Section 5.2: Organizing and Graphing Categorical Data

Objective: Create a frequency table.

Data is being collected all the time by businesses, governments, and researchers. The data can range from small to quite large. We need to be able to better understand the nature of the data. Organizing it helps! In this section, we will organize qualitative data. Because qualitative data is classified into categories, we will refer to this data as categorical data.

Definition. *A **frequency table** shows how data are divided among several categories (or classes) by listing the categories along with the number (frequency) of data values in each of them.*

Example. Coffee Shop
A local coffee shop keeps a list of types of drinks that their customers order each hour. Below is the data from 50 drinks sold an hour before closing on a recent Tuesday.

Coffee	Espresso	Coffee	Tea	Espresso
Tea	Coffee	Tea	Coffee	Espresso
Espresso	Tea	Espresso	Espresso	Espresso
Espresso	Espresso	Coffee	Espresso	Tea
Coffee	Soda	Espresso	Coffee	Coffee
Espresso	Tea	Espresso	Soda	Tea
Coffee	Espresso	Coffee	Tea	Espresso
Coffee	Soda	Coffee	Coffee	Espresso
Soda	Espresso	Tea	Espresso	Coffee
Coffee	Espresso	Coffee	Espresso	Tea

A frequency table provides a useful way to organize this data. There are four different categories of drinks they sell (Coffee, Espresso, Soda, and Tea). They are in the first column of the table. The second column contains the counts of each type sold that hour.

Drink	Frequency
Coffee	16
Espresso	20
Soda	4
Tea	10

Objective: Compute relative frequencies.

An additional column can be added to the table to give us a better understanding of the data. This column is called a relative frequency. It can be found by dividing the frequency count for a category by the sum of all frequency counts.

$$\textbf{Relative frequency for a category } = \frac{\textbf{frequency of a category}}{\textbf{sum of all frequencies}}$$

Drink	Frequency	Relative Frequency
Coffee	16	$\frac{16}{50} = 0.32$
Espresso	20	$\frac{20}{50} = 0.40$
Soda	4	$\frac{4}{50} = 0.08$
Tea	10	$\frac{10}{50} = 0.20$

If we turn the relative frequency into a percent we get a better understanding of the sales that hour.

$$\textbf{Percentage for a category } = \frac{\textbf{frequency of a category}}{\textbf{sum of all frequencies}} \times \textbf{100\%}$$

Drink	Frequency	Relative Frequency	Percentage
Coffee	16	$\frac{16}{50} = 0.32$	$0.32 \times 100\% = 32\%$
Espresso	20	$\frac{20}{50} = 0.40$	$0.40 \times 100\% = 40\%$
Soda	4	$\frac{4}{50} = 0.08$	$0.08 \times 100\% = 8\%$
Tea	10	$\frac{10}{50} = 0.20$	$0.20 \times 100\% = 20\%$

Initially it was easy to see that Espresso is the most popular drink, but looking at the relative frequency or percentage gives us a better idea of how it relates to the other drink choices. The manager can use information such as this to better stock the shop.

Objective: Create a bar graph.

Nothing makes a report look better than a nice graph. In addition to creating frequency tables, an analyst might want to create a graph of categorical data. There are many different types of graphs. **Bar graphs** are probably the most commonly used graphs and they are used to compare things between different groups.

