Chapter 2

Frequency Distributions and Graphs

توزيعات تكرارية الرسوم البيانية

Chapter 2

Outline:

- 2-1 Organizing Data تنظيم البيانات
- 2-2 Histograms, Frequency Polygons, and Ogives مدرج مضلع
- 2-3 Other Types of Graphs Summary

Chapter 2

Objectives:
After completing this chapter, you should be able to

- Organize data using a frequency distribution.
- Represent data in frequency distributions graphically, using histograms, frequency polygons, and ogives.
- Represent data using bar graphs, Pareto charts, time series graphs, pie graphs, and dot plots.
- Draw and interpret a stem and leaf plot.

Chapter 2

The purpose of this chapter

This chapter explains how to organize data by

- Convert raw data into a data array.
- Construct:
  - Frequency distributions.
  - A relative frequency distribution.
  - A cumulative (relative) frequency distribution.
- Present the data by constructing charts and graphs. The charts and graphs illustrated here are **histograms, frequency polygons, ogives, pie graphs, pareto charts, and time series graphs.**
- A graph that combines the characteristics of a frequency distribution and a histogram, called **a stem and leaf plot** is also explained $\cdots$

2-1 Organizing Data

Raw data is the data collected in original form.

البيانات الخام:- البيانات التي جمعت بشكلها الاصلي

Example:

Suppose a researcher wished to do a study on the ages of the 50 wealthiest people in the world. The researcher first would have to get the data on the ages of the people. They are called **raw data** and are listed next.

بيانات خام

45	46	64	57	85
92	51	71	54	48
27	66	76	55	69
54	44	54	75	46
61	68	78	61	83
88	45	89	67	56
81	58	55	62	38
55	56	64	81	38
49	68	91	56	68
46	47	83	71	62

دراسة حول عدد ساعات استخدام الجوال

10 ، 20 ، 0 ، 1 ، 5 ، 4

2

Row data

2-1 Organizing Data

تنظيم البيانات

When the raw data are organized into a table, which is called frequency distribution, the frequency will be the number of values in a specific class of the distribution.

جدول

Frequency distribution is the organization of raw data in a table form, using classes and frequencies.

تنظم البيانات في جدول التوزيع التكراري وتعني اي تكرارات وفئات

For more information see page 42 -43 of Bluman's Book

frequency : عدد مرات تكرار فئه معينه

جدول توزيع تكراري Frequency Distributions

فئات (classes)	frequency (تكرارات)
اقل من 5 سنوات	10 عدد مرات تكرار الاجابه المختلفه
من 5 الى 10 سنوات	60
من 10 الى 15 سنة	20
اكثر من 15 سنة	10
	100 مجموع الاجابات

2-1 Organizing Data

Two types of frequency distributions that are most often used are the categorical frequency distribution and the grouped frequency distribution.

1. Categorical frequency distributions is constructed for qualitative data used for data that can be placed in specific categories, such as **nominal** or **ordinal level** data.
2. Group frequency distributions is constructed for quantitative data used when the range of values in a data set is large (continuous data). The data must be grouped into classes that are more than one unit in width, e.g., 58-64.

1) categorical

classes	frequency
black	5
white	10
red	7
blue	15
other	3

2)

classes	frequency
0-20	3
20-40	7
40-60	2
<60	1

2) group frequency distribution

2-1 Categorical Frequency Distributions

Example 2-1 (Distribution of Blood Types) page 43

Twenty-five army inductees were given a blood test to determine their blood type. The data set is

اجري شخص دم لـ 25 ملجند لمعرفة نوع زمرة الدم

Row data

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

classes	frequency
A	5
B	7
O	9
AB	4

Construct a categorical frequency distribution?

The frequency distribution table has four columns:

1. the first column for Class, 2. the second column for Tallies, 3. the third column for Frequency, and 4. the fourth column for Percent.

نسبه مئويه لكل فئه

فئه

Class	Tally	Frequency	Percent%
A	𝍸	5	20
B	𝍸 ||	7	28
O	𝍸 ||||	9	36
AB	||||	4	16
Total		25	100%

المجموع ←

حساب النسبه المئويه للفئه = $\frac{\text{تكرار الفئة}}{\text{العدد الكلي للتكرار}} \times 100\%$

النسبه مئويه لزمرة الدم A

نسبه A $= \frac{5}{25} \times 100 = 20\%$

نسبه B $= \frac{7}{25} \times 100 = 28\%$

2-1 Categorical Frequency Distributions

Example 2-1 (Distribution of Blood Types) page 43

Class	Tally	Frequency	Percent
A	𝍸	5	20
B	𝍸 | |	7	28
O	𝍸 | | | |	9	36
AB	| | | |	4	16
Total		25	100%

$$\% = \frac{f}{n} \times 100 = \frac{5}{25} = .2 \times 100 = 20$$

For more information see page 43-44 of Bluman's Book

$\% = \frac{f}{n}$ ⇐ تكرار الفئة
⇐ عدد التجارب الكلي

سيتكون جدول التوزيع التكراري النوعي من الاعمدة

class	Tally	frequency	percent %
Total		n	100%

2-1 Categorical Frequency Distributions

Example 2.2

New restaurant wants to explore the impression of visitors about the customer service. Sample of 20 visitors selected randomly and their impression noted as below

Good	Excellent	Excellent	Excellent	Good
Bad	Good	Good	Excellent	Bad
Good	Excellent	Excellent	Good	Good
Bad	Good	Excellent	Excellent	Excellent

Construct a categorical frequency distribution?

Class	Tally	Frequency	Percent
Excellent	|||| |||	9	45
Good	|||| ||	8	40
Bad	|||	3	15
Total		20	100%

2-1 Categorical Frequency Distributions

Solution:

Class	Tally	Frequency	Percent
Excellent	|||| ||||	9	45
Good	|||| |||	8	40
Bad	|||	3	15
Total		20	100%

Solution:

Class	Tally	Frequency	Percent
Excellent	𝍸 ||||	9	45
Good	𝍸 |||	8	40
Bad	|||	3	15
Total		20	100%

2-1 Grouped Frequency Distributions

هداول التوزيعات التكرارية (مجموعات) مبوبة

2- Group frequency distributions are constructed for quantitative data. used when the range of values in a data set is large (continuous data). The data must be grouped into classes that are more than one unit in width, e.g., 58-64.

- The lower class limit represents the smallest data value that can be included in a class, e.g., 58 for the class limit 58-64.

- The upper class limit represents the largest value that can be included in the class, e.g., 64 for the class limit 58-64.

2-1Grouped Frequency Distributions

The class boundaries are used to separate the classes so that there are no gaps in the frequency distribution and they can be found by the following steps:

1- make sure the number of decimal digits are equal in the lower and upper limit if not add zero to the one that has less decimal digit. (if exist) e.g., 58-64.3 $\Longrightarrow$ 58.0-64.3

$58.00 - 0.05 = 57.95$ $\quad$ $64.30 + 0.05 = 64.35$

2- add zero after the last decimal digit in the lower and upper limit. e.g., 58.0-64.3 $\Longrightarrow$ 58.00-64.30

3- subtracting 5 from the last digit in the lower class limit and adding 5 to the last digit in the upper class limit. e.g., 58.00-64.30$\Longrightarrow$ 57.95-64.35

For more information see page 44-45 of Bluman's Book

2-1 Grouped Frequency Distributions

Steps For Constructing a Grouped Frequency Distributions

Step 1: Determine the number of classes k.

- Chose an appropriate number. Or
- Round output of Sturge's rule, $K = 1 + 3.33 \times log(n)$. Or
- Use the formula: $n < 2^k$.

Step 2: Determine the classes

- Find the (Range) by: finding the difference between the Highest value (H) and Lowest value (L) in the data set.

$$R = max - min = H - L.$$

- Determine the width of classes by: dividing the range(R) over the number of classes(K) and **round up**.

$$W = \frac{range}{\text{number of classes}} = \frac{R}{K}.$$

- Determine the starting point for the first class.

  By select the smallest number in the data or a convenient number that is smaller than the lowest data value.

2-1 Grouped Frequency Distributions

- Calculate the lower class limits. By repeatedly adding the class width to the prior class limit.
- Calculate the upper class limits. Subtract one unit from the lower limit of the second class to get the upper limit of the first class

Step 2: Find the tallies.

Step 3: Find the frequencies from the tallies or the percentages (relative frequencies).

2-1 Grouped Frequency Distributions

In constructing the class limit we must take into a count the following:

- The classes must be mutually exclusive.

age
10-20
20-30
30-40
40-50

10-19
20-29
30-39
40-49

- The class must be continuous. There should be no gaps in the frequency distribution.

age
10-20
21-31
43-53

- The class must be exhaustive. There should be enough classes to accommodate all the data.
- the class must be equal in width.

2-1Grouped Frequency Distributions

اشيء جدول توزيع تكراري حلوه هه ج صناق

Example 2-2:Record High Temperatures (page 47)

The following data represent the record high temperatures for each of the 50 states.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Construct a grouped frequency distribution for the data using 7 classes?

The procedure for constructing a grouped frequency distribution for numerical data follows.

فئة الحدود الفعلية تكرار مركز الفئة

Class	Class boundary	frequency	Class Mid point

- Calculate the (Range): R = Highest value (H) - Lowest value (L) = 134-100=34.
- Given the number of classes, the class width:

$$\text{Class Width} = \frac{\text{range}(=R)}{\text{number of classes}(=k)} = \frac{34}{7} = 4.9 \approx 5.$$

2-1 Grouped Frequency Distributions

Example 2-2: Record High Temperatures (page 47)

The following data represent the record high temperatures for each of the 50 states.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Construct a grouped frequency distribution for the data using 7 classes?

The procedure for constructing a grouped frequency distribution for numerical data follows.

① تحديد اعلى وادنى قيمة و مدى القيم

المدى :- Range = اعلى قيمة - اقل قيمة

$$Range = Max - Min$$

$Min = 100$ $\quad$ $Max = 134$

$$Range = 134 - 100 = 34$$

② تحديد عدد الفئات number of classes (K)

① يمكن ان يكون عدد الفئات قريب بالسؤال (5-15)

يجب ان لا يكون كبير جدا او صغير جدا

② اذ لم يقيد بالسؤال نستخدم القواعد التالية

$$K = 1 + 3.33 \log(n)$$

$$K = 1 + 3.33 \log(50) = 6.6 \approx 7$$

③ استخدم العلاقة

$n < 2^K$

$50 < 2^K$

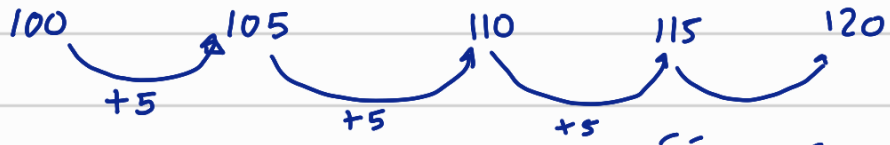
$50 < 2^7$ $\quad$ يجوز ان ناخذ

✓ $50 < 128$

③ عرض الفئة Class width = W

$$W = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{R}{K} = \frac{34}{7} = 4.9 \approx 5$$

④ نبدأ بتكوين الجدول وتحديد الفئات



والتعديل على الفئات حتى لا تكون متداخلة

- خرج ١ من كل فئة
- اجمع ١ للحد الأقل

classes	classes
100 - 105	100 - 104
105 - 110	105 - 109
110 - 115	110 - 114
115 - 120	115 - 119
120 - 125	120 - 124
125 - 130	125 - 129
130 - 135	130 - 134

⑤ نكون جدول التوزيع التكراري

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114.2	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Temperature (Class limits)	Tally	# of states (Frequency)
100-104	||	2
105-109	𝍸 |||	8
110-114	𝍸 𝍸 𝍸 |||	18
115-119	𝍸 𝍸 |||	13
120-124	𝍸 ||	7
125-129	|	1
130-134	|	1
Total		50

⑥ كتابة الحدود الفعلية (حدود خاصة للتخلص من الفجوات)

Class boundary

عدد صحيح

L — U

-0.5 +0.5

حدود الفئة 5 — 9 | 5 — 8.2

-0.5 +0.5

5.00 — 8.20

حدود فعلية 4.5 — 9.5 | -0.05 0.05 +

4.95 — 8.25

100 — 104

100 − 0.5 = 99.5 ← 100 — 104 → 104 + 0.5 = 104.5

Temperature (Class limits)	Class Boundaries	# of states (Frequency)
100-104	99.5 − 104.5	2
105-109	⋮	8
110-114	⋮	18
115-119	⋮	13
120-124	⋮	7
125-129	⋮	1
130-134		1
Total		50

Temperature	Class Boundaries
100-104	99.5-104.5
105-109	104.5-109.5
110-114	109.5-114.5
115-119	114.5-119.5
120-124	119.5-124.5
125-129	124.5-129.5
130-134	129.5-134.5

102

Mid Point مركز الفئة (7)

$$\text{Class Mid Point} = \frac{\text{lower limit} + \text{upper limit}}{2}$$

$$= \frac{100 + 104}{2} = 102$$

Temperature	Class Boundaries	Class Midpoints	(Frequency)
100-104	99.5-104.5	102	2
105-109	104.5-109.5	107	8
110-114	109.5-114.5	112	18
115-119	114.5-119.5	117	13
120-124	119.5-124.5	122	7
125-129	124.5-129.5	127	1
130-134	129.5-134.5	132	1
Total			50

Example 2-2:Record High Temperatures (page 47)

The following data represent the record high temperatures for each of the 50 states.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Construct a grouped frequency distribution for the data using 7 classes?

The procedure for constructing a grouped frequency distribution for numerical data follows.

- Calculate the (Range):

  R = Highest value (H)- Lowest value (L)=134-100=34.

- Given the number of classes, the class width:

$$\text{Class Width} = \frac{\text{range(=R)}}{\text{number of classes(=k)}} = \frac{34}{7} = 4.9 \approx 5.$$

2-1Grouped Frequency Distributions

- Determine the class limits (boundaries); a Lower limit, which is the **smallest** data value that can be included in the class, and an upper limit, which represents the **largest** data value that can be included in the class.

