Chapter Two

Frequency Distributions and Graphs

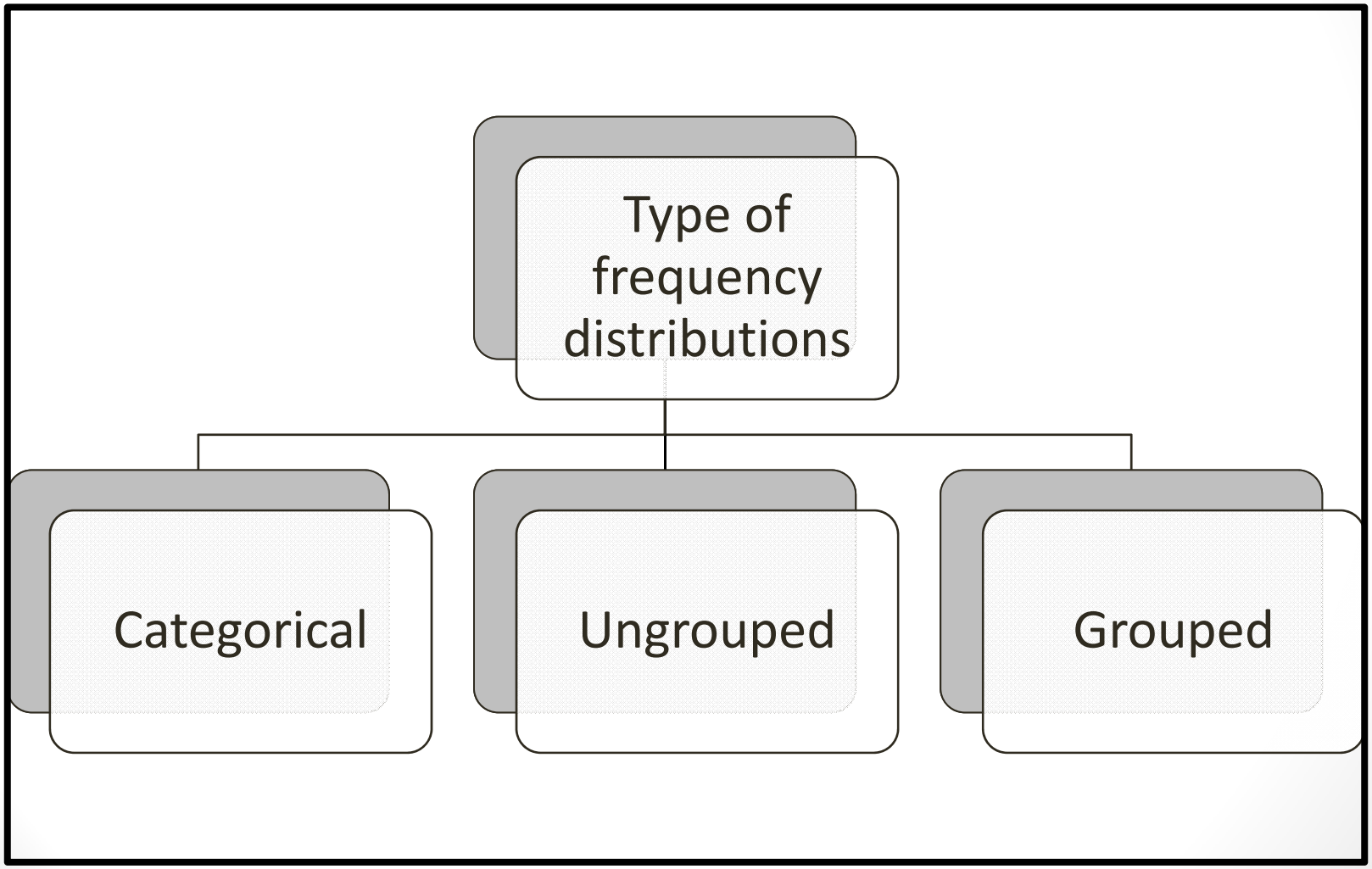
2-1 Organizing Data

- **Raw data** is the data when are collected in original form.

- 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.

- A **frequency distribution** is the organization of raw data in a table form, using classes and frequencies.

2-1 Organizing Data



2-2 Categorical Frequency Distributions

- **Categorical Frequency Distributions** are used for data that can be placed in specific categories, such as **nominal** or **ordinal** level data.

$$n = \sum f$$

$$Percent = \frac{f}{\sum f} * 100$$

2-2 Categorical Frequency Distributions

- Example: Blood type frequency distribution of 28 patients

Class	Frequency	Percent
A	6	$\frac{6}{28} * 100 = 21\%$
B	8	$\frac{8}{28} * 100 = 29\%$
O	11	$\frac{11}{28} * 100 = 39\%$
AB	3	$\frac{3}{28} * 100 = 11\%$
$\sum$	28	100%

2-3 Ungrouped Frequency Distributions

- **Ungrouped frequency distributions** are used for data that can be enumerated and when the range of values in the data set is small (discrete data).

2-3 Ungrouped Frequency Distributions

- Example: Number of patients in the waiting rooms of 16 clinics within a hospital at a specific time.

Class	Frequency	Cumulative Frequency	Percent
4	8	8	$\frac{8}{16} * 100 = 50\%$
5	3	8+3=11	$\frac{3}{16} * 100 = 19\%$
8	5	11+5=16	$\frac{5}{16} * 100 = 31\%$
$\sum$	16	-	100%

2-4 Grouped Frequency Distributions

- **Grouped frequency distributions** are 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., 24 – 30.

- The **lower class limit** represents the smallest data value that can be included in a class, e.g., 24 for the class limit 24 – 30.

- The **upper class limit** represents the largest value that can be included in the class, e.g., 30 for the class limit 24 – 30.

