Mathematics 13 : Elementary Statistics

Unit 1 : Section 4

Example 2

Example 2

Twelve residents from the city of Rocklin were randomly selected and asked "How many TVs are in your household?".

The following data were obtained:

   2, 3, 3, 1, 2, 5, 3, 4, 1, 2, 4, and 3

Construct a frequency distribution table with this data and describe the result.

Example 2

The Number of TVs in Rocklin Households

Number of TVs	Frequency
1	2
2	3
3	4
4	2
5	1

n = 12

Example 2

Description:

Most, which was 4, of the 12 Rocklin households sampled had 3 TVs. This was closely followed by 3 other households which had 2 TVs each. Next, there was an equal number, with frequency of 2, of 1 and 4 TV households. Finally, just one household reported to have 5 TVs.

Example 2

This description, although accurate, was not expressed very clearly and seemed somewhat convoluted.

The objective in descriptive statistics is to organize and present data in an accurate, useful, and informative manner.

Data that reaches the ordinal level or above can be further examined to identify particular patterns in the distribution.

Example 2

In this sample, the highest frequency occurs in the center of the distribution. As the data values move away from the center, both before and after, the frequencies decline in a symmetric fashion.

In statistics, this type of pattern exhibited in data is referred to as bell-shaped.

Example 2

Description:

For this sample, the number of TVs in Rocklin households resembled a bell-shaped distribution.

Learning and understanding the language of statistics allows us to make clear, concise, and constructive descriptions.

Lesson 13 :

Histogram

Histogram

A histogram is a bar graph constructed from quantitative data where the data values are arranged, individually or in groups, in numerical order along the horizontal axis and either the frequency or the proportion is presented along the vertical axis.

Histogram

In a histogram, the height of the bar communicates the information regarding the parameter or statistic of interest.

Also, adjacent bars are drawn touching each other in order to convey that the data is arranged in a precise order where one data value follows the next.
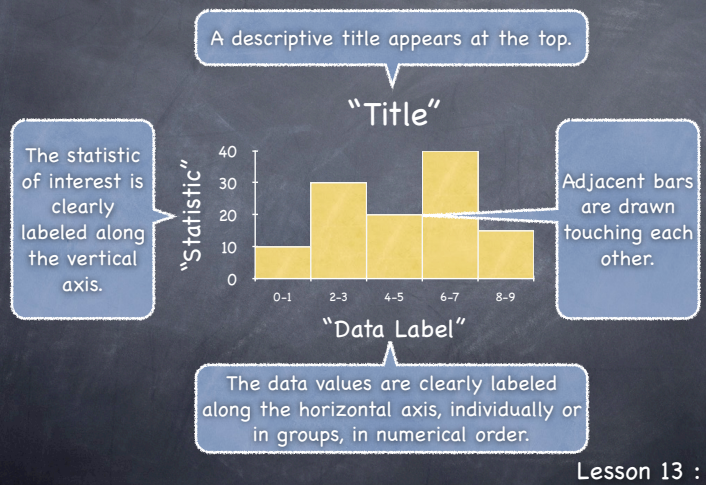
Histogram

Since each data value can belong to one and only one bar in the histogram, it is essential to avoid any ambiguity as to which data value belongs to which specific bar when labeling the horizontal axis.
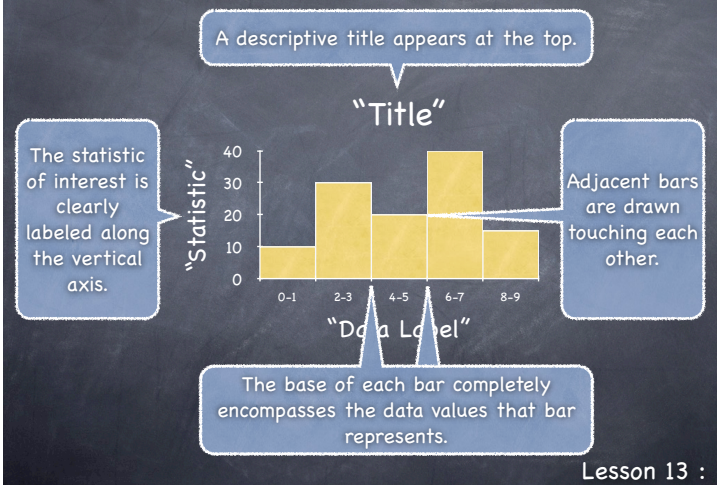
This is accomplished by drawing the bars in such a way that the base of each bar completely encompasses the data values that bar represents.

