

Exercises 2-1

1. List five reasons for organizing data into a frequency distribution?

- To organize data
- identification of patterns
- Statistical Analysis
- Data Interpretation
- Comparison

2. Name the three types of frequency distributions, and explain when each should be used.

1. Ungrouped frequency distribution: List each individual value and its frequency.

It is used for small datasets with distinct, individual values.

2. Grouped frequency distribution: Organizes data into intervals (bins) and shows frequencies for each interval. It is used for large datasets or continuous data, to simplify and summarize the distribution.

3. Cumulative frequency distribution

It tracks the total count of values up to each interval or point. It is used to determine the accumulation of frequencies and understand data distribution over intervals.

+ 4. Relative Frequency distribution

It displays the proportion or percentage of data points in each interval to the total number of data points. It is used to show the relative importance of each interval or value within the dataset.

3. Find the class boundaries, midpoints and widths for each class.

Classes	Lower boundary	Upper boundary	Midpoints	Width
12-18	12	18	$\frac{12+18}{2} = 15$	$18-12=6$
56-74	56	74	65	6
685-705	685	705	700	10
13.6-14.7	13.6	14.7	14.15	1.1
2.15-3.93	2.15	3.93	3.04	1.78

1. Find the range of the data: Range = Max - Min = P.T.D

3. Find the class boundaries, midpoints, and widths for each class.

→ Class Boundaries

$$\bullet \text{Lower Boundary} = \text{Lower Limit} - 0.5$$

$$\bullet \text{Upper Boundary} = \text{Upper Limit} + 0.5$$

→ Class Midpoint

$$\bullet \text{Midpoint} = \frac{\text{Lower Limit} + \text{Upper Limit}}{2}$$

→ Class Width

$$\bullet \text{Class width} = \text{Upper Limit} - \text{Lower Limit}$$

Class Interval	Class Boundaries	Midpoint	Class width
12 - 18	11.5 - 18.5	15	$18.5 - 11.5 = 7$
56 - 74	55.5 - 74.5	65	$74.5 - 55.5 = 19$
69.5 - 70.5	694.5 - 705.5	700	11
13.6 - 14.7	13.55 - 14.75	14.15	1.2
2.15 - 3.93	2.10 - 3.98	3.04	1.88

→ Class Boundaries Adjustment

1. Continuous Data: Typically uses a 0.5 unit adjustment because it's based on integer intervals.

2. Decimal Data: Adjust the boundaries according to the precision of the data. For data in tenths (0.1), a 0.05 adjustment is more appropriate.

4. How many classes should frequency distribution have?

Why should the class width be an odd number?

• Five to twenty classes.

• The class width should be an odd number so that the midpoint will have the same place value as the data. This will guarantee that the class midpoints are integers instead of decimals. The classes must be all inclusive or exhaustive.

→ Inclusive: Every data point falls into one of the class intervals without overlap or gaps.

→ Exhaustive: All possible values in the dataset are covered by class intervals, with no values left out.

• Classes must be inclusive (cover all data points) and exhaustive (cover all possible values).

(2)

5. Shown here are four frequency distributions. Each is incorrectly constructed. State the reason why.

a. Class Frequency

Class	Frequency
27-32	1
33-38	0
39-44	6
45-49	4
50-55	2

Ans: The intervals are not covering all possible data values without overlap or gaps. The upper boundary of one class should be exactly the lower boundary of the next class interval to ensure that all values are covered and no values are left out.
 → class width is not uniform

The intervals have gaps between them (e.g., 32 and 33) and are not continuous.

- When the class width is not uniform, it means that the intervals in frequency distribution are of varying sizes. This lack of consistency can lead to difficulties in analyzing and comparing the data accurately.

b. Class Frequency

Class	Frequency
5-9	1
9-13	2
13-17	5
17-20	6
20-24	3

Ans: Class limits overlap. The intervals 5-9 and 9-13 overlap at 9. Similarly 13-17 and 17-20 overlap at 17.
 → class width is not uniform
 $9-5 = 4$
 $13-9 = 4$
 $17-13 = 4$
 $20-17 = 3$
 $24-20 = 4$

c. Class Frequency

123 - 127	3
128 - 132	7
138 - 142	2
143 - 147	9

Ans: A class is missing: 133 - 137.

d. Class Frequency

9-13	1
14-18	6
20-25	2
26-30	5
31-32	9

Ans: Class width is not uniform

6. What are open-ended frequency distribution? Why are they necessary?

Open-ended frequency distributions are those where one or both class intervals do not have a defined upper or lower limit. This means the interval extends indefinitely in one direction.

Examples:

- Lower open-ended: Below 10
- Upper open-ended: Above 100
- Both End Open: Below 5 and Above 50

- They are useful for handling data that includes extreme values or outliers that do not fit neatly into predefined intervals.
- They prevent the need for creating excessively detailed intervals, which might not be practical or useful for summary purposes.

Examples:

Salary Range	Frequency
Below \$30,000	5
\$30,000 - \$70,000	50
Above \$70,000	3

7. Trust in Internet Information A survey was taken on how much trust people place in the information they read on the Internet. Construct a categorical frequency distribution for the data. A = trust in everything they read, M = trust in most of what they read, H = trust in about one-half of what they read, S = trust in a small portion of what they read.
(Based on the information from the UELA Internet report)

M	M	M	A	H	M	S	M	H	M
S	M	M	M	M	A	M	M	A	M
M	M	H	M	M	M	H	M	H	M
A	M	m	M	H	M	m	M	M	M

Class	Tally	Frequency
A		4
S		2
H	HHH	6
M	HHH HHH HHH HHH HHH HHH	28

8. State Gasoline Tax The state gas tax in cents per gallon for 25 states is given below. Construct a grouped frequency distribution below and a cumulative frequency distribution with 5 classes.

7.5	16	23.5	17	22
21.5	19	20	27.1	20
22	20.7	17	28	20
23	18.5	28.3	24	31
14.5	28.9	18	30	21.5

Step 1: Determine the range

$$\text{Lowest value} = 7.5 \quad \text{Max} = 31.5$$

$$\text{Range} = 31.5 - 7.5 = 24$$

Step 2: Determine class width

$$\text{Class width} = \frac{\text{Range}}{\text{No. of classes}} = \frac{24}{5} = 4.8 \approx 5$$

$$\rightarrow \text{class width} = \frac{\text{Range}}{k^5}, \text{ where } 2^k \geq n$$

$k = \text{classes}$, $n = \text{No. of observation}$

$$\text{Let } k = 5, \quad 2^5 \geq 25 \text{ True.}$$

$$\text{So No. of classes} = k = 5..$$

Class Interval	F	F _c
7.5 - 12.5	1	1
12.6 - 17.5	4	5
17.6 - 22.5	10	15
22.6 - 27.5	6	21
27.6 - 32.5	11	25

Class Interval	F	F _c
7.5 - 12.4	1	1
12.5 - 17.4	4	5
17.5 - 22.4	10	15
22.5 - 27.4	6	21
27.5 - 32.4	11	25

Class Interval	F	F _c
7.5 - 12.5	1	1
12.6 - 17.6	4	5
17.7 - 22.7	10	15
22.8 - 27.8	6	21
27.9 - 32.9	4	25

9. Weights of the NBA's Top 50 Players Listed are the weights of the NBA's top 50 players. Construct a grouped frequency distribution and a cumulative frequency distribution with 8 classes. Analyze the results in terms of peakes, extreme values, etc.

240	210	220	260	250	185	230	270
165	285	205	230	280	210	220	210
250	265	230	210	240	245	225	180
215	235	245	250	215	210	185	240
260	210	190	260	230	190	210	230
325	230	175	240	185	225	202	215
225	260						

$$\text{Range} = 325 - 165 = 160$$

$$\text{Class width} = \frac{160}{8} = 20$$

Class Interval	Tally	F	F _c
165 - 185		4	4
186 - 206		5	10
207 - 227		5	24
228 - 248		5	38
249 - 269		5	47
270 - 280		1	48
281 - 311		1	49
312 - 332		1	50

(5)

10. Stories in the World's Tallest Buildings The number of stories in each of the world's 30 tallest buildings is listed below. Construct a grouped frequency distribution and a cumulative frequency distribution with 7 classes.

88	88	110	88	80
79	85	80	100	60
54	60	75	64	105
69	102	78	70	55
90	77	55	75	55
56	71	70	65	72

$$\text{Class width} = \frac{110 - 54}{7} = \frac{56}{7} = 8$$

Class Interval	F	F _c
54 - 62	1	1
63 - 71	5	12
72 - 80	9	21
81 - 89	11	25
90 - 98	1	26
99 - 107	3	28
108 - 116	1	30

11. GRE Scores at Top-Ranked Engineering Schools The average quantitative GRE scores for the top 30 graduate schools of engineering are listed. Construct a frequency distribution and a cumulative frequency distribution with 5 classes.

767	770	761	760	771
763	760	747	766	754
780	750	746	764	769
768	776	771	756	770
771	771	778	766	762
758	757	753	758	746

Class Interval	F	F _c
746 - 753	5	5
754 - 761	8	13
762 - 769	7	20
770 - 776	8	28
777 - 784	2	30

$$\text{Class width} = \frac{780 - 746}{5} = \frac{34}{5} = 6.8 \approx 7$$

12. Airline Passengers The number of passengers (in thousands) for leading U.S. passenger airlines in 2001 indicated below. Use the data to construct a grouped frequency distribution and a cumulative frequency distribution with a reasonable number of classes and comment on the shape of the distribution.

91,570	81,066	55,373
40,551	16,280	13,658
13,170	11,731	10,024
7,041	6,406	5,830
5,427	70,786	42,400
86,755	14,869	13,447
21,118	10,420	9,122
12,632	6,362	5,585
6,954		

Class Interval	F	F _c
5000 - 14,999	16	16
15000 - 24999	12	18
25000 - 34999	0	18
35000 - 44999	2	20
45000 - 54999	0	20
55000 - 64999	1	21
65000 - 74999	1	22
75000 - 84999	1	23
85000 - 94999	2	25

$$\rightarrow \text{Class width} = \frac{91570 - 5427}{17229} = 5$$

$$2^k \geq n \rightarrow 2^5 \geq 25$$

K = Number of classes

n = number of observations

$$\rightarrow \text{Class width} = \frac{\text{Range}}{K} = \frac{86143}{8} \approx 10,768 \text{ Round to convenient number: } 10,000$$

13. Ages of Declaration of Independence Singers

Construct a grouped frequency distribution and a cumulative frequency distribution for the data using 7 classes.

41	47	39	50	39	60
44	39	40	34	52	32
44	60	42	50	46	42
35	48	31	55	42	46
35	45	53	50	45	62
54	40	35	31	38	
52	50	30	68	33	
63	27	34	42	70	
43	46	27	63	33	
46	34	50	48	36	

Class Interval	F	F _c
27 - 33	8	8
34 - 40	14	22
41 - 47	14	36
48 - 54	12	48
55 - 61	2	50
62 - 68	3	53
69 - 75	2	55

$$\text{Class width} = \frac{70 - 27}{7} = \frac{43}{7} = 6.14$$

(7)

14. Online Gambling Below are listed the number of players playing a free online game at various times of the day. Construct a grouped frequency distribution and a cumulative frequency distribution with 6 classes.

