DUM 2413 STATISTICS & PROBABILITY

CHAPTER 2
DESCRIPTIVE STATISTICS

PREPARED BY:
DR. CHUAN ZUN LIANG; DR. NORATIKAH ABU; DR. SITI ZANARIAH SATARI
FACULTY OF INDUSTRIAL SCIENCES & TECHNOLOGY
chuanzl@ump.edu.my; atikahabu@ump.edu.my; zanariah@ump.edu.my

CONTENT

2.1 • DATA ORGANISATION AND FREQUENCY DISTRIBUTION

2.2 • TYPES OF GRAPH

2.3 • SUMMARY STATISTICS (DATA DESCRIPTION)

2.4 • EXPLOROTARY DATA ANALYSIS

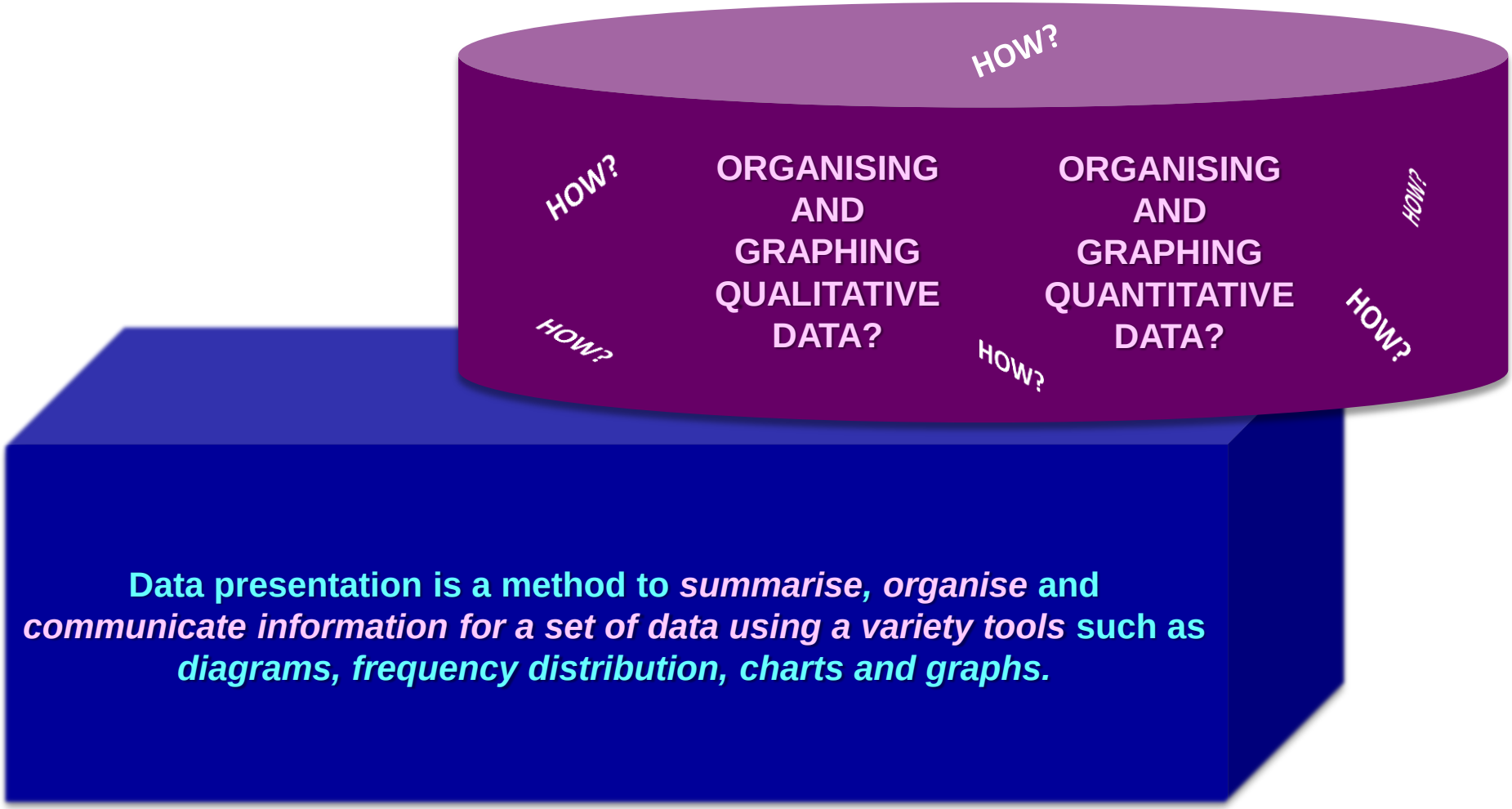
EXPECTED OUTCOMES

- Able to organise and represent qualitative and quantitative data using an appropriate analysis tool
- Able to differentiate between the grouped and ungrouped data
- Able to summarise the data using non-graphical and graphical exploratory data analysis tools
- Able to apply Chebyshev's Theorem in applications

