

Statistics



4.1

Population, Sample, and Data

Objectives

- Construct a frequency distribution
- Construct a histogram
- Construct a pie chart

Population, Sample, and Data

The field of **statistics** can be defined as the science of collecting, organizing, and summarizing data in such a way that valid conclusions and meaningful predictions can be drawn from them.

The first part of this definition, “collecting, organizing, and summarizing data,” applies to **descriptive statistics**.

The second part, “drawing valid conclusions and making meaningful predictions,” describes **inferential statistics**.



Population versus Sample

Population versus Sample

Who will become the next president of the United States?

Population versus Sample

Who will become the next president of the United States?

During election years, political analysts spend a lot of time and money trying to determine what percent of the vote each candidate will receive.

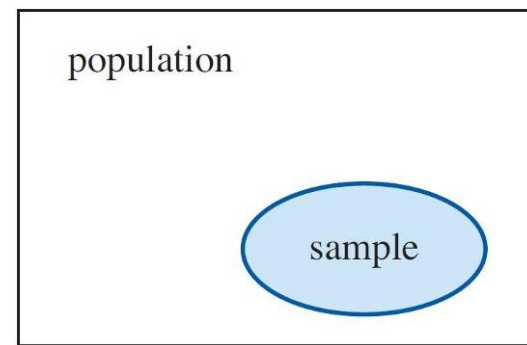
However, because there are over 175 million registered voters in the United States, it would be virtually impossible to contact each and every one of them and ask, “Whom do you plan on voting for?”

Population versus Sample

Analysts select a smaller group of people, determine their intended voting patterns, and project their results onto the entire body of all voters.

Because of time and money constraints, it is very common for researchers to study the characteristics of a small group in order to estimate the characteristics of a larger group.

In this context, the set of all objects under study is called the **population**, and any subset of the population is called a **sample** (see Figure 4.1).



Population versus sample.

Figure 4.1

Population versus Sample

So we collect data from a smaller, more manageable sample.

Population versus Sample

So we collect data from a smaller, more manageable sample.

Once we have collected these data, we can summarize by calculating various descriptive statistics, such as the *average value*.

Population versus Sample

So we collect data from a smaller, more manageable sample.

Once we have collected these data, we can summarize by calculating various descriptive statistics, such as the *average value*.

Inferential statistics, then, deals with drawing conclusions (hopefully, valid ones!) about the *population*, based on the descriptive statistics of the sample data.

Population versus Sample

Sample data are collected and summarized to help us draw conclusions about the *population*.

Population versus Sample

Sample data are collected and summarized to help us draw conclusions about the *population*.

A good sample is representative of the population from which it was taken.

Population versus Sample

Sample data are collected and summarized to help us draw conclusions about the *population*.

A good sample is representative of the population from which it was taken.

If the sample is not representative, the conclusions concerning the population might not be valid.

Population versus Sample

Sample data are collected and summarized to help us draw conclusions about the *population*.

A good sample is representative of the population from which it was taken.

If the sample is not representative, the conclusions concerning the population might not be valid.

The most difficult aspect of inferential statistics is obtaining a representative sample.

Conclusions are only as reliable as the sampling process and information will usually change from sample to sample. 15



Frequency Distributions

Frequency Distributions

Phase 1 of any statistical study: the collection of data.

Each element in a set of data is referred to as a ***data point***. When data are first collected, the data points might show no apparent patterns or trends.

Phase 2, phase of descriptive statistics: organize the data to summarize it and detect any trends.

The most common way to *organize raw data* is to create a ***frequency distribution***, a table that lists each data point along with the number of times it occurs (its ***frequency***).

Example 1 – *Creating a Frequency Distribution: Single Values*

While bargaining for their new contract, the employees of **2 Dye 4 Clothing** asked their employers to provide daycare service as an employee benefit.

Examining the personnel files of the company's fifty employees, the management recorded the number of children under six years of age that each employee was caring for.

Example 1 – *Creating a Frequency Distribution: Single Values* cont'd

The following results were obtained:

0	2	1	0	3	2	0	1	1	0
0	1	1	2	4	1	0	1	1	0
2	1	0	0	3	0	0	1	2	1
0	0	2	4	1	1	0	1	2	0
1	1	0	3	5	1	2	1	3	2

Example 1 – *Creating a Frequency Distribution: Single Values* cont'd

The following results were obtained:

0	2	1	0	3	2	0	1	1	0
0	1	1	2	4	1	0	1	1	0
2	1	0	0	3	0	0	1	2	1
0	0	2	4	1	1	0	1	2	0
1	1	0	3	5	1	2	1	3	2

Let's organize the data: create a *frequency distribution*.

Example 1 – *Creating a Frequency Distribution: Single Values* cont'd

The following results were obtained:

0	2	1	0	3	2	0	1	1	0
0	1	1	2	4	1	0	1	1	0
2	1	0	0	3	0	0	1	2	1
0	0	2	4	1	1	0	1	2	0
1	1	0	3	5	1	2	1	3	2

Let's organize the data:
create a *frequency
distribution*.