3807	3446	3194	3898	3344
3628	2313	3739	2101	3647
3640	2537	3886	1525	
3839	2037	3698	2311	

$$\text{Class width} = \frac{3807 - 1525}{6} = 397$$

Let's choose a reasonable number 400.

Class Interval	F	F _c
1500 - 1800	1	1
1801 - 2100	2	3
2101 - 2400	3	6
2401 - 2700	0	6
2701 - 3000	3	9
3001 - 3300	8	17
3301 - 3600	1	18

15. Presidential Vetoes

Use the data to construct a grouped frequency distribution with 5 classes. What is challenging about this set of data?

44	170	42	82
39	44	6	86
37	632	280	181
21	30	43	66
31	78	44	37

$$\text{Class width} = \frac{632 - 6}{20} = \frac{626}{20} = 31.3$$

(d) class width = 30

→ There is an outlier 632 that affects the whole distribution of the data set.

Class Interval	F	C.I	F
5 - 35	4	563 - 583	0
36 - 66	10	584 - 624	0
67 - 97	2	625 - 655	1
98 - 128	0		
129 - 159	0		
160 - 190	2		
191 - 221	0		
222 - 252	1		
253 - 283	0		
284 - 314			
315 - 345			
346 - 376			
377 - 407	0		
408 - 438			
439 - 469			
470 - 500			
501 - 531			
532 - 562	0		

(8)

16. U.S National Park Acreage

Construct a grouped frequency distribution for the data using 8 classes.
also cumulative frequency.

41	66	233	775	169
36	338	233	236	64
183	61	13	308	77
520	77	27	217	5
680	462	106	52	52
505	84	75	265	402
186	70	132	28	220
760	143	46	539	

$$C.W = \frac{775 - 5}{8} = \frac{770}{8} = 96.25$$

Let's choose 100 for C.W

Class interval	F	F _c
5 - 105	17	17
106 - 206	6	23
207 - 307	6	29
308 - 408	3	32
409 - 509	2	34
510 - 610	2	36
611 - 711	1	37
712 - 812	2	39

17. Height of Alaskan Volcanoes

Construct a grouped frequency distribution and a cumulative frequency distribution for the data using 10 classes.

4265	3545	4025	7050	11413
3490	5370	4885	5030	6830
4450	5775	3845	7545	8450
3895	10140	6050	10265	6965
150	8185	7295	2015	8085
5315	2845	6720	3465	1880
2860	4450	2759	9430	
7985	7840	3540	11070	
5710	885	8860	7015	

$$\text{Range} = 11413 - 150 = 11263$$

$$\text{Class W} = \frac{11263}{10} = 1126.3$$

Let's choose 1130

Class Interval	F	F _c
150 - 1280	2	2
1281 - 2411	2	4
2412 - 3542	6	10
3543 - 4673	7	17
4674 - 5804	7	24
5805 - 6935	4	28
6936 - 8066	6	34
8067 - 9197	3	37
9198 - 10328	3	40
10329 - 11459	2	42

18. Home Run Record Breakers

Construct a grouped frequency distribution and a cumulative frequency distribution for each player, using 8 classes.

McGwire							Sosa						
306	310	310	430	408	368		371	350	430	410	370	433	
420	340	460	410	390	500		430	434	370	430	430	434	
440	410	380	360	470	458		440	410	410	460	414	440	
350	527	380	550	430	385		400	430	410	370	480	370	
478	420	390	420	420	400		400	410	380	340	410	390	
425	370	480	390	377	460		370	420	410	415	433	344	
430	388	423	410	385	460		350	380	380	366	483		
360	410	480	350	510	380		430	380	390	400	405		
450	430	461	430	430	410		500	430	450	440	480		
470	440	400	390	341	440		364	420	350	410	420		
510	430	450	482	380			365	380	390	400	388		
420	380	470	398	370			400	420	360	368	364		

$$C.W = \frac{550 - 366}{7} = \frac{244}{7} = 34.8 \approx 35$$

Let's choose $C.W = 30$

Class Interval	Tally	F	f_2
306 - 336	I	1	1
337 - 367	I	6	7
368 - 398		15	22
399 - 429		17	39
430 - 460		20	59
461 - 491	I	6	65
492 - 522		3	68
523 - 553		2	70

(McGwire)

Class Interval	Tally	F	f_2
344 - 364		8	8
365 - 385		15	23
386 - 406		8	31
407 - 427		15	46
428 - 448		13	59
449 - 469		3	62
470 - 490		3	65
491 - 511	I	1	66

(Sosa)

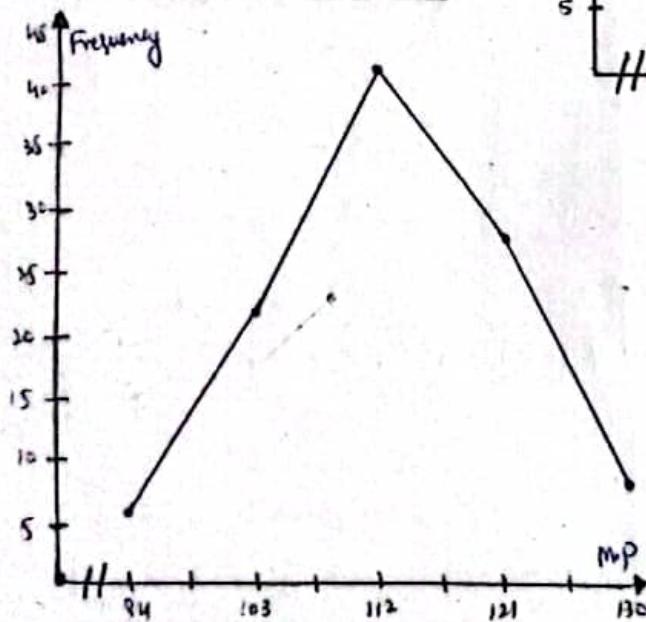
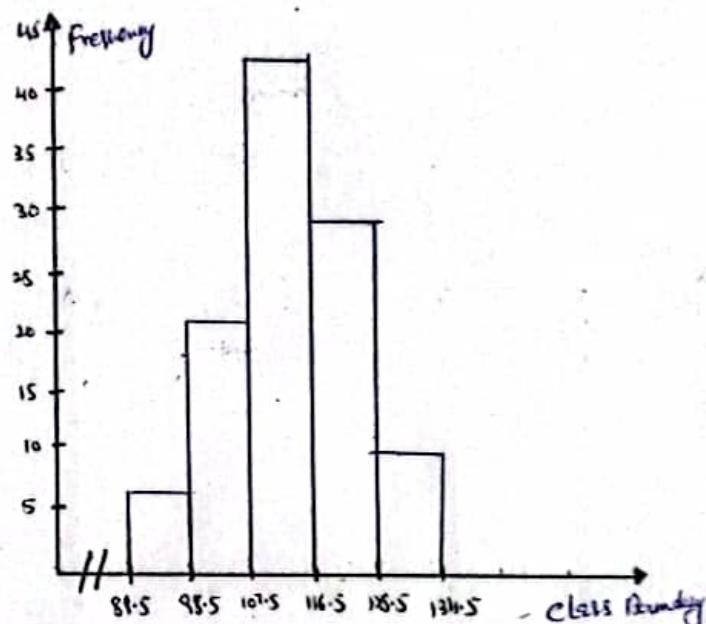
Exercise 2-2 Histograms, Frequency Polygons, and Ogives

1. Do Students Need Summer Development? For 108 randomly selected college applicants, the following frequency distribution for entrance exam scores was obtained. Construct a histogram, frequency polygon, and ogive for the data.

Class Limits	Frequency
90 - 98	6
98 - 107	22
108 - 116	43
117 - 125	28
126 - 134	9

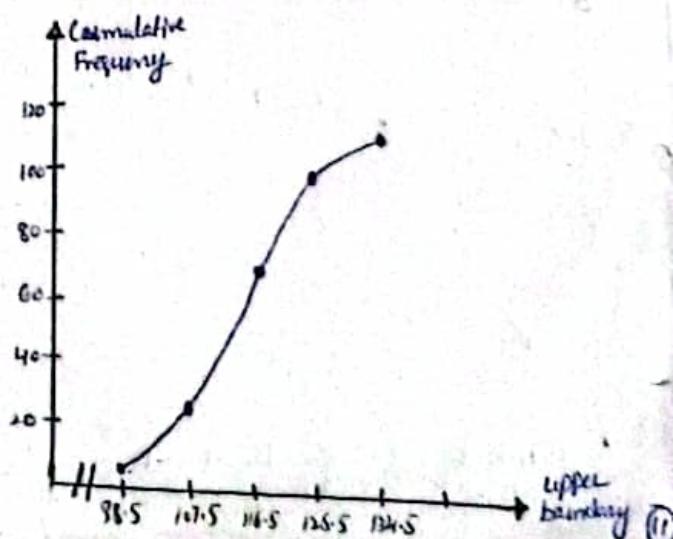
Applicants who score above 107 need not enroll in a summer development program. In this group, how many students do not have to enroll in the development program?

Class Boundaries	Frequency
89.5 - 98.5	6
98.5 - 107.5	22
107.5 - 116.5	43
116.5 - 125.5	28
125.5 - 134.5	9



Class Boundaries	Middle Points	F	F _c
89.5 - 98.5	94	6	6
98.5 - 107.5	103	22	28
107.5 - 116.5	112	43	71
116.5 - 125.5	121	28	99
125.5 - 134.5	130	9	108

Class Boundaries	Middle Points	F	F _c
89.5 - 98.5	94	6	6
98.5 - 107.5	103	22	28
107.5 - 116.5	112	43	71
116.5 - 125.5	121	28	99
125.5 - 134.5	130	9	108