2.1
DATA ORGANISATION AND FREQUENCY DISTRIBUTION

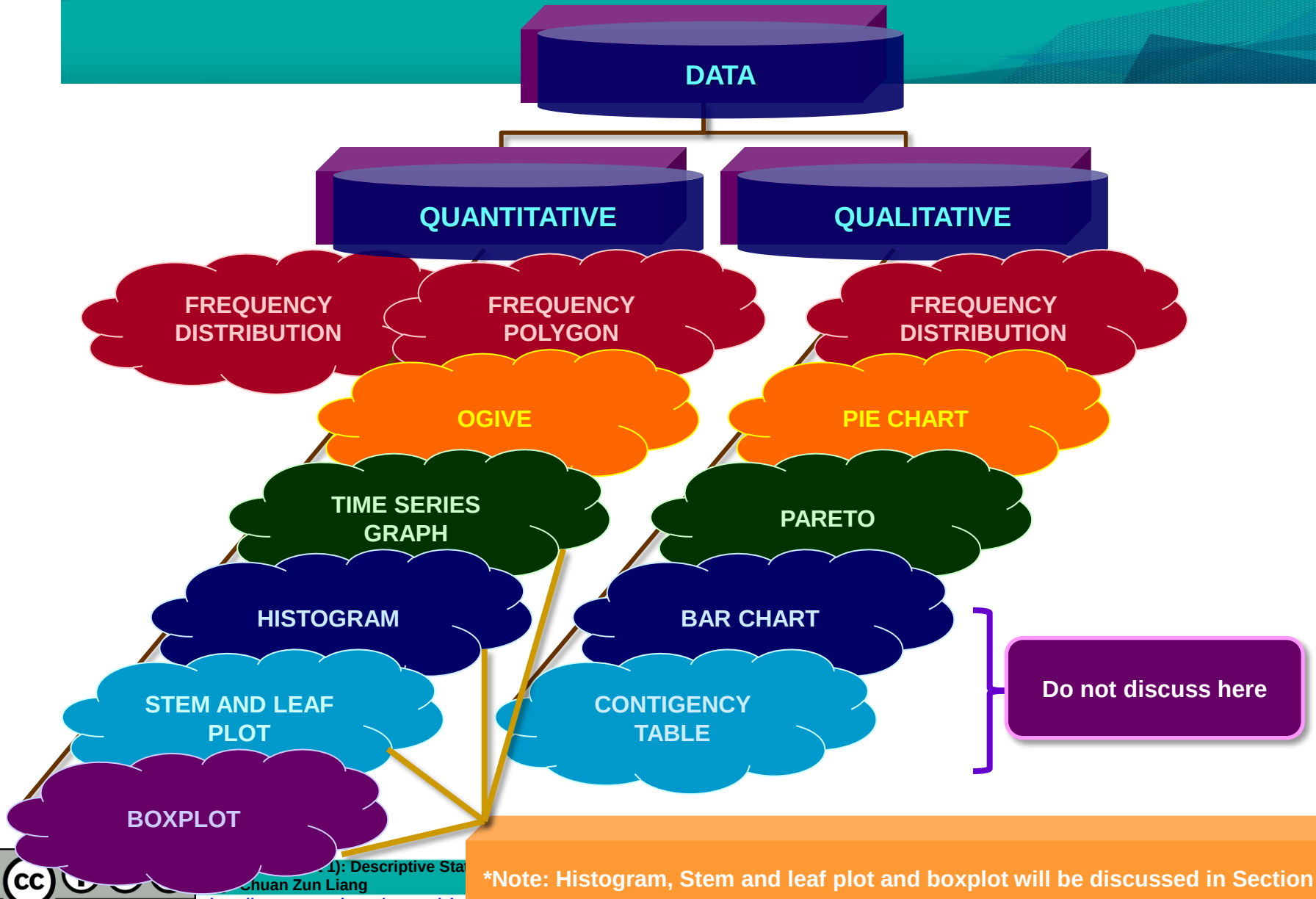
2.2
TYPES OF GRAPH

DATA PRESENTATION



Data presentation is a method to *summarise, organise* and *communicate information for a set of data using a variety tools* such as *diagrams, frequency distribution, charts and graphs.*

ORGANIZING AND GRAPHING DATA



*Note: Histogram, Stem and leaf plot and boxplot will be discussed in Section 2.4

ORGANISING AND GRAPHING QUALITATIVE DATA

FREQUENCY DISTRIBUTION

A frequency distribution represents the tabulation of data according all of several categories partitioned from a data set.

EXAMPLE: The blood groups of A, B, AB, and O of randomly selected from a blood bank.

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

Blood Group	Frequency
A	20
B	5
AB	3
O	22

FREQUENCY DISTRIBUTION

EXAMPLE 2.1

50 road traffic accidents categorised according the types of vehicles involved are listed below, where L denotes lorry, M1 denotes motorcycle, M2 denotes motorcar, and O denotes other causes. Construct a frequency distribution to summarise these data according the types of vehicles involved in road traffic accidents.

M1	M1	M1	L	L	M2	M2	M2	M2	M2
O	O	M2	M2	M2	L	L	M1	M1	M1
L	M1	M2	O	M1	M1	M1	M1	M1	M1
M1	M2	M1	M1	M2	L	L	M1	M1	L
L	M1	M1	M1	M2	M1	M1	O	O	O