Here is a bar chart of our coffee shop data. The categories are along the horizontal axis and the frequency counts correspond to the height of the bars.
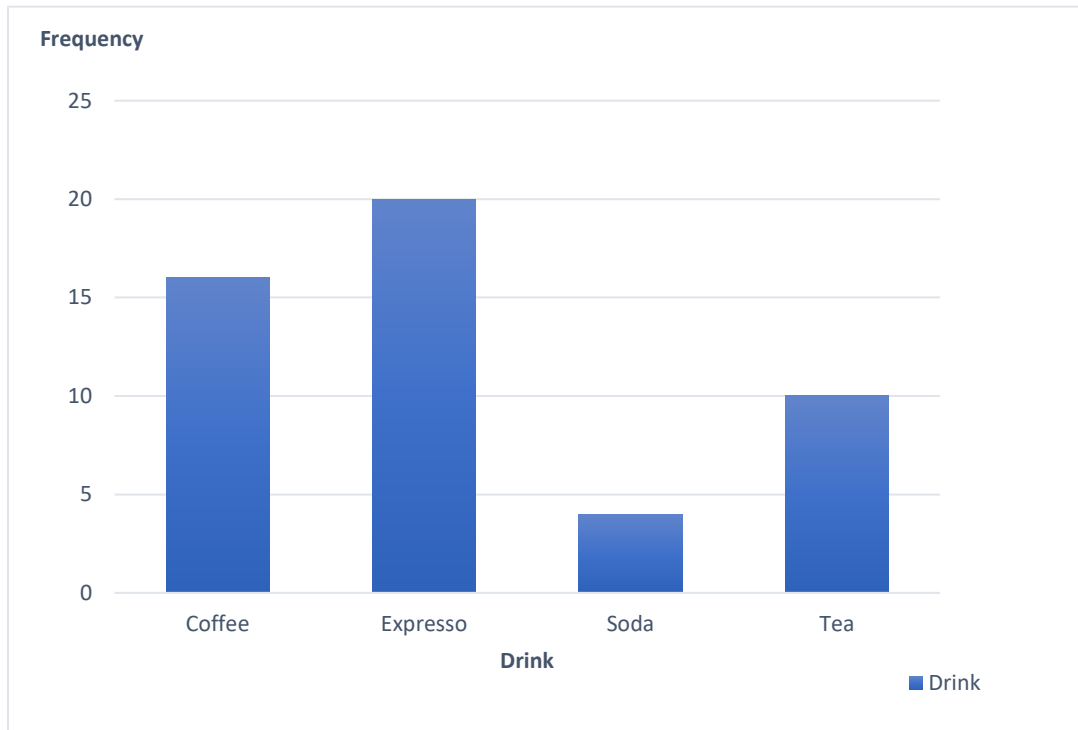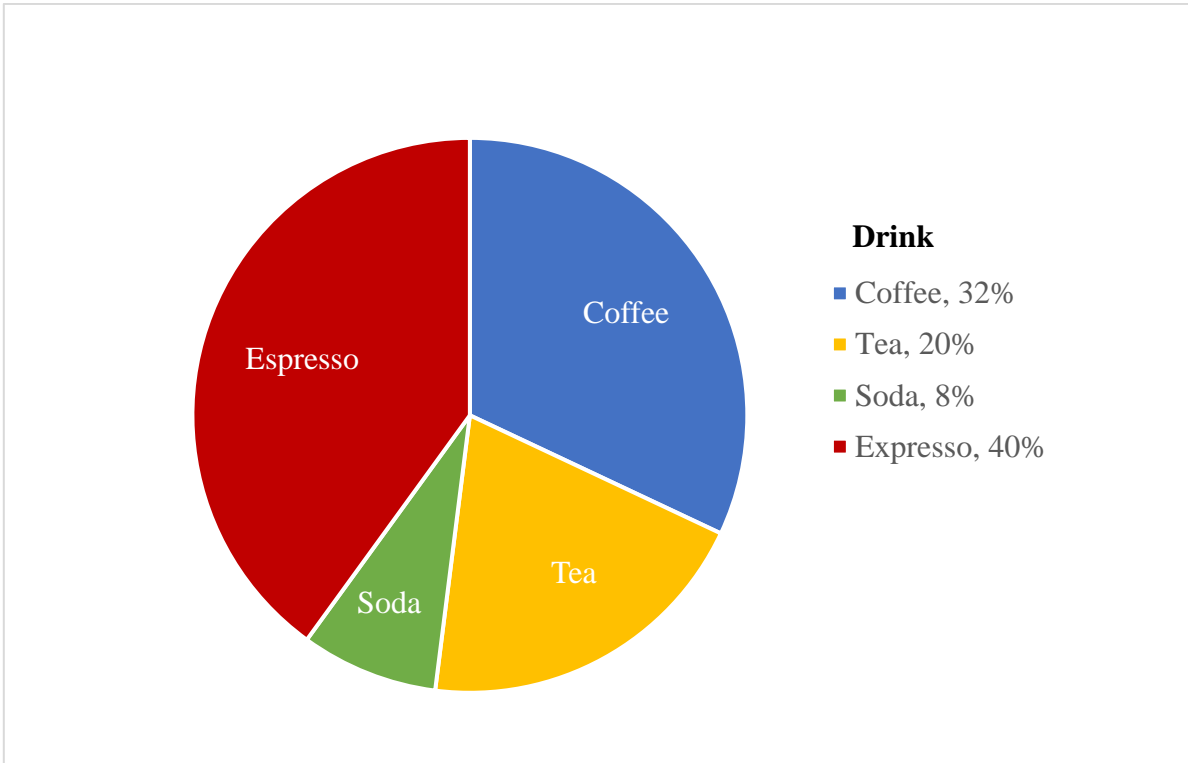