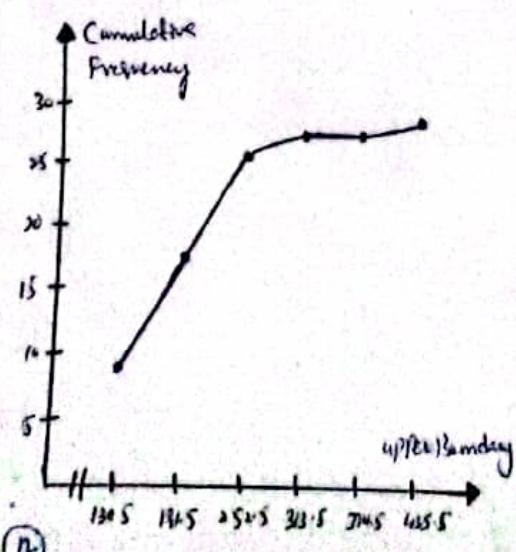
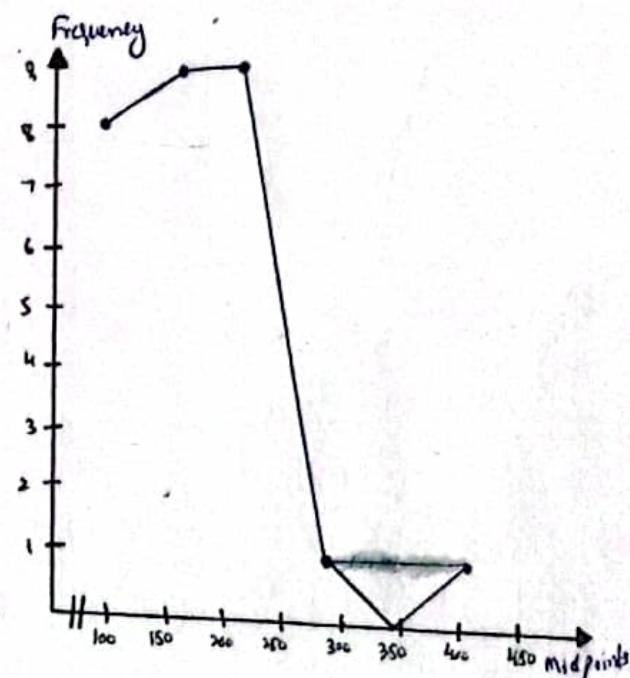
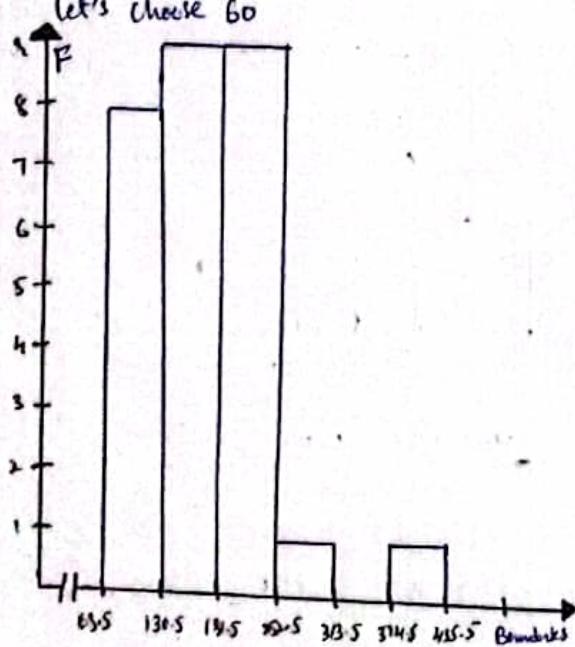
2. Number of College Faculty The number of faculty listed for a variety of private colleges which offer only Bachelor's degree is listed below. Use these data to construct a frequency distribution with 6 classes, a histogram, a frequency polygon, and an ogive. Discuss the shape of this distribution. What proportion of schools have 180 or more faculty?

165	218	138	77
10	207	155	204
176	225	83	116
221	128	135	135
221	206	82	
110	154	389	
162	214	224	
161	370	120	

$$C.W = \frac{389 - 70}{6} = \frac{319}{6} \approx 54$$

let's check 60

Limits	Boundaries	Middle Points	F	F _c
70 - 130	69.5 - 130.5	100	8	8
131 - 191	130.5 - 191.5	161	9	17
192 - 252	191.5 - 252.5	222	9	26
253 - 313	252.5 - 313.5	283	1	27
314 - 374	313.5 - 374.5	344	0	27
375 - 435	374.5 - 435.5	405	1	28

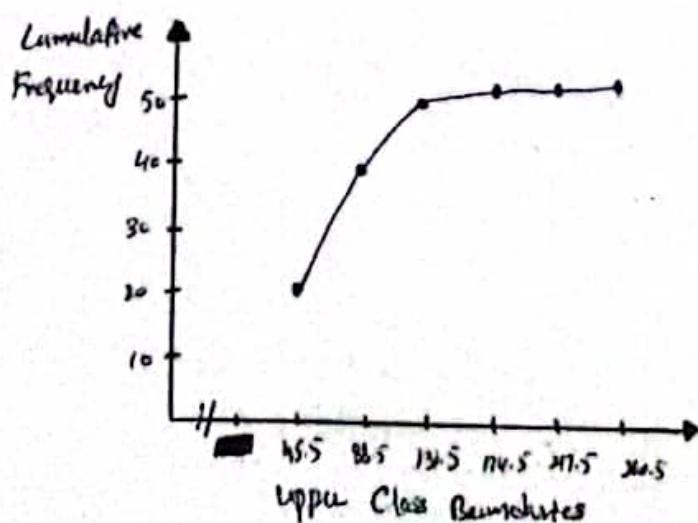
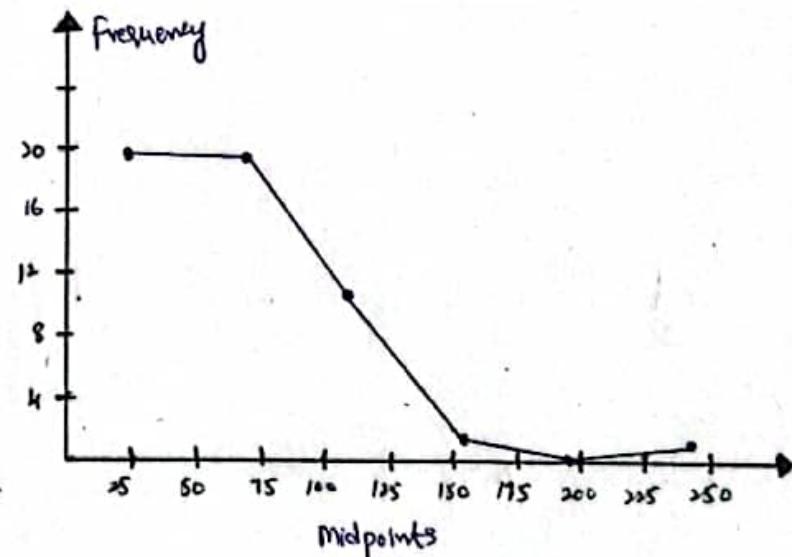
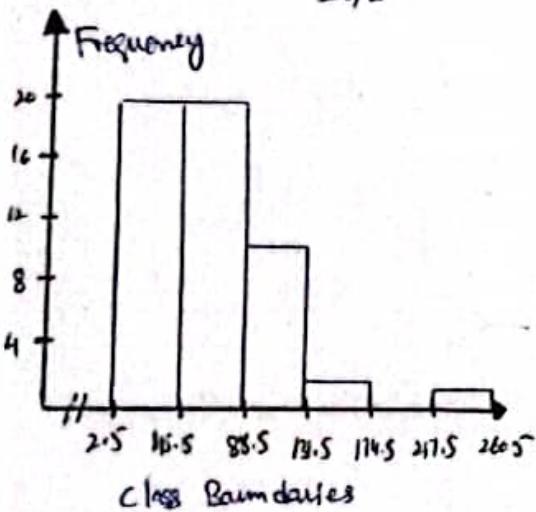


3. Counties, Divisions, or Parishes for 50 States The number of counties, division or parishes for each of the 50 states is given below. Use the data to construct a grouped frequency distribution with 6 classes, a histogram, a frequency polygon, and an ogive. Analyze the distribution.

67	15	58	8	159
102	82	105	64	23
83	82	56	16	21
62	53	77	67	46
95	29	95	55	23
27	75	64	67	5
44	98	120	16	14
87	114	93	10	33
100	88	36	5	66
254	14	38	72	3

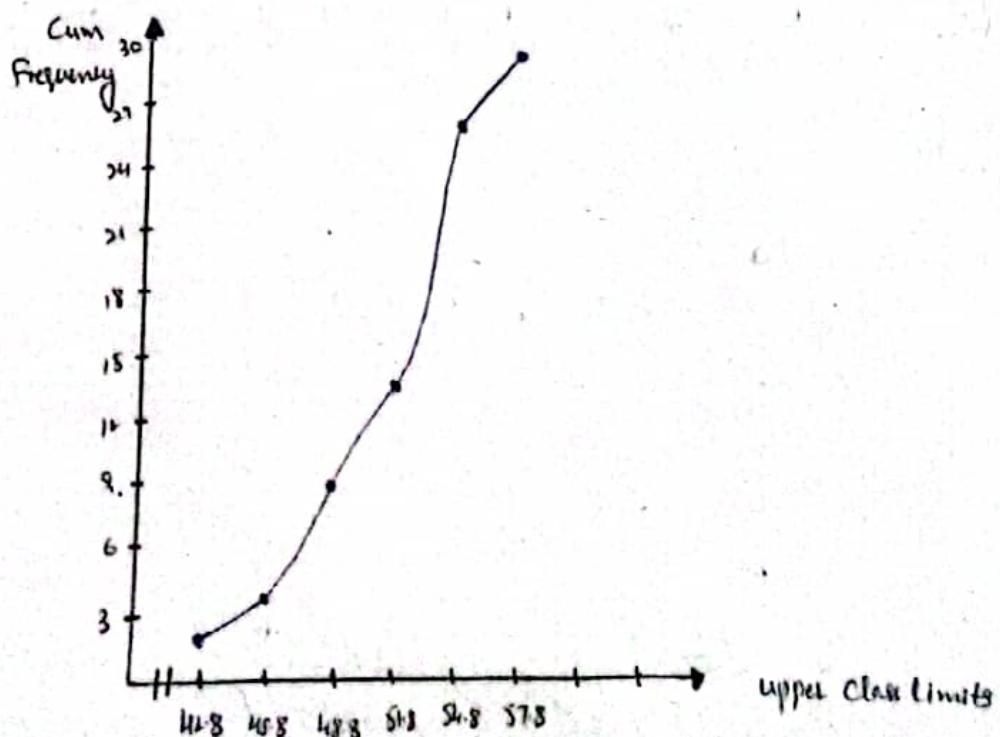
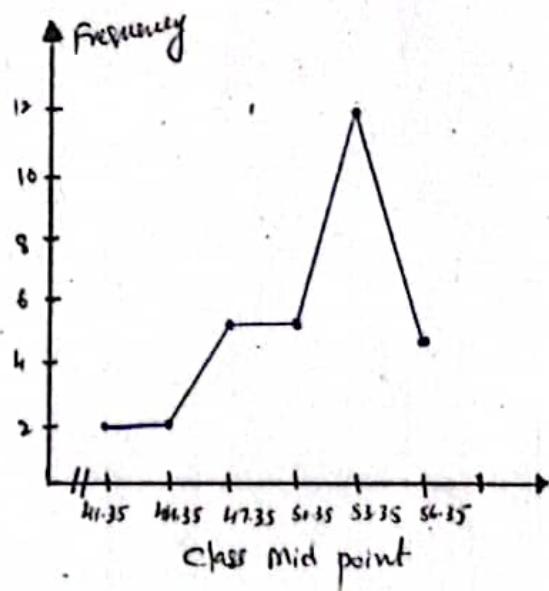
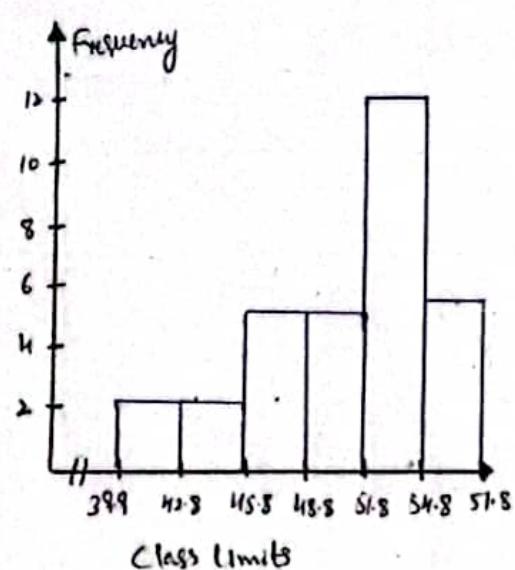
Limits	Boundaries	Midpoints	F	Fe
3-45	2.5-45.5	24	19	18
46-88	45.5-88.5	67	18	38
89-131	88.5-131.5	110	10	48
132-174	131.5-174.5	153	1	49
175-217	174.5-217.5	196	0	48
218-260	217.5-260.5	239	1	50