Slide 1

A descriptive title appears at the top.

"Title"

The statistic of interest is clearly labeled along the vertical axis.

"Statistic"

40
30
20
10
0

0–1 2–3 4–5 6–7 8–9

Adjacent bars are drawn touching each other.

"Data Label"

The data values are clearly labeled along the horizontal axis, individually or in groups, in numerical order.

Lesson 13 :

Slide 2

Histogram

A descriptive title appears at the top.

"Title"

The statistic of interest is clearly labeled along the vertical axis.

"Statistic"

40
30
20
10
0

0–1 2–3 4–5 6–7 8–9

Adjacent bars are drawn touching each other.

"Data Label"

The base of each bar completely encompasses the data values that bar represents.

Lesson 13 :

Slide 3

Histogram

A histogram is an alternative graphical representation of the same information that could be presented in a distribution table.

As such, most of the same factors that guide the construction of a distribution table are also considered in the construction of a histogram.

Lesson 13 :

Slide 4

Histogram

For instance,

The data values could be displayed individually or in groups.

The first bar must contain (not necessarily start at) the minimum data value.

The last bar must contain (not necessarily end at) the maximum data value.

Lesson 13 :

Slide 5

Histogram

The scale should be consistent throughout the histogram. That is, the size of the increment from bar to bar should be the same.

The resulting histogram should not have too few nor too many bars. In general, histograms containing anywhere from 4 to 24 bars are reasonable.

Lesson 13 :

Slide 6

Histogram

Each and every data value must belong to one and only one of the bars in the histogram.

There is no single acceptable solution for grouping the data values. As long as the selected grouping produces a histogram that clearly reveals the distribution of the data, that solution is considered acceptable.

Lesson 13 :

Histogram

A histogram is the preferred graphical method of displaying the distribution of the data whenever the data under examination is quantitative, and it is important to the issue under investigation to illustrate any underlying patterns in the data.

Example 1

Twelve residents from the city of Rocklin were randomly selected and asked "How many TVs are in your household?".
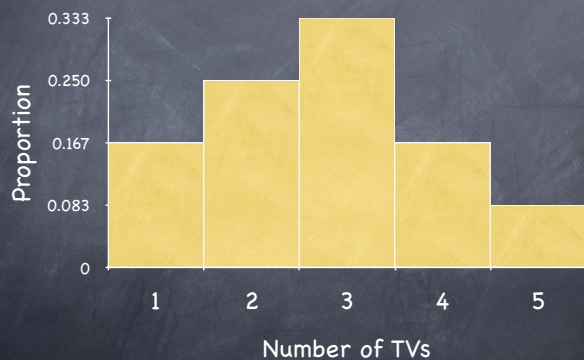
The following data were obtained:

  2, 3, 3, 1, 2, 5, 3, 4, 1, 2, 4, and 3

Construct a proportion histogram with this data and describe the result.

Example 1

The Number of TVs in Rocklin Households

Example 1

Description:

For this sample, the number of TVs in Rocklin households resembled a bell-shaped distribution.

Lesson 14 :

Stem-and-Leaf Plot

Stem-and-Leaf Plot

A stem-and-leaf plot incorporates the data values themselves in the construction of the graph.

The digits making up the data values are separated. This separation in the digits is illustrated in the graph with a vertical line. The digits to the left of the separation line are called the stems. The stems are listed vertically in numerical order.

Stem-and-Leaf Plot

The digits to the right of the separation line are called the leafs. The leafs are listed horizontally in numerical order.

Each and every data value is utilized in the construction of a stem-and-leaf plot. In order to distinguish one data value from the next, a noticeable space is displayed between consecutive leaf values.

Stem-and-Leaf Plot
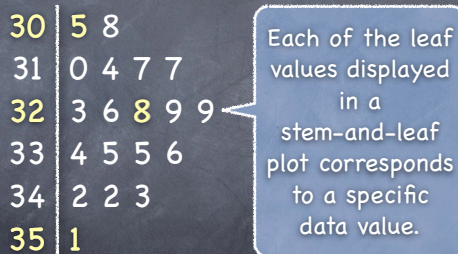
A descriptive title appears at the top.