Number of Children under Six
0
1
2
3
4
5

Example 1 – Creating a Frequency Distribution: Single Values cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
0 1 1 2 4 1 0 1 1 0
2 1 0 0 3 0 0 1 2 1
0 0 2 4 1 1 0 1 2 0
1 1 0 3 5 1 2 1 3 2

Let's organize the data:
create a *frequency
distribution*.

Number of Children under Six	Tally
0	+++ ++
1	+++ ++ +++
2	+++
3	
4	
5	

Example 1 – Creating a Frequency Distribution: Single Values cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
0 1 1 2 4 1 0 1 1 0
2 1 0 0 3 0 0 1 2 1
0 0 2 4 1 1 0 1 2 0
1 1 0 3 5 1 2 1 3 2

Let's organize the data:
create a *frequency distribution*.

Number of Children under Six	Tally	Frequency
0		16
1		18
2		9
3		4
4		2
5		1
		$n = 50$

Example 1 – Creating a Frequency Distribution: Single Values

cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
0 1 1 2 4 1 0 1 1 0
2 1 0 0 3 0 0 1 2 1
0 0 2 4 1 1 0 1 2 0
1 1 0 3 5 1 2 1 3 2

Number of Children under Six	Tally	Frequency
0		16
1		18
2		9
3		4
4		2
5		1
		$n = 50$

Let's organize the data:
create a *frequency distribution*.

The composition of a frequency distribution is often easier to see if the frequencies are converted to percents, especially if large amounts of data are being summarized.

Example 1 – Creating a Frequency Distribution: Single Values

cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
0 1 1 2 4 1 0 1 1 0
2 1 0 0 3 0 0 1 2 1
0 0 2 4 1 1 0 1 2 0
1 1 0 3 5 1 2 1 3 2

Number of Children under Six	Tally	Frequency
0		16
1		18
2		9
3		4
4		2
5		1
		$n = 50$

Let's organize the data:
create a *frequency distribution*.

*The **relative frequency** of a data point is the frequency of the data point expressed as a percent of the total number of data points (that is, made relative to the total)*

Example 1 – Creating a Frequency Distribution: Single Values cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
0 1 1 2 4 1 0 1 1 0
2 1 0 0 3 0 0 1 2 1
0 0 2 4 1 1 0 1 2 0
1 1 0 3 5 1 2 1 3 2

Let's organize the data:
create a *frequency distribution*.

Number of Children under Six	Tally	Frequency	Relative Frequency
0		16	$\frac{16}{50} = 0.32 = 32\%$
1		18	$\frac{18}{50} = 0.36 = 36\%$
2		9	$\frac{9}{50} = 0.18 = 18\%$
3		4	$\frac{4}{50} = 0.08 = 8\%$
4		2	$\frac{2}{50} = 0.04 = 4\%$
5		1	$\frac{1}{50} = 0.02 = 2\%$
		$n = 50$	total = 100%

*The **relative frequency** of a data point is the frequency of the data point expressed as a percent of the total number of data points (that is, made relative to the total)*

Example 1 – Creating a Frequency Distribution: Single Values cont'd

The following results were obtained:

0 2 1 0 3 2 0 1 1 0
 0 1 1 2 4 1 0 1 1 0
 2 1 0 0 3 0 0 1 2 1
 0 0 2 4 1 1 0 1 2 0
 1 1 0 3 5 1 2 1 3 2

Let's organize the data:
 create a *frequency distribution*.

Number of Children under Six	Tally	Frequency	Relative Frequency
0		16	$\frac{16}{50} = 0.32 = 32\%$
1		18	$\frac{18}{50} = 0.36 = 36\%$
2		9	$\frac{9}{50} = 0.18 = 18\%$
3		4	$\frac{4}{50} = 0.08 = 8\%$
4		2	$\frac{2}{50} = 0.04 = 4\%$
5		1	$\frac{1}{50} = 0.02 = 2\%$
		$n = 50$	total = 100%

The raw data have now been organized and summarized.

Example 1 – Solution

cont'd

- we can see that about one-third of the employees have no need for child care (**32%**), while

- the remaining two-thirds (**68%**) have at least one child under 6 years of

age who would benefit from company-sponsored daycare

- the most common trend (that is, the data point with the highest relative frequency for the fifty employees) is having one child (36%).

Number of Children under Six	Tally	Frequency	Relative Frequency
0		16	$\frac{16}{50} = 0.32 = 32\%$
1		18	$\frac{18}{50} = 0.36 = 36\%$
2		9	$\frac{9}{50} = 0.18 = 18\%$
3		4	$\frac{4}{50} = 0.08 = 8\%$
4		2	$\frac{2}{50} = 0.04 = 4\%$
5		1	$\frac{1}{50} = 0.02 = 2\%$
		$n = 50$	total = 100%



Grouped Data

Grouped Data

When raw data consist of only a few distinct values (for instance, the data in Example 1, which consisted of only the numbers 0, 1, 2, 3, 4, and 5), we can easily organize the data and determine any trends by listing each data point along with its frequency and relative frequency.

Grouped Data

When raw data consist of only a few distinct values (for instance, the data in Example 1, which consisted of only the numbers 0, 1, 2, 3, 4, and 5), we can easily organize the data and determine any trends by listing each data point along with its frequency and relative frequency.