SOLUTION

Types of vehicles	Frequency
M1	23
M2	12
L	9
O	6

ORGANISING AND GRAPHING QUALITATIVE DATA

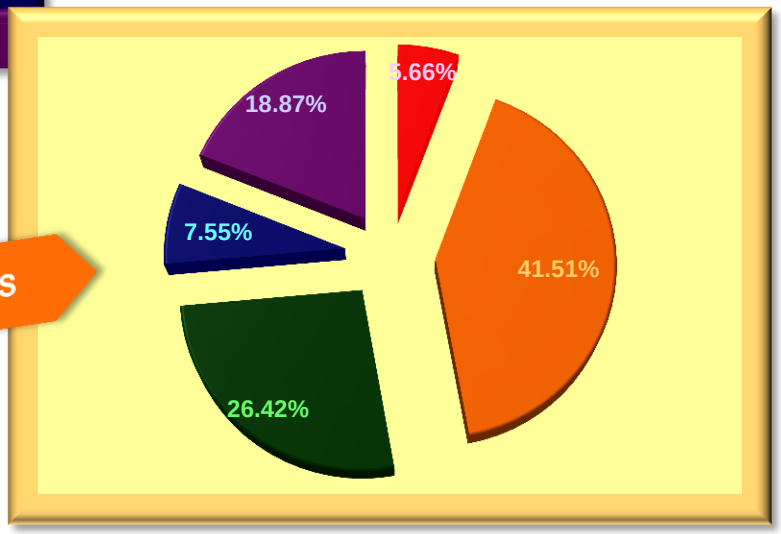
PIE CHART

A pie chart represents the proportional to the frequency counted according all of several categories partitioned from a data set as slices of a circle.

EXAMPLE: The total number of undergraduate majors from the 4 local universities

College	Number	Percentage
Agriculture	1500	5.66%
Art and Sciences	11000	41.51%
Business Administration	7000	26.42%
Education	2000	7.55%
Engineering	5000	18.87%

PIE CHARTS



EXAMPLE 2.2

A researcher from UMP interviewed 1540 employees of a company regarding their job satisfaction. The responses are summarised in table below.

Job satisfaction	Highly satisfied	Moderately satisfied	Little satisfied	Not satisfied
Number of employess	370	640	370	160

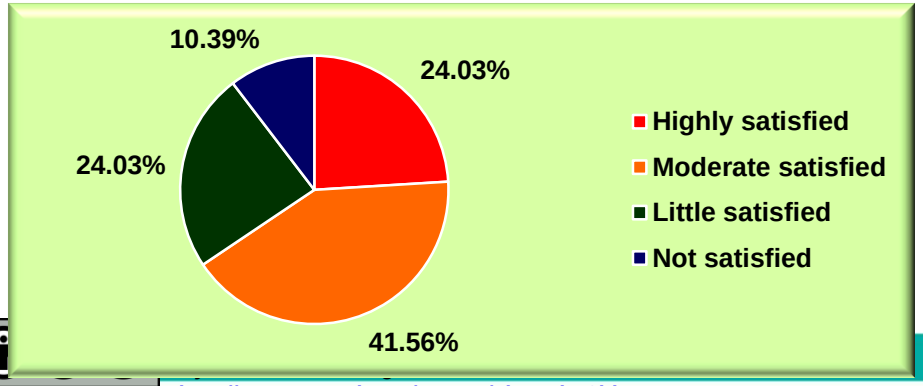
(i) Identify the variable and type of variable involved in this study.

Job stress; Qualitative

(ii) Based on the answer in (i), state the level of measurement of the variable.

Ordinal-level

(ii) Draw a chart to represent the contents in the table.



ORGANISING AND GRAPHING QUALITATIVE DATA

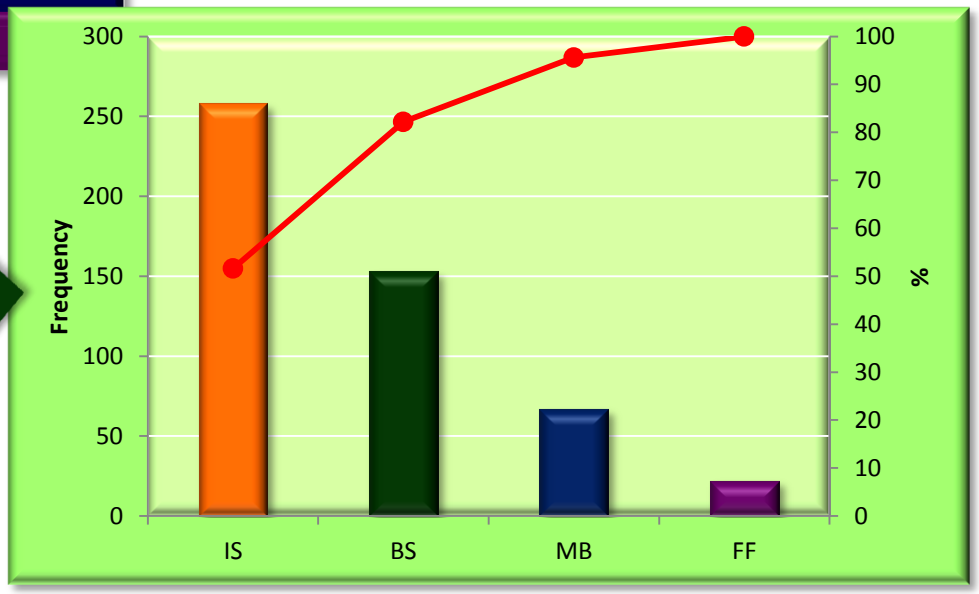
PARETO CHART

A Pareto chart is a chart constitute both bars and a line graph. The bars represent the frequency according all of several categories partitioned from a data set in descending order. Meanwhile, the line graph represents the cumulative frequencies.