"Title"

30	5 8
31	0 4 7 7
32	3 6 8 9 9
33	4 5 5 6
34	2 2 3
35	1

The stems are listed vertically in numerical order.

The leafs are listed horizontally in numerical order with a noticeable space displayed between consecutive leaf values.
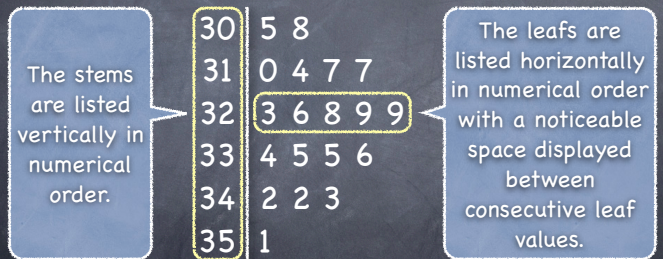
The vertical line represents the separation of the digits in the data values.

Stem-and-Leaf Plot

"Title"

30	5 8
31	0 4 7 7
32	3 6 8 9 9
33	4 5 5 6
34	2 2 3
35	1

Each of the leaf values displayed in a stem-and-leaf plot corresponds to a specific data value.

Stem-and-Leaf Plot

One advantage to the stem-and-leaf plot is the data values themselves are retained in the graph.

One disadvantage to the stem-and-leaf plot is the construction of the graph becomes more and more problematic as the sample size increases.

Stem-and-Leaf Plot

There are several factors to consider when constructing a stem-and-leaf plot.

One factor is the location of the separation point within the data value digits. This location must be in the same position for each and every data value in the graph.

Stem-and-Leaf Plot

Another factor to consider is the grouping of potential leaf values.

For instance,

the 10 potential leaf values

0 1 2 3 4 5 6 7 8 9

can be grouped together evenly in four different arrangements.

Stem-and-Leaf Plot

A group of 10 leaf values across 1 stem

```
1│0 1 2 3 4 5 6 7 8 9
```

A group of 5 leaf values across 2 stems

```
2│0 1 2 3 4
2│5 6 7 8 9
```

Stem-and-Leaf Plot

A group of 2 leaf values across 5 stems

```
5│0 1
5│2 3
5│4 5
5│6 7
5│8 9
```

Stem-and-Leaf Plot

A group of 1 leaf value across 10 stems

```
10│0
10│1
10│2
10│3
10│4
10│5
10│6
10│7
10│8
10│9
```

Stem-and-Leaf Plot

An appropriate choice for the location of the separation point and the grouping arrangement of potential leaf values produces an informative stem-and-leaf plot that clearly reveals the distribution of the data.

Too few stem values can conceal crucial details. Too many can dilute the essential elements of the distribution.

Example 1

Twelve residents from the city of Rocklin were randomly selected and asked "How many TVs are in your household?".

The following data were obtained:

2, 3, 3, 1, 2, 5, 3, 4, 1, 2, 4, and 3

Construct a stem-and-leaf plot with this data and describe the result.

Example 1

The Number of TVs in
Rocklin Households

```
1.│0 0
2.│0 0 0
3.│0 0 0 0
4.│0 0
5.│0
```

Example 1

The Number of TVs in
Rocklin Households

```
0 | 1 1
0 | 2 2 2
0 | 3 3 3 3
0 | 4 4
0 | 5
```

Example 1

Description:

For this sample, the number of TVs in Rocklin households resembled a bell-shaped distribution.

Lesson 10 : Distribution Table

Example 3

Example 3

Use the age data (Q4) from our Sierra College Elementary Statistics Student Survey to construct a percentage distribution table and describe the result.

Example 3

Q4. What is your age? _____

The minimum data value belongs to the first row of the table.

Age
17
53

The maximum data value belongs to the last row of the table.

Example 3

Q4. What is your age?

If every age from 17 to 53 is listed individually, the table would be 37 rows long. This is too many rows for a reasonable distribution table. In this case, the data values need to be grouped.