$$C.I.U = \frac{254-3}{6} = \frac{251}{6} \approx 41.83$$



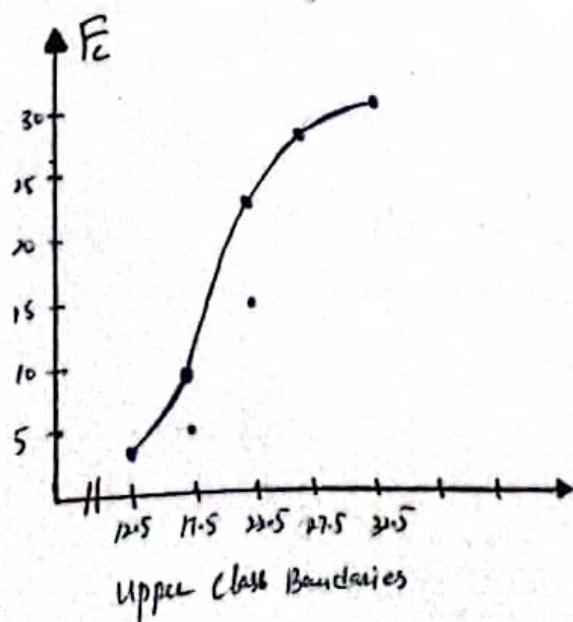
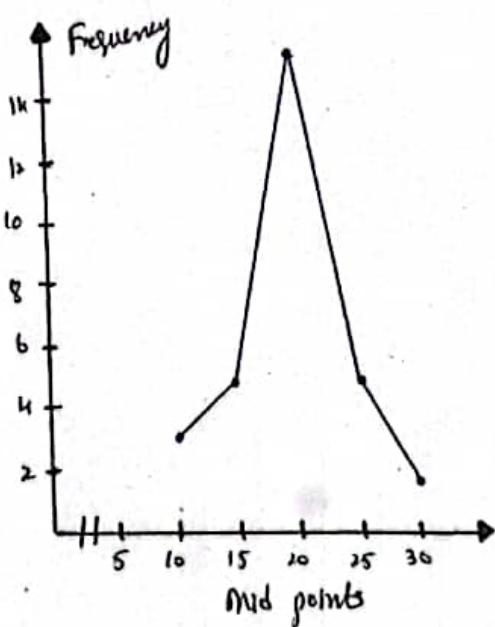
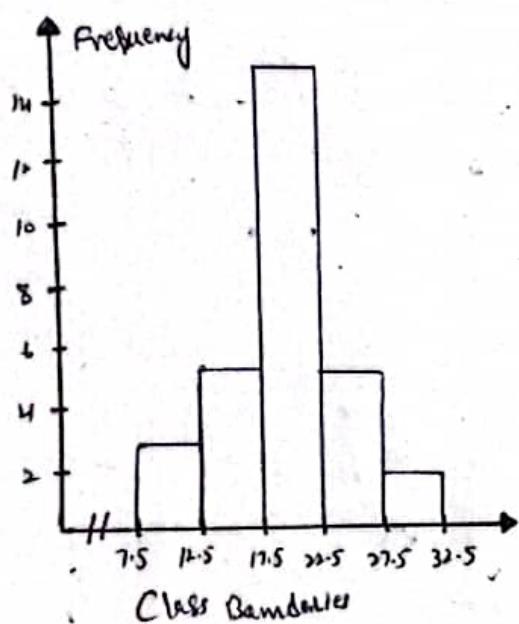
4. NFL Salaries Construct a histogram, a frequency polygon, and a cumulative frequency curve for the data; and comment on the shape of the distribution.

Class Limits	Frequency	Cum Frequency	Mid points
39.9 - 42.8	2	2	41.35
42.9 - 45.8	2	4	44.35
45.9 - 48.8	5	9	47.35
48.9 - 51.8	5	14	50.35
51.9 - 54.8	12	26	53.35
54.9 - 57.8	5	31	56.35



5. Automobile Fuel Efficiency. Thirty automobiles were tested for fuel efficiency, in miles per gallon (mpg). The following frequency distribution was obtained. Construct a histogram, a frequency polygon, and an ogive for data.

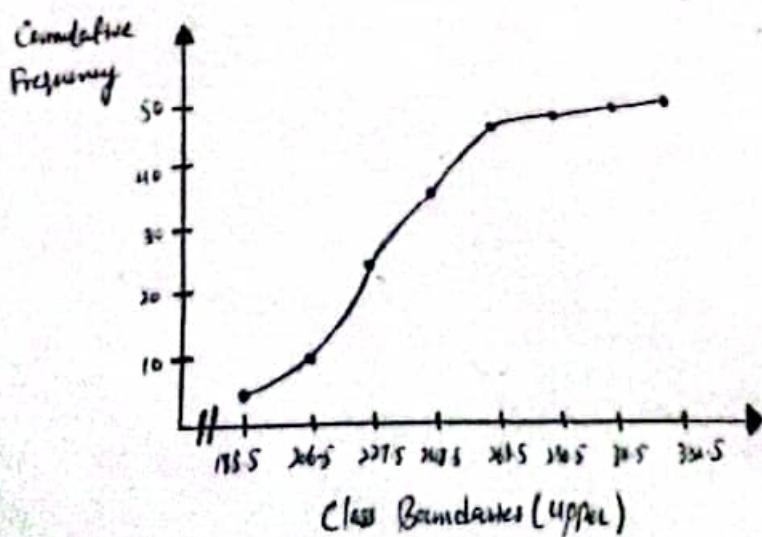
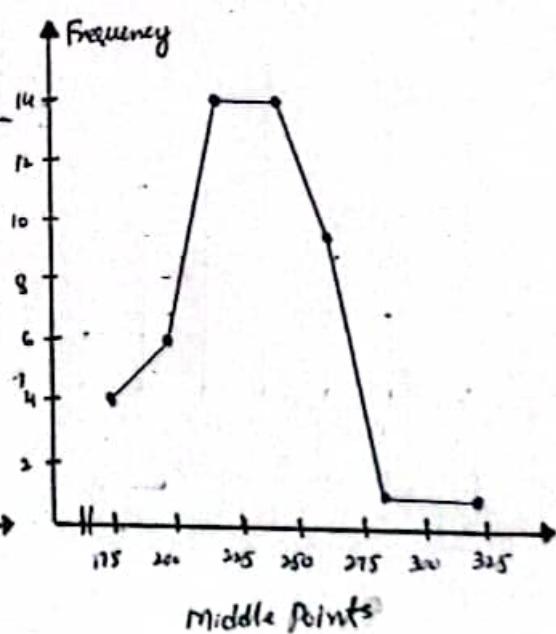
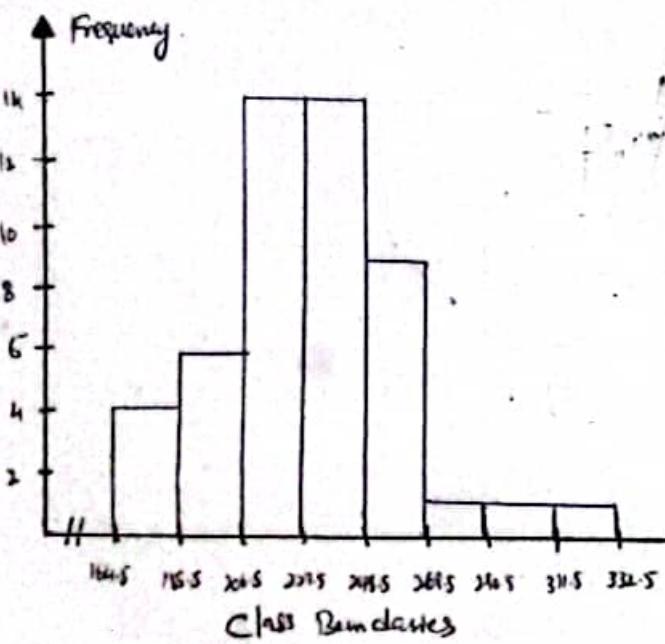
Class Boundaries	Frequency	F_c	Mid points
7.5 - 12.5	3	3	10
12.5 - 17.5	5	8	15
17.5 - 22.5	15	23	20
22.5 - 27.5	5	28	25
27.5 - 32.5	2	30	30



(13)

6. Construct a frequency histogram, a frequency polygon, and an ogive for the data in Exercise 8 in Section 2-1. Analyze the results. (8 classes)

Class Interval	Class Boundaries	Frequency	F_c	Mid Points
165 - 185	164.5 - 185.5	4	4	175
186 - 206	185.5 - 206.5	6	10	196
207 - 227	206.5 - 227.5	14	24	217
228 - 248	227.5 - 248.5	14	38	238
249 - 269	248.5 - 269.5	8	47	258
270 - 290	269.5 - 280.5	1	48	280
291 - 311	280.5 - 311.5	1	49	301
312 - 332	311.5 - 332.5	1	50	322



(16)

7. Air Quality Standards The number of days that selected U.S. metropolitan areas failed to meet acceptable air quality standard is shown below for 1998 and 2003. Construct grouped frequency distributions and a histogram for each set of data, and compare your results.