EXAMPLE: The number of defective T-shirts manufactured in a factory

Defective Cause	Frequency
Improperly sized (IS)	258
Bad seam (BS)	153
Missing button (MB)	67
Fabric flaw (FF)	22

PARETO CHARTS

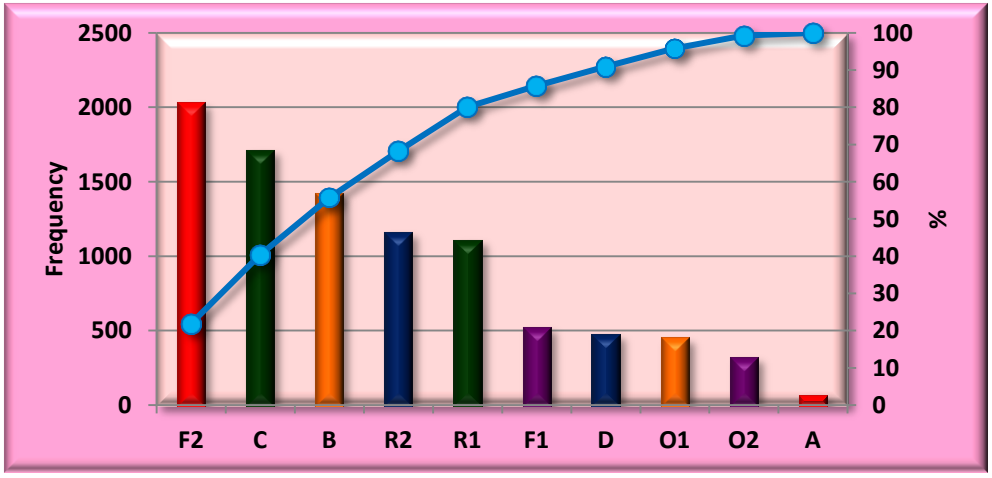


EXAMPLE 2.3

The table listing the number of consumer complaints against the airline services according to complaint category.

Complaint Category		Frequency	Complaint Category		Frequency
Advertising	(A)	68	Flight problems	(F2)	2031
Baggage	(B)	1421	Over-sales	(O1)	454
Customer Service	(C)	1715	Refunds	(R1)	1106
Disability	(D)	477	Reservations/ticketing/boarding	(R2)	1159
Fares	(F1)	523	Other	(O2)	322

(i) Construct a Pareto chart depicting this information.



(ii) What complaints would you recommend the airlines pay attention to correcting if they want to have the most effect on the overall number of complaints?