  - Lower limits we can consider the **lowest value** to be the starting point for the lowest class limits, i.e., use 100 as the lower limit of the first class then **repeatedly add the class width** to get the lower limit of the next six classes, i.e., 105, 110, 115, 120, 125, 130.
  - Upper limits **subtract one unit** from the lower limits of the second class until the seventh class to get the upper limit of the first class until the sixth class. Finally, use the **largest value** as the upper limit of the final class.
    i.e., 104,109, 114, 119, 124, 129.

2-1Grouped Frequency Distributions

- Step 1:The limits of the classes are:

$$\begin{aligned} L &= \underline{100 - 104} \\ & \quad 105 - 109 \\ l_3 &= \underline{110 - 114} = u_3 \\ & \quad 115 - 119 \\ & \quad 120 - 124 \\ & \quad 125 - 129 \\ & \quad 130 - 134 = H \end{aligned}$$

- Step 2: create a table with three columns, the first column is for Temperature (classes), the second column for counting, and the third column is for the number(#) of states (frequencies).
- Step 3: tally data then delete column 2.

2-1Grouped Frequency Distributions

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Temperature (Class limits)	Tally	# of states (Frequency)
100-104	||	
105-109	𝍸 |||	
110-114	𝍸 𝍸 𝍸 |||	
115-119	𝍸 𝍸 |||	
120-124	𝍸 ||	
125-129	|	
130-134	|	
Total		50

2-1Grouped Frequency Distributions

The class boundaries for the grouped frequency distribution. It is used so that there are no gaps in the frequency distribution for our example is given by

$$l_i - 0.5, u_i + 0.5$$

Temperature (Class limits)	Class Boundaries	# of states (Frequency)
100-104		2
105-109		8
110-114		18
115-119		13
120-124		7
125-129		1
130-134		1
Total		50

The class width for a class in a frequency distribution is found by one of the following:

- usually, grouped frequency distribution consist of equal class widths.
- Subtracting the lower (or upper) class limit of one class from the lower (or upper) class limit of the next class.
- Subtracting the lower (or upper) class boundary of one class from the lower (or upper) class boundary of the next class.
- Subtracting the lower class limit of one class from the upper class limit of the same class then add 1 to the last digit.
- subtracting the lower class boundary of one class from the upper class boundary of the same class.

Temperature	Class Boundaries
100-104	99.5-104.5
105-109	104.5-109.5
110-114	109.5-114.5
115-119	114.5-119.5
120-124	119.5-124.5
125-129	124.5-129.5
130-134	129.5-134.5

The class midpoint مرکز، نقطه

Sometime, we need to calculate the **midpoints** of each class. The class midpoint is found by adding the lower and upper limits then divide by 2 or adding the lower and upper boundaries then divide by 2.

$$\text{class midpoint} = \frac{\text{lower limit} + \text{upper limit}}{2}$$

$$\text{or class midpoint} = \frac{\text{lower boundary} + \text{upper boundary}}{2}$$

For our example:

$$\textbf{class \#}i\ \textbf{midpoint} = \frac{l_i + u_i}{2}$$

2-1Grouped Frequency Distributions

Temperature	Class Boundaries	Class Midpoints	(Frequency)
100-104	99.5-104.5	102	2
105-109	104.5-109.5	107	8
110-114	109.5-114.5	112	18
115-119	114.5-119.5	117	13
120-124	119.5-124.5	122	7
125-129	124.5-129.5	127	1
130-134	129.5-134.5	132	1
Total			50

2-1 Cumulative frequency distribution التوزيع التكراري المتجمع

A cumulative frequency distribution: is a distribution that shows the number of data values less than or equal to each upper boundary.

جميع القيم التي تساوي او اقل من الحد العلوي للفئة

The values are found by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class.

This gives an ascending cumulative frequency. The **last cumulative frequency must be equal to n**.

n = اخر عدد في جدول التكرار المتجمع

classes	f	Cum.f	
		0	Less than 6
0-4	2	2	Less than 4
5-9	3	5	Less than 9
10-14	6	11	Less than 14
	11		

Example 2-2 (page 49)

Temp. (Class)	Class boundaries	Frequency		Cumulative Frequency
			Less than 99.5	0
100 – 104	99.5 – 104.5	2	Less than 104.5	2
105 – 109	104.5 – 109.5	8	Less than 109.5	10
110 – 114	109.5 – 114.5	18	Less than 114.5	28
115 – 119	114.5 – 119.5	13	Less than 119.5	41
120 – 124	119.5 – 124.5	7	Less than 124.5	48
125 – 129	124.5 – 129.5	1	Less than 129.5	49
130 – 134	129.5 – 134.5	1	Less than 134.5	50

Example 2-2 (page 49)

Temp. (Class)	Class boundaries	Frequency		Cumulative Frequency
		Keep Adding!	Less than 99.5	0
100 – 104	99.5 – 104.5	2	Less than 104.5	2
105 – 109	104.5 – 109.5	8	Less than 109.5	10
110 – 114	109.5 – 114.5	18	Less than 114.5	28
115 – 119	114.5 – 119.5	13	Less than 119.5	41
120 – 124	119.5 – 124.5	7	Less than 124.5	48
125 – 129	124.5 – 129.5	1	Less than 129.5	49
130 – 134	129.5 – 134.5	1	Less than 134.5	$50 = n$

$P = \frac{f}{n} \times 100$

0

$\frac{2}{50} \times 100 = 4$

- If the sample size and the (cumulative) percentage of a specific class are given, then the corresponding (cumulative) frequency of this class can be calculated by

قانون حساب التكرار المجمع في هذا الصف بمعلومية النسبة

$$f_i = \frac{P_i}{100} \times n$$

$P = \frac{f}{n} \times 100\%$

- If both the (cumulative) frequency and the (cumulative) percentage of a specific class are given, then the sample size can be calculated by

قانون لحساب حجم العينة n

$$n = \frac{f_i}{P_i} \times 100$$

لنسبة التكرار المجمع

Temperature	Class Boundaries		Cum. Freq.	Percent
		Less than 99.5	0	
100-104	99.5-104.5	Less than 104.5	2	4
105-109	104.5-109.5	Less than 109.5	10	20
110-114	109.5-114.5	Less than 114.5	28	56
115-119	114.5-119.5	Less than 119.5	41	82
120-124	119.5-124.5	Less than 124.5	48	96
125-129	124.5-129.5	Less than 129.5	49	98
130-134	129.5-134.5	Less than 134.5	50	100

لتحديد حجم العينة

$f = 2 \quad P = 4$

$n = \frac{2}{4} \times 100$

$= 50$

اختر صفا يحتوي التكرار النسبي P_i يجب أن يكون 100

Relative Frequency Distribution

Relative Frequency Distribution

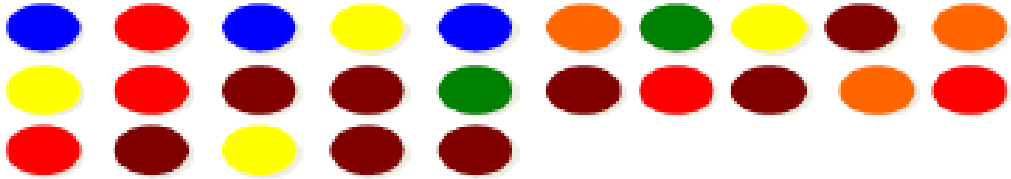
Used to compare datasets of different sizes.

1.Sum the total number of observations across all classes of the frequency distribution.

2.Divide the frequency for each class by the total number of obs.

Examples A bag of colored balls contains 25 balls

Raw Data:



Statistical Table:

Relative frequency

$$R = \frac{f}{n}$$

$$P = \frac{f}{n} \times 100$$

Color	Tally	f	Cum.f	R.f = f/n
Red		5	5	5/25 = 0.2
Blue		3	8	3/25 = 0.12
Green		2	16	
Orange		3	13	
Brown		8	21	
Yellow		4	25	
		25		

Relative Frequency Distribution

Relative Frequency Distribution

Color	Frequency	Relative Frequency
Red	5	5/25 = .20
Blue	3	3/25 = .12
Green	2	2/25 = .08
Orange	3	3/25 = .12
Brown	8	8/25 = .32
Yellow	4	4/25 = .16
Sum	25	1

Cumulative Frequency Distribution

Color	Frequency	Cumulative Frequency
Red	5	5
Blue	3	8
Green	2	10
Orange	3	13
Brown	8	21
Yellow	4	25

2-1 Ungrouped frequency distributions

توزيعات تكرارية (بدون فئات)

class	qas
red	0-4
blue	5-9

Ungrouped frequency distributions - can be used for data that can be enumerated and when the range of values in the data set is not large. لحين استخدامه عندما يكون مدى البيانات غير كبير

Examples - number of miles your instructors have to travel from home to campus, number of girls in a 4-child family etc.

Class	Frequency
5 km	24
10 km	16
15 km	10

Ungrouped frequency distributions

Example 2-3: Hours of Sleep (page 49)

The data shown represent the number of hours 30 college students said they sleep per night. Construct and analyze a frequency distribution.

8	6	6	8	5	7
7	8	7	6	6	7
9	7	7	6	8	10
6	7	6	7	8	7
7	8	7	8	9	8

Range = 10 − 5 = 5

Classes	f	Class boundary
5	1	4.5 − 5.5
6	7	5.5 − 6.5
7	11	6.5 − 7.5
8	8	7.5 − 8.5
9	2	8.5 − 9.5
10	1	9.5 − 10.5

Solution:

Step 1: the Range (R)= *Highestvalue*(*H*) – *Lowestvalue*(*L*)

Step 2: Tally the data.

Ungrouped frequency distributions

Example 2-3: Hours of Sleep (page 49)

Class limits	Tally	Frequency
5	|	1
6	𝍸 ||	7
7	𝍸 𝍸 |	11
8	𝍸 |||	8
9	||	2
10	|	1

Step 3: From the class boundaries.

Class limits	Class boundaries	Frequency
5	4.5 – 5.5	1
6		7
7		11
8		8
9		2
10		1

Ungrouped frequency distributions

Example 2-3: Hours of Sleep (page 49)

Step 4: The cumulative frequencies are.

Class limits	Class boundaries	Frequency		Cumulative frequency
			Less than 4.5	
5	4.5-5.5	1	Less than 5.5	1
6	5.5-6.5	7	Less than 6.5	8
7	6.5-7.5	11	Less than 7.5	19
8	7.5-8.5	8	Less than 8.5	27
9	8.5-9.5	2	Less than 9.5	29
10	9.5-10.5	1	Less than 10.5	30

Ungrouped frequency distributions

Example 2-3: Hours of Sleep (page 49)

Class limits	Class boundaries	Frequency		Cumulative frequency
			Less than 4.5	0
5	4.5-5.5	1	Less than 5.5	1
6	5.5-6.5	7	Less than 6.5	8
7	6.5-7.5	11	Less than 7.5	19
8	7.5-8.5	8	Less than 8.5	27
9	8.5-9.5	2	Less than 9.5	29
10	9.5-10.5	1	Less than 10.5	30

see page 49-50 of Bluman's Book

frequency distributions

	Nominal	Ordinal	Discrete	Continuous
Categorical frequency distribution	✓	✓		class: red, blue
Grouped frequency distribution		class: 0-4, 5-9		✓ (large range)
Ungrouped frequency distribution		class: 1, 2, 3	✓	✓ (small range)

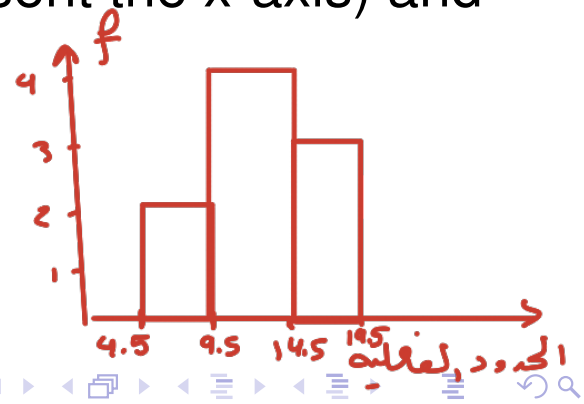
22 Histograms

One of the most Common Graphs in Research that summarize the data

The histogram displays the continuous data that are organized in a grouped frequency distribution by using vertical bars of various heights to represent the frequencies.