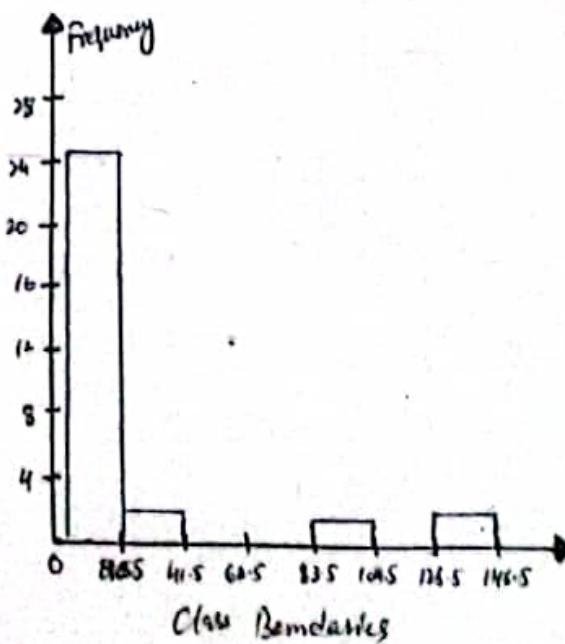
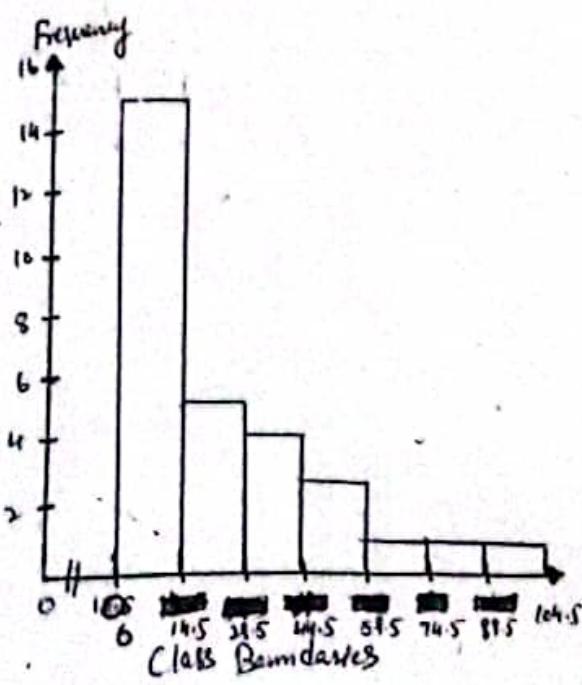
1998							2003						
43	76	51	14	0	10		10	11	14	20	15	6	
20	0	5	11	61	25		17	0	5	18	12	4	
38	0	56	8	0	9		31	5	88	1	1	16	
14	5	37	14	95	20		14	18	20	8	138	22	
23	12	33	0	3	45		13	10	20	20	20	12	

$$C.M = \frac{95-0}{7} = 13.57 \approx 14$$

$$C.M = \frac{138-0}{7} = 19.7 \approx 20$$

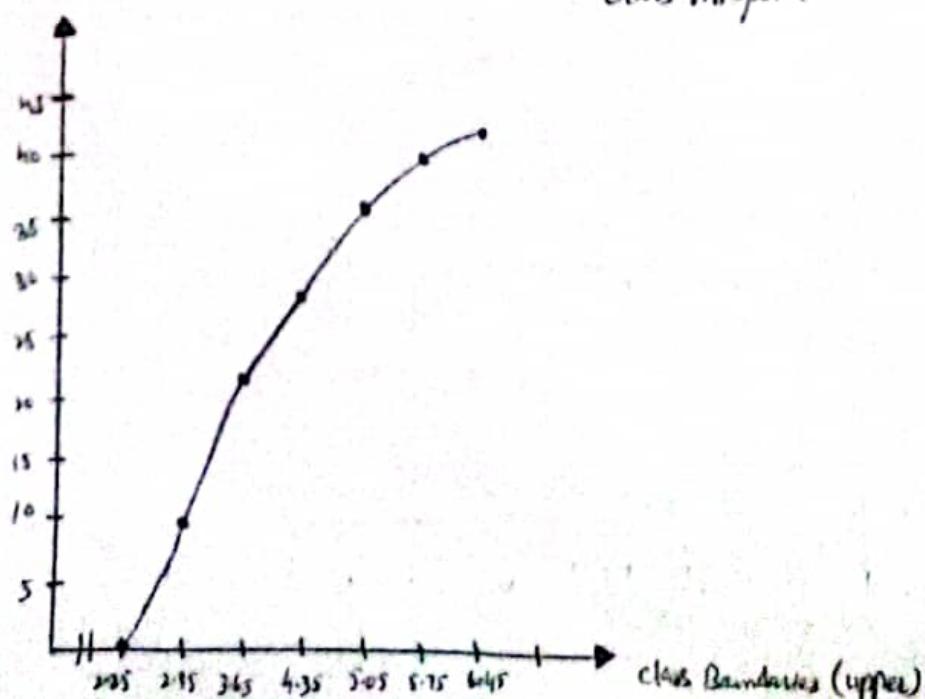
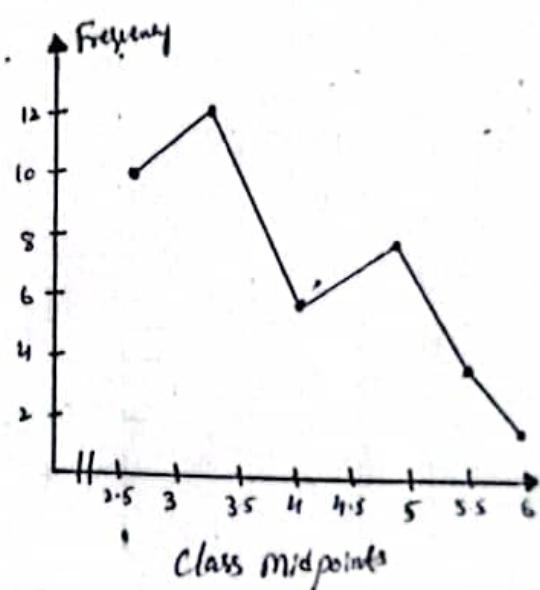
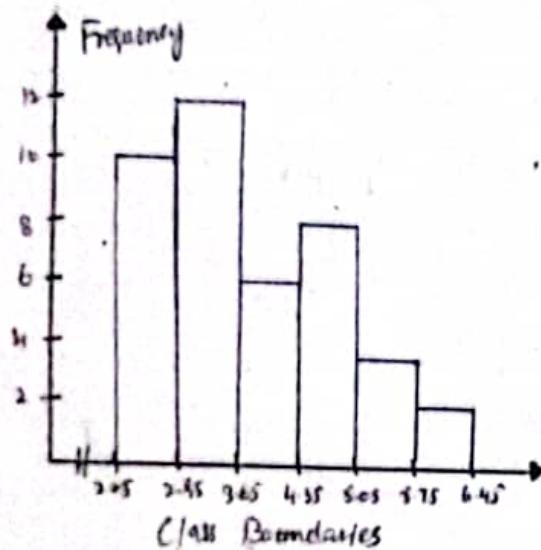
Class Limits	Tally	F
0-14	HHH	15
15-29		5
30-44	.	4
45-59		3
60-74		1
75-89		1
90-104		1

Class Limits	Tally	F
0-20	HHHH 	25
21-41		2
42-62		0
63-83		0
84-104		1
105-125		0
126-146		3



8. How Quick Are Dogs? In a study of reaction times of dogs to a specific stimulus, an animal trainer obtained the following data, given in seconds. Construct a histogram, a frequency polygon, and an ogive for the data; Analyse the results

Class limits	frequency	Class Boundaries	f_c	Mid points
2.3 - 2.9	10	2.25 - 2.95	10	2.6
3.0 - 3.6	12	2.95 - 3.65	22	3.3
3.7 - 4.3	6	3.65 - 4.35	28	4
4.4 - 5.0	8	4.35 - 5.05	36	4.7
5.1 - 5.7	4	5.05 - 5.75	40	5.4
5.8 - 6.4	2	5.75 - 6.45	42	6.05



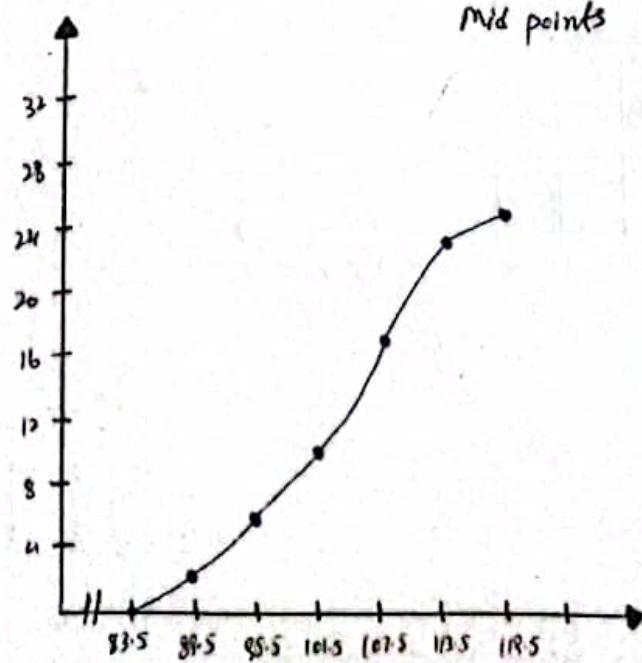
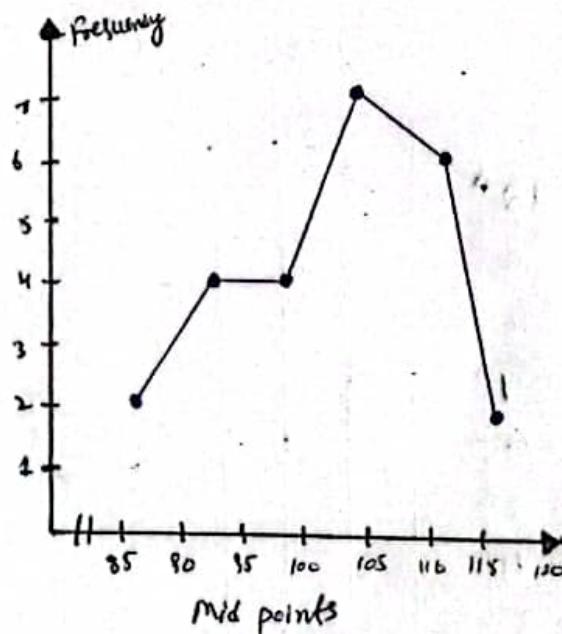
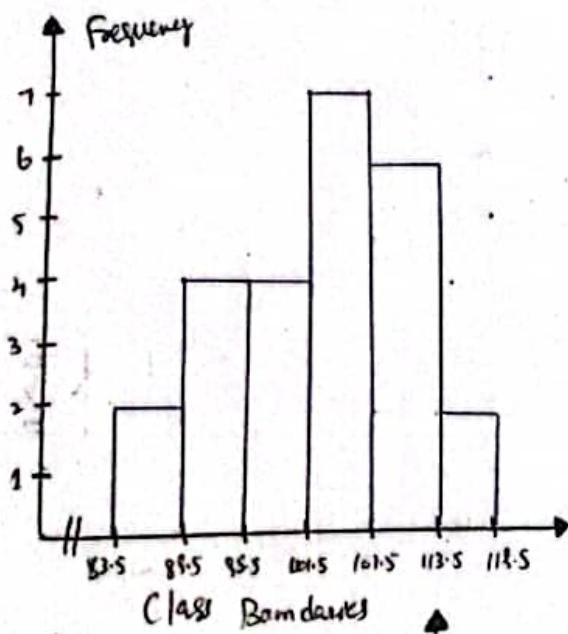
Q. Quality of Health Care The scores of health care quality as calculated by a professional risk management company are listed on the next page for selected states. Use the data to construct a frequency distribution, a histogram, a frequency polygon, and an ogive.