However, when the raw data consist of many nonrepeated data points, listing each one separately does not help us to see any trends the data set might contain.

In such cases, it is useful to group the data into intervals or classes and then determine the frequency and relative frequency of each group rather than of each data point.

Example 2 – *Creating a Frequency Distribution: Grouped Data*

Keith Reed is an instructor for an acting class offered through a local arts academy. The class is open to anyone who is at least 16 years old. Forty-two people are enrolled; their ages are as follows:

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Organize the data by creating a frequency distribution.

Listing each distinct data point and its frequency might not summarize the data well enough for us to draw conclusions. Instead, we will work with grouped data.

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Step 2: decide on number of groups (usually 4 to 8 groups).
Let's do 6 groups!

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Step 2: decide on number of groups (usually 4 to 8 groups).
Let's do 6 groups!

Step 3: find the beginning and end of all groups.

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Step 2: decide on number of groups (usually 4 to 8 groups).
Let's do 6 groups!

Step 3: find the beginning and end of all groups.
The *first group* begins with **16** (*smallest value*).

Example 2 – Creating a Frequency Distribution: Grouped Data

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Step 2: decide on number of groups (usually 4 to 8 groups).
Let's do 6 groups!

Step 3: find the beginning and end of all groups.

The *first group* begins with **16** (*smallest value*).

The *second group* will begin with **16** (smallest value) + *step*, where

$$\text{step} = \frac{\text{range}}{\text{number of groups}}, \text{ rounded off.}$$

width

Example 2 – Creating a Frequency Distribution: Grouped Data

26	16	21	34	45	18	41	38	22
48	27	22	30	39	62	25	25	38
29	31	28	20	56	60	24	61	28
32	33	18	23	27	46	30	34	62
49	59	19	20	23	24			

Step 1: find the *range*, *largest age - smallest age*:
 $62 - 16 = 46$ years.

Step 2: decide on number of groups (usually 4 to 8 groups).
Let's do 6 groups!

Step 3: find the beginning and end of all groups.

The *first group* begins with **16** (*smallest value*).

The *second group* will begin with **16** (smallest value) + *step*, where

$$\text{step} = \frac{\text{range}}{\text{number of groups}}, \text{ rounded off.}$$
$$\text{width} / \text{step} = \frac{46}{6} = 7.6666 \dots \approx 8$$

Example 2 – Creating a Frequency Distribution: Grouped Data

26 16 21 34 45 18 41 38 22
48 27 22 30 39 62 25 25 38
29 31 28 20 56 60 24 61 28
32 33 18 23 27 46 30 34 62
49 59 19 20 23 24

$x = \text{Age (years)}$
$16 \leq x < 24$
$24 \leq x < 32$
$32 \leq x < 40$
$40 \leq x < 48$
$48 \leq x < 56$
$56 \leq x < 64$

range = $62 - 16 = 46$ years.

6 groups

first group begins with **16** (smallest value).

second group begins with $16 + 8 = 24$

$$\text{width/step} = \frac{46}{6} = 7.6666 \dots \approx 8$$

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26 16 21 34 45 18 41 38 22
 48 27 22 30 39 62 25 25 38
 29 31 28 20 56 60 24 61 28
 32 33 18 23 27 46 30 34 62
 49 59 19 20 23 24

$x = \text{Age (years)}$	Tally
$16 \leq x < 24$	
$24 \leq x < 32$	
$32 \leq x < 40$	
$40 \leq x < 48$	
$48 \leq x < 56$	
$56 \leq x < 64$	

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26 16 21 34 45 18 41 38 22
 48 27 22 30 39 62 25 25 38
 29 31 28 20 56 60 24 61 28
 32 33 18 23 27 46 30 34 62
 49 59 19 20 23 24

$x = \text{Age (years)}$	Tally	Frequency
$16 \leq x < 24$		11
$24 \leq x < 32$		13
$32 \leq x < 40$		7
$40 \leq x < 48$		3
$48 \leq x < 56$		2
$56 \leq x < 64$		6
		$n = 42$

Example 2 – *Creating a Frequency Distribution: Grouped Data*

26 16 21 34 45 18 41 38 22
 48 27 22 30 39 62 25 25 38
 29 31 28 20 56 60 24 61 28
 32 33 18 23 27 46 30 34 62
 49 59 19 20 23 24

$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$		11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$		13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$		7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$		6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%

Example 2 – *Creating a Frequency Distribution: Grouped Data*

$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$		11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$		13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$		7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$		6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%

Example 2 – *Creating a Frequency Distribution: Grouped Data*

$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$		11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$		13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$		7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$		6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%

Trends we observe:

- ages from 24 to 32 are most common (31% is the highest relative frequency), and
- ages from 48 to 56 are least common (5% is the lowest)
- Also, over half the people enrolled (57%) are from 16 to 32 years old.