Rules when constructing a bar graph

1. The height of each bar represents the frequency or relative frequency for that category.
2. The bars should be of the same "width."
3. The bars should not overlap.
4. Each piece of data should belong to only one category.

Objective: Create a pie chart.

Like bar graphs, pie charts are very common to graph categorical data. **Pie charts** show how the size of the category relates to the whole group. Pie charts are great for showing percentages. Below is the coffee shop pie chart. Notice how the percentages correspond to the size of the pie pieces.

Drink
- Coffee, 32%
- Tea, 20%
- Soda, 8%
- Expresso, 40%

Rules when constructing a pie chart

1. Always include the relative frequency or percentage.
2. Include labels, either as a legend or directly on pie.

5.2 Practice

1. Twenty-four students answered a survey about pet preferences. Their responses are below.

Cat	Guinea pigs	Guinea pigs	Cat	Rabbit	Dog	Guinea pigs	Dog
Dog	Cat	Dog	Dog	Guinea pigs	Dog	Cat	Rabbit
Dog	Rabbit	Cat	Guinea pigs	Dog	Cat	Rabbit	Dog

   a) Construct a frequency table for this data.
   b) Draw a bar graph.
   c) How many students participated in this survey?
   d) What percent of students like dogs?

2. In a software engineering class, the professor asked his students to name their favorite programming language. Their replies are listed in the table below.

Java	Lisp	Perl	Java	Perl	Perl	Perl	C++	Perl	Java
Perl	Java	Java	Perl	Java	Lisp	Java	Perl	Java	Lisp
Perl	C++	C++	Perl	C++	Perl	Java	C++	Perl	C++

a) Construct a frequency table for this data.
b) Draw a bar graph.
c) How many students participated in this survey?
d) What percent of students like C++?

3. The following frequency table represents the number of new HIV/AIDS cases in the US in 2008 according to race/ethnicity. What percent of the new cases were Hispanic/Latino?

Race/Ethnicity	Number of HIV/AIDS Cases
American Indian/Alaskan Native	228
Asian	451
Black/African American	21,443
Hispanic/Latino	7,461
Native Hawaiian/other Pacific Islander	47
White	12,534

4. A school district performed a study to find the main causes leading to its students dropping out of school. Fifty cases were analyzed and a primary cause was assigned to each case. The results for the fifty cases are listed below. What percent of students drop out due to family problems?

Causes to drop out of school	Frequency
Unexcused absences	12
Illness	16
Family problems	14
Other causes	8

5. Relative frequencies allow us to compare groups. Here is the 2008 new HIV/AIDS cases in the US separated by sex.

| Males		Females	
Race/Ethnicity	Frequency	Race/Ethnicity	Frequency
Black	14,247	Black	7,196
White	10,563	White	1,971
Hispanic	5,906	Hispanic	1,555
Other	565	Other	161

   a) Compute the relative frequencies for each sex.
   b) Write a few sentences explaining the trend of new cases in 2008 using what you learned in part a.

6. The table below represents 360 books grouped by their category:

| Book Category	Frequency
Science Fiction	150
Romance	125
Adventure	50
Horror	35

Which pie chart best represents this table?