118.2	108.8	105.3	99.8	92.5
114.6	103.8	103.7	98.1	91.0
113.1	107.7	103.2	96.6	90.0
111.9	107.0	102.8	95.7	87.1
110.0	106.7	101.6	93.6	83.1

$$C.W = \frac{118.2 - 83.1}{6} = \frac{35.1}{6} = 5.85 \cong 5.9$$

Let's choose C.W = 5.8

Units	Boundaries	Mid-P	F	F _c
82.1 - 87	83.5 - 88.5	86.05	2	2
87.1 - 95	88.5 - 98.5	92.05	4	6
95.1 - 101	98.5 - 101.5	98.05	4	10
101.1 - 107	101.5 - 107.5	104.05	7	17
107.1 - 113	107.5 - 113.5	110.5	6	23
113.1 - 118	113.5 - 119.5	116.5	2	25

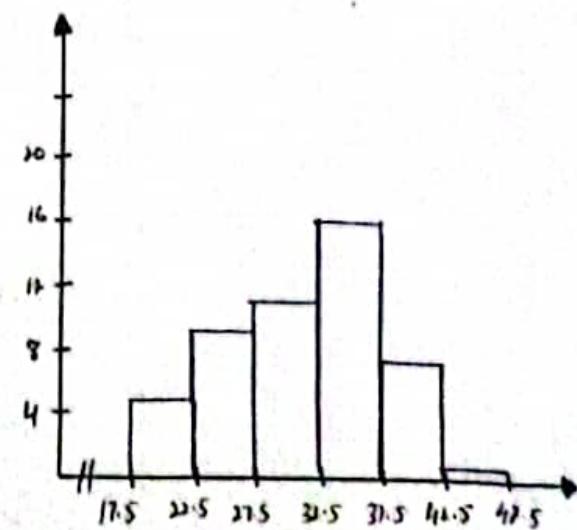
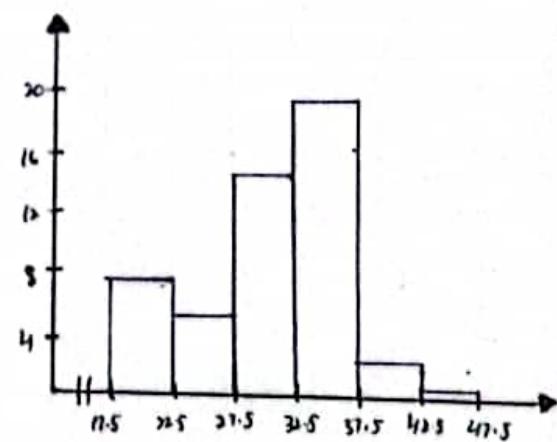
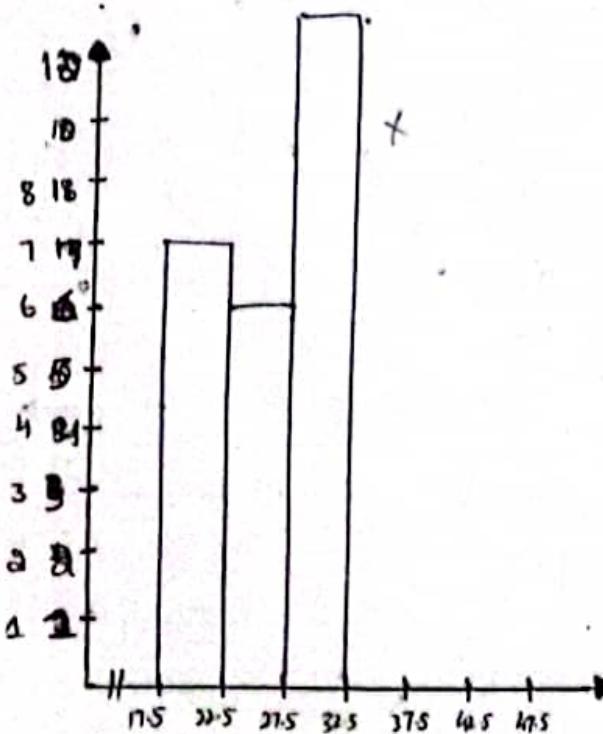


(13)

10. Making the Grade

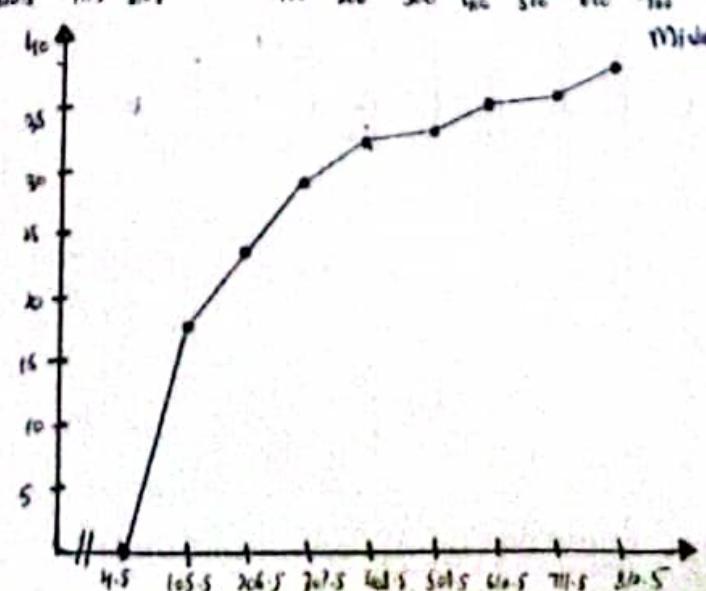
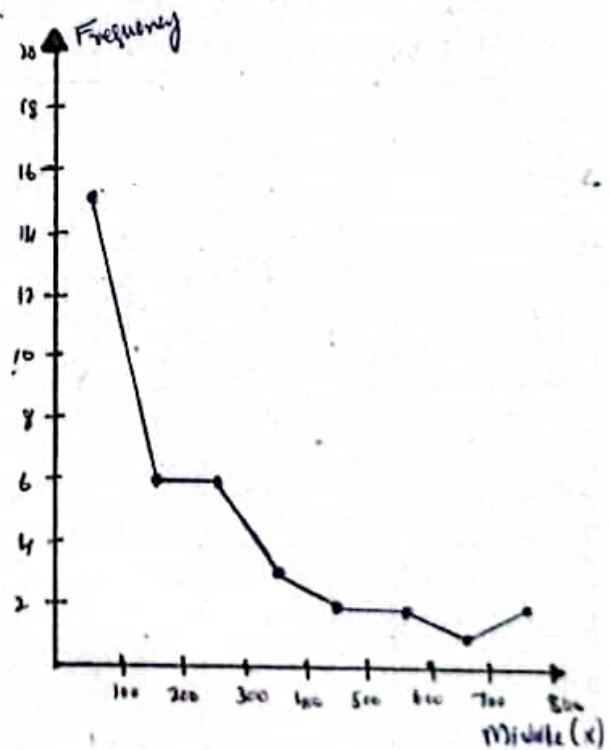
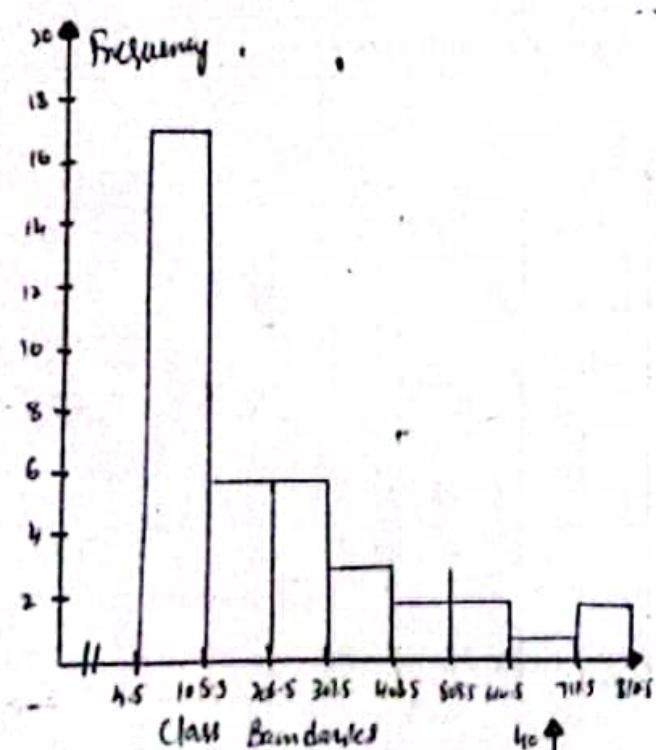
Draw histograms for each, and decide if there is any difference in the performance of the students in the subjects. The frequency dist shown indicate the percentages of public school students in fourth grade reading and mathematics who performed at or above the required proficiency levels for the 50 states in the U.S.

Class	Reading Frequency	Math-Freq
17.5 - 22.5	7	5
22.5 - 27.5	6	9
27.5 - 32.5	14	11
32.5 - 37.5	19	16
37.5 - 42.5	3	8
42.5 - 47.5	1	1



11. Construct a histogram, a frequency polygon, and an ogive for the data in Exercise 1b in Section 2-1, and analyze the results.

Class interval	Class Boundaries	Frequency	C = Frey (F_2)	Middle Points
5 - 105	4.5 - 105.5	17	17	55
106 - 206	105.5 - 206.5	6	23	156
207 - 307	206.5 - 307.5	6	28	257
308 - 408	307.5 - 408.5	3	32	358
409 - 509	408.5 - 509.5	2	34	458
510 - 610	509.5 - 610.5	2	36	560
611 - 711	610.5 - 711.5	1	37	661
712 - 812	711.5 - 812.5	2	38	762



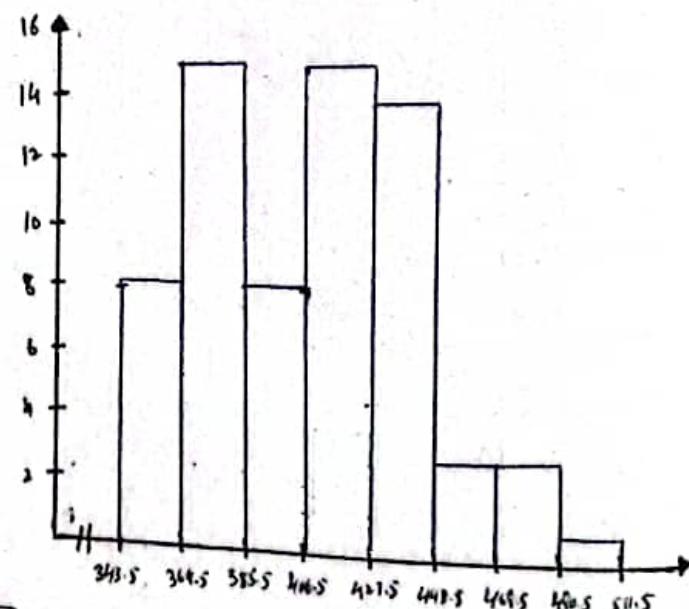
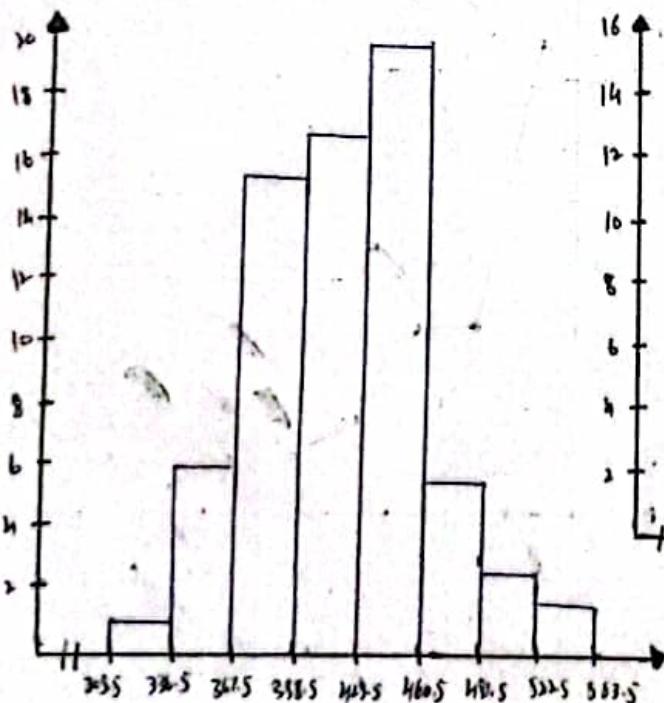
12. For the data in Exercise 18 in Section 2-1, construct a histogram for the home run distances for each player and compare them. Are they basically the same, or are there any noticeable differences?