2-4 Grouped 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., 24.4 – 30 → 24.4 – 30.0

    4.4 – 7.93 → 4.40 – 7.93

    6 – 9 → 6 – 9

2-4 Grouped Frequency Distributions

- 2- add zero after the last decimal digit in the lower and upper limit.

  e.g., 24.4 – 30.0 → 24.40 – 30.00

  4.40 – 7.93 → 4.400 – 7.930

  6 – 9 → 6.0 – 9.0

- 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., 24.40 – 30.00 → 24.35 – 30.05

  4.400 – 7.930 → 4.395 – 7.935

  6.0 – 9.0 → 5.5 – 9.5

2-4 Grouped Frequency Distributions

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

2-4 Grouped Frequency Distributions

- 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.

e.g., 24 – 30 →

$$\text{class midpoint} = \frac{24 + 30}{2} = 27$$

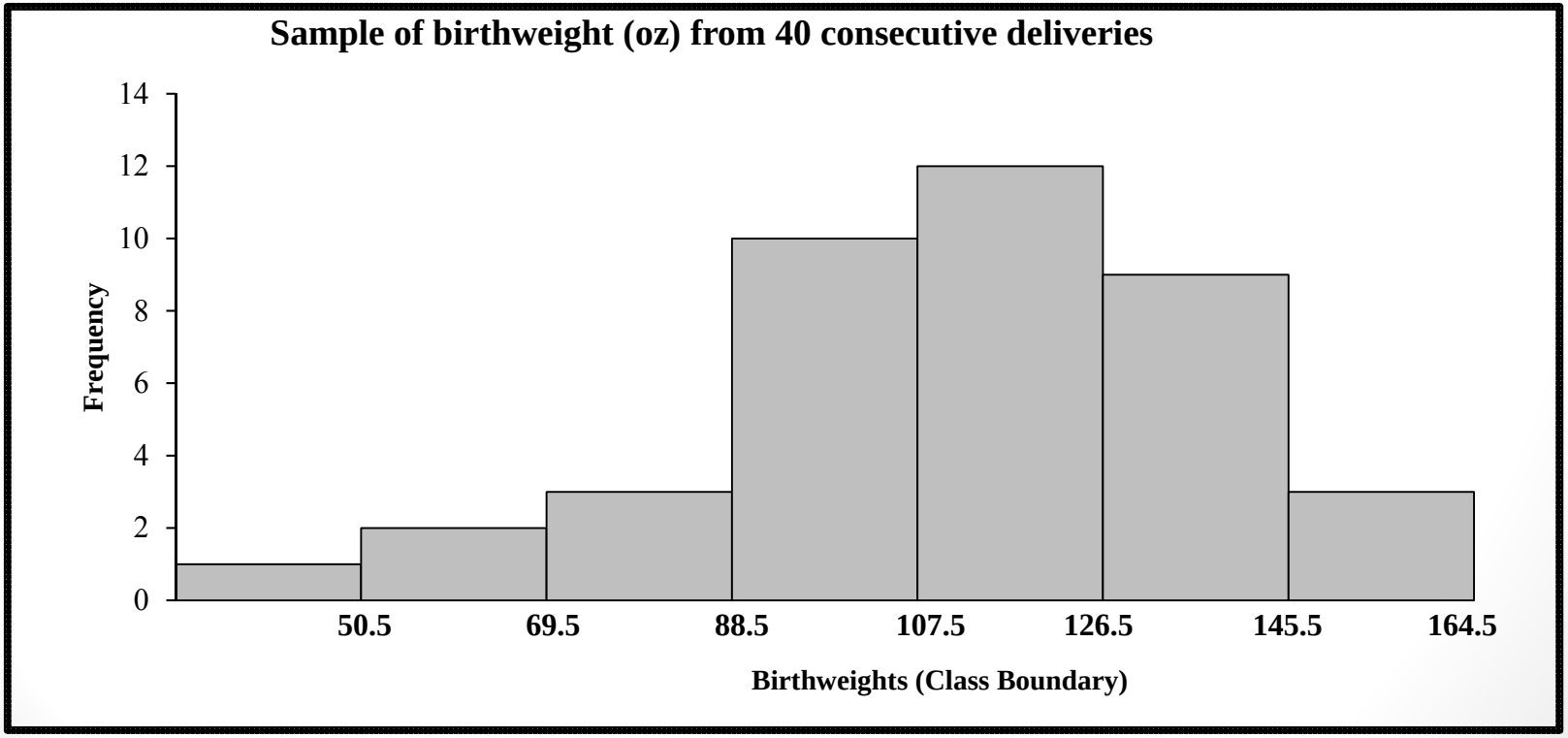
2-4 Grouped Frequency Distributions

- Example: Sample of birth weight (oz) from 40 consecutive deliveries.

Class limit	Frequency	Relative Frequency	Class boundaries	Class midpoint	Percent		Cumulative frequency	Cumulative Relative Frequency
32 – 50	1	$\frac{1}{40} = 0.025$	31.5 – 50.5	41	2.5%	Less than 50.5	1	0.025
51 – 69	2	0.050	50.5 – 69.5	60	5%	Less than 69.5	1+2=3	0.025+0.050=0.075
70 – 88	3	0.075	69.5 – 88.5	79	7.5%	Less than 88.5	3+3=6	0.075+0.075=0.150
89 – 107	10	0.250	88.5 – 107.5	98	25%	Less than 107.5	6+10=16	0.150+0.250=0.400
108 – 126	12	0.300	107.5 – 126.5	117	30%	Less than 126.5	16+12=28	0.400+0.300=0.700
127 – 145	9	0.225	126.5 – 145.5	136	22.5%	Less than 145.5	28+9=37	0.700+0.225=0.925
146 – 164	3	0.075	145.5 – 164.5	155	7.5%	Less than 164.5	37+3=40	0.925+0.075=1
Total	40	1	-	-	100%	-	-	-