Grouped Data

Constructing a Frequency Distribution

1. If the raw data consist of many different values, create intervals and work with grouped data. If not, list each distinct data point. [When working with grouped data, choose from four to eight intervals. Divide the range (high minus low) by the desired number of intervals, round off this answer to be consistent with the data, and then add the result to the lowest data point to find the beginning of the second group.]
2. Tally the number of data points in each interval or the number of times each individual data point occurs.
3. List the frequency of each interval or each individual data point.
4. Find the relative frequency by dividing the frequency of each interval or each individual data point by the total number of data points in the distribution. The resulting decimal can be expressed as a percent.



Histograms

Histograms

When data are grouped in intervals, they can be depicted by a **histogram**, a bar chart that shows how the data are distributed in each interval.

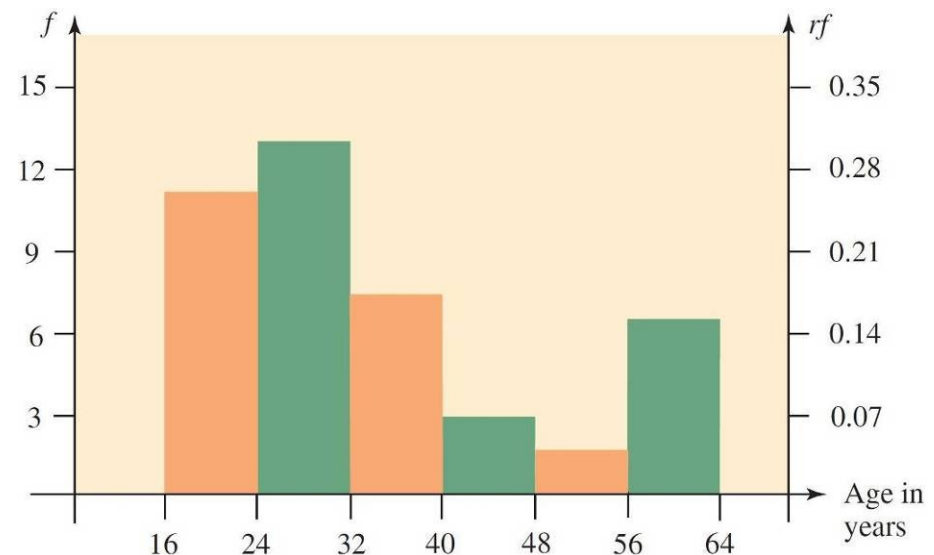
$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$		11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$		13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$		7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$		6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%

Ages of the people in Keith Reed's acting class.

Histograms

When data are grouped in intervals, they can be depicted by a **histogram**, a bar chart that shows how the data are distributed in each interval.

$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$		11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$		13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$		7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$		6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%



Ages of the people in Keith Reed's acting class.

Histograms

When data are grouped in intervals, they can be depicted by a **histogram**, a bar chart that shows how the data are distributed in each interval.

To construct a histogram, mark off the class limits on a horizontal axis. If each interval has equal width, we draw two vertical axes; the axis on the left exhibits the frequency of an interval, and the axis on the right gives the corresponding relative frequency.

We then draw a rectangle above each interval; the height of the rectangle corresponds to the number of data points contained in the interval.

Histograms

What happens if the intervals do not have equal width? For instance, suppose the ages of the people in Keith Reed's acting class are those given in the frequency distribution shown in Figure 4.5.

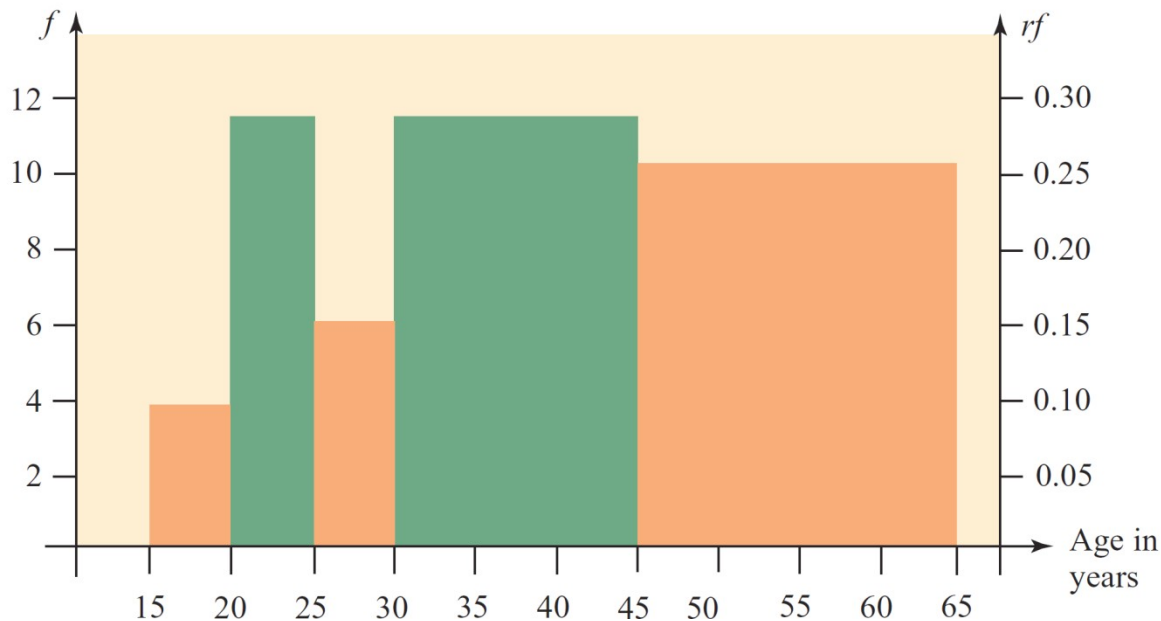
$x = \text{Age (years)}$	Frequency	Relative Frequency	Class Width
$15 \leq x < 20$	4	$\frac{4}{42} \approx 10\%$	5
$20 \leq x < 25$	11	$\frac{11}{42} \approx 26\%$	5
$25 \leq x < 30$	6	$\frac{6}{42} \approx 14\%$	5
$30 \leq x < 45$	11	$\frac{11}{42} \approx 26\%$	15
$45 \leq x < 65$	10	$\frac{10}{42} \approx 24\%$	20
	$n = 42$	Total = 100%	