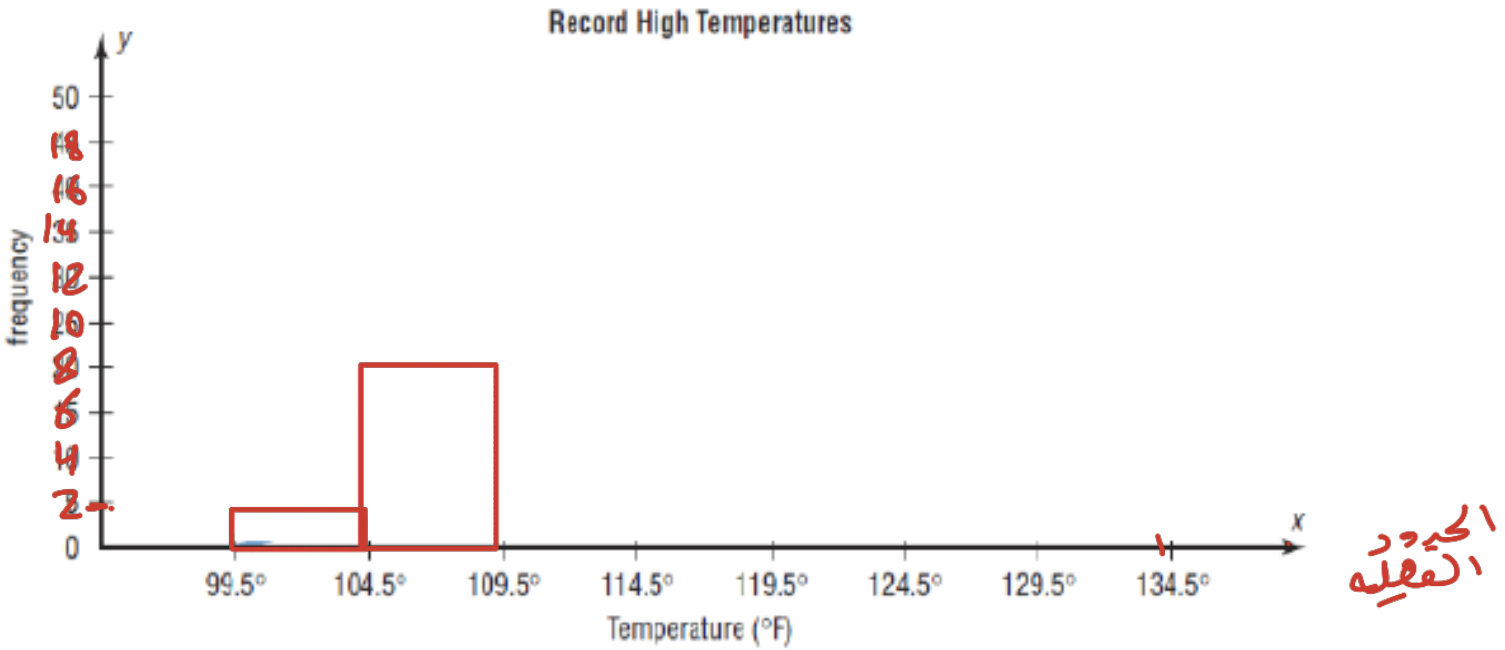
To construct a histogram, we need two columns from the frequency distribution of the data, class boundaries (represent the x-axis) and the frequency (represent y-axis).

class boundary	f
4.5 – 9.5	2
9.5 – 14.5	4
14.5 – 19.5	3



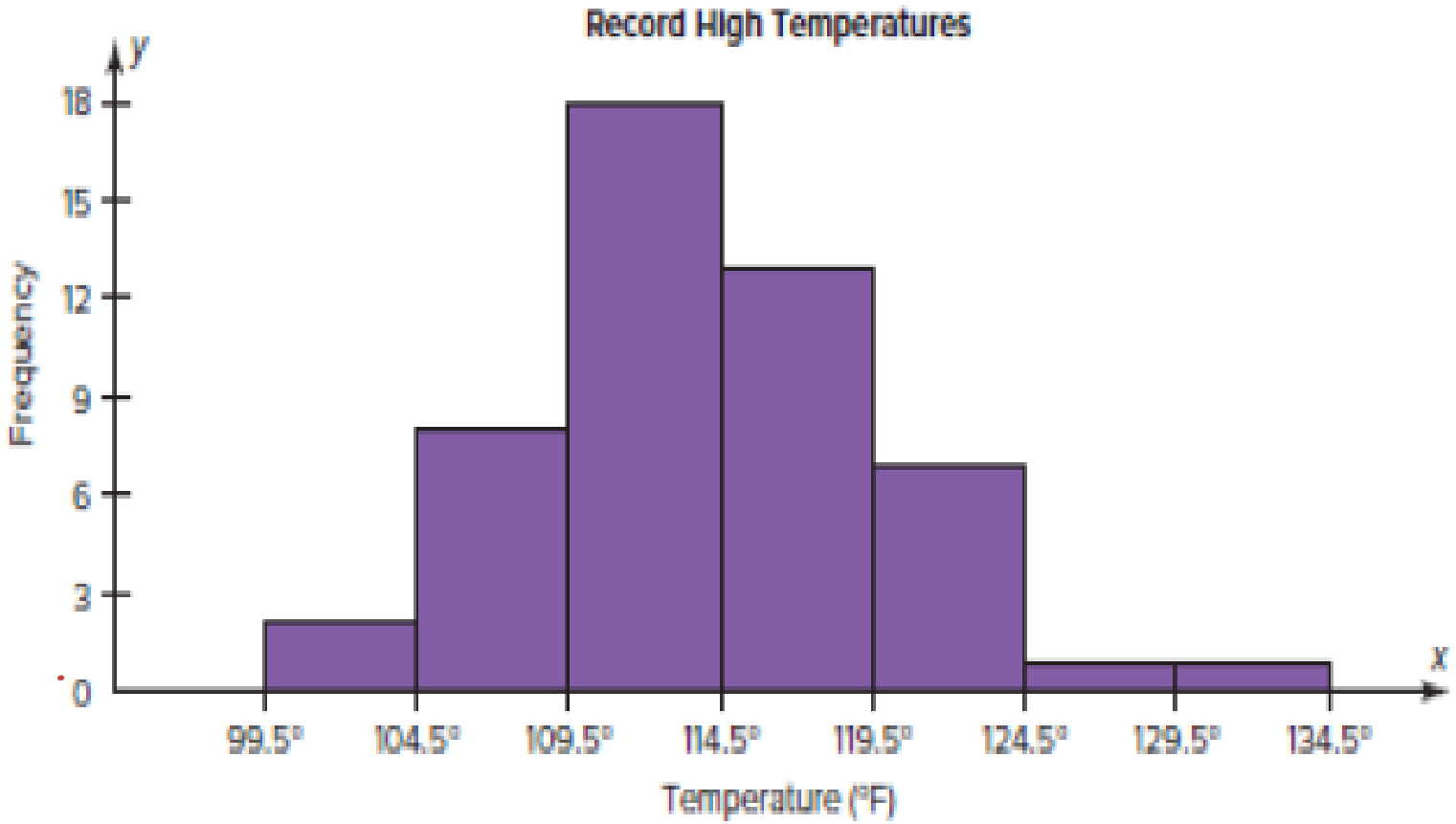
EXAMPLE 2-4 Record High Temperatures Page (57-58)

Class Boundaries	Frequency
99.5-104.5	2
104.5-109.5	8
109.5-114.5	18
114.5-119.5	13
119.5-124.5	7
124.5-129.5	1
129.5-134.5	1



EXAMPLE 2-4 Record High Temperatures Page (57-58)

FIGURE 2–1 Histogram for Example 2–4

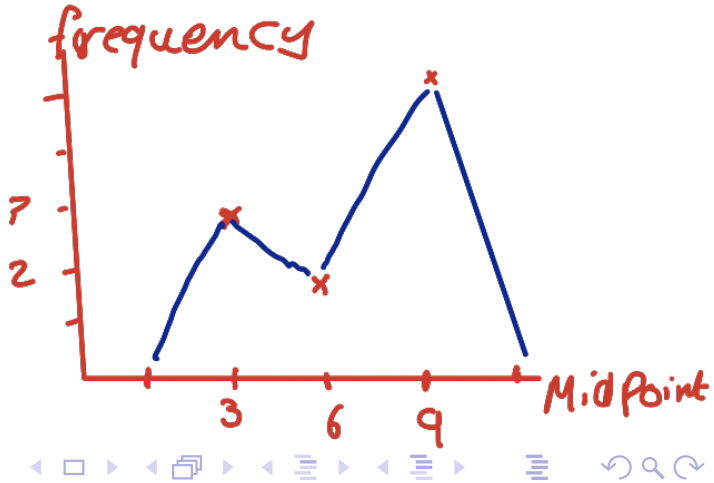


2-2 The Frequency Polygon مضلع تكراري

The frequency polygon: خط يوصل النقاط
Is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.

To construct a frequency polygon, we need to find a new column of class midpoints and use it instead of the class boundaries in the x-axis and the frequency in y- axis.

Midpoint	freq
3	3
6	2
9	5

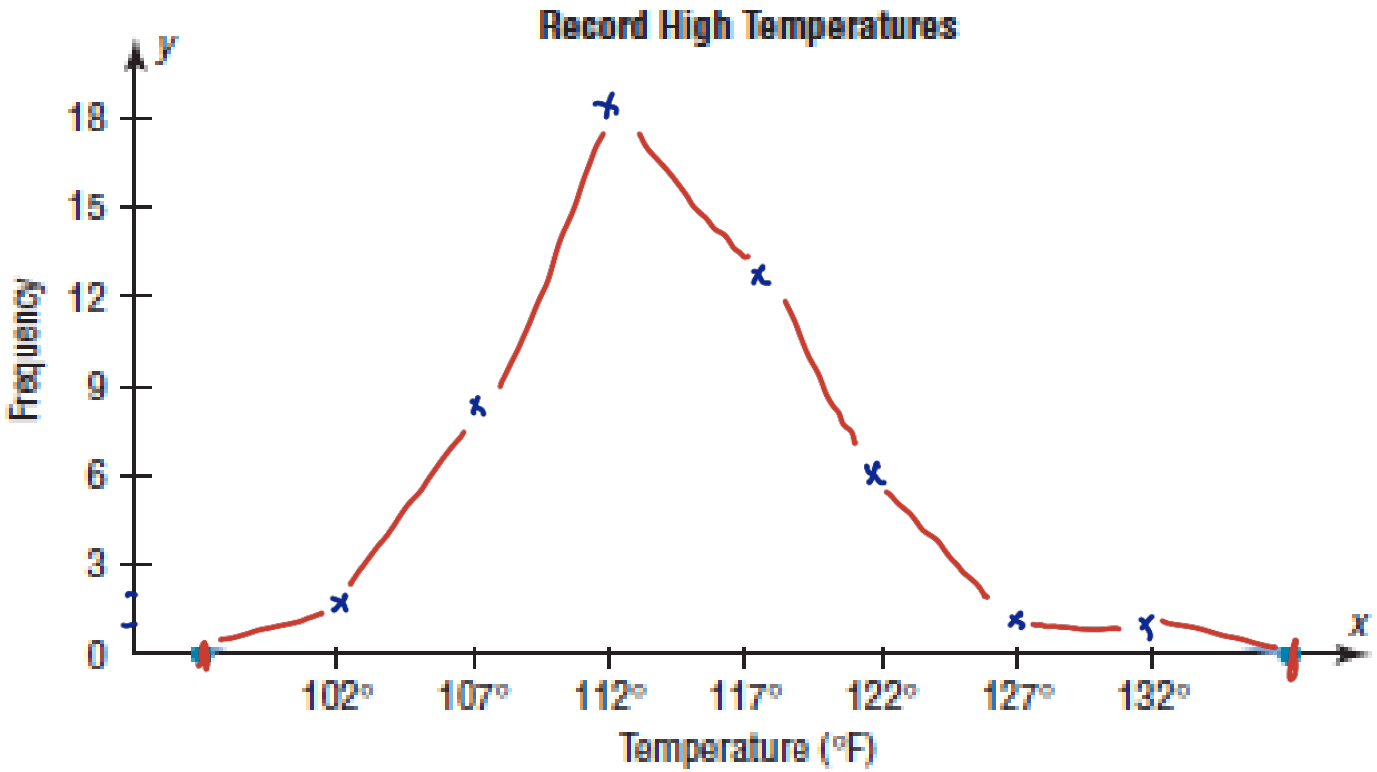


The Frequency Polygon

EXAMPLE 25 Record High Temperatures Page (58-59)

Temperature	Class Midpoints	(Frequency)
100-104	102	2
105-109	107	8
110-114	112	18
115-119	117	13
120-124	122	7
125-129	127	1
130-134	132	1
Total		50

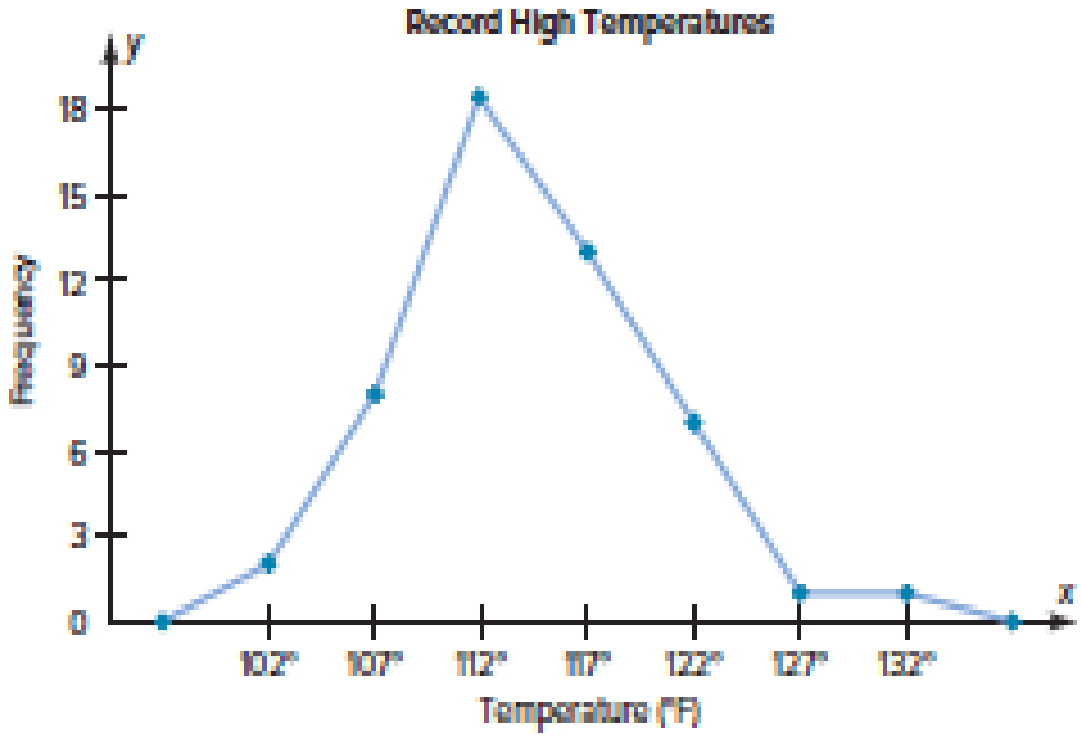
Midpoint $\frac{100+104}{2} = 102$



The Frequency Polygon

EXAMPLE 25 Record High Temperatures Page (58-59)

FIGURE 2–2
Frequency Polygon for Example 2–5



2-2 The Ogive

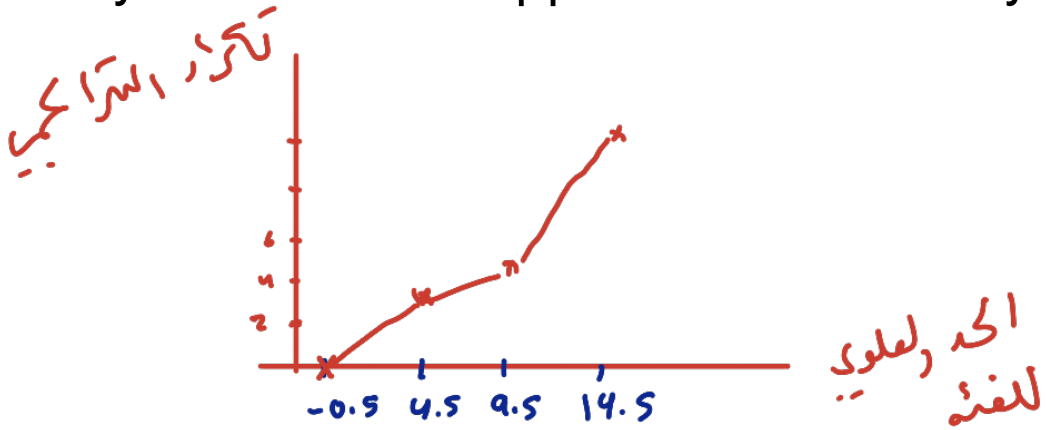
مضلع تراكمي

class	class boundar	f	acu.f	
			0	less than -0.5
0 4	-0.5 4.5	3	3	less than 4.5
5-9	4.5-9.5	2	5	less than 9.5
10-14	9.5-14.5	5	10	less than 14.5

Ogive: Serves as visual presentation of cumulative frequency. Displays the continuous data that are organized in a grouped frequency distribution.