(a)   (b)   (c)
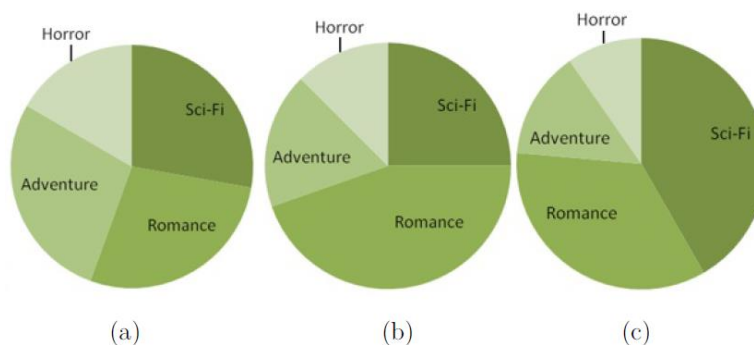
7. The bar chart below describes the day of the week workers called in sick for workers at a company. What is the relative frequency for Monday?
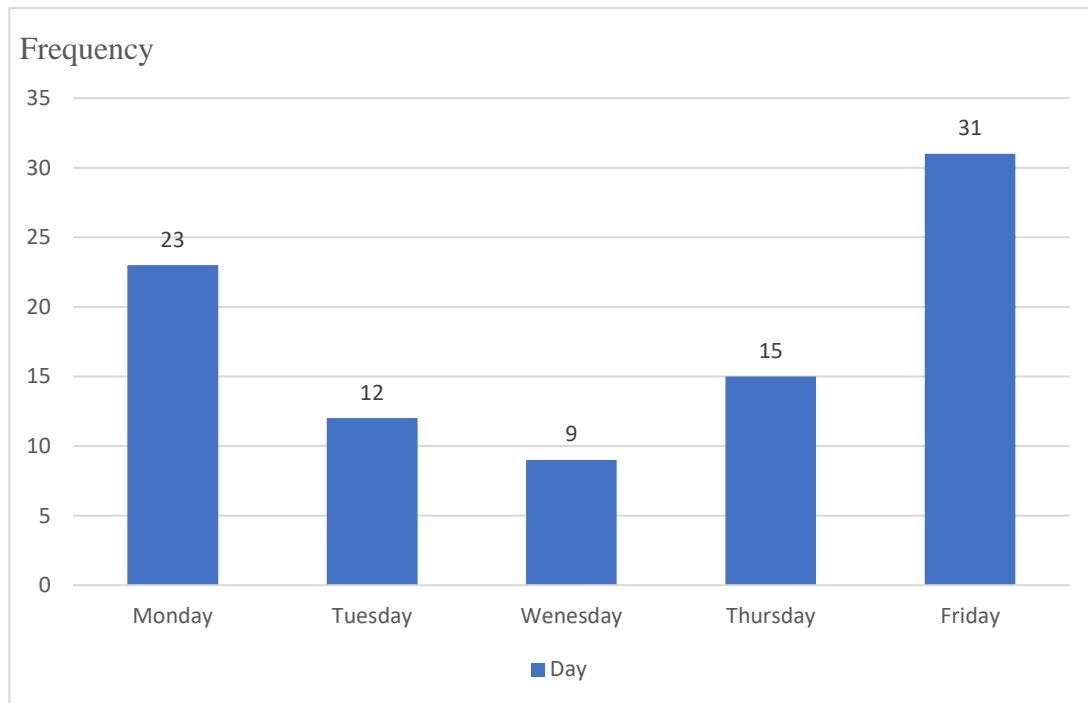
8. The bar chart below show the blood groups of O, A, B, and AB of a group of forty randomly selected blood donors. How many donors have a blood group of O?