Flight problems

ORGANISING AND GRAHPING QUANTITATIVE DATA

FREQUENCY DISTRIBUTION

FREQUENCY DISTRIBUTION

TYPE

UNGROUPED DATA

Data are given in as individual points

EXAMPLE

Number of television sets	Frequency
0	2
1	13
2	18
3	0
4	10
5	2
Total	45

GROUPED DATA

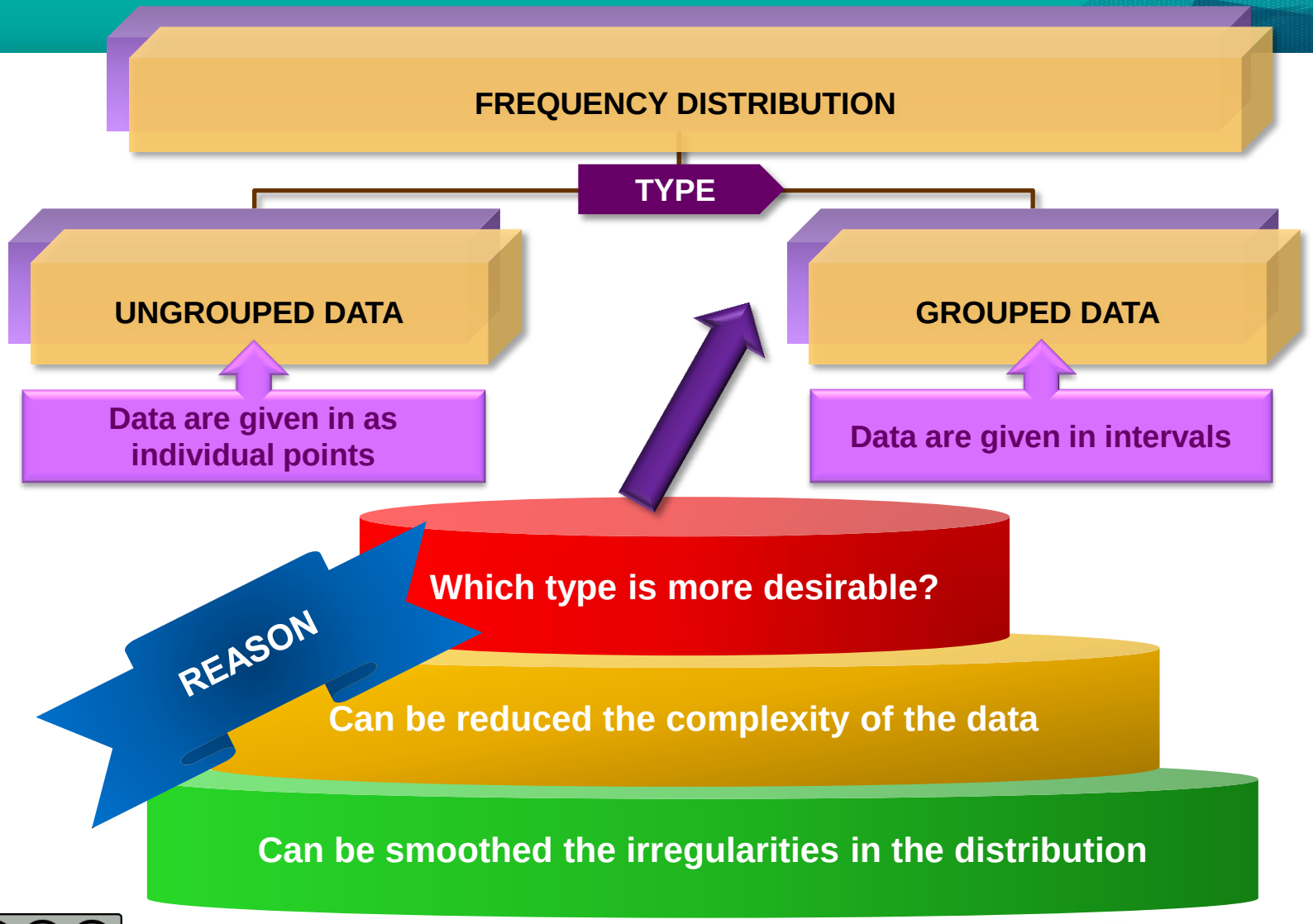
Data are given in intervals

EXAMPLE

Exam Score	Frequency
90-99	7
80-89	5
70-79	15
60-69	4
50-59	5
40-49	1
Total	37

ORGANISING AND GRAHPING QUANTITATIVE DATA

FREQUENCY DISTRIBUTION



ORGANISING AND GRAHPING QUANTITATIVE DATA

FREQUENCY DISTRIBUTION (GROUPED DATA)

STEP 1

Determine the number of classes using Sturges' formula, $k = 1 + 3.322 \log_{10} n$

Note:
This is one of the rules can be used to determine the number of classes.

STEP 2

Calculate the class width, $\text{Class width} = \dfrac{\text{maximum} - \text{minimum}}{k}$ (round-up)

STEP 3

Choose either minimum data value or a convenient value below the minimum data value as the first lower class limit.

STEP 4

Using the first lower class limit and the class width, list the other lower class limits.

STEP 5

List the lower class limits in a vertical column and then enter the upper class limits.

STEP 6

Take each individual data value and put a tally mark in the appropriate class. Add the tally marks to find the total frequency for each class.

EXAMPLE 2.4
(UNGROUPED DATA)

A lecturer conducted a study for her statistics students. One of the variable of her study is the heights of students. The table below shows the last digits of the heights of 36 students. Construct a frequency distribution with 10 categories.

0	1	6	5	5	5	0	3	3
5	5	8	8	0	5	5	5	0
8	0	2	5	3	1	9	5	5
5	5	4	6	5	0	0	5	0

SOLUTION

LAST DIGITS	FREQUENCY
0	8
1	2
2	1
3	3
4	1
5	15
6	2
7	0
8	3
9	1

EXAMPLE 2.5
(GROUPED DATA)

The table below shows the number of pairs of footwear sold on 50 consecutive weeks for a retail store. Construct a frequency distribution for these data.

62	21	4	26	7	38	64	12	38	45
6	33	55	62	48	49	7	9	41	21
30	31	3	25	57	48	8	18	43	72
23	5	8	37	31	31	39	65	53	4
75	17	14	61	50	51	38	36	40	56

Minimum

Maximum

SOLUTION

STEP 1

Number of classes, $k = 1 + 3.322 \log_{10} 50 = 6.6440 \approx 7$ classes (round-up)

STEP 2

Class width $= \dfrac{\text{maximum} - \text{minimum}}{k} = \dfrac{75 - 3}{7} = 10.2857 \approx 11$ (round-up)

EXAMPLE 2.5-CONTINUE
(GROUPED DATA)

62	21	4	26	7	38	64	12	38	45
6	33	55	62	48	49	7	9	41	21
30	31	3	25	57	48	8	18	43	72
23	5	8	37	31	31	39	65	53	4
75	17	14	61	50	51	38	36	40	56

STEP 3 & 4 | STEP 5 | STEP 6

Class	Tally	Frequency
3-13	𝍸 𝍸 I	11
14-24	𝍸 I	6
25-35	𝍸 II	7
36-46	𝍸 𝍸	10
47-57	𝍸 IIII	9
58-68	𝍸	5
69-79	II	2

ORGANISING AND GRAPHING QUANTITATIVE DATA

FREQUENCY POLYGON

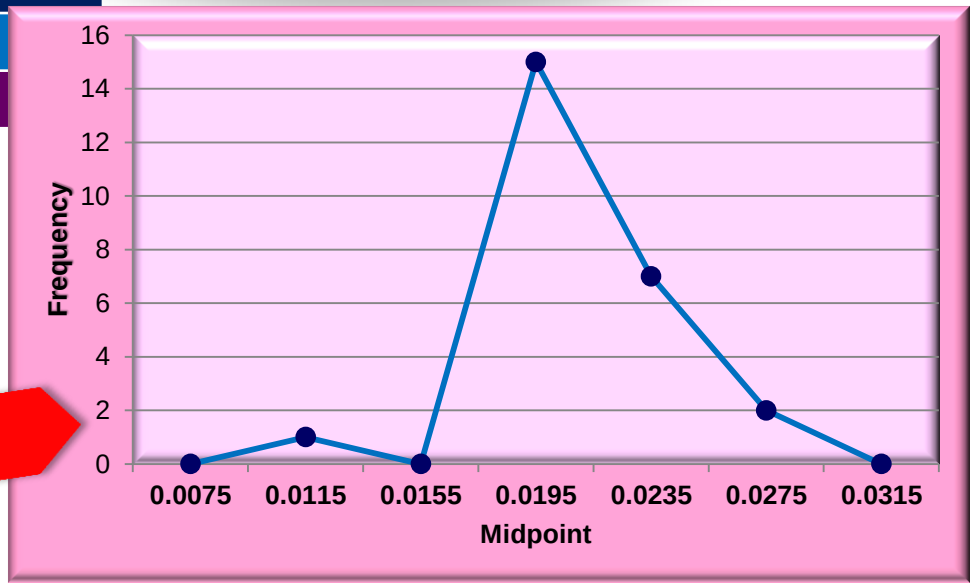
A frequency polygon is a line graph which joining the midpoints of the upper edges of the rectangles in a histogram. Similar like histogram, a frequency polygon can illustrate the shapes of distribution.