Age
17
53

Example 3

There is no single acceptable solution for grouping the data values.

As long as the selected grouping produces a table that clearly reveals the distribution of the data, that solution is considered acceptable.

There are several factors to consider when deciding on how to group the data values.

Example 3

For instance,

The first row must contain (not necessarily start at) the minimum data value.

The last row must contain (not necessarily end at) the maximum data value.

The scale should be consistent throughout the table. That is, the size of the increment from row to row should be the same.

Example 3

The resulting distribution table should not have too few nor too many rows. In general, tables containing anywhere from 4 to 24 rows are reasonable.

Each and every data value must belong to one and only one of the groups in the table.

Example 3

Age of Sierra College
Elementary Statistics Students

Age	Percentage
15 – 19	20.5%
20 – 24	59.1%
25 – 29	9.1%
30 – 34	2.3%
35 – 39	2.3%
40 – 44	4.5%
45 – 49	0.0%
50 – 54	2.3%

Example 3

The majority, 59.1%, of the Sierra College Elementary Statistics students surveyed were 20 to 24 years old. About 20.5% were around 15 to 19 years old while just under half that much, 9.1%, were 25 to 29 years old. There were students that ranged in age from about 30 to nearly 54 years old, but this older age group only made up 11.4% of this sample.

Example 3

In the language of statistics, distributions that are asymmetric (that is, not symmetric or lopsided) are called skewed.

The ages of Sierra College Elementary Statistics students is skewed (asymmetric). The majority of students were 20 to 24. Students younger than 20 were only younger by a few years. Students older than 24 were older by several years.

Example 3

When the data values extend much further in the increasing (or positive) direction than the decreasing (or negative) direction, the distribution is said to be positively skewed.

When the data values extend much further in the decreasing (or negative) direction than the increasing (or positive) direction, the distribution is said to be negatively skewed.

Example 3

Description:

The age of Sierra College Elementary Statistics Students has a positively skewed shaped distribution.
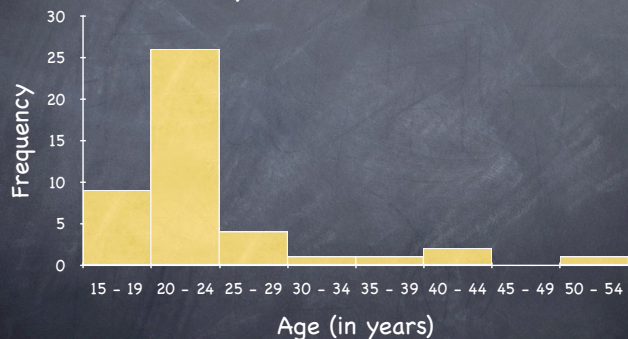

Lesson 13 : Histogram

Example 2

Example 2

Use the age data (Q4) from our Sierra College Elementary Statistics Student Survey to construct a frequency histogram and describe the result.

Example 2

Age of Sierra College Elementary Statistics Students

Example 2

Description:

The age of Sierra College Elementary Statistics Students has a positively skewed shaped distribution.

Example 2

Realizing that the distribution of the data is skewed can provide some insight into important aspects of the data.

For instance, skewed distributions could be an indication of the presence or absence of various constraints or restrictions on potential data values.

Example 2

If the data values do not extend very far in one direction, it may be due to the presence of a nearby limiting factor.

If the data values do extend very far in one direction, it may be due to either the nonexistence of any limiting factor or the limiting factor being faraway from the data.

Example 2

In this sample, the age of Sierra College Elementary Statistics students did not extend very far in the younger direction.

That is probably due to the fact that students need to progress through several years of education before they are sufficiently prepared to enroll in an Elementary Statistics course at Sierra College.

Example 2

On the other hand, the age of Sierra College Elementary Statistics students did extend far in the older direction.

This is because there is no upper restriction on age as to when a student can enroll in an Elementary Statistics course at Sierra College. As long as the student is prepared to take the course, it is never to late to learn.

Lesson 14 : Stem-and-Leaf Plot

Example 2

Example 2

Use the age data (Q4) from our Sierra College Elementary Statistics Student Survey to construct a stem-and-leaf plot and describe the result.