How to construct a Ogive for the data:

To construct a Ogive, we need to use cumulative distribution for the y-axis and the upper class boundary for the x- axis.



Example 2-6:Record High Temperatures (page 59)

Construct an ogive for the frequency distribution described in Example 2-4.

Temperature		Cum. Freq.
	Less than 99.5	0
100-104	Less than 104.5	2
105-109	Less than 109.5	10
110-114	Less than 114.5	28
115-119	Less than 119.5	41
120-124	Less than 124.5	48
125-129	Less than 129.5	49
130-134	Less than 134.5	50

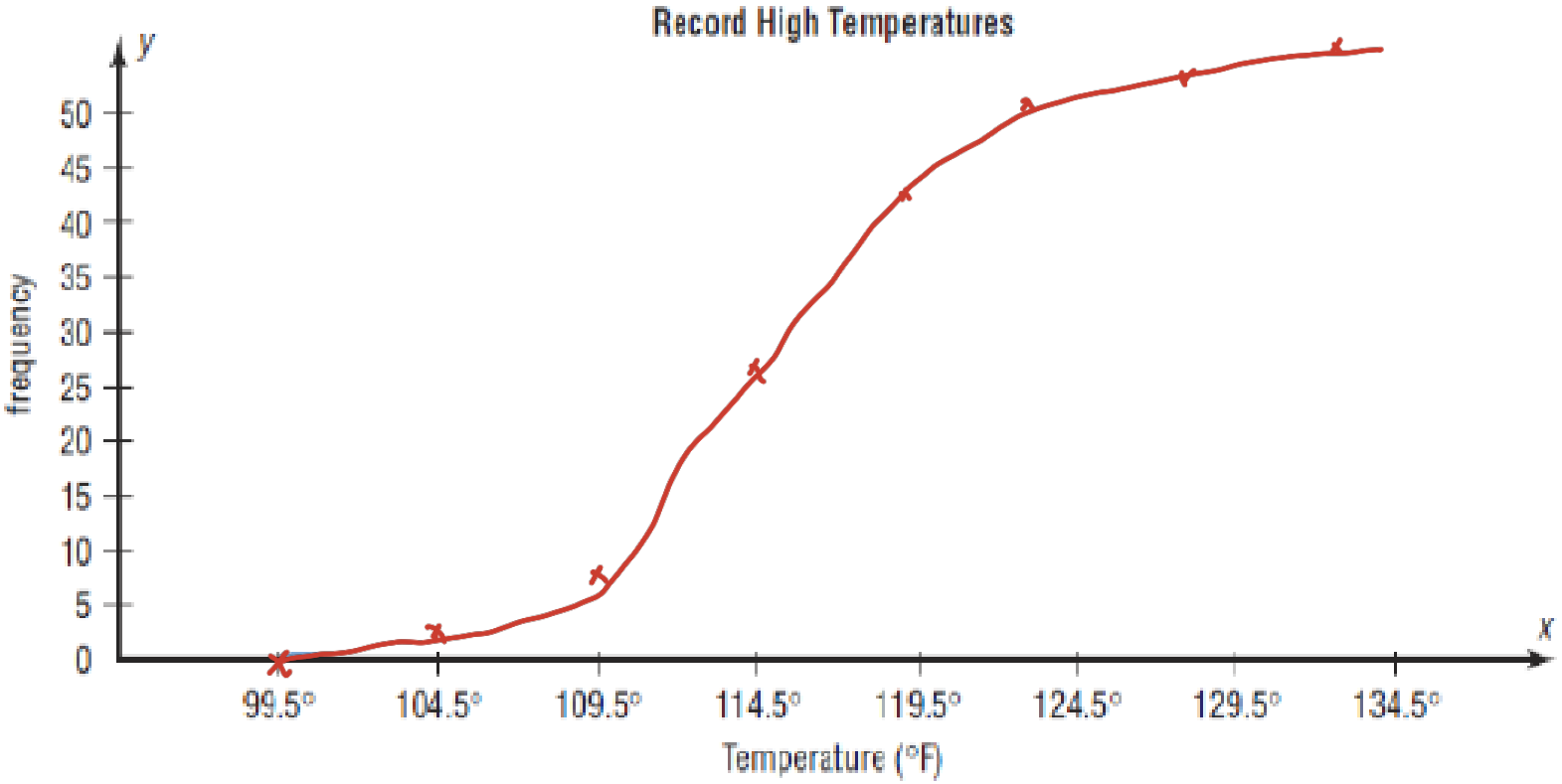


FIGURE 2–3
Plotting the Cumulative Frequency for Example 2–6

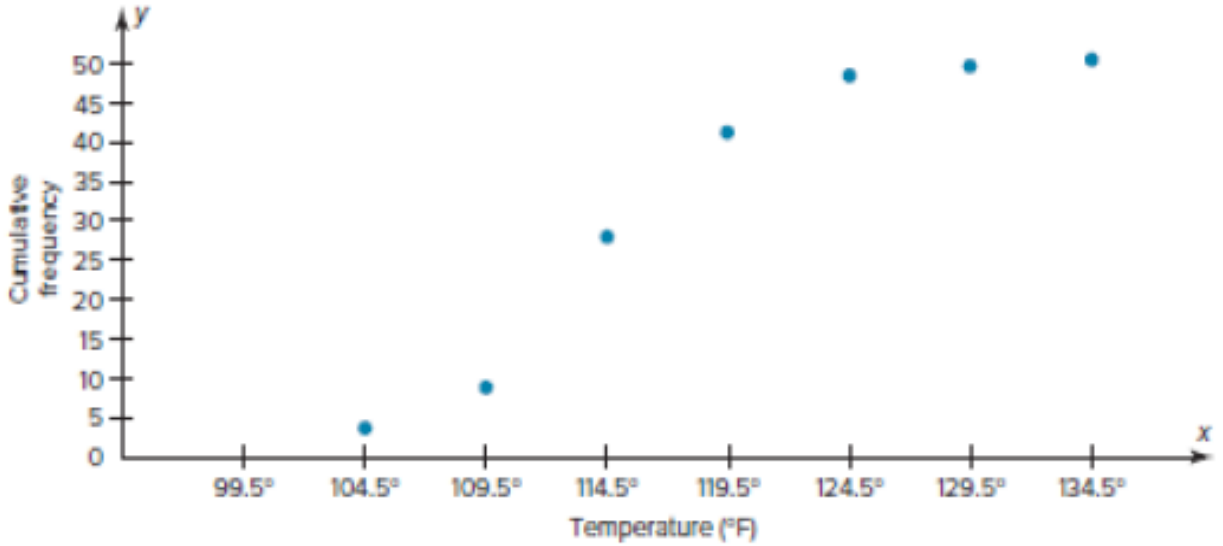
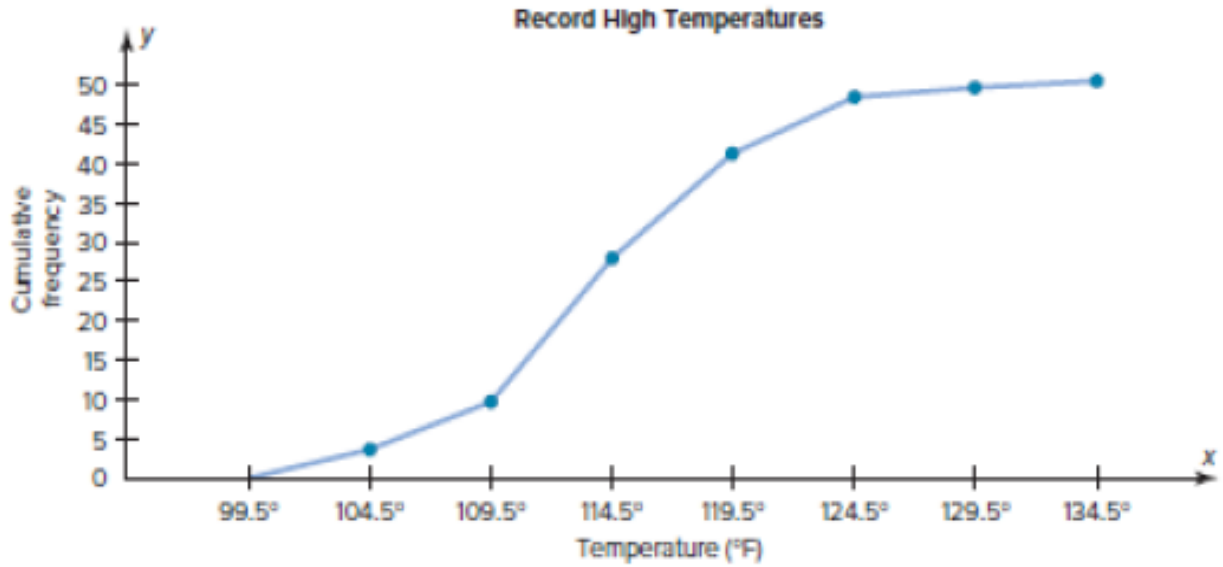


FIGURE 2–4
Ogive for Example 2–6



For more information see page 59-60 of Bluman's Book

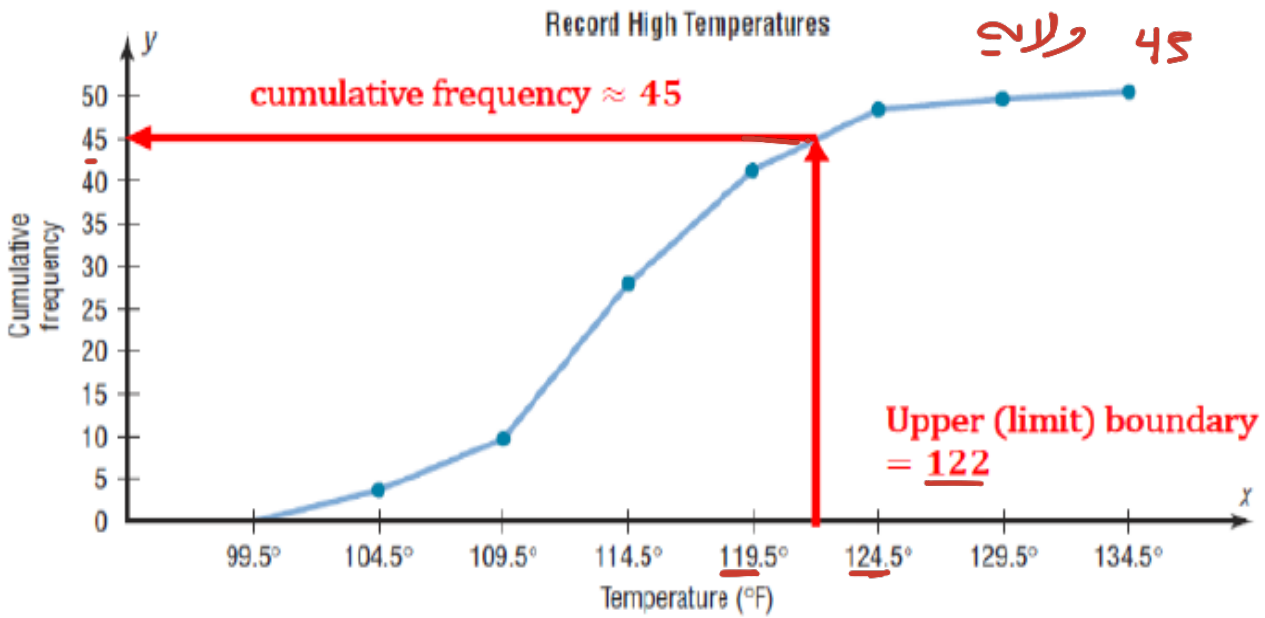
2-2 The Ogive

The main purpose of using ogive

The ogive is mainly used to visually represent how many cumulative frequency (percentage) are approximately below a certain upper class boundary and vice versa.