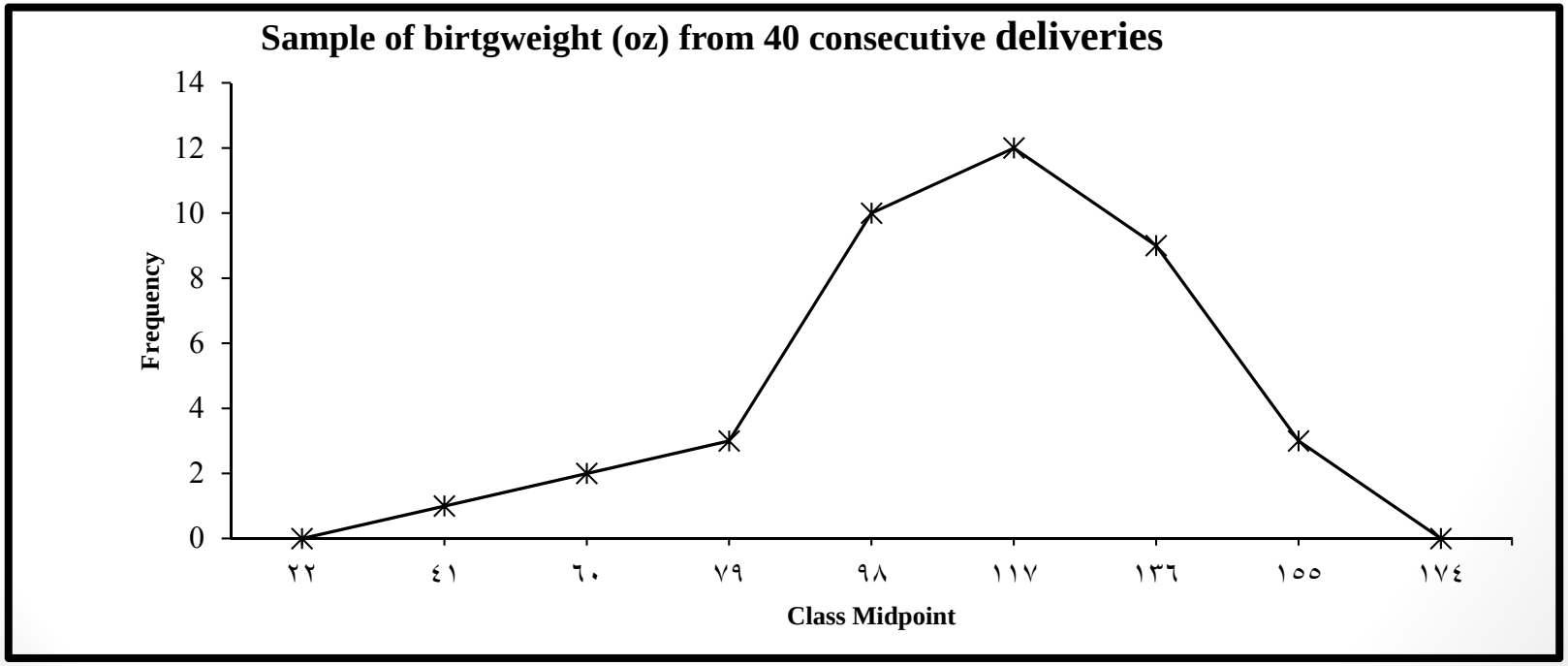
2-5 The Most Common Graphs

- 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.