EXAMPLE: Tar in Non-filtered cigarettes

Tar (g)	Frequency
0.010-0.013	1
0.014-0.017	0
0.018-0.021	15
0.022-0.025	7
0.026-0.029	2

Use Grouped Frequency Distribution

FREQUENCY POLYGON



EXAMPLE 2.6

A team of biologists measured the height of 200 plants as summarised in the table below. Based on these data, construct a frequency polygon.

Height(cm)	46-50	51-55	56-60	61-65	66-70
Frequency	23	64	57	32	24

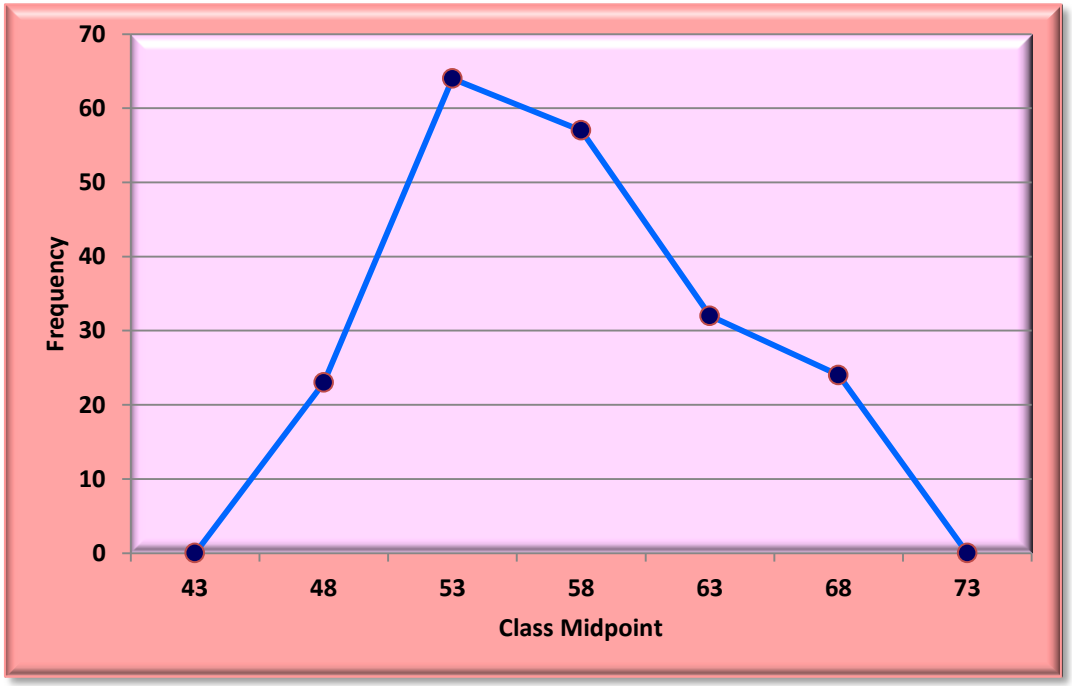
SOLUTION

Height (cm)	46-50	51-55	56-60	61-65	66-70
Midpoint	$\frac{(46+50)}{2}=48$	$\frac{(51+55)}{2}=53$	$\frac{(56+60)}{2}=58$	$\frac{(61+65)}{2}=63$	$\frac{(66+70)}{2}=68$
Frequency	23	64	57	32	24

EXAMPLE 2.6-CONTINUE

SOLUTION

Height (cm)	41-45	46-50	51-55	56-60	61-65	66-70	71-75
Midpoint	43	48	53	58	63	68	73
Frequency	0	23	64	57	32	24	0



ORGANISING AND GRAPHING QUANTITATIVE DATA

OGIVE (CUMULATIVE FREQUENCY GRAPH)

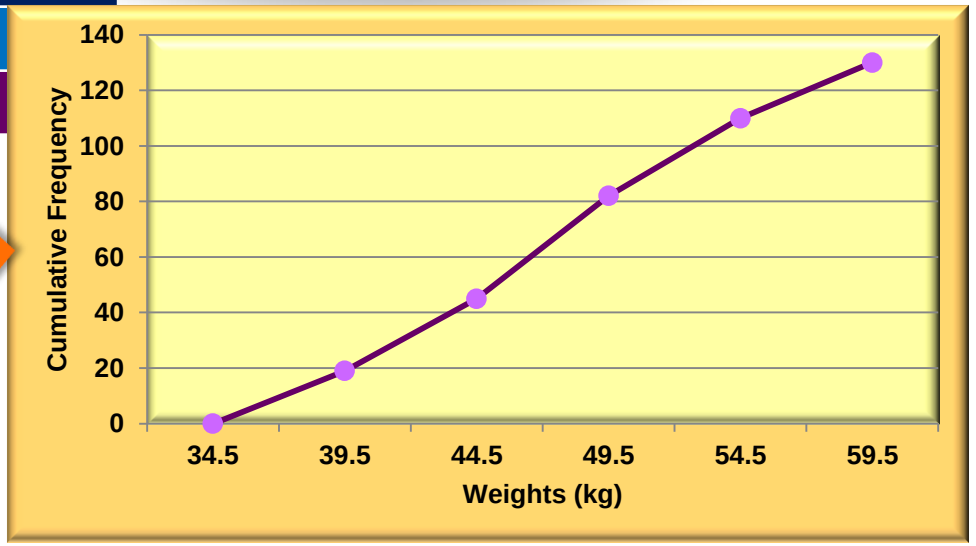
An ogive, also known as a cumulative frequency graph is a line graph that depicts cumulative frequencies, which use the lower or upper class boundaries