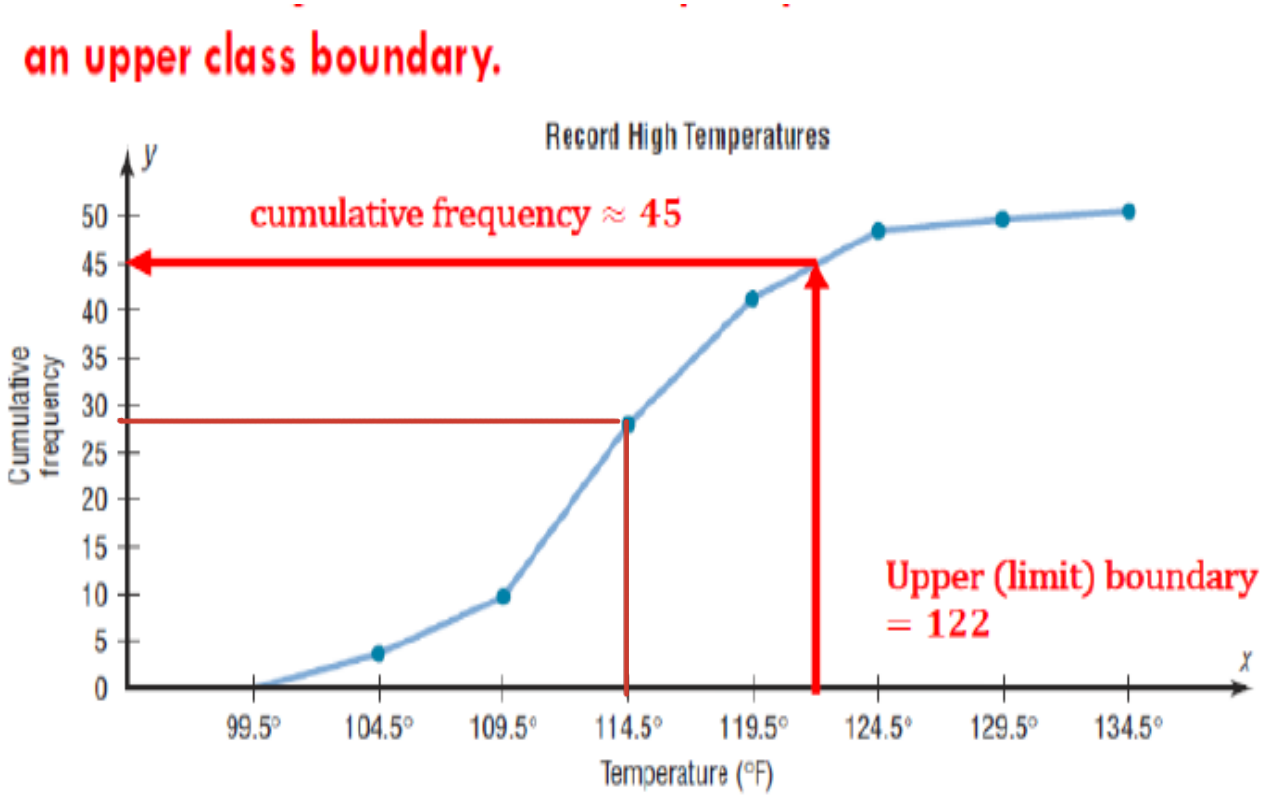
Case1: Getting the cumulative frequency based on an upper class boundary.

an upper class boundary.



2-2 The Ogive

How many state reached high temperatures of 114.5^o F and less?

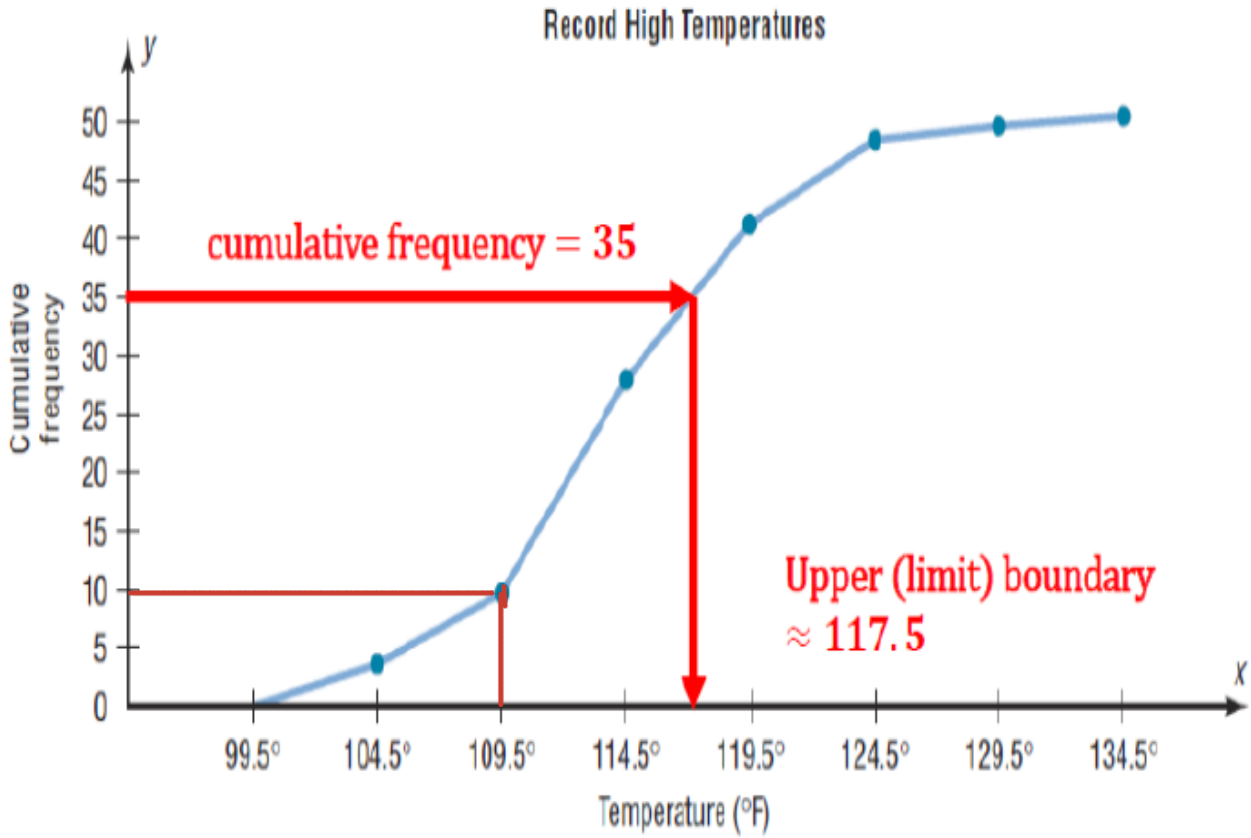


2-2 The Ogive

Case 2: Getting the upper class boundary based on a cumulative frequency.

What is the maximum temperatures that at most 10 states reached ?

ما اعلى درجه حرارة سجلت على الاقل ١٥ ولايات

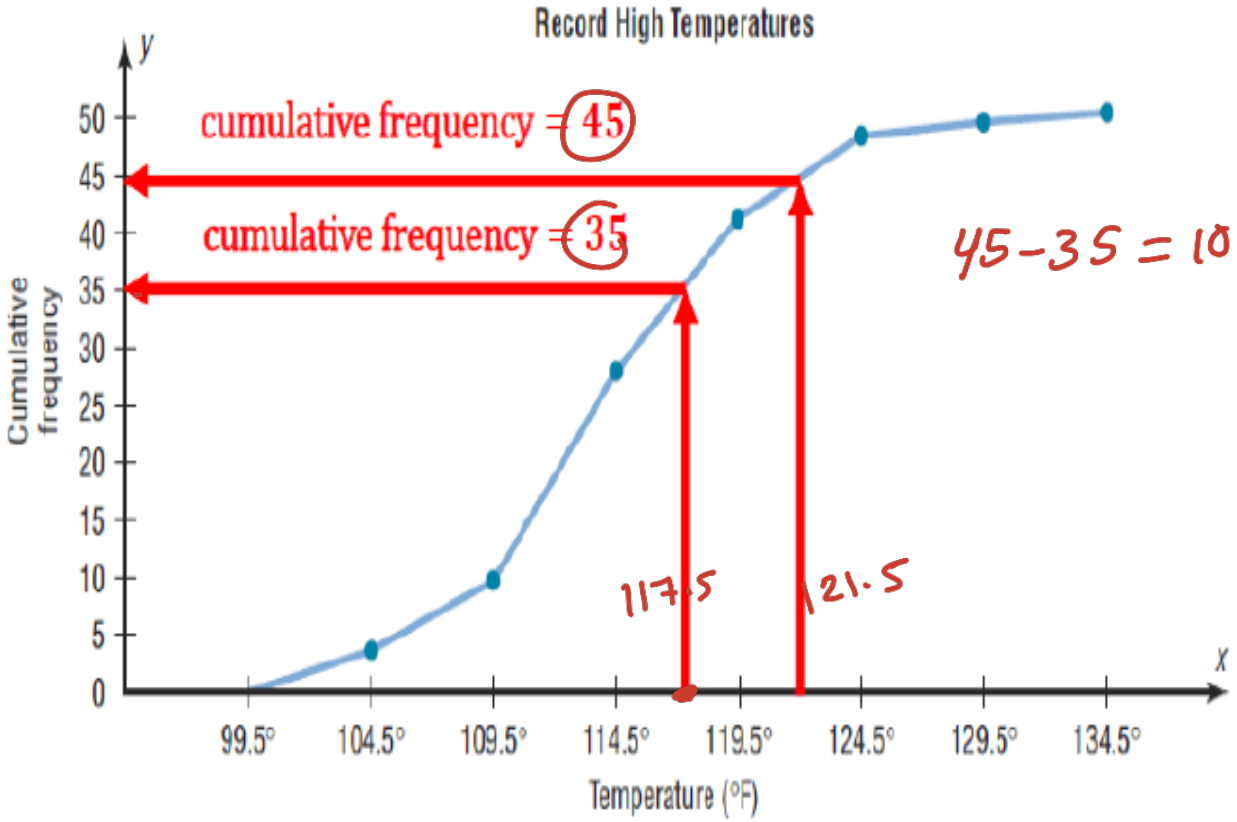


2-2 The Ogive

Case 3: Getting the cumulative frequencies based on two upper class boundaries.

How many state reached a maximum temperatures between 117.5^o F and 121.5^o ?

حاحد الولايات التي درجه الحراره فيها
بين 117.5 و 121.5



Required cumulative frequency = 45 − 35 = 10

Example(IQ scores)

The following data set represents the IQ scores for 40 students selected randomly from Umm Al-Qura university.

60	110	121	101	91	105	93
101	103	105	75	78	90	69
85	107	99	82	95	71	83
92	90	90	115	94	100	102
138	95	80	85	84	70	77
91	99	64	80	96		

Construct a grouped frequency distribution using 8 classes?

Solution:

Step 1: the Range (R)= $Highestvalue(H) - Lowestvalue(L)$

$$R = 138 - 60 = 78$$

$$\text{width}=R/K, \quad W = 78/8 \approx 9.7 \text{ (round up) } W = 10.$$

Example IQ scores (continued)
Step 2: Find the tallies.

Step 3: Find the frequencies from the tallies.

Class limits	Tally	Frequency
60-69	𝍸	3
70-79	𝍸	5
80-89	𝍸 ‖	7
90-99	𝍸 𝍸 ‖‖	13
100-109	𝍸 ‖‖	8
110-119	‖	2
120-129	|	1
130-139	|	1
Total		40

2-1Constructing class boundaries from class limits

Class boundaries: are connected classes that established from the disconnected class limits.

Use class limits in IQ scores example to construct class boundaries ?

Class limits	Class Boundaries	Frequency
60-69		3
70-79		5
80-89		7
90-99		13
100-109		8
110-119		2
120-129		1
130-139		1
Total		25

2-1Cumulative frequency distribution

Example IQ scores (continued):

Find the cumulative frequency distribution of the data ?

A cumulative frequency distribution: can be stablished by finding the sum of frequency for values less than each upper class boundary.

Class limits	Class Boundaries	Frequency	Less then	Cumulative Frequencies
			Less then 59.5	
60-69	59.5-69.5	3	Less then 69.5	
70-79	69.5-79.5	5	Less then 79.5	
80-89	79.5-89.5	7	Less then 89.5	
90-99	89.5-99.5	13	Less then 99.5	
100-109	99.5-109.5	8	Less then 109.5	
110-119	109.5-119.5	2	Less then 119.5	
120-129	119.5-129.5	1	Less then 129.5	
130-139	129.5-139.5	1	Less then 139.5	

2-1Cumulative frequency distribution

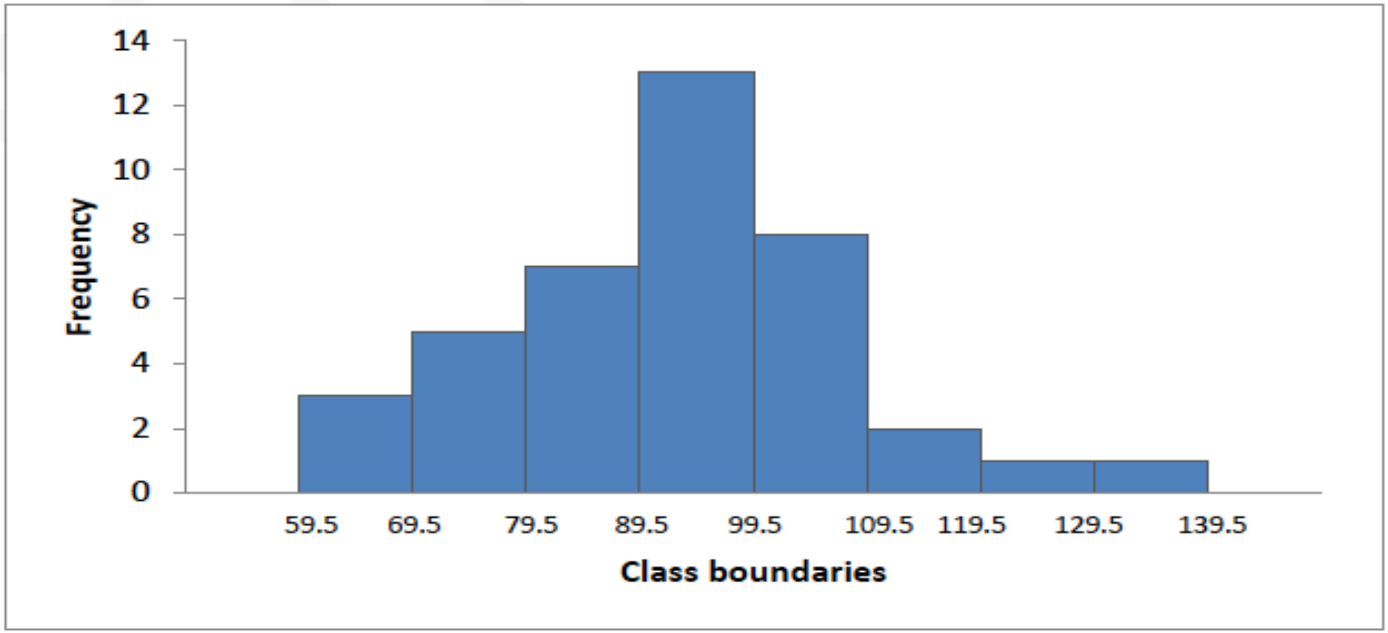
Example IQ scores (continued):