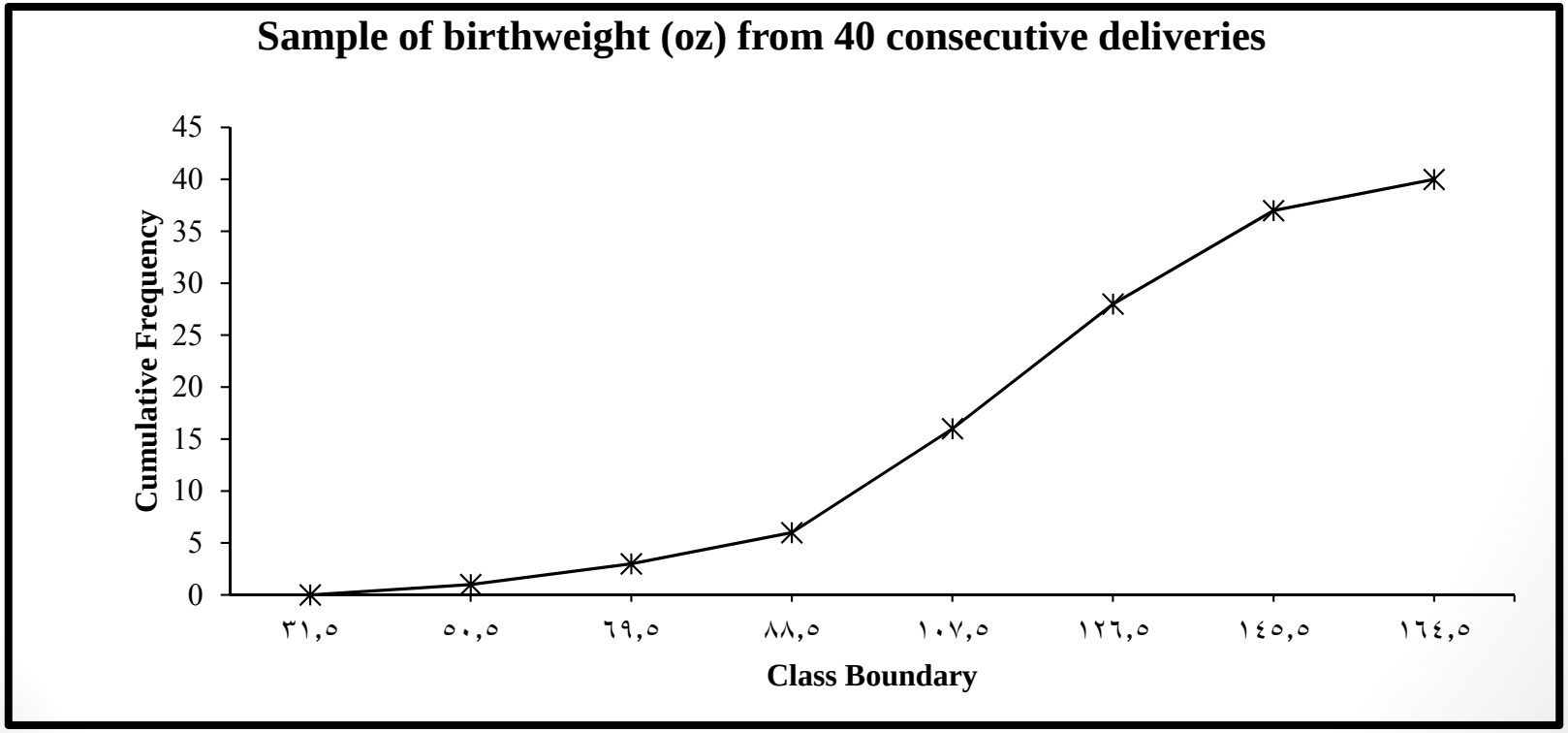
2-5 The Most Common Graphs

- The **frequency polygon** displays the **continuous** data that are organized in a grouped frequency distribution by using lines that connect points plotted for the frequencies at the midpoints of the classes.



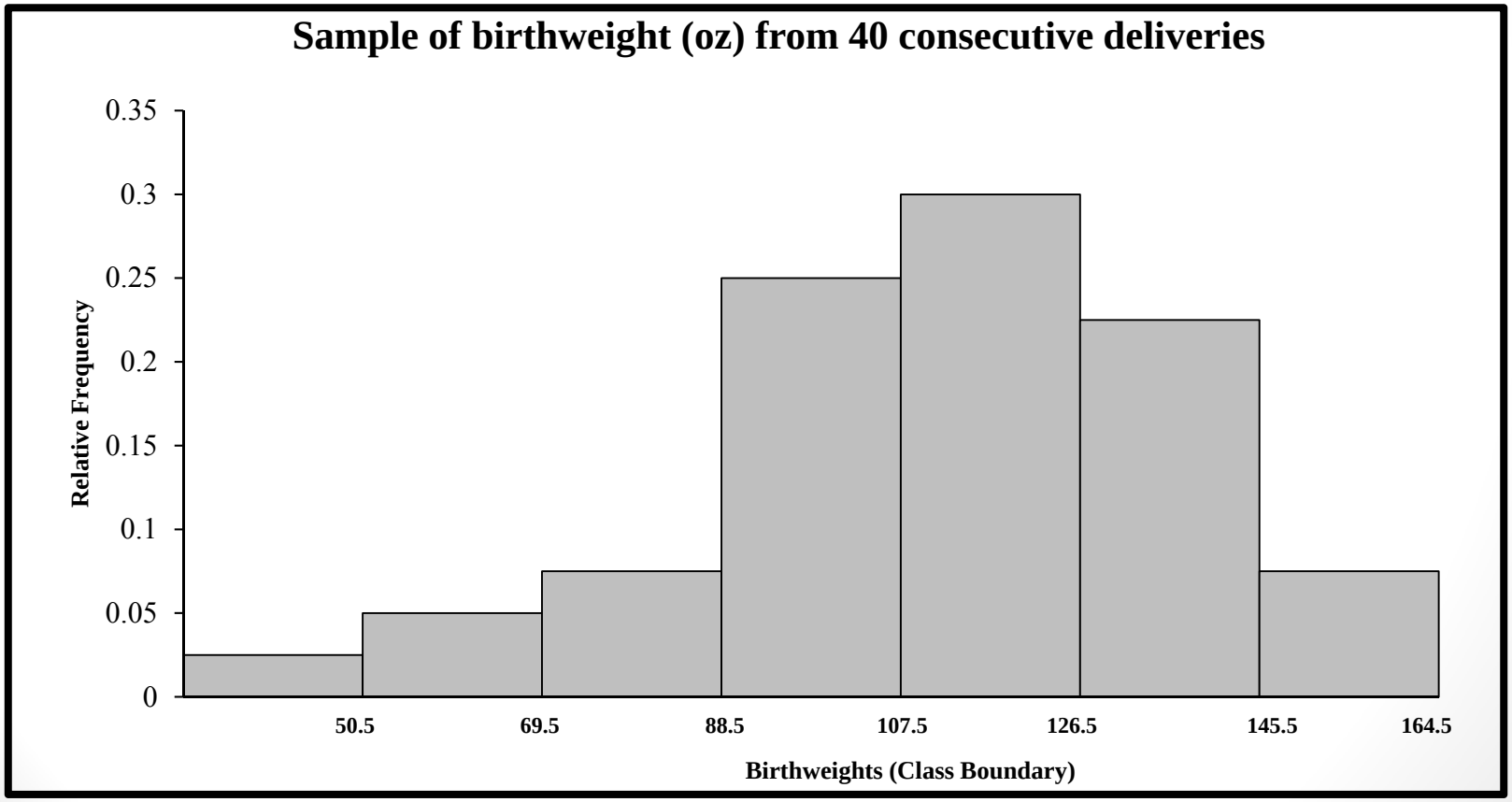
2-5 The Most Common Graphs

- The **cumulative frequency graph** or **ogive** displays the **continuous** data that are organized in a grouped frequency distribution.