McGwire

Limits	Boundaries	Middle-point	F	F _c
306 - 336	305.5 - 336.5	321	1	1
337 - 367	336.5 - 367.5	352	6	7
368 - 398	367.5 - 398.5	383	15	22
399 - 429	398.5 - 429.5	414	17	39
430 - 460	429.5 - 460.5	445	20	58
461 - 491	460.5 - 491.5	476	6	65
492 - 522	491.5 - 522.5	507	3	68
523 - 553	522.5 - 553.5	538	2	70

Sosa

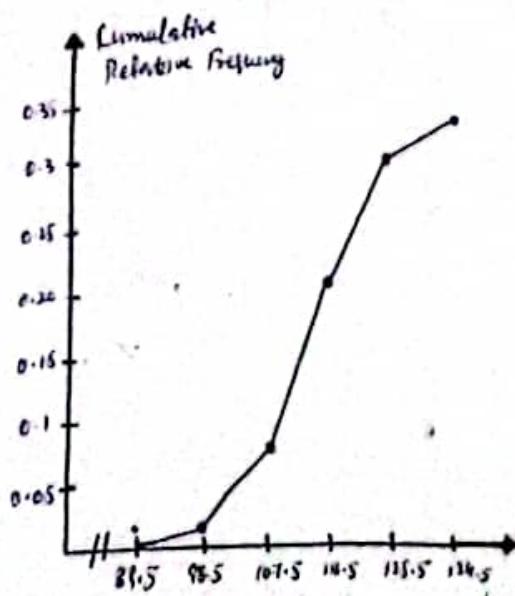
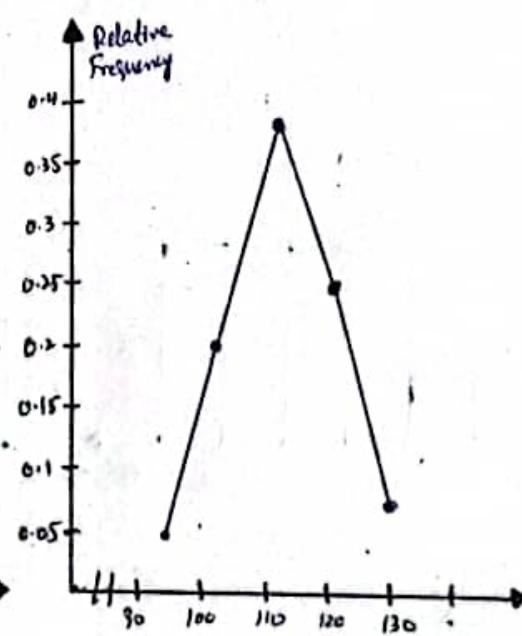
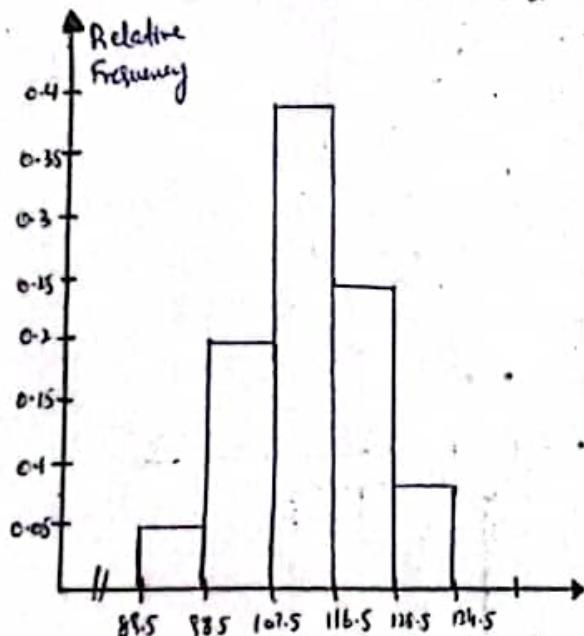
Limits	Boundaries	M. P	F	F _c
344-364	343.5 - 364.5	354	8	8
365-385	364.5 - 385.5	375	15	23
386-406	385.5 - 406.5	396	8	31
407-427	406.5 - 427.5	417	15	46
428-448	427.5 - 448.5	438	13	58
449-469	448.5 - 469.5	459	3	62
470-490	469.5 - 480.5	480	3	65
491-511	490.5 - 511.5	501	1	66



13. For the data in Exercise 1 in this Section, construct a histogram, a frequency polygon, and an ogive. (using relative frequencies) What proportion of the applicants needs to enroll in the summer development program?

Limits	Boundaries	F	F_c	Mid-P	Relative-F	Relative F_c
90 - 98	88.5 - 98.5	6	6	94	0.05	0.01
98 - 107	98.5 - 107.5	22	28	103	0.2	0.08
108 - 116	107.5 - 116.5	43	71	112	0.39	0.22
117 - 125	116.5 - 125.5	28	99	121	0.25	0.31
126 - 134	125.5 - 134.5	8	108	130	0.08	0.34

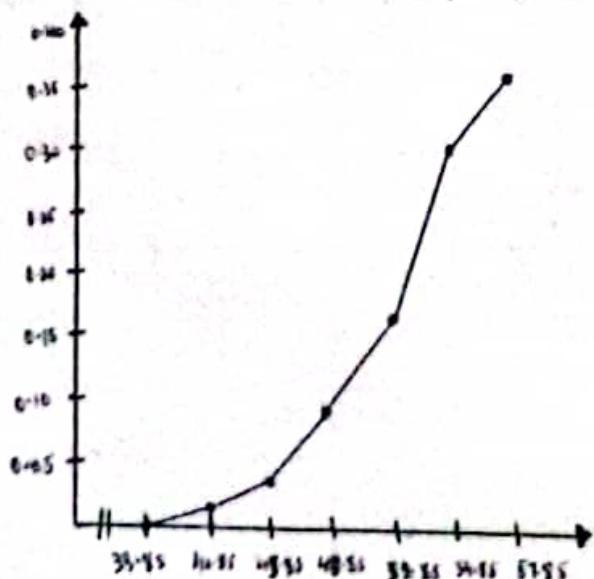
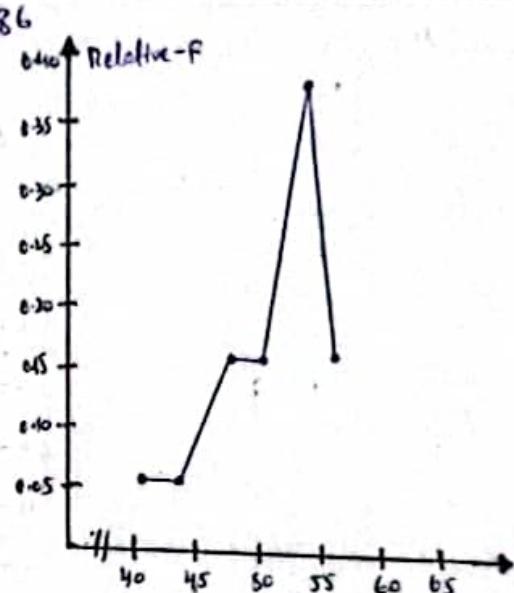
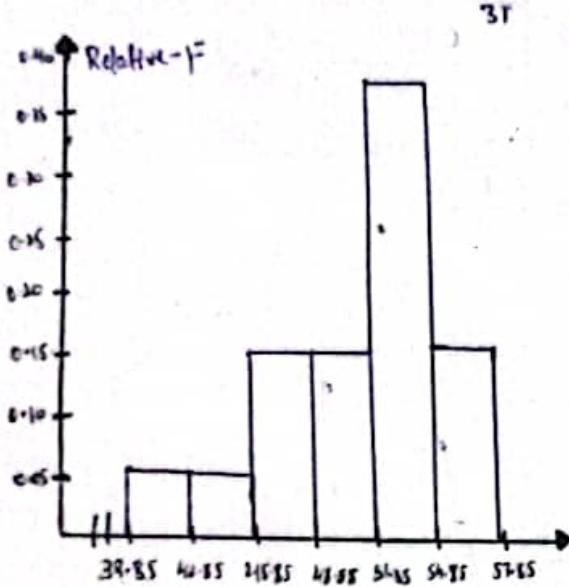
108. 312



(23)