Solution:

Class limits	Class Boun.	Frequency	Less then	Cum. Freq.
			Less then 59.5	0
60-69	59.5-69.5	3	Less then 69.5	3
70-79	69.5-79.5	5	Less then 79.5	8
80-89	79.5-89.5	7	Less then 89.5	15
90-99	89.5-99.5	13	Less then 99.5	28
100-109	99.5-109.5	8	Less then 109.5	36
110-119	109.5-119.5	2	Less then 119.5	38
120-129	119.5-129.5	1	Less then 129.5	39
130-139	129.5-139.5	1	Less then 139.5	40

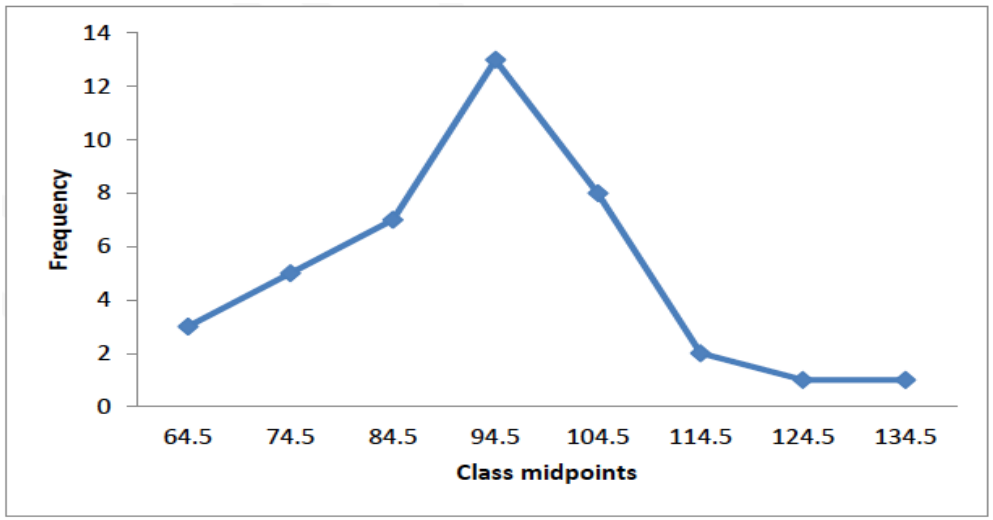
Example Construct a histogram to represent the data for the IQ scores.

Class Boundaries	Frequency
59.5-69.5	3
69.5-79.5	5
79.5-89.5	7
89.5-99.5	13
99.5-109.5	8
109.5-119.5	2
119.5-129.5	1
129.5-139.5	1



Example Construct a frequency polygon, to represent the data for the IQ scores example.

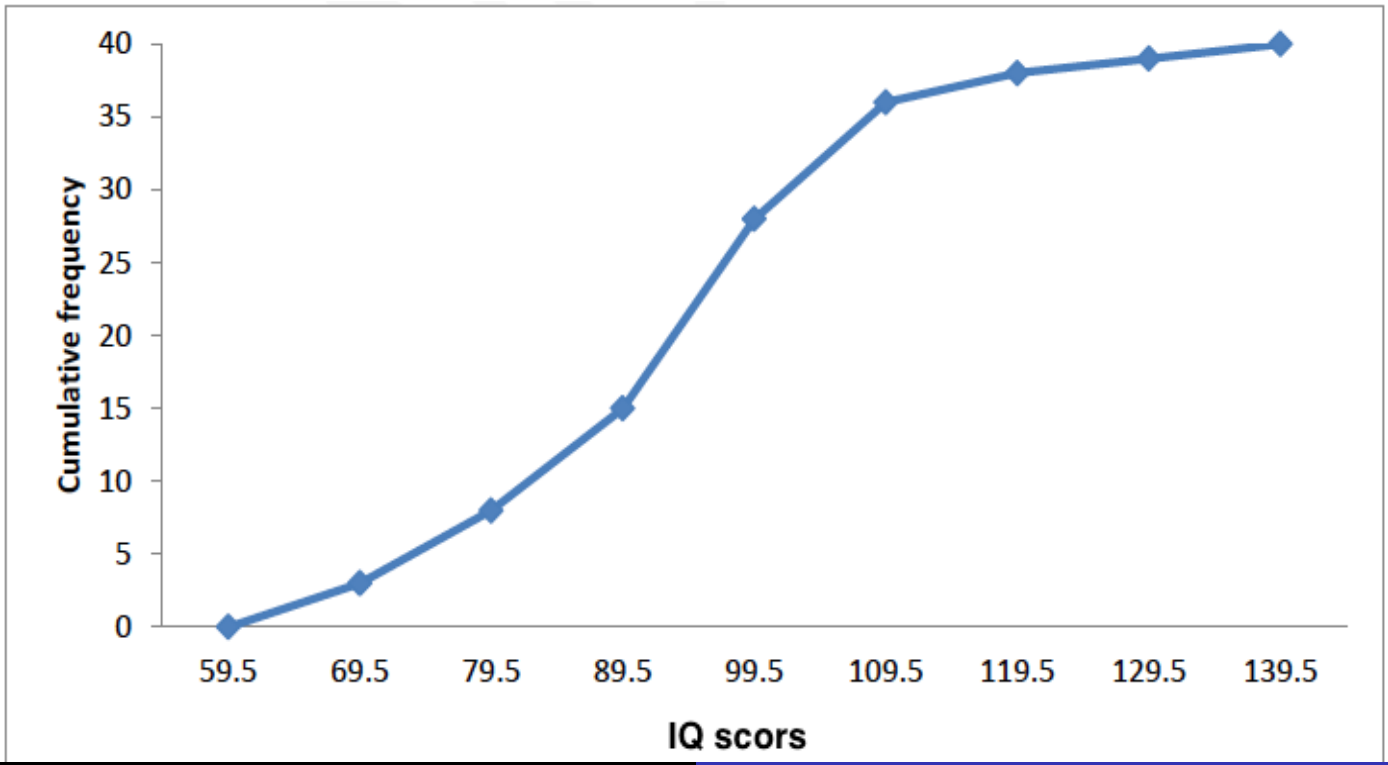
Class Boundaries	Midpoints	Frequency
59.5-69.5	64.5	3
69.5-79.5	74.5	5
79.5-89.5	84.5	7
89.5-99.5	94.5	13
99.5-109.5	104.5	8
109.5-119.5	114.5	2
119.5-129.5	124.5	1
129.5-139.5	134.5	1



Example IQ scores (continued):

Construct a ogive, to represent the data for the IQ scores .

Less then	Cumulative Frequencies
Less then 59.5	0
Less then 69.5	3
Less then 79.5	8
Less then 89.5	15
Less then 99.5	28
Less then 109.5	36
Less then 119.5	38
Less then 129.5	39
Less then 139.5	40



2-3 Other Types of Graphs (page 74-75)

1-Bar Graph

The bar chart displays the data by using vertical bars of various heights to represent the frequencies of discrete or categorical variables.

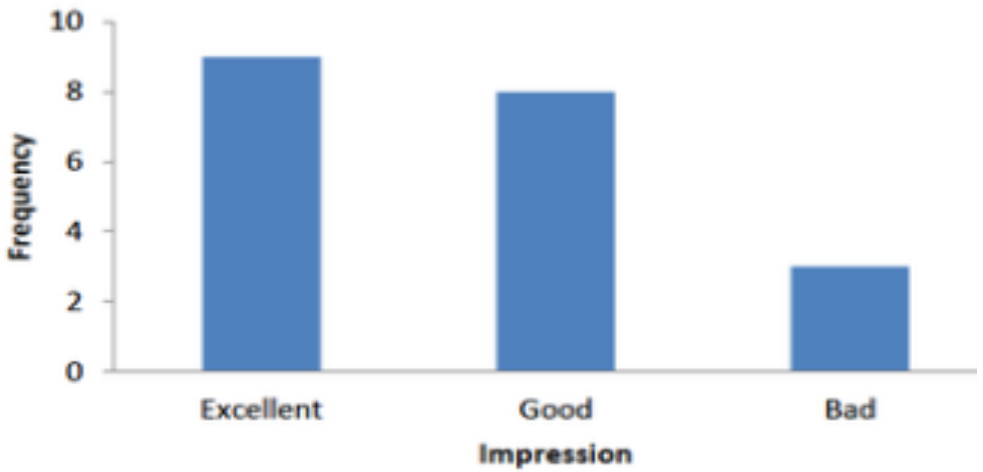
Example : Draw a bar graph to represent the impression of the visitors about the new sandwich.

Class	Frequency
Excellent	9
Good	8
Bad	3
Total	20

2-3 Other Types of Graphs (page 74-75)

1-Bar Graph

Example : Draw a bar graph to represent the impression of the visitors about the new sandwich.



simple

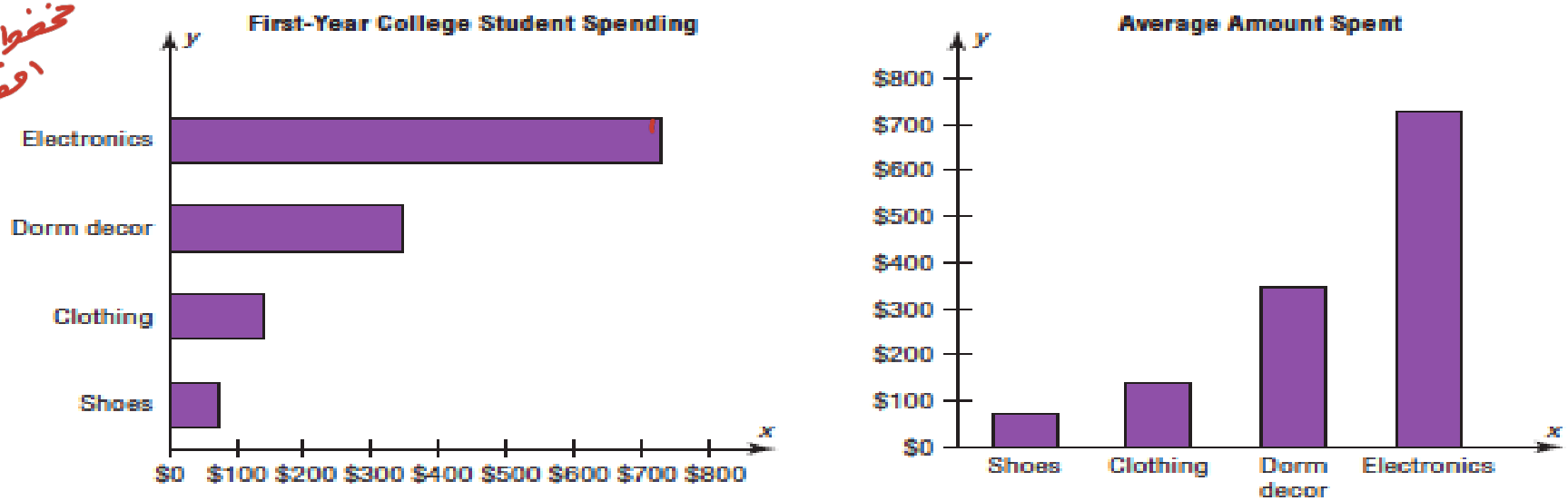
Example 2-8 page 76 College Spending for First-Year Students

The table shows the average money spent by first-year college students. Draw a horizontal and vertical bar graph for the data.

Electronics	$728
Dorm decor	344
Clothing	141
Shoes	72

Source: The National Retail Federation.

FIGURE 2–9 Bar Graphs for Example 2–8



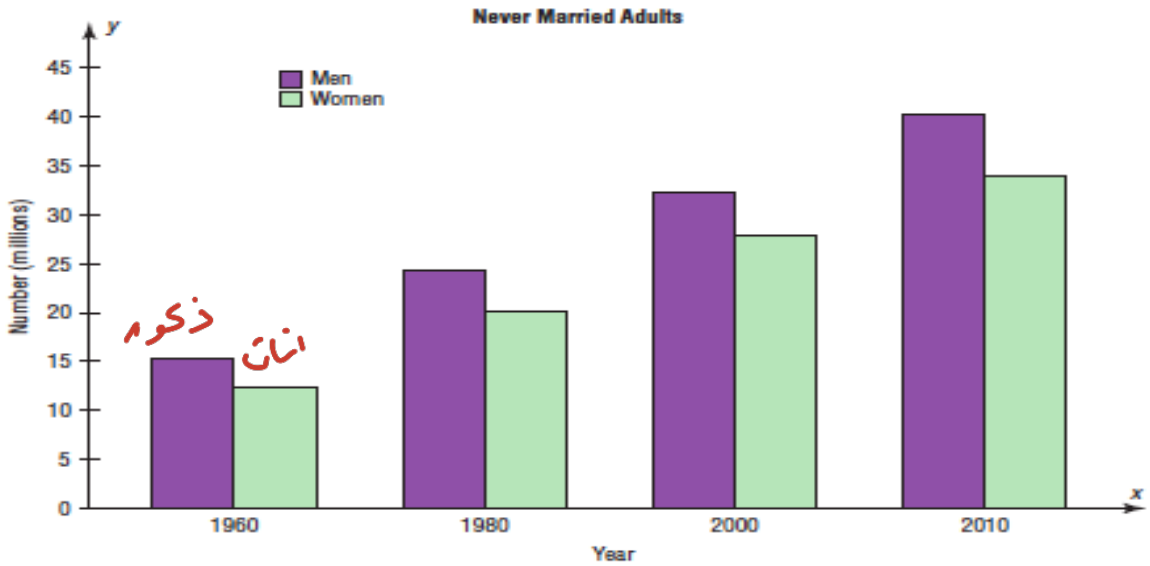
The graphs show that first-year college students spend the most on electronic equipment.

Bar graphs

The bar graphs can also be used to compare data for two or more groups. These types of bargraphs are called compound bar graphs. Consider the following data for the number (in millions) of never married adults in the United States.