Example 2

Age of Sierra College
Elementary Statistics Students

```
1 | 778888999
2 | 000000000001111122222222334
2 | 5699
3 | 4
3 | 8
4 | 03
4 |
5 | 3
```

Example 2

Description:

The age of Sierra College Elementary Statistics Students has a positively skewed shaped distribution.

> Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

Use the courses enrolled data (Q7) from our Sierra College Elementary Statistics Student Survey to construct a frequency distribution table.

Exercise 2

Use the courses enrolled data (Q7) from our Sierra College Elementary Statistics Student Survey to construct a proportion histogram and describe the result.

Exercise 3

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

Use this data to complete the distribution table given below and describe the result.

Cholesterol (in mg)	Frequency	Percentage
170 – 174		

Exercise 4

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

Use this data to construct a frequency histogram with an initial grouping of 170 to 174 mg of cholesterol.

Exercise 5

University Hospital conducted a study in order to describe the age (in weeks) at which infants learn to crawl. Consent forms collected from the parents of all the children born at the hospital last year who agreed to participate in future research projects were separated by the gender of the child. A random sample of 12 girls and a random sample of 12 boys were selected from the available consent forms. The parents of the randomly selected children were contacted, and the age at which their child learned to crawl was confirmed. The following data were obtained as a result.

    Girls :   25, 23, 36, 27, 32, 31, 37, 40, 33, 29, 38, 34

    Boys :   39, 28, 47, 29, 35, 44, 35, 24, 36, 27, 25, 36

Construct a stem-and-leaf plot with this data and describe the result.

Exercise 6

Use the running time data collected in the Movie Database Sample to construct a stem-and-leaf plot.

> Your solutions should be clear, complete, and sufficiently detailed in order to demonstrate your understanding and communicate your reasoning and method of solving the problem.

Exercise 1

Use the courses enrolled data (Q7) from our Sierra College Elementary Statistics Student Survey to construct a frequency distribution table.
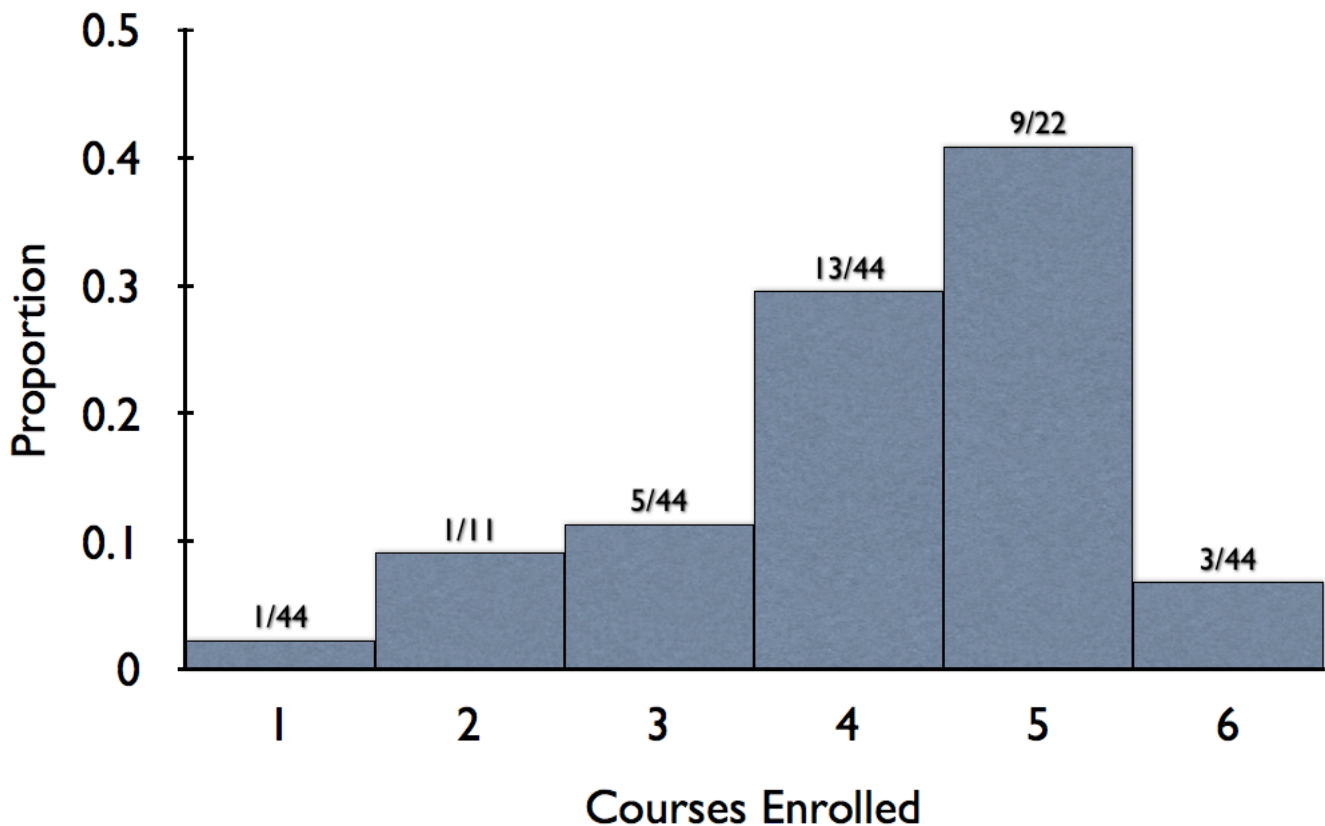
The Number of Courses Sierra College Elementary Statistics Students Surveyed were Enrolled in this Semester