EXAMPLE: The distribution of weights of metals

Weights (kg)	Frequency
35-39	19
40-44	26
45-49	37
50-54	28
55-59	20

Use Grouped Frequency Distribution

Ogive



CAUTION

CLASS LIMITS ≠ CLASS BOUNDARIES

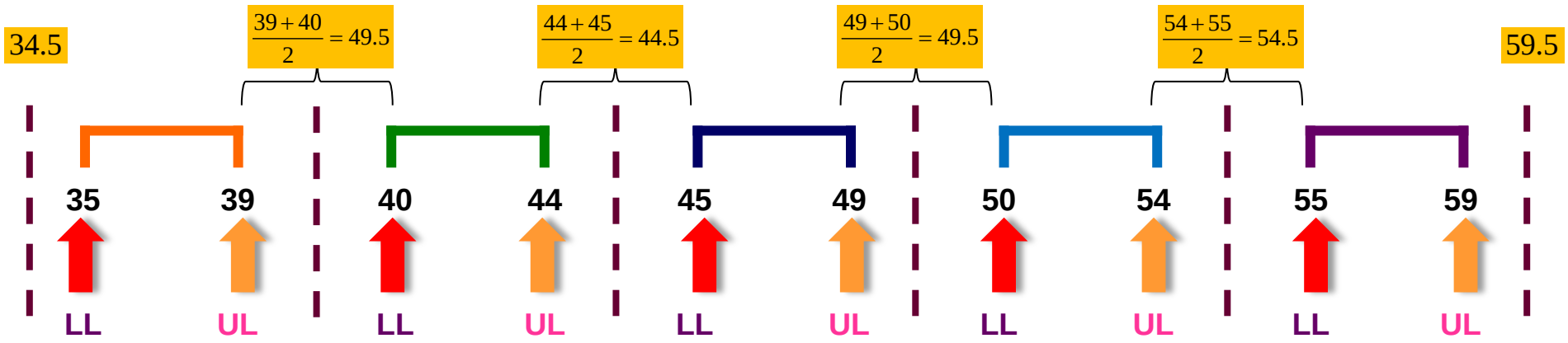
CLASS LIMITS

- Lower Limits (LL): smallest number that can belong to the different classes.
- Upper Limits (UL): largest number that can belong to the different classes.

CLASS BOUNDARIES

- Class boundaries: numbers used to separate the classes, but without the gaps created by class limits.

CLASS BOUNDARIES



*NOTE: The class width always calculated based on the distance between class boundary

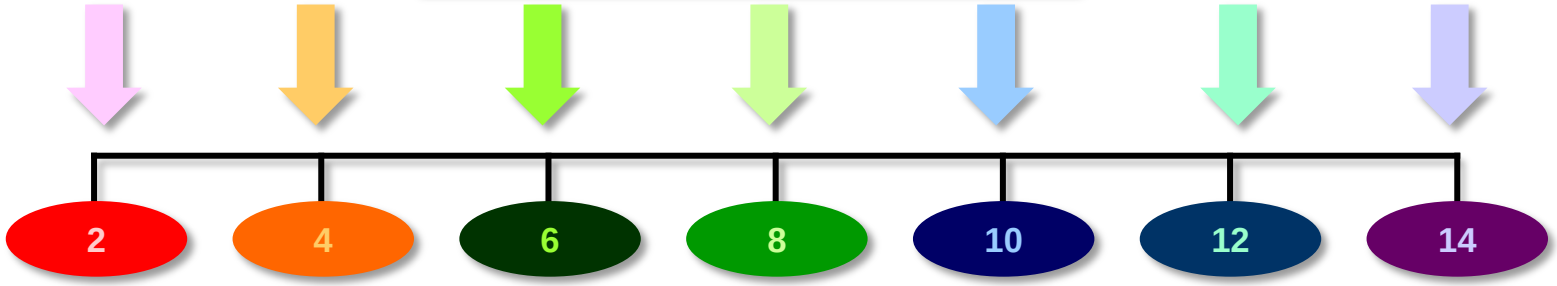
EXAMPLE 2.7

A manager interested to study the effectiveness of a particular training onto the productivity of his staffs. To purse his objective, he collected the productivity of his staffs after attended the training as recorded in the table below. Construct an ogive for the data below.

Productivity (units per hours)	Number of workers
2-4	5
4-6	10
6-8	18
8-10	32
10-12	23
12-14	12