Year	Males	Females
1960	15.3	12.3
1980	24.2	20.2
2000	32.3	27.8
2010	40.2	34.0

FIGURE 2–10
Example of a Compound Bar Graph



2-3 Other Types of Graphs (page 74-75)

2-Pareto chart

A pareto chart is used to represent a frequency distribution for categorical variable. The frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

2-Pareto chart

Example 2-9 page 77 Traffic Congestion

الساعات اللي بيضيعها الزحام بالزحمة

The data shown consist of the average number of hours that a commuter spends in traffic congestion per year in each city. Draw and analyze a Pareto chart for the data.

City	Hours
Atlanta	52
Boston	64
Chicago	61
New York	74
Washington,D.C.	82

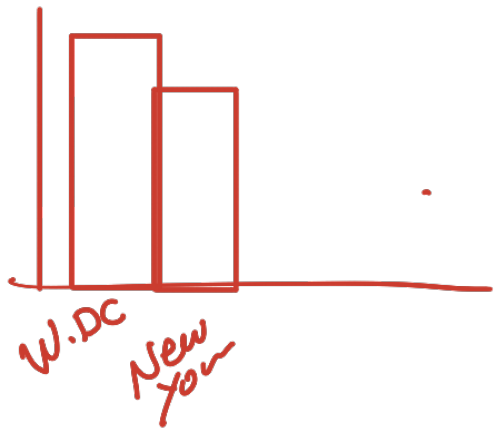


FIGURE 2–11 Pareto Chart for Example 2–9

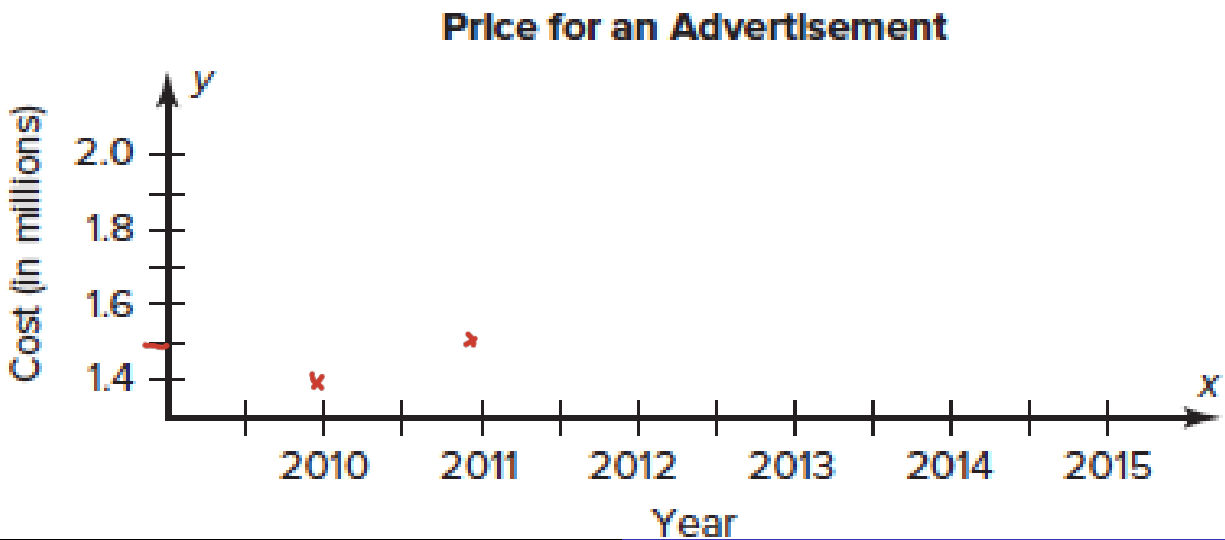
3-Time Series Graph

A time series graph represents data that occur over a specific period of time.

Example 2-10 page 78 Price of an Advertisement for the Academy Awards Show

The data show the average cost (in millions of dollars) of a 30-second television ad on the Academy Awards show. Draw and analyze a time series graph for the data.

Year	2010	2011	2012	2013	2014	2015
Cost	1.40	1.55	1.61	1.65	1.78	1.90



3-Time Series Graph

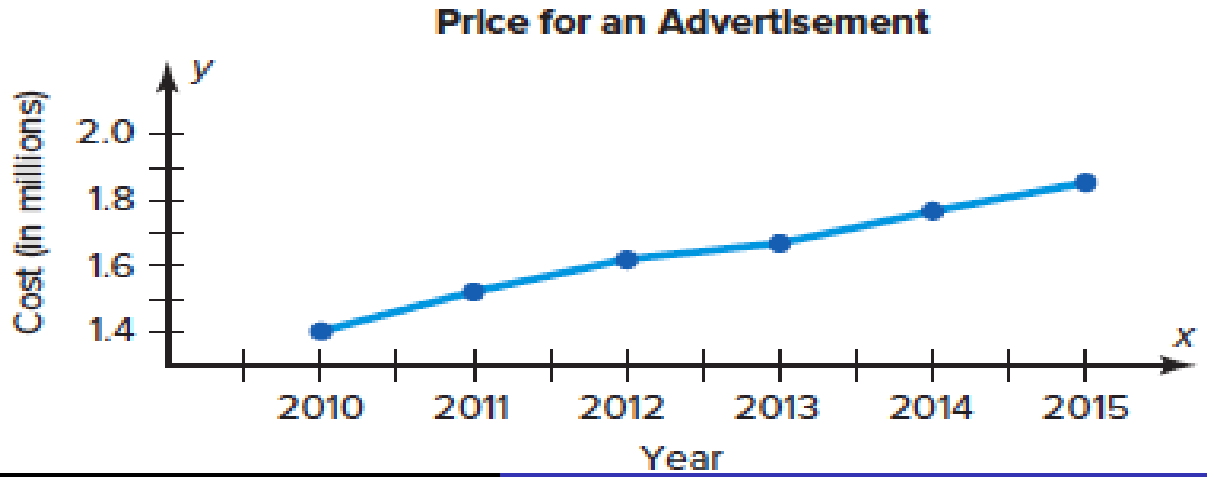
A time series graph represents data that occur over a specific period of time.

Example 2-10 page 78 Price of an Advertisement for the Academy Awards Show

The data show the average cost (in millions of dollars) of a 30-second television ad on the Academy Awards show. Draw and analyze a time series graph for the data.

Year	2010	2011	2012	2013	2014	2015
Cost	1.40	1.55	1.61	1.65	1.78	1.90

FIGURE 2–12 Figure for Example 2–10



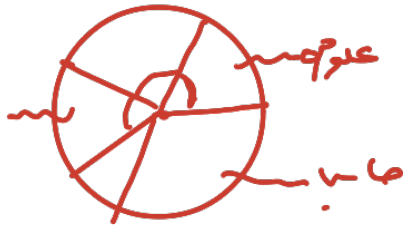
4-Pie Graph مخطط القطاع او الدائري

Pie Graph is a circle that is divided into sections according to the percentage of frequencies in each **category** of the distribution.

The degree of a class in a pie chart is defined as

$$\frac{f}{n} \times 360^o$$

درجة الزاوية لكل فئة



Example 2-11 page 80 Super Bowl Snack Foods

Snack	Pounds (frequency)	
Potato chips	11.2 million	$\frac{11.2}{30} \cdot 360^\circ = 134^\circ$
Tortilla chips	8.2 million	$\frac{8.2}{30} \cdot 360^\circ = 98^\circ$
Pretzels	4.3 million	$\frac{4.3}{30} \cdot 360^\circ = 52^\circ$
Popcorn	3.8 million	$\frac{3.8}{30} \cdot 360^\circ = 46^\circ$
Snack nuts	2.5 million	$\frac{2.5}{30} \cdot 360^\circ = 30^\circ$
	Total $n = 30.0$ million	Total 360°

$\frac{11.2}{30} \times 360$

4-Pie Graph

Example 2-11 page 80 Super Bowl Snack Foods

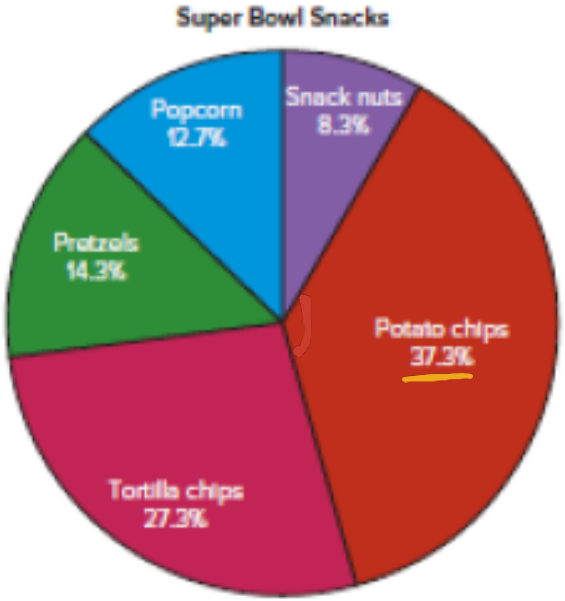
This frequency distribution shows the number of pounds of each snack food eaten during the Super Bowl. Construct a pie graph for the data.

Snack	Pounds (frequency)
Potato chips	11.2 million
Tortilla chips	8.2 million
Pretzels	4.3 million
Popcorn	3.8 million
Snack nuts	2.5 million
	Total $n = 30.0$ million

Source: USA TODAY Weekend

See page 80-81

FIGURE 2–14 Pie Graph for Example 2–11



134

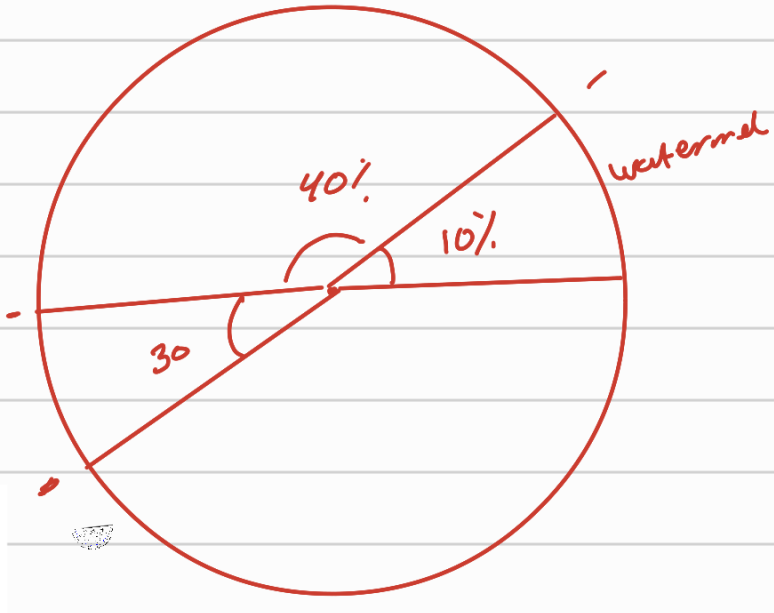
Pie Graph

class	frequency	angle	P = f/n x100
Watermelons	5	$\frac{5}{50} \times 360 = 36$	10%
Pears	20	$\frac{20}{50} \times 360 = 144$	40%
Apples	15	$\frac{15}{50} \times 360 = 108$	30%
Oranges	10	$\frac{10}{50} \times 360 = 72$	20%
	n = 50	36	

الزاوية = التكرار / مجموع التكرارات x360

$angle = \frac{f}{n} \times 360$

180 90 45



5-Stem-and-Leaf Plot

Stem and Leaf plot: is a data plot that uses part of a data value as the stem, the most significant digit (i.e. the tens), and the other part of the data value as the leaf, the less significant digits (the units), to form groups or classes.

It has the advantage over grouped frequency distribution of retaining the actual data while showing them in a graphic form.

19 18 11 25 26

Stem-and-Leaf Plot: how to construct the plot:

- Spilt each observation into two parts consisting of a stem and leaf.
- The number of leaves recorded opposite each stem is summarized under the frequency column .

s	l
1	1 8 9
2	5 6

Example 2-14 page 84 Out Patient Cardiograms

At an outpatient testing center, the number of cardiograms performed each day for 20 days is shown. Construct a stem and leaf plot for the data?

25	31	20	32	13
14	43	02	57	23
36	32	33	32	44
32	52	44	51	45

Stem	Leaf
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

stem	leaf
0	2
1	3 4
2	0 3 5
3	1 2
4	3

Stem	Leaf
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

The stem and leaf plot is similar to a horizontally flipped histogram.
By joining the stem and leaves, we notice that the Minimum is 02, while the Maximum is 57.
Consider the following numbers 1403 and 102.

- The stem of 1403 is 140 and the leaf is 3.
- The stem of 102 is 10 and the leaf is 2.

Advantage of Stem-and-Leaf Diagrams

Once a frequency distribution or histogram of continuous data is created, the raw data is lost (unless reported with the frequency distribution), however, the raw data can be retrieved from the stem-and-leaf plot.