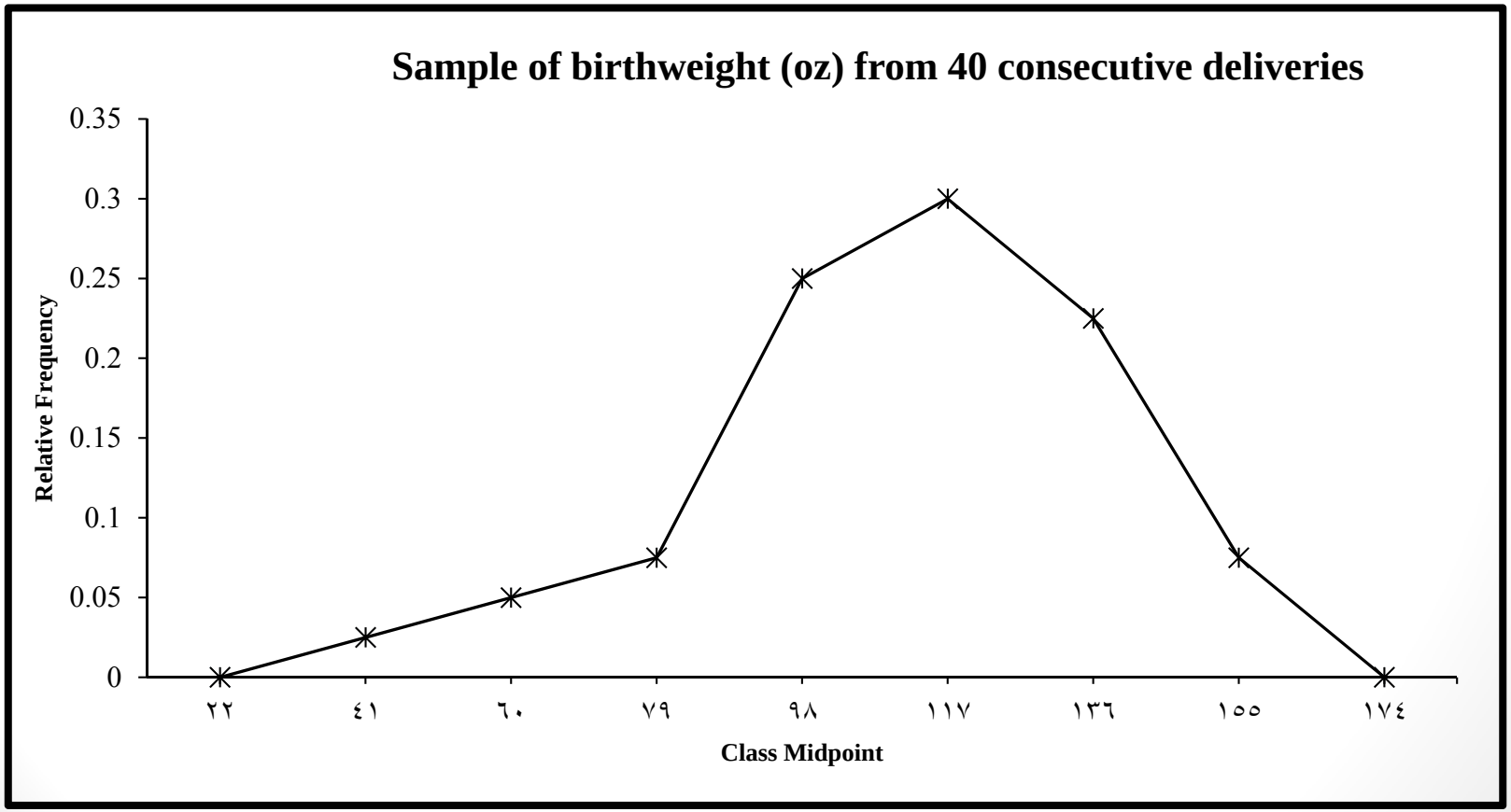
2-6 Relative frequency graph

- The **histogram**



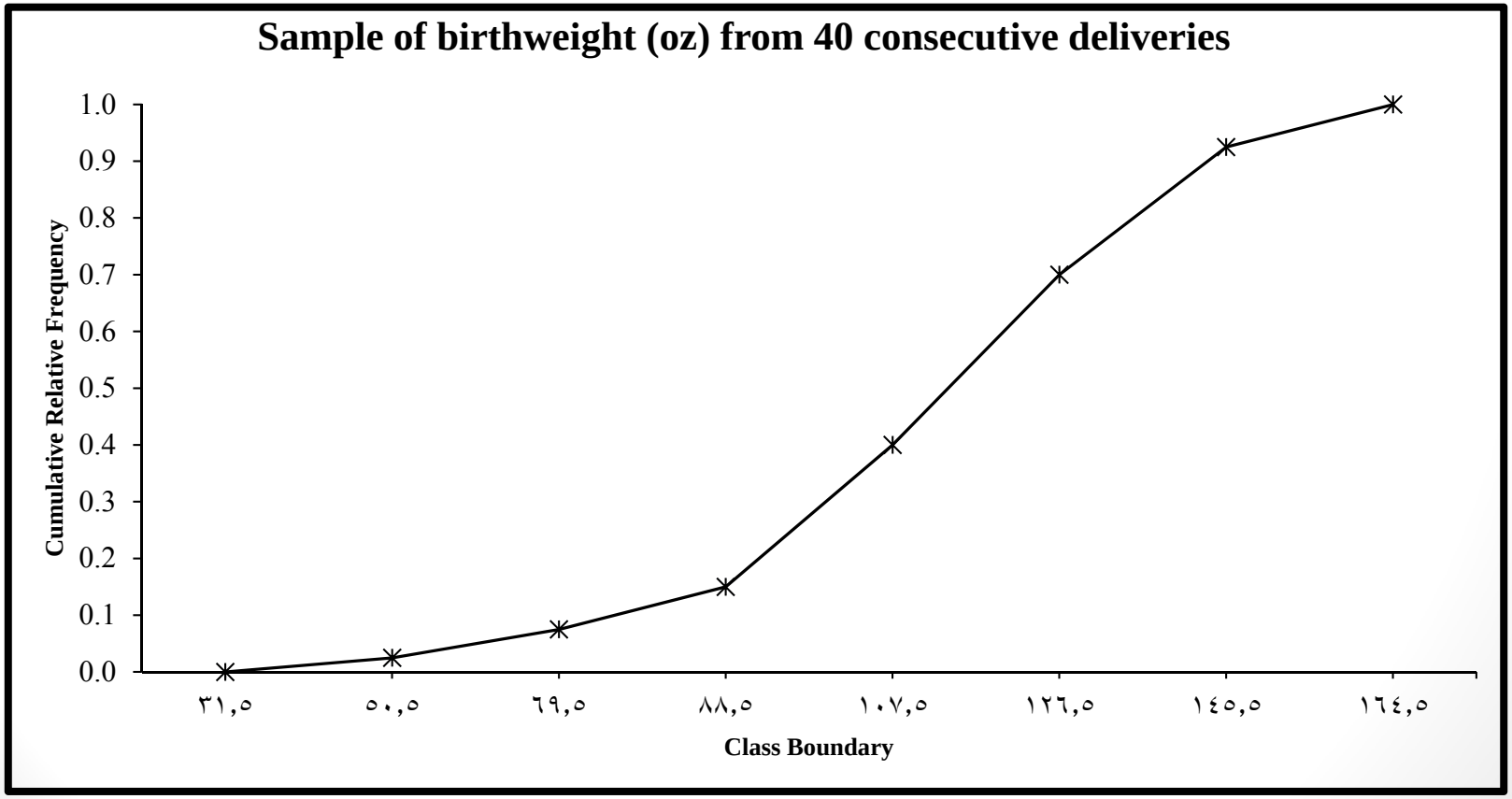
2-6 Relative frequency graph

- The **frequency polygon**



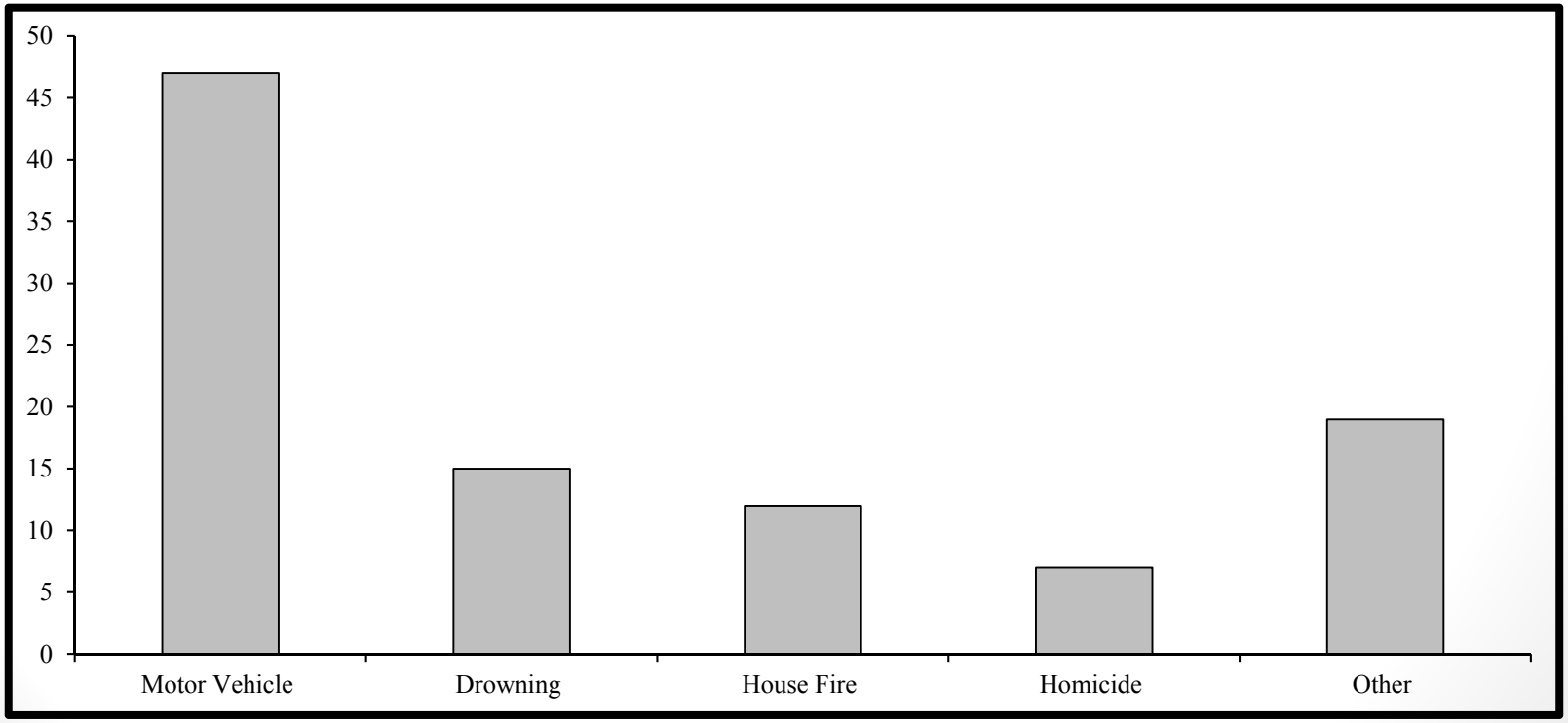