Courses Enrolled	Frequency
1	1
2	4
3	5
4	13
5	18
6	3

Exercise 2

Use the courses enrolled data (Q7) from our Sierra College Elementary Statistics Student Survey to construct a proportion histogram and describe the result.

The Number of Courses Sierra College Elementary Statistics Students Surveyed were Enrolled in this Semester

The number of courses Sierra College Elementary Statistics students surveyed were enrolled in this semester displayed a negatively skewed shaped distribution.

Exercise 3

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

Use this data to complete the distribution table given below and describe the result.

The Amount of Cholesterol in a Standard Large Egg

Cholesterol (in mg)	Frequency	Percentage
170 – 174	1	2.8%
175 – 179	5	13.9%
180 – 184	12	33.3%
185 – 189	9	25.0%
190 – 194	7	19.4%
195 – 199	2	5.6%

Based on the USDA sample, the amount of cholesterol in a standard large egg follows a bell-shaped distribution.
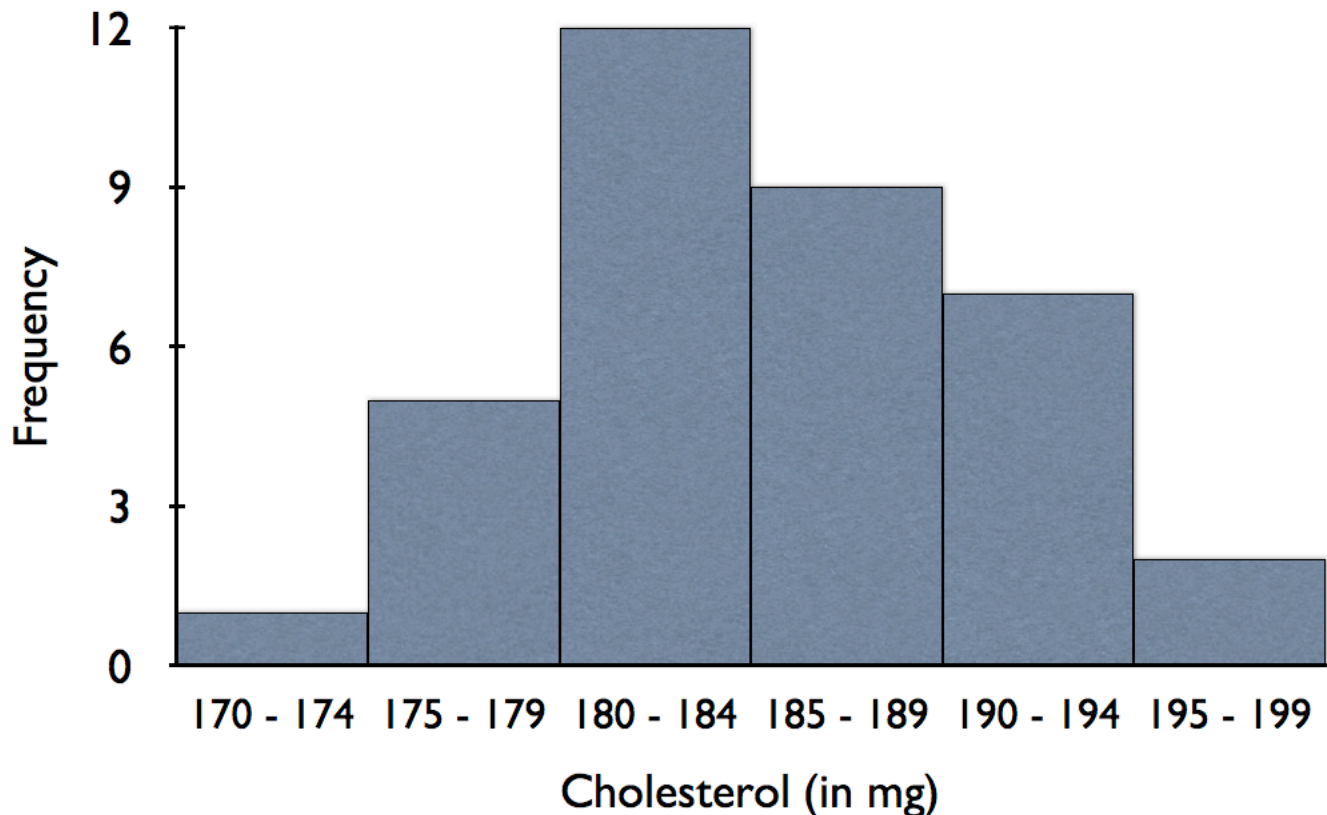
Exercise 4

