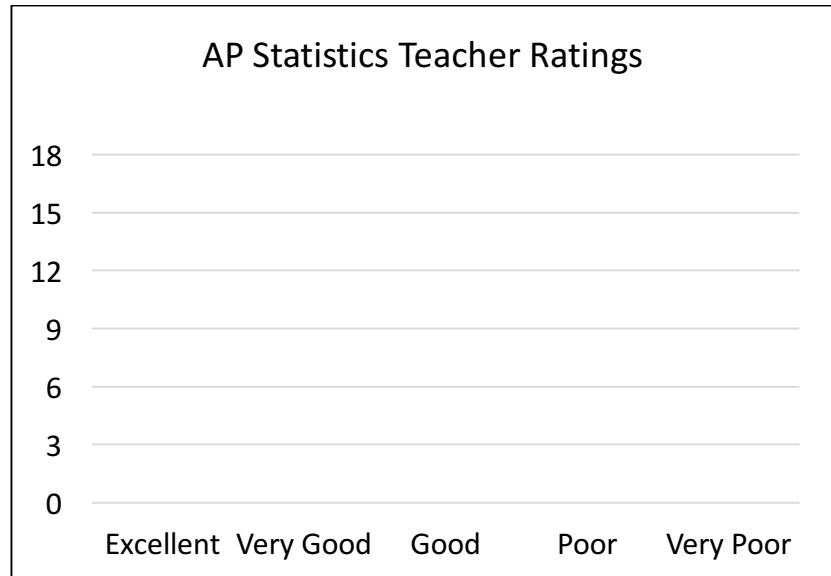
1. Assume 500 teens were interviewed and the results are shown in the bar chart below:



How many teens in the sample were very familiar with 90s popular music?

2. Suppose 50 AP Statistics students were randomly selected, and each student was asked to rate his or her teacher. Shown below is the frequency table summarizing the results. Construct a bar chart of the data.

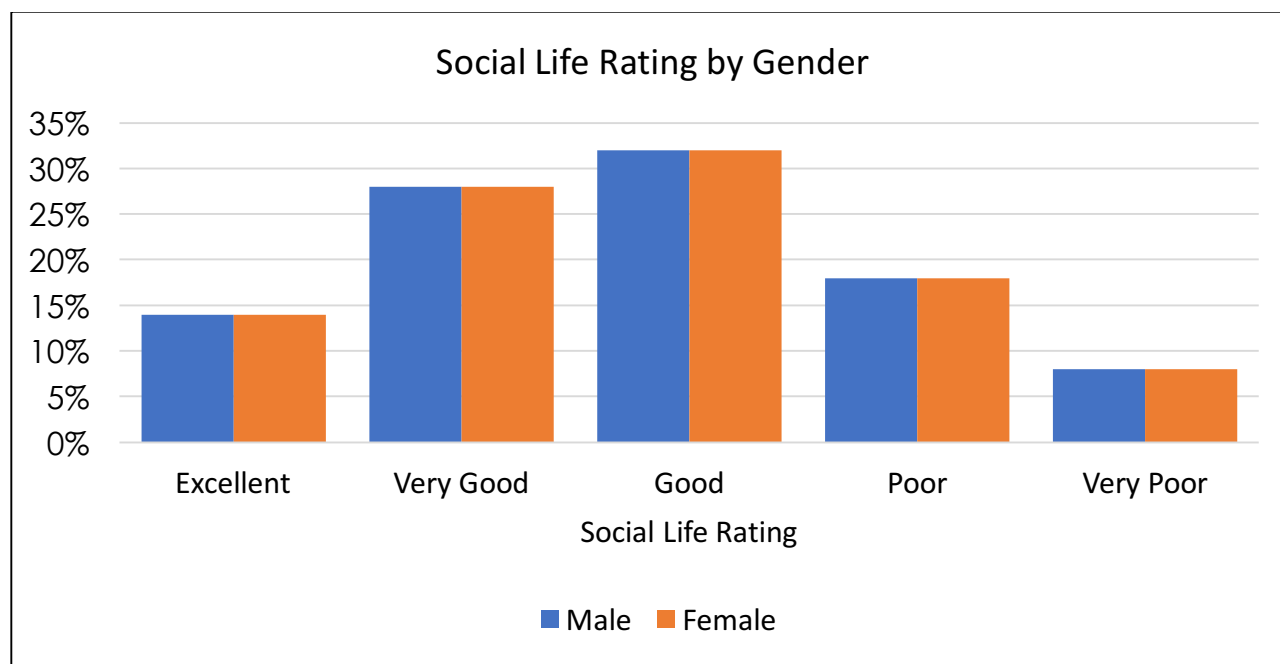
| <b><u>Frequency Table</u></b> |           |
|-------------------------------|-----------|
| Rating                        | Frequency |
| Excellent                     | 15        |
| Very Good                     | 14        |
| Good                          | 10        |
| Poor                          | 6         |
| Very Poor                     | 5         |
| Total                         | 50        |