2-6 Relative frequency graph

- The **cumulative frequency graph** or **ogive**



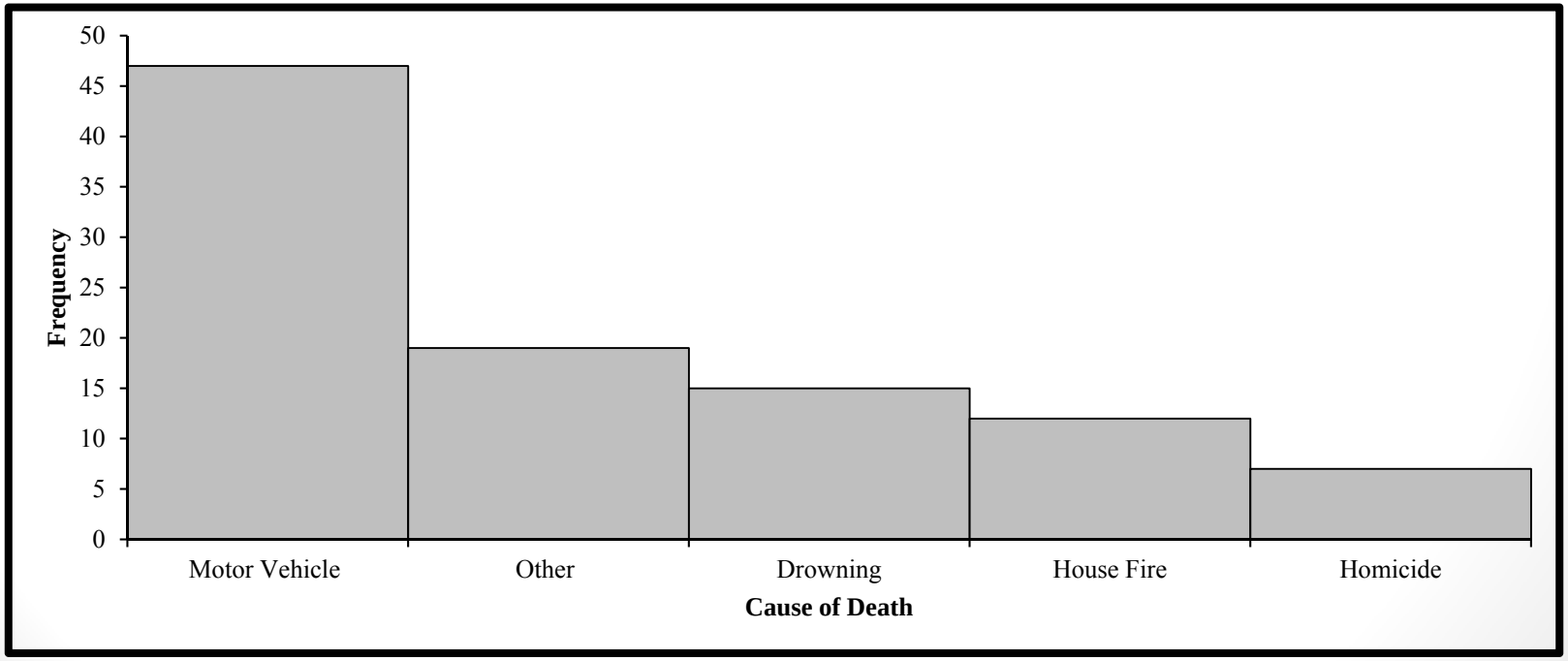
2-7 Other Types of Graphs

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



2-7 Other Types of Graphs