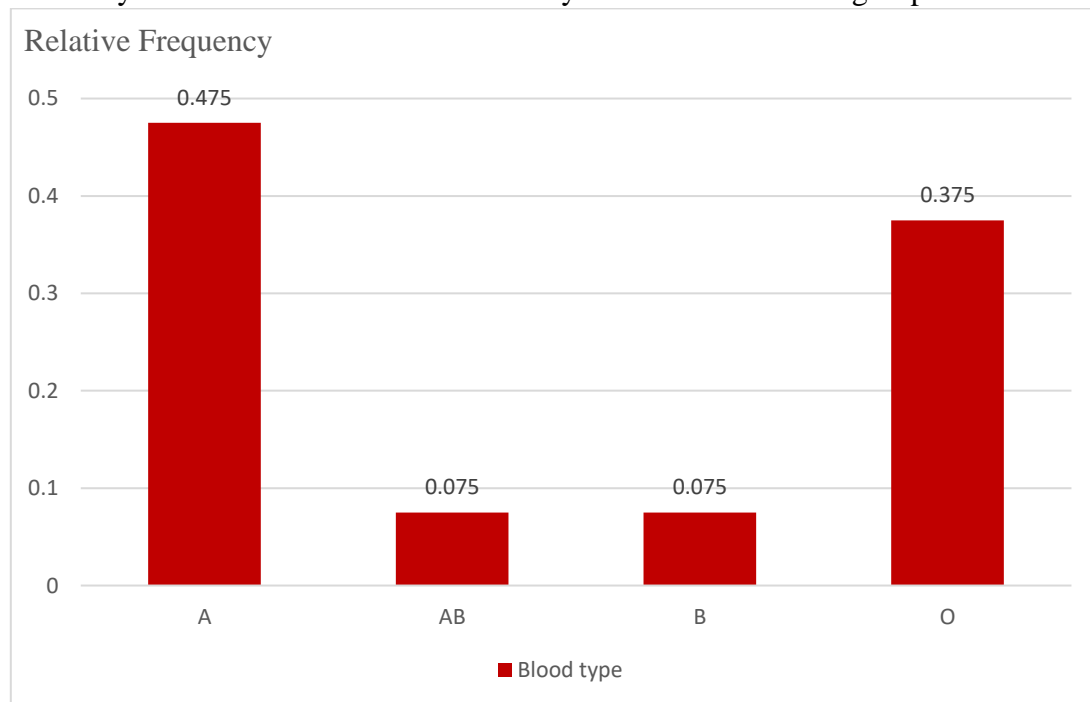
9. The pie chart below shows the student responses to a survey asking them about their favorite holiday. Use the graph to find the percent of students who answered "4th of July".
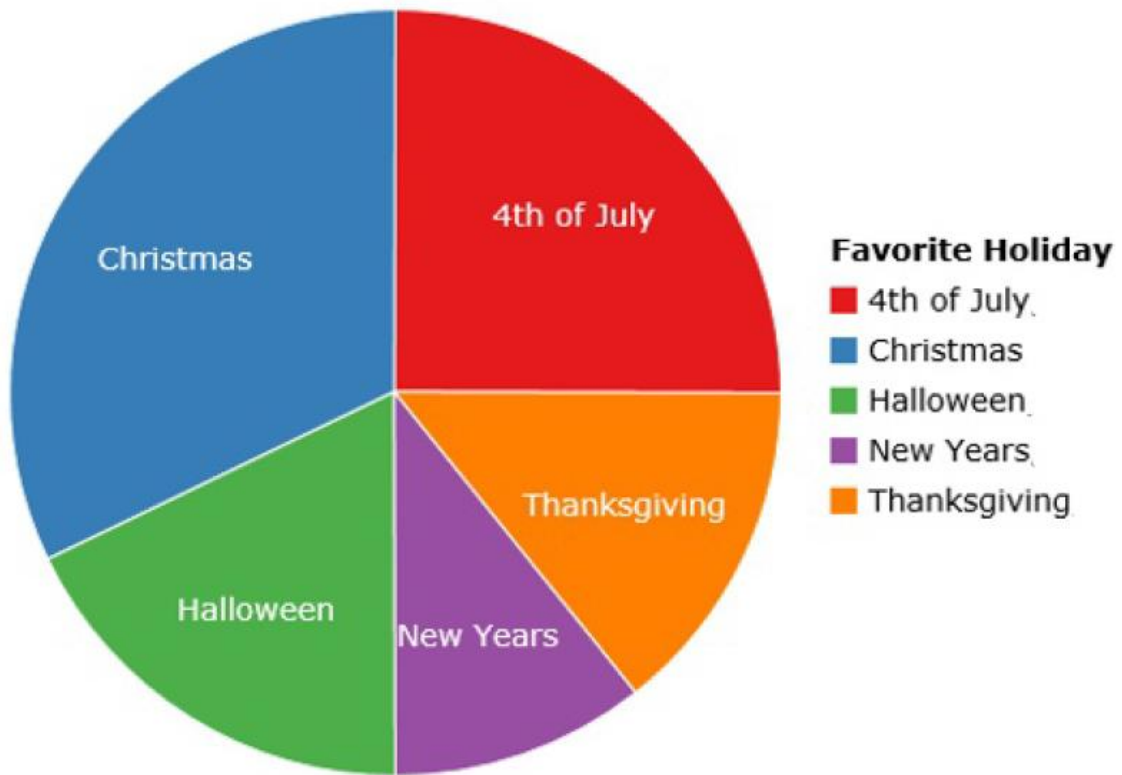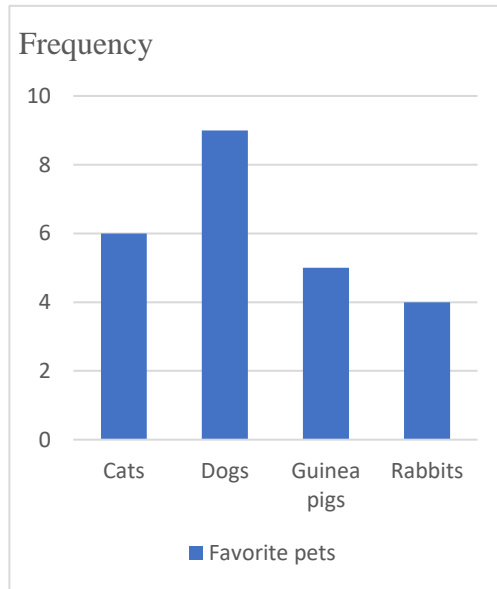
5.2 Answers