Frequency distribution of age.

Figure 4.5

Histograms

With frequency and relative frequency as the vertical scales, the histogram depicting this new distribution is given in Figure 4.6.

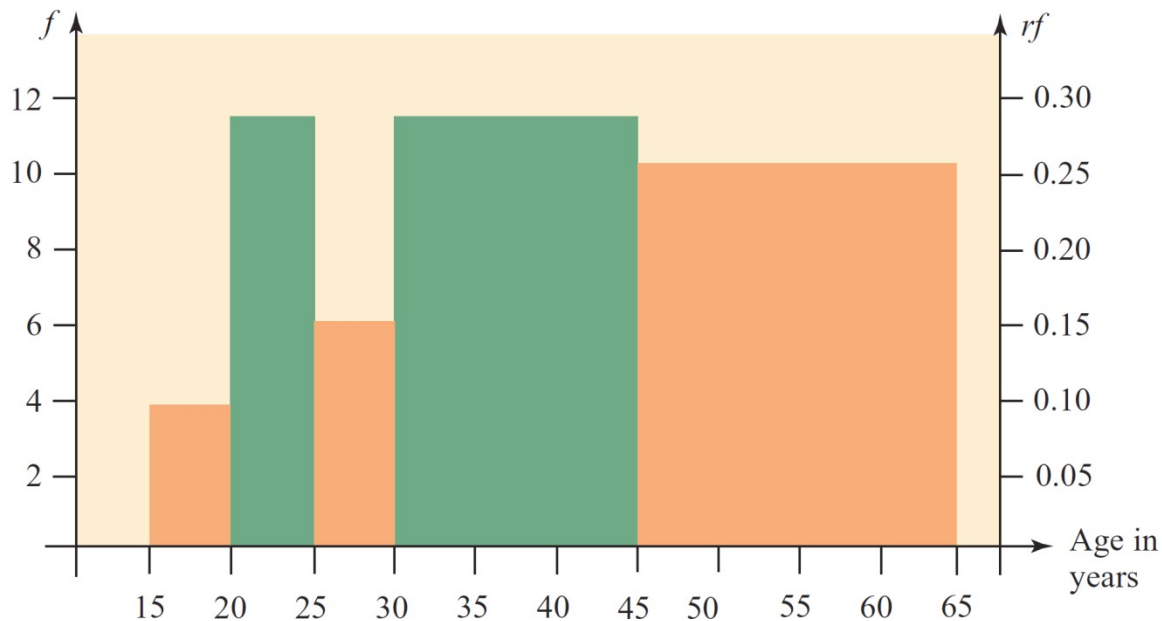


Why is this histogram misleading?

Figure 4.6

Histograms

With frequency and relative frequency as the vertical scales, the histogram depicting this new distribution is given in Figure 4.6.



An issue:

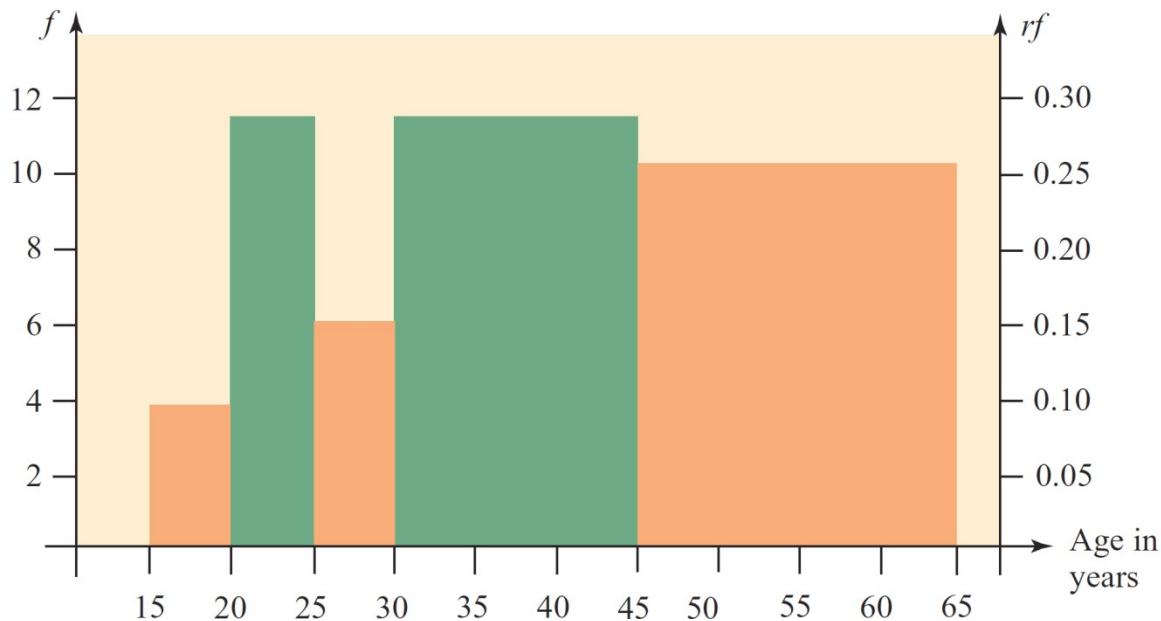
the rectangle over the interval from 45 to 65 appears to be larger than the rectangle over the interval from 30 to 45, yet this interval contains fewer data than the interval from 30 to 45.