A nutritionist working for the United States Department of Agriculture (USDA) randomly selected three cartons of eggs from all of the available cartons of standard large eggs at a neighborhood grocery store. Each egg in the randomly selected cartons had their nutritional content analyzed. The data provide here are the amounts of milligrams of cholesterol in each of the sampled eggs.

186, 188, 179, 180, 192, 186, 183, 177, 184, 178, 191, 174,
189, 176, 190, 188, 196, 187, 184, 184, 192, 194, 198, 183,
183, 181, 187, 190, 186, 176, 183, 185, 191, 180, 184, 182

Use this data to construct a frequency histogram with an initial grouping of 170 to 174 mg of cholesterol.

The Amount of Cholesterol in a Standard Large Egg

Exercise 5

University Hospital conducted a study in order to describe the age (in weeks) at which infants learn to crawl. Consent forms collected from the parents of all the children born at the hospital last year who agreed to participate in future research projects were separated by the gender of the child. A random sample of 12 girls and a random sample of 12 boys were selected from the available consent forms. The parents of the randomly selected children were contacted, and the age at which their child learned to crawl was confirmed. The following data were obtained as a result.

Girls :   25, 23, 36, 27, 32, 31, 37, 40, 33, 29, 38, 34

Boys :   39, 28, 47, 29, 35, 44, 35, 24, 36, 27, 25, 36

Construct a stem-and-leaf plot with this data and describe the result.

Age (in weeks) at which Infants Learn to Crawl

```
2 | 3  4
2 | 5  5  7  7  8  9  9
3 | 1  2  3  4
3 | 5  5  6  6  6  7  8  9
4 | 0  4
4 | 7
```


The age (in weeks) at which infants learn to crawl appears to have a double bell-shaped distribution where the peak of the first bell occurs in the 25 to 29 week interval and the peak of the second bell occurs in the 35 to 39 week interval.

Exercise 6

Use the running time data collected in the Movie Database Sample to construct a stem-and-leaf plot.

<div align="center">

Running Time (in minutes) of Movies in the Movie Database Sample

</div>

```
 6 | 9
 7 | 8
 8 | 1   8
 9 | 1   1   2   4   5   8   8
10 | 0   0   0   1   3   5   5   5   8
11 | 0   0   1   3   5   5   7   8   8   9
12 | 0   1   2   4   4   5   5   5   6   6   7   8   8
13 | 1   3   4   5   5   6   8
14 | 1   2
15 | 2   4
16 | 1   2   9
17 | 5   7
18 |
19 | 4
```