$2 \le x < 4$

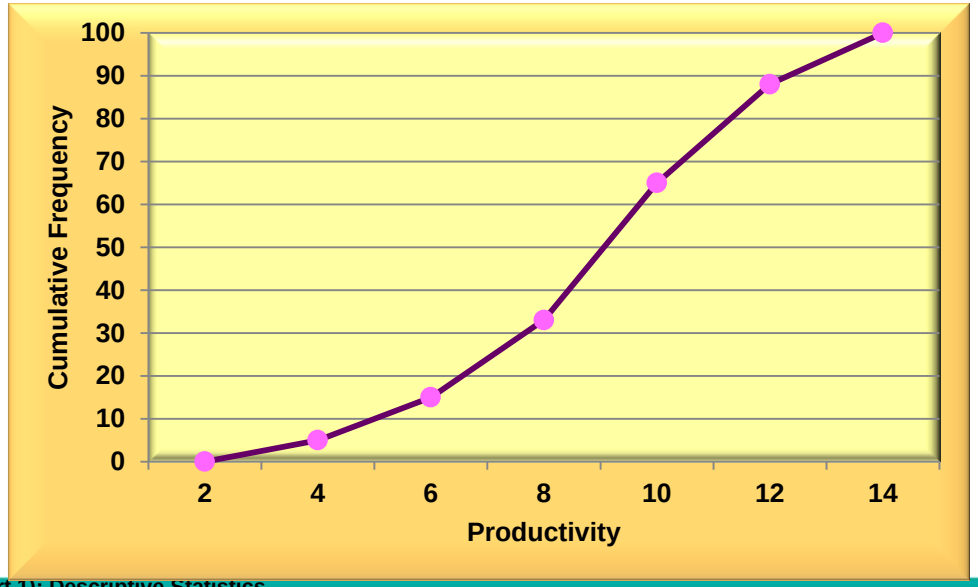
CLASS BOUNDARIES



EXAMPLE 2.7

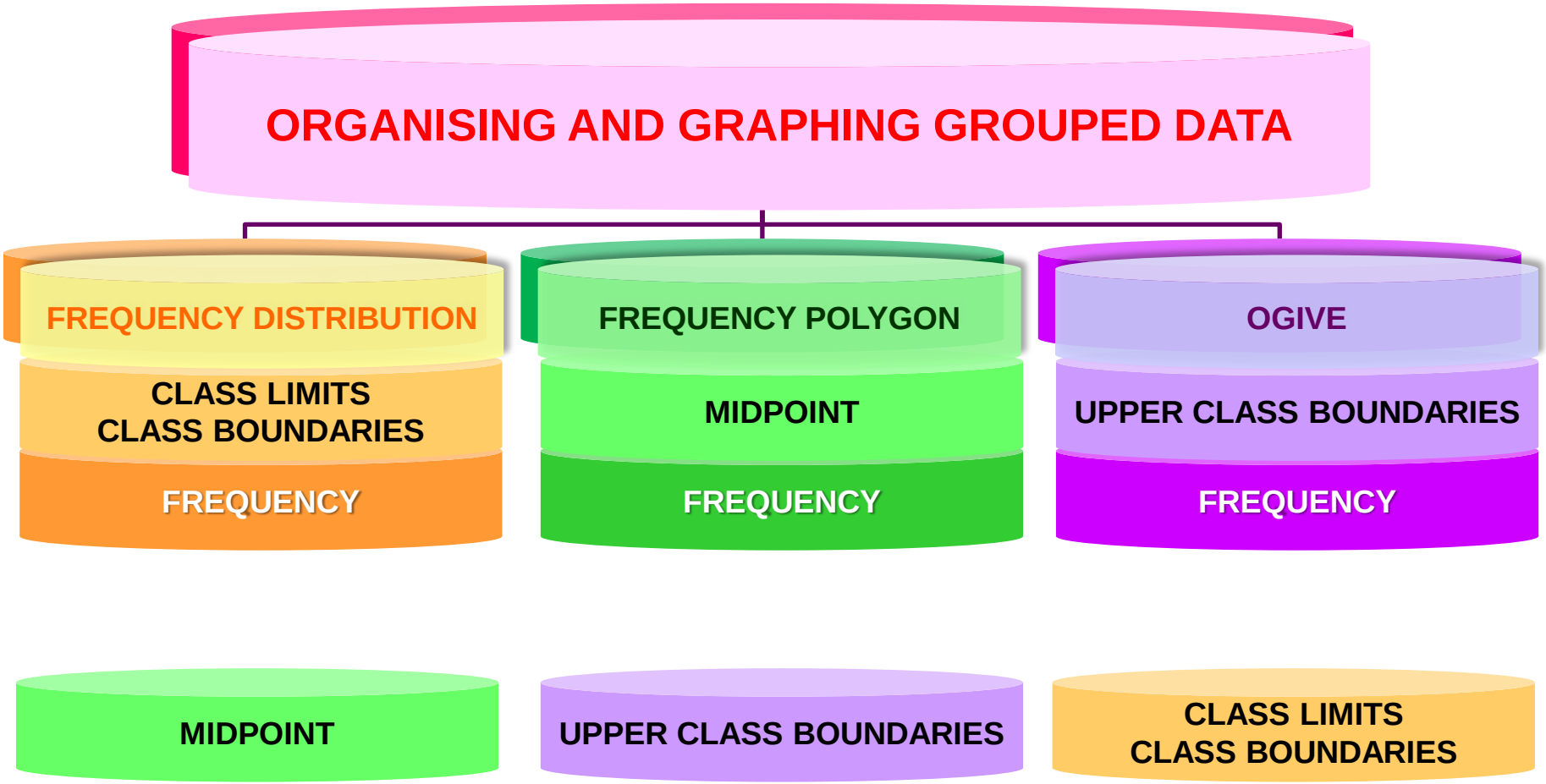
SOLUTION

Productivity	Number of workers	Upper Class Boundary	Cumulative Frequency
0-2	0	2	0
2-4	5	4	0 + 5 = 5
4-6	10	6	5 + 10 = 15
6-8	18	8	15 + 18 = 33
8-10	32	10	33 + 32 = 65
10-12	23	12	65 + 23 = 88
12-14	12	14	88 + 12 = 100



SUMMARY

FREQUENCY DISTRIBUTION, FREQUENCY POLYGON AND OGIVE



THANK YOU

END OF CHAPTER 2 (PART 1)