Why is this histogram misleading?

Figure 4.6

Histograms

With frequency and relative frequency as the vertical scales, the histogram depicting this new distribution is given in Figure 4.6.



Why is this histogram misleading?

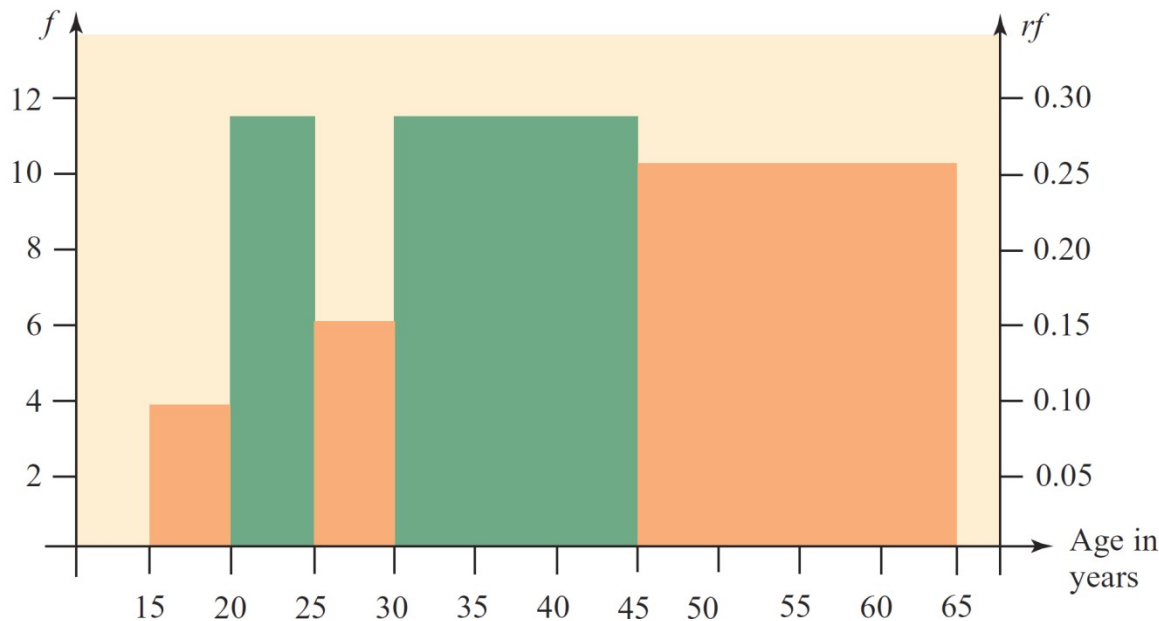
Figure 4.6

An issue:

This is misleading; rather than comparing the heights of the rectangles, our eyes naturally compare the areas of the rectangles.

Histograms

With frequency and relative frequency as the vertical scales, the histogram depicting this new distribution is given in Figure 4.6.



Why is this histogram misleading?

Figure 4.6

A solution:

to make an accurate comparison, *the areas of the rectangles must correspond to the relative frequencies of the intervals.*

This is accomplished by utilizing the **density** of each interval.



Histograms and Single-Valued Classes



Histograms and Single-Valued Classes

When a set of data consists of only a few distinct values, it may be advantageous to consider each distinct value to be a “class” of data; that is, we utilize single-valued classes of data.

Example 4 – *Constructing a Histogram: Single Values*

A sample of high school seniors was asked, “How many television sets are in your house?” The frequency distribution in Figure 4.13 summarizes the data. Construct a histogram using single-valued classes of data.

Number of Television Sets	Frequency
0	2
1	13
2	18
3	11
4	5
5	1

Frequency distribution.

Figure 4.13

Example 4 – *Solution*

Rather than using intervals of grouped data, we use a single value to represent each class.

Each class has the same width = 1, hence we construct a combined frequency and relative frequency histogram.