	Nominal	Ordinal	Discrete	Continuous
Bar Chart	✓	✓	✓	
Time Series			**Independent variable:** time **Dependent variable:** discrete or continuous	
Pie Chart	✓	✓		
Stem-and-leaf plot			✓	✓

Home work

Exercises 2-1: page 51-52 (from 5-12, 14,16,17).
Exercises 2-2: page 67 (11)
Exercises 2-3: page 91 (11)
Chapter Quiz: page 105-106 (from 8-11, from 12-17)

مثال صفحة 50

اكتب جدول توزيع تكراري يحتوي على 8 فئات

① H = highest value $\quad R = H - L$

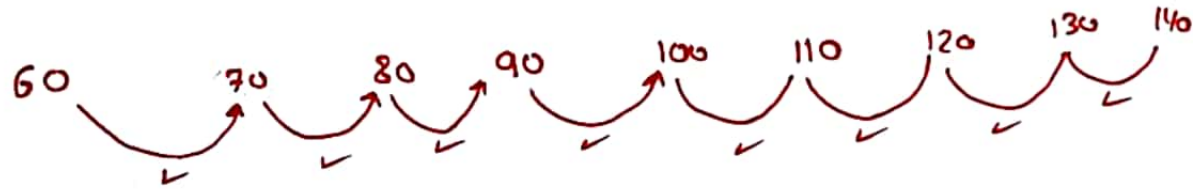
L = Lowest value

$H = 138 \qquad L = 60 \qquad R = 138 - 60 = 78$

② Width (عرض الفئة)

$$W = \frac{R}{K} = \frac{78}{8} = 9.7 \approx 10$$

③



Class limit	Tally	frequency	Class boundaries	Xm
60 - 69	///	3	59.5 - 69.5	64.5
70 - 79	////	5	69.5 - 79.5	74.5
80 - 89	//// //	7	79.5 - 89.5	84.5
90 - 99	//// //// ///	13	89.5 - 99.5	94.5
100 - 109	//// ///	8	99.5 - 109.5	104.5
110 - 119	//	2	109.5 - 119.5	114.5
120 - 129	/	1	119.5 - 129.5	124.5
130 - 139	/	1	129.5 - 139.5	134.5
Total		40		

find the class boundaries, Class midpoint cummulative frequency distribution and represent the Data in histogram, polygon, ogive

$$X_m = \frac{60+69}{2} = 64.5$$

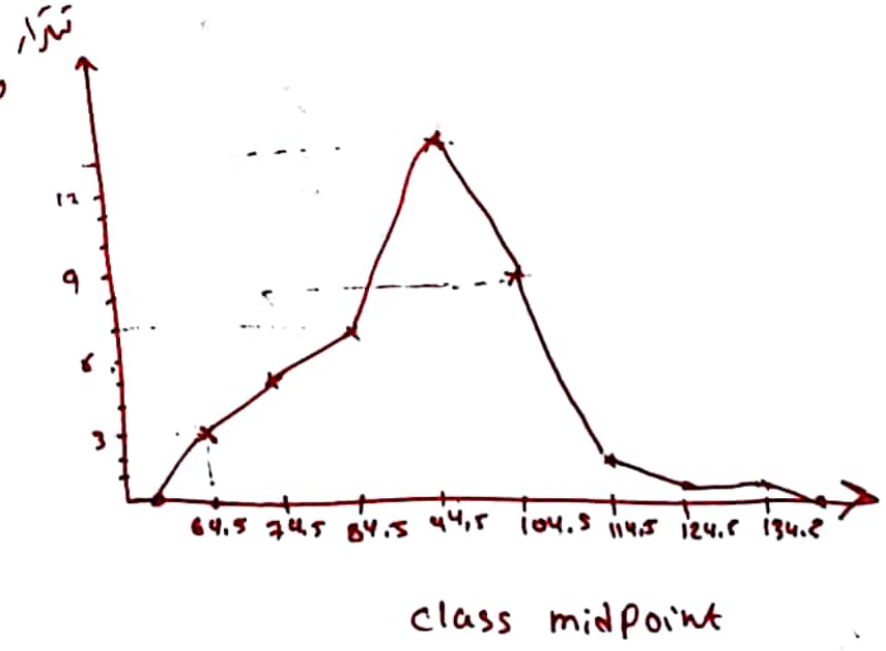
freq	class boundy	less than 59.5 = 0	Cum - freq
3	59.5 - 69.5	less than 69.5	3
5	69.5 - 79.5	⟨ ⟨ 79.5	8
7	79.5 - 89.5	⟨ ⟨ 89.5	15
13	89.5 - 99.5	⟨ ⟨ 99.5	28
8	99.5 - 109.5	⟨ ⟨ 109.5	36
2	109.5 - 119.5	⟨ ⟨ 119.5	38
1	119.5 - 129.5	⟨ ⟨ 129.5	39
1	129.5 - 139.5	⟨ ⟨ 139.5	40

histogram

class boundary	frequency

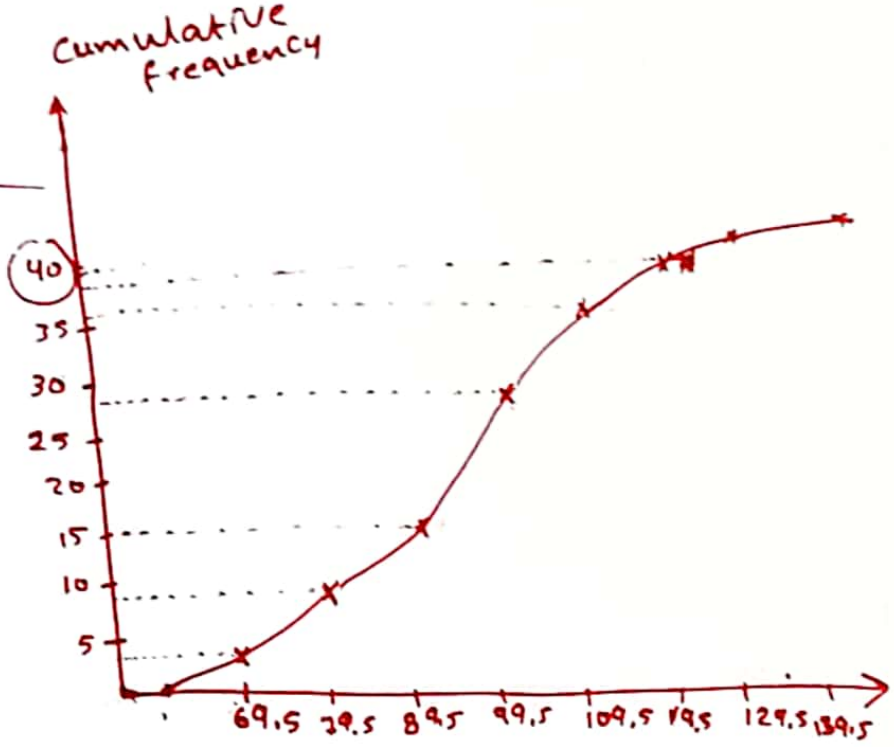
Polygon

Xm	frequency
64.5	3
74.5	5
84.5	7
94.5	13
104.5	8
114.5	2
124.5	1
134.5	1



Ogive

upper limit	Cum - Freq
69.5	3
79.5	8
89.5	15
99.5	28
109.5	36
119.5	38
129.5	39
139.5	40



2-1 Grouped Frequency Distributions

Example IQ scores (continued)
Step 2: Find the tallies.

Step 3: Find the frequencies from the tallies.

Class limits	Tally	Frequency
60-69	𝍸	3
70-79	𝍸	5
80-89	𝍸 ‖	7
90-99	𝍸 𝍸 ‖‖	13
100-109	𝍸 ‖‖	8
110-119	‖	2
120-129	|	1
130-139	|	1
Total		40

2-1 Constructing class boundaries from class limits

Class boundaries: are connected classes that established from the disconnected class limits.

Use class limits in IQ scores example to construct class boundaries ?

Class limits	Class Boundaries	Frequency
60-69		3
70-79		5
80-89		7
90-99		13
100-109		8
110-119		2
120-129		1
130-139		1
Total		25

2-1 Cumulative frequency distribution

Example IQ scores (continued):

Find the cumulative frequency distribution of the data ?

A cumulative frequency distribution: can be stablished by finding the sum of frequency for values less than each upper class boundary.

Class limits	Class Boundaries	Frequency	Less then	Cumulative Frequencies
			Less then 59.5	
60-69	59.5-69.5	3	Less then 69.5	
70-79	69.5-79.5	5	Less then 79.5	
80-89	79.5-89.5	7	Less then 89.5	
90-99	89.5-99.5	13	Less then 99.5	
100-109	99.5-109.5	8	Less then 109.5	
110-119	109.5-119.5	2	Less then 119.5	
120-129	119.5-129.5	1	Less then 129.5	
130-139	129.5-139.5	1	Less then 139.5	

2-1 Cumulative frequency distribution

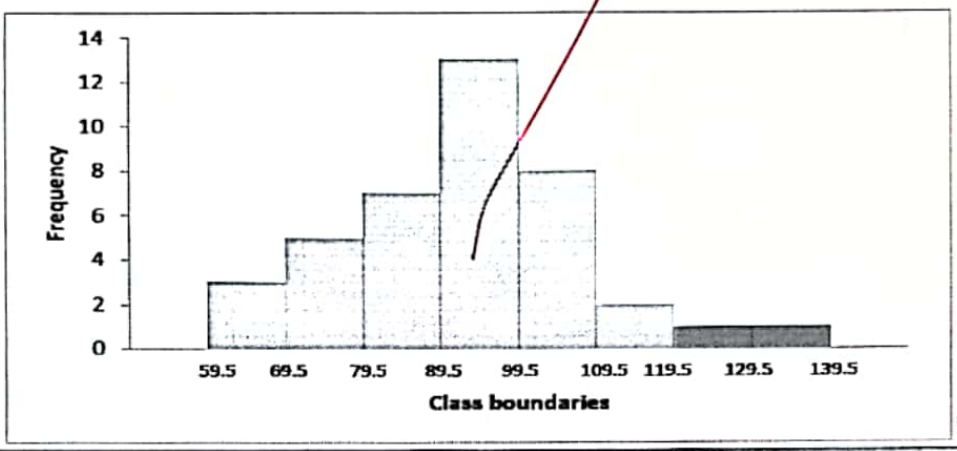
Example IQ scores (continued):

Solution:

Class limits	Class Boun.	Frequency	Less then	Cum. Freq.
			Less then 59.5	0
60-69	59.5-69.5	3	Less then 69.5	3
70-79	69.5-79.5	5	Less then 79.5	8
80-89	79.5-89.5	7	Less then 89.5	15
90-99	89.5-99.5	13	Less then 99.5	28
100-109	99.5-109.5	8	Less then 109.5	36
110-119	109.5-119.5	2	Less then 119.5	38
120-129	119.5-129.5	1	Less then 129.5	39
130-139	129.5-139.5	1	Less then 139.5	40

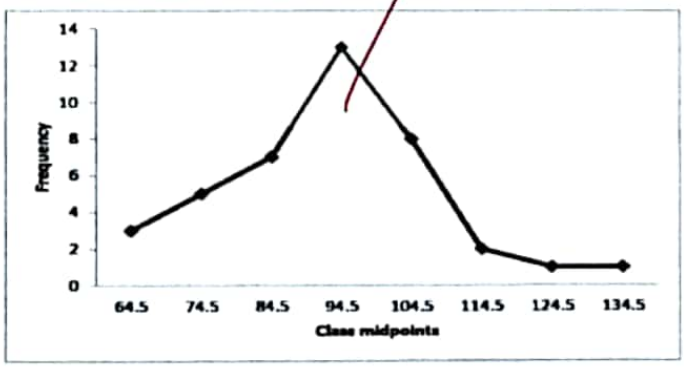
Example Construct a histogram to represent the data for the IQ scores.

Class Boundaries	Frequency
59.5-69.5	3
69.5-79.5	5
79.5-89.5	7
89.5-99.5	13
99.5-109.5	8
109.5-119.5	2
119.5-129.5	1
129.5-139.5	1



Example Construct a frequency polygon, to represent the data for the IQ scores example.

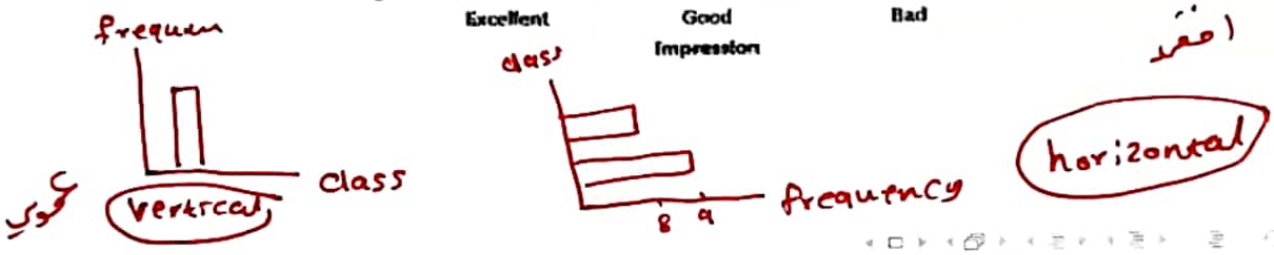
Class Boundaries	Midpoints	Frequency
59.5-69.5	64.5	3
69.5-79.5	74.5	5
79.5-89.5	84.5	7
89.5-99.5	94.5	13
99.5-109.5	104.5	8
109.5-119.5	114.5	2
119.5-129.5	124.5	1
129.5-139.5	134.5	1



2-3 Other Types of Graphs (page 74-75)

1-Bar Graph

Example : Draw a bar graph to represent the impression of the visitors about the new sandwich.

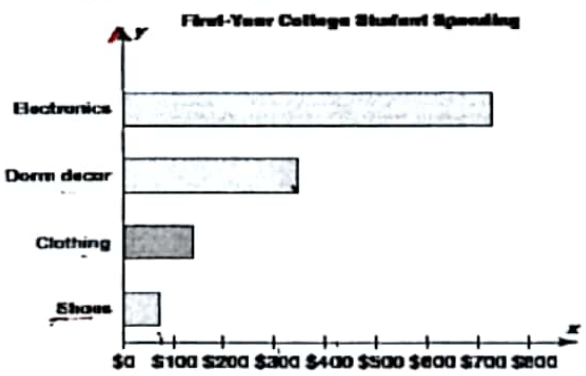


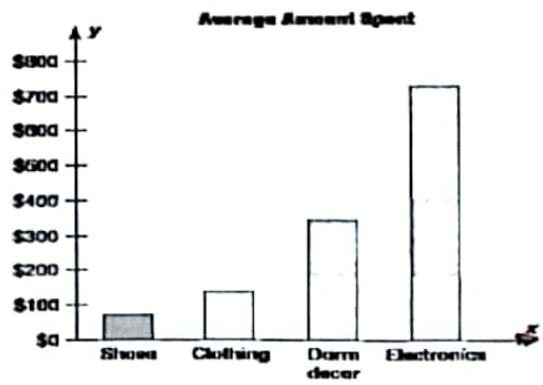
Example 2-8 page 76 College Spending for First-Year Students
The table shows the average money spent by first-year college students. Draw a horizontal and vertical bar graph for the data.

Electronics	$728
Dorm decor	344
Clothing	141
Shoes	72

Source: The National Retail Federation.

FIGURE 2–9 Bar Graphs for Example 2–8





The graphs show that first-year college students spend the most on electronic equipment.