- 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.



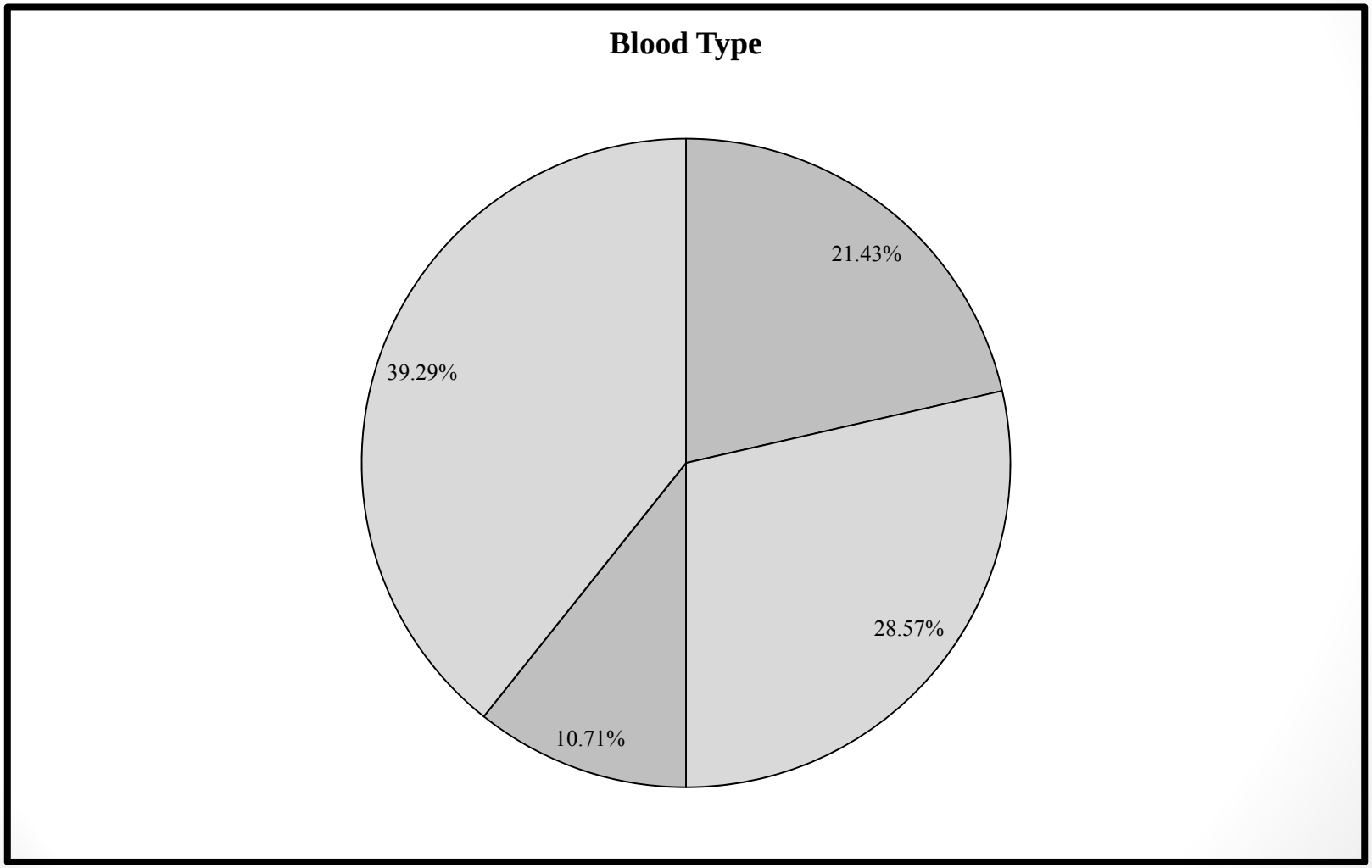
2-7 Other Types of Graphs

- The **pie graph** is a circle that is divided into sections according to the percentage of frequencies in each **category** of the distribution.