Number of Television Sets	Frequency	Relative Frequency
0	2	$2/50 = 4\%$
1	13	$13/50 = 26\%$
2	18	$18/50 = 36\%$
3	11	$11/50 = 22\%$
4	5	$5/50 = 10\%$
5	1	$1/50 = 2\%$
	$n = 50$	Total = 100%

Calculating relative frequency.

Figure 4.14

Example 4 – Solution

Rather than using intervals of grouped data, we use a single value to represent each class.

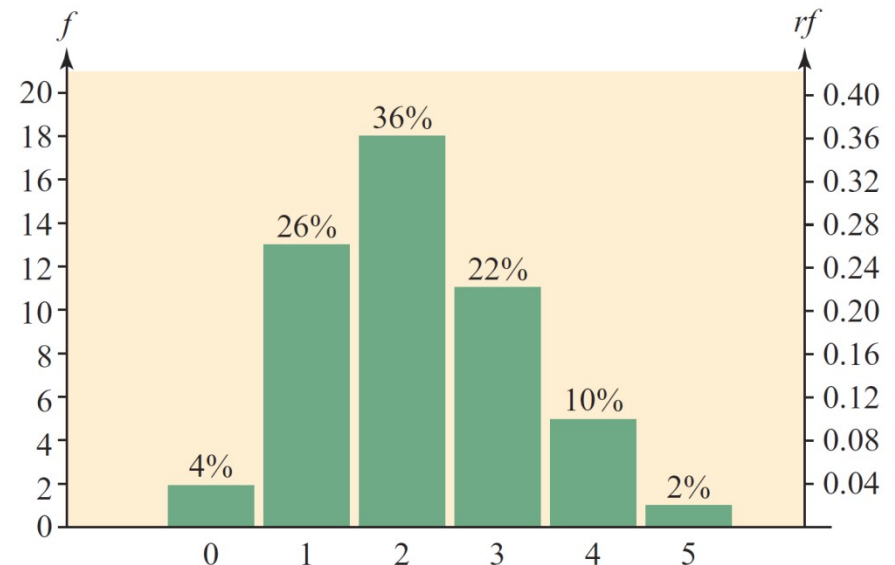
Each class has the same width = 1, hence we construct a combined frequency and relative frequency histogram.

We write the single value at the midpoint of the base of each rectangle.

Number of Television Sets	Frequency	Relative Frequency
0	2	$2/50 = 4\%$
1	13	$13/50 = 26\%$
2	18	$18/50 = 36\%$
3	11	$11/50 = 22\%$
4	5	$5/50 = 10\%$
5	1	$1/50 = 2\%$
	$n = 50$	Total = 100%

Calculating relative frequency.

Figure 4.14



Number of television sets.

Figure 4.15



Pie Charts

Pie Charts

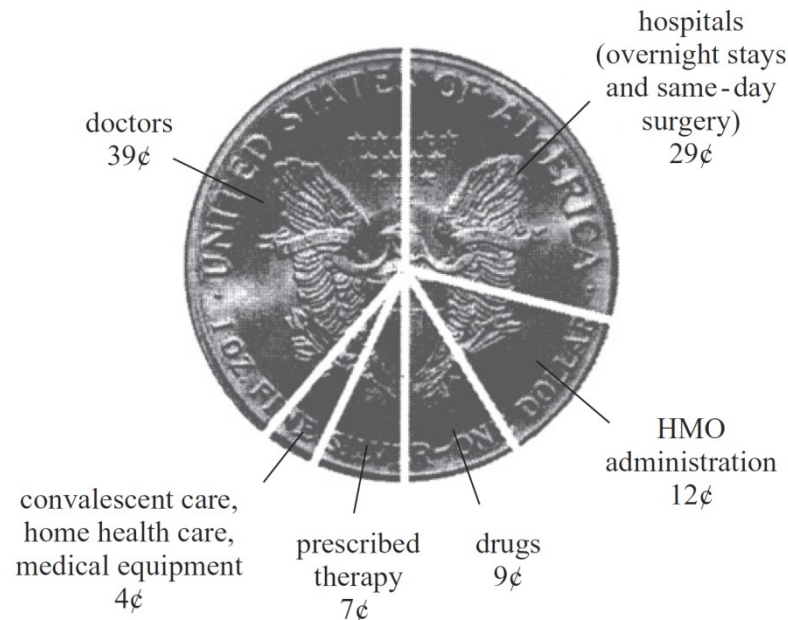
Many statistical studies involve ***categorical data***—those which is grouped according to some common feature or quality.

One of the easiest ways to summarize categorical data is through the use of a ***pie chart***.

A pie chart shows how various categories of a set of data account for certain proportions of the whole.

Pie Charts

Financial incomes and expenditures are invariably shown as pie charts, as in Figure 4.16.



How a typical medical dollar is spent.

Figure 4.16

Pie Charts

To draw the “slice” of the pie representing the relative frequency (percentage) of the category, the appropriate *central angle* must be calculated.

Since a complete circle comprises 360 degrees, we obtain the required angle *by multiplying 360° times the relative frequency* of the category.

Example 5 – Constructing a Pie Chart

What type of academic degree do you hope to earn? The different types of degrees and the number of each type of degree conferred in the United States during the 2010–11 academic year is given in Figure 4.17. Construct a pie chart to summarize the data.