1. a)

Favorite pets	Frequency
Rabbits	4
Cats	6
Dogs	9
Guinea pigs	5
Total	24

b)

Frequency

Favorite pets

c) 24 students participated in the survey

d) 37.5% of students like dogs

2. a)

Programming language	Frequency
C++	6
Java	9
Lisp	3
Perl	12
Total	30

b)
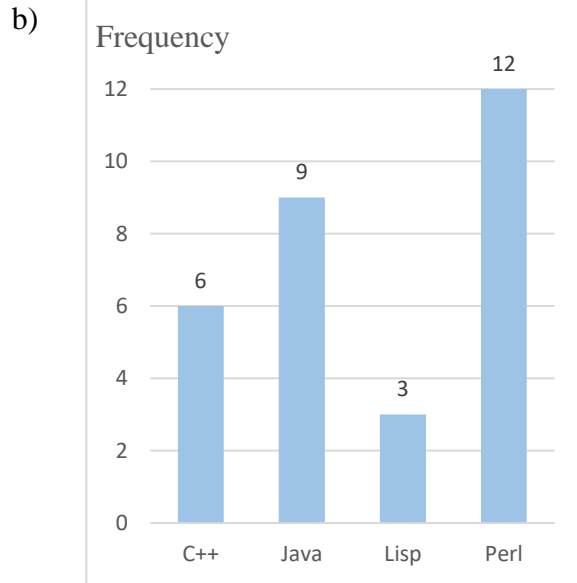
Frequency

c) 30 students participated in the survey

d) 20% of students like C++

3. 17.7%

4. 28%

5. a)

Males	
Race/Ethnicity	**Relative Frequency**
Black	0.455
White	0.338
Hispanic	0.189
Other	0.018

Females	
Race/Ethnicity	**Relative Frequency**
Black	0.661
White	0.181
Hispanic	0.143
Other	0.015

b) When you are comparing new cases of HIV among men and women their relative frequencies for race are nothing alike. With both males and females, the majority of the cases are with Blacks and Whites. Females have two thirds of new cases just among Blacks. The epidemic of HIV is very different according to race for men and women.

6. Pie chart (c)

7. 0.2556

8. 15

9. 25%