$$\boldsymbol{Degree} = \frac{\boldsymbol{f}}{\sum \boldsymbol{f}} * \mathbf{360}$$

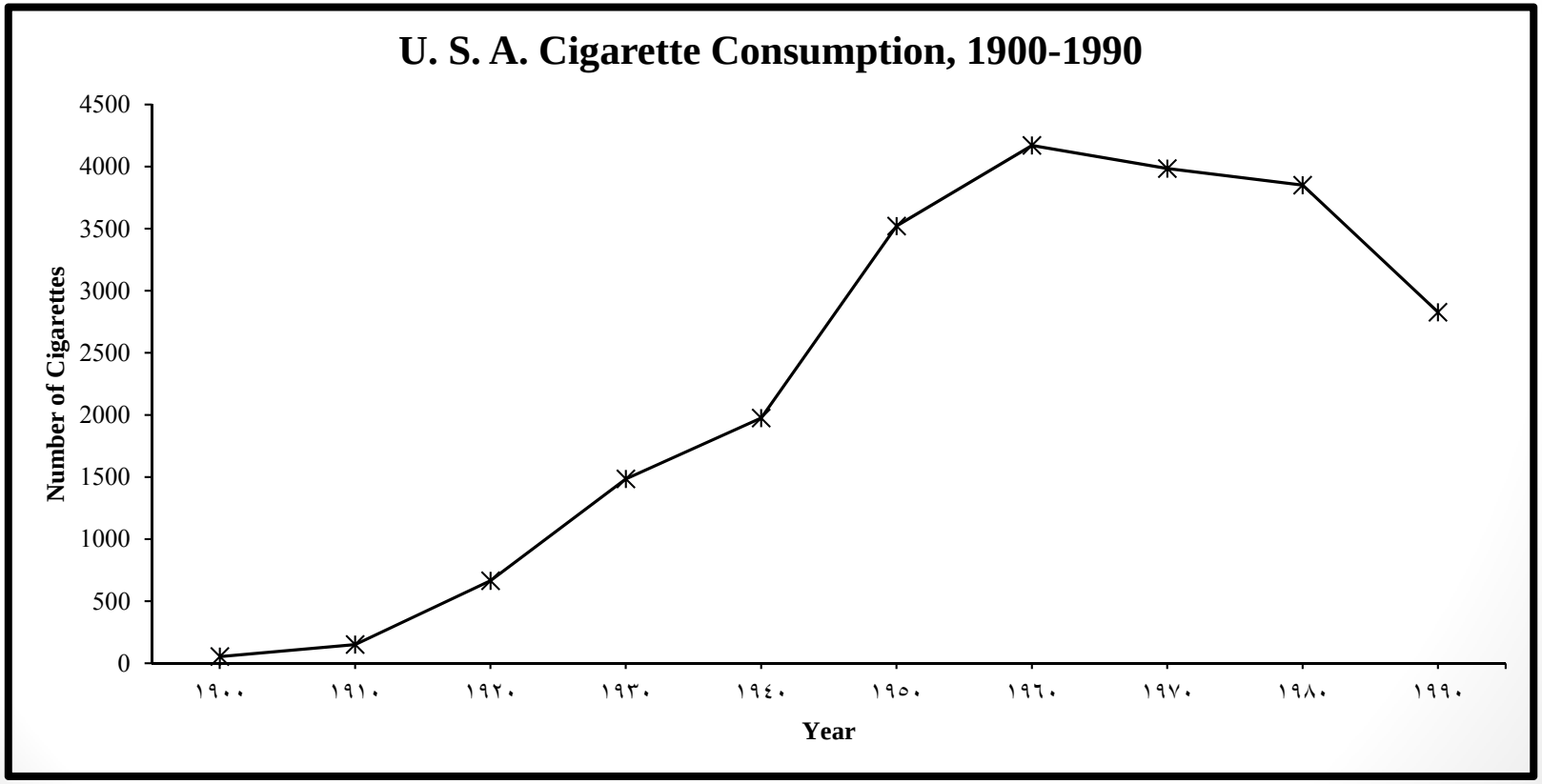
Class	Frequency	Percent	Degree
A	6	$\frac{6}{28} * 100 = 21.43\%$	$\frac{6}{28} * 360 = 77.14°$
B	8	$\frac{8}{28} * 100 = 28.57\%$	$\frac{8}{28} * 360 = 102.86°$
O	11	$\frac{11}{28} * 100 = 39.29\%$	$\frac{11}{28} * 360 = 141.43°$
AB	3	$\frac{3}{28} * 100 = 10.71\%$	$\frac{3}{28} * 360 = 38.57°$
Total	28	100%	360°

2-7 Other Types of Graphs



2-7 Other Types of Graphs

- The **time series graph** represents data that occur over a specific period of time.



2-7 Other Types of Graphs

- A **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.

2-7 Other Types of Graphs

3	2
4	
5	8
6	7
7	
8	3 5 6
9	2 3 4 4 5 6 8
10	0 4 5 8
11	2 3 5 6 8 8
12	0 2 3 4 4 7 8
13	2 2 2 4 8
14	0 1 6
15	5
16	1