Type of Degree	Frequency (thousands)
Associate's	942
Bachelor's	1,716
Master's	731
Doctorate	164
Total	3,553

Academic degrees conferred, 2010–11.
Source: National Center for Education Statistics.

Figure 4.17

Example 5 – Solution

cont'd

Find the relative frequency of each category and multiply it by 360° to determine the appropriate central angle.

Type of Degree	Frequency (thousands)	Relative Frequency	Central Angle
Associate's	942	$\frac{942}{3,553} \approx 0.265$	$0.265 \times 360^\circ = 95.40^\circ$
Bachelor's	1,716	$\frac{1,716}{3,553} \approx 0.483$	$0.483 \times 360^\circ = 173.88^\circ$
Master's	731	$\frac{731}{3,553} \approx 0.206$	$0.206 \times 360^\circ = 74.16^\circ$
Doctorate	164	$\frac{164}{3,553} \approx 0.046$	$0.046 \times 360^\circ = 16.56^\circ$
	$n = 3,553$	Sum = 1,000	Total = 360°

Calculating relative frequency and central angles.

Figure 4.18

Example 5 – Solution

cont'd

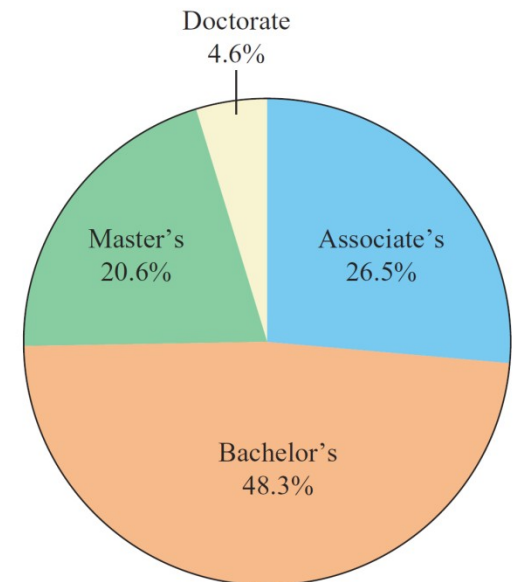
Find the relative frequency of each category and multiply it by 360° to determine the appropriate central angle.

Then use a protractor to lay out the angles and draw the “slices.”

Type of Degree	Frequency (thousands)	Relative Frequency	Central Angle
Associate's	942	$\frac{942}{3,553} \approx 0.265$	$0.265 \times 360^\circ = 95.40^\circ$
Bachelor's	1,716	$\frac{1,716}{3,553} \approx 0.483$	$0.483 \times 360^\circ = 173.88^\circ$
Master's	731	$\frac{731}{3,553} \approx 0.206$	$0.206 \times 360^\circ = 74.16^\circ$
Doctorate	164	$\frac{164}{3,553} \approx 0.046$	$0.046 \times 360^\circ = 16.56^\circ$
	$n = 3,553$	Sum = 1,000	Total = 360°

Calculating relative frequency and central angles.

Figure 4.18



Academic degrees conferred, 2010–11.

Figure 4.19

Each slice of the pie should contain its relative frequency, expressed as a percent.

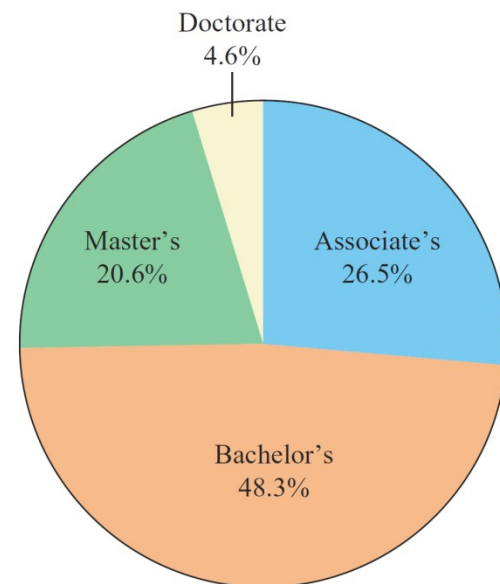
Example 5 – Solution

cont'd

Find the relative frequency of each category and multiply it by 360° to determine the appropriate central angle.

Then use a protractor to lay out the angles and draw the “slices.”

Type of Degree	Frequency (thousands)	Relative Frequency	Central Angle
Associate's	942	$\frac{942}{3,553} \approx 0.265$	$0.265 \times 360^\circ = 95.40^\circ$
Bachelor's	1,716	$\frac{1,716}{3,553} \approx 0.483$	$0.483 \times 360^\circ = 173.88^\circ$
Master's	731	$\frac{731}{3,553} \approx 0.206$	$0.206 \times 360^\circ = 74.16^\circ$
Doctorate	164	$\frac{164}{3,553} \approx 0.046$	$0.046 \times 360^\circ = 16.56^\circ$
	$n = 3,553$	Sum = 1,000	Total = 360°



Academic degrees conferred, 2010–11.

The name of each category can be written directly on the slice, or, if the names are too long, a legend consisting of various shadings may be used.