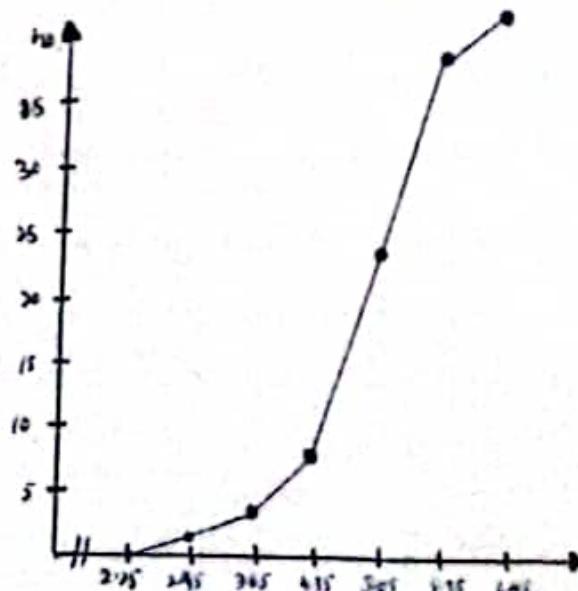
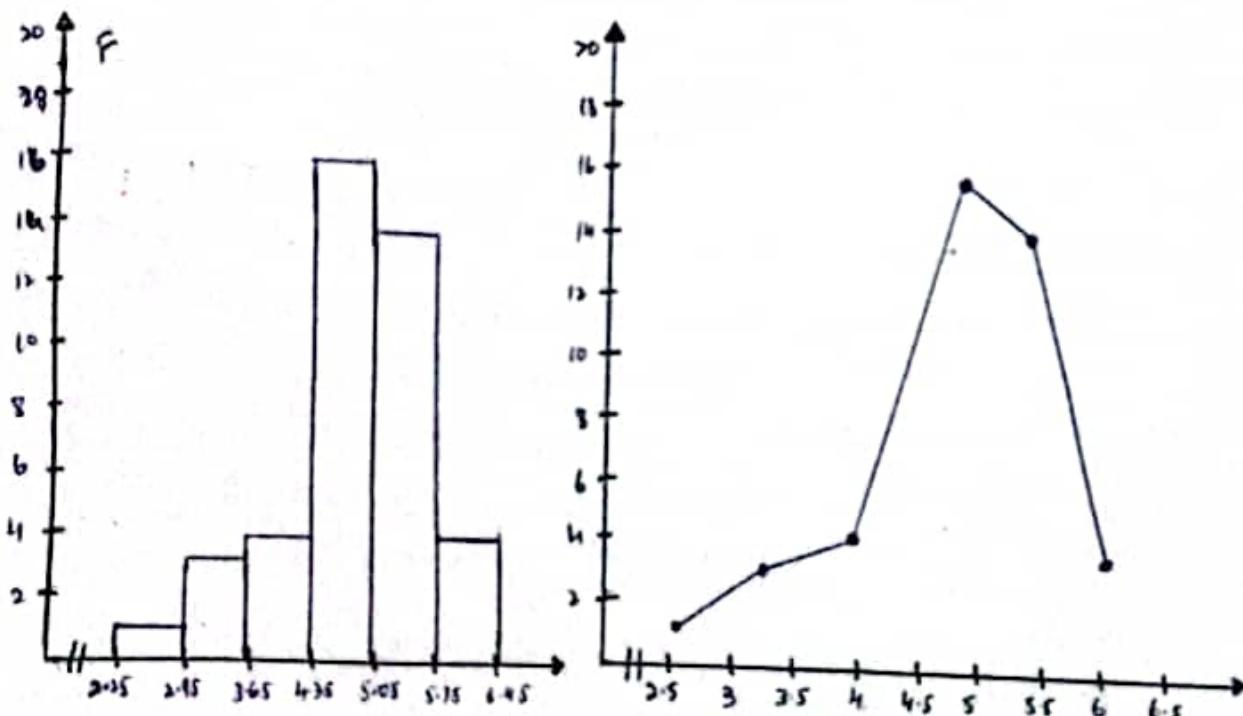
14. For the data for 2003 in Exercise 14 in this section, construct a histogram, a frequency polygon, and an ogive, using relative frequencies.

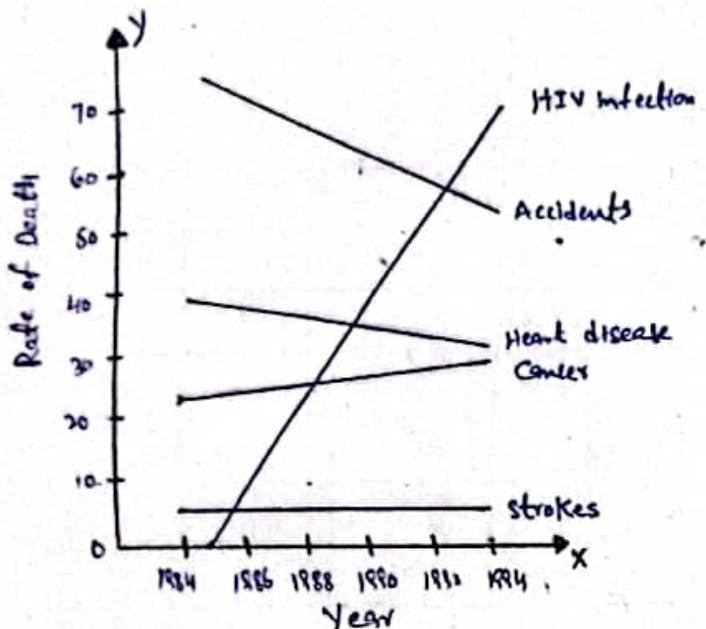
Units	Boundaries	F	F_c	Mid Point	Relative F	Relative F_c
39.9 - 42.8	39.95 - 42.85	2	2	41.35	0.06	0.02
42.9 - 45.8	42.85 - 45.85	2	4	44.35	0.06	0.04
45.9 - 48.8	45.85 - 48.85	5	9	47.35	0.16	0.10
48.9 - 51.8	48.85 - 51.85	5	14	50.35	0.16	0.16
51.9 - 54.8	51.85 - 54.85	12	26	53.35	0.38	0.30
54.9 - 57.8	54.85 - 57.85	5	31	56.35	0.16	0.36



18. How Quick Are Older Dogs? The animal trainer in Exercise 8 in this section selected another group of dogs who were much older than the first group and measured their reaction times to the same stimulus. Construct a histogram, a frequency polygon, and an ogive for the data.

Limits	Boundaries	F	F_c	Mid Points
2.3 - 2.9	2.25 - 2.95	1	1	2.6
3.0 - 3.6	2.95 - 3.65	3	4	3.3
3.7 - 4.3	3.65 - 4.35	4	8	4.0
4.4 - 5.0	4.35 - 5.05	16	24	4.7
5.1 - 5.7	5.05 - 5.75	14	38	5.4
5.8 - 6.4	5.75 - 6.45	4	42	6.1





3. What are the variables in the graph?

:- Rate and Year.

2. Are the variables qualitative or quantitative?

:- Quantitative because they are numerical.

3. Are the variables discrete or continuous?

:- The variable are continuous because the rate of death can take any value within a range, and years are represented on a continuous timeline.

4. What type of graph was used to display the data?

:- A line graph was used to display the data.

5. Could a Pareto chart be used to display the data?

:- No, Pareto chart is used with Qualitative data.

6. Could a pie chart be used to display the data?

:- NO, a pie chart is used to show proportions. (Qualitative data).

7. List some typical uses for, the time series chart?

Time Series charts are used to show how things change over time.

They are used to track patterns, like Stock prices, weather patterns, Sales over time, and health data.

Exercises 2-3

1. Women's Softball Champions The NCAA Women's Softball Division 1 champions since 1982 are listed below. Use the data to construct a pareto chart and a vertical bar graph.

82 UCLA	84 Arizona
83 Texas A&M	85 UCLA
84 UCLA	86 Arizona
85 UCLA	87 Arizona
86 Cal St-Fullerton	88 Fresno State
87 Texas A&M	89 UCLA
88 UCLA	90 Oklahoma
89 UCLA	91 Arizona
90 UCLA	92 California
91 Arizona	93 UCLA
92 UCLA	94 UCLA
93 Arizona	95 Michigan

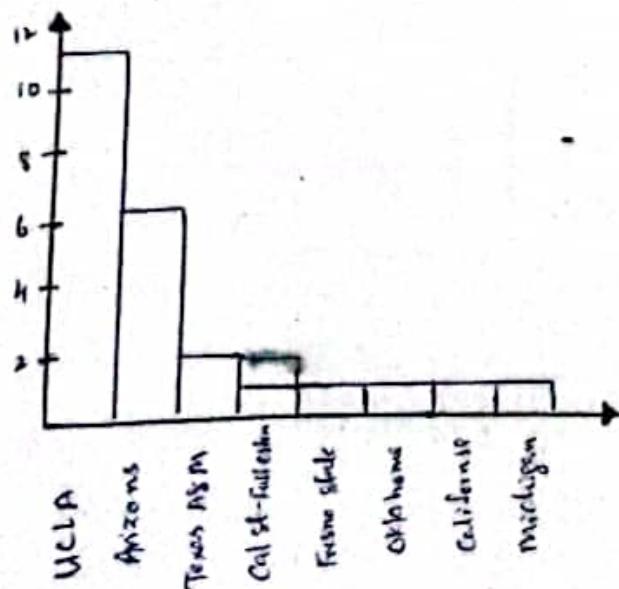
Class/Category	Tally	F
UCLA		5
Texas A&M		2
Cal St-Fullerton		1
Arizona		4
Fresno State		1
Oklahoma		1
California		1
Michigan		1

→ Pareto Chart: Used to show frequencies for nominal or qualitative variables

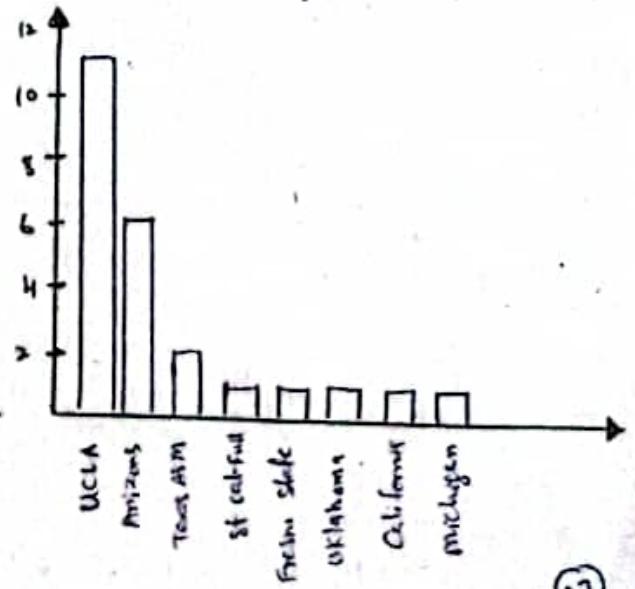
→ Bar Graph: A bar graph represents the data by using vertical or horizontal bars whose height or lengths represent the frequencies of data.

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

Pareto Chart

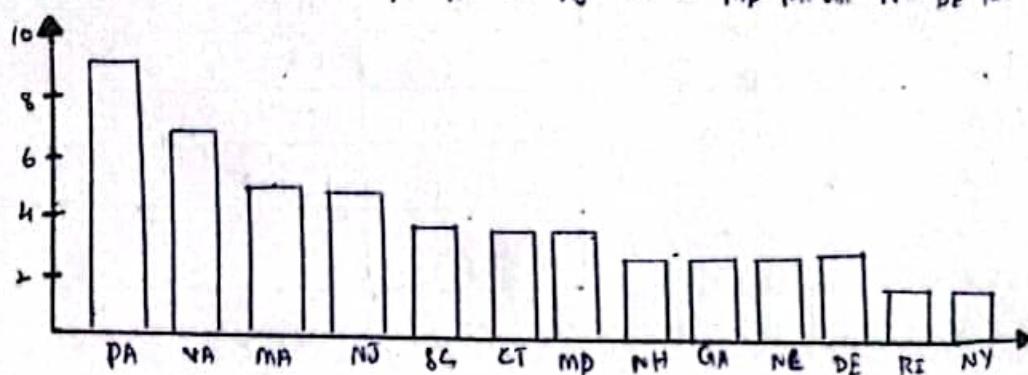