### 3.05

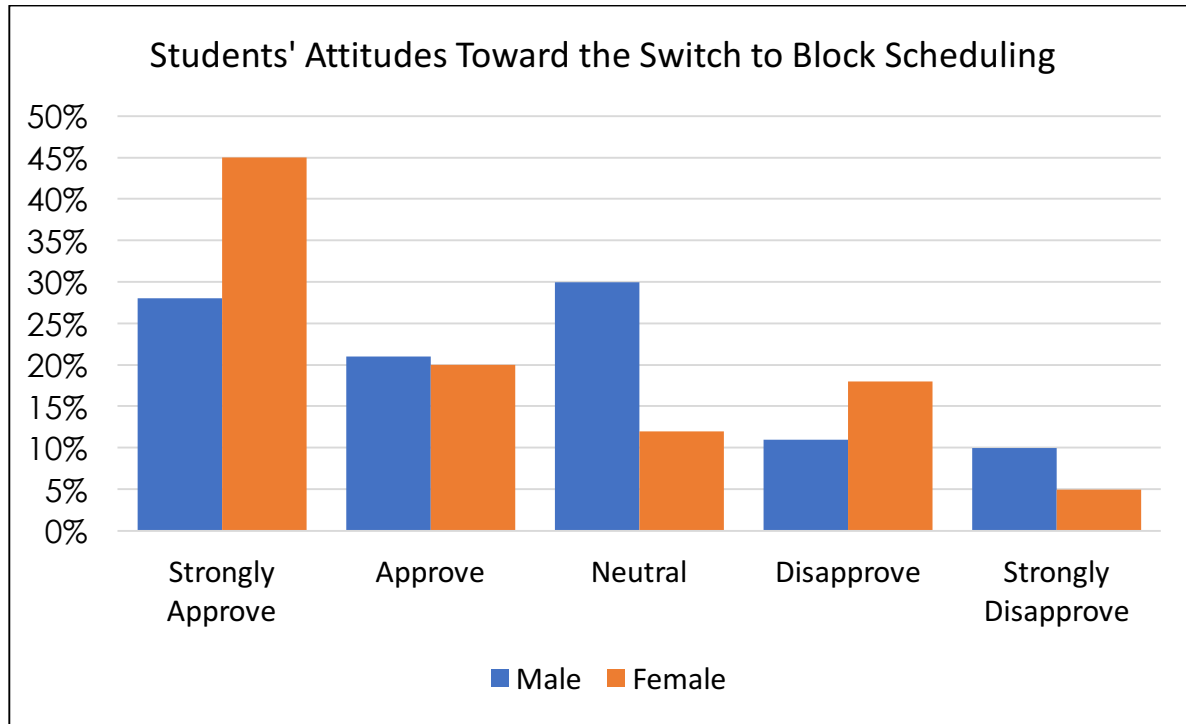
#### Comparing Distributions

1. Suppose 100 people in the 26–35 age group were randomly selected, and each person was asked to rate his or her social life. If Social Life Rating is independent of Gender, a side-by-side bar chart of the conditional distributions would look like the one below.



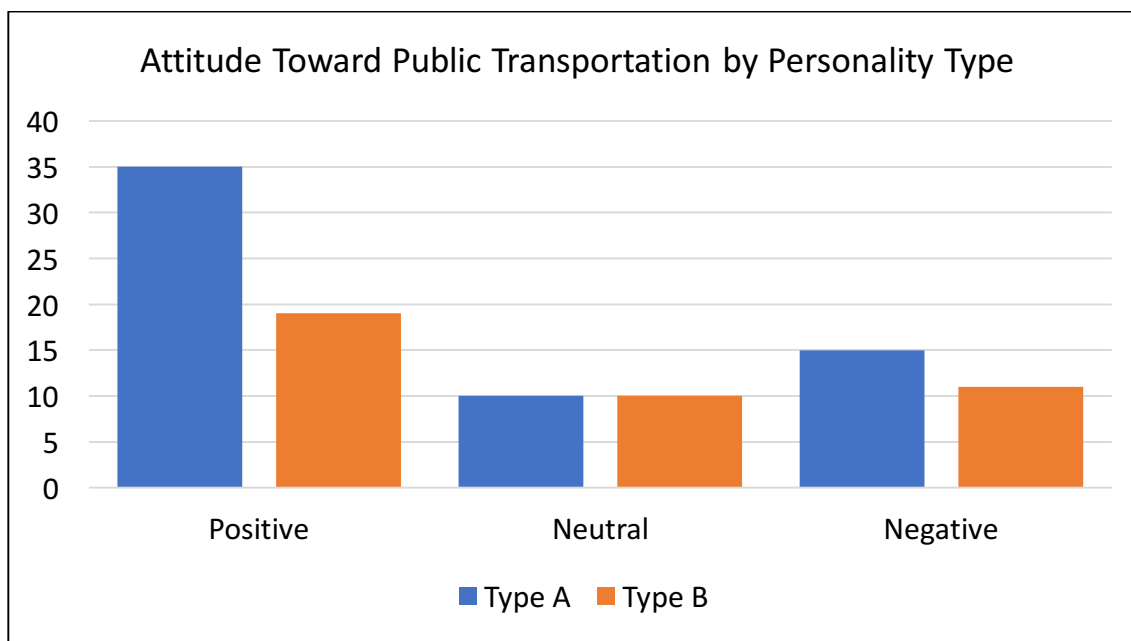
What conclusion can be drawn about the data above?

2. Now, let's consider students' attitudes toward a switch to block scheduling. Suppose 500 students were randomly selected: 50% males and 50% females. The bar graph displaying the results of the survey is shown below.



What conclusions can be drawn?

3. Suppose 100 people were interviewed and classified according to their attitudes toward public transportation in their city, as well as their personality types. The results are shown in the bar chart below.

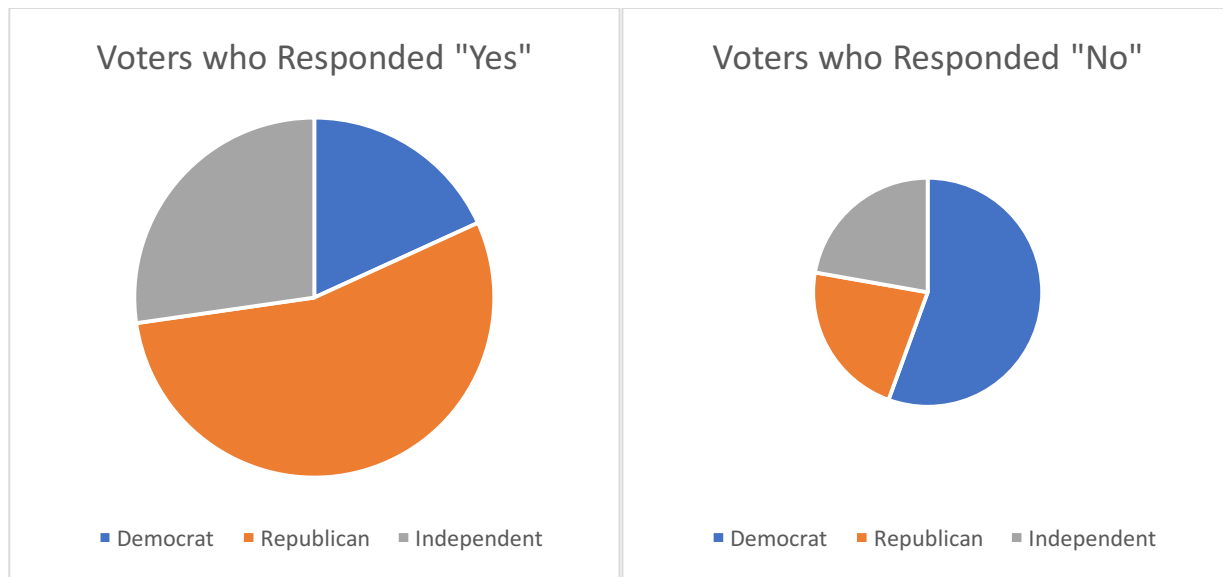


Which of the following statements is correct?

- A. Of the people interviewed, 35% have a positive attitude toward public transportation.
- B. There were 50 people with a neutral attitude toward public transportation.
- C. Forty people with Type B personality were interviewed.
- D. People with Type A personality tend to have a negative attitude toward public transportation.
- E. Of the people interviewed, 50% have Type A personality.

4. Suppose a random sample of 100 voters was taken. Each voter was asked his or her political affiliation and whether he or she favored a certain political issue. Their responses are cross-classified in the contingency table and pie charts below.

| Party Affiliation | Favor Issue |    | Total |
|-------------------|-------------|----|-------|
|                   | Yes         | No |       |
| Democrat          | 10          | 25 | 35    |
| Republican        | 30          | 10 | 40    |
| Independent       | 15          | 10 | 25    |
| Total             | 55          | 45 | 100   |



What conclusions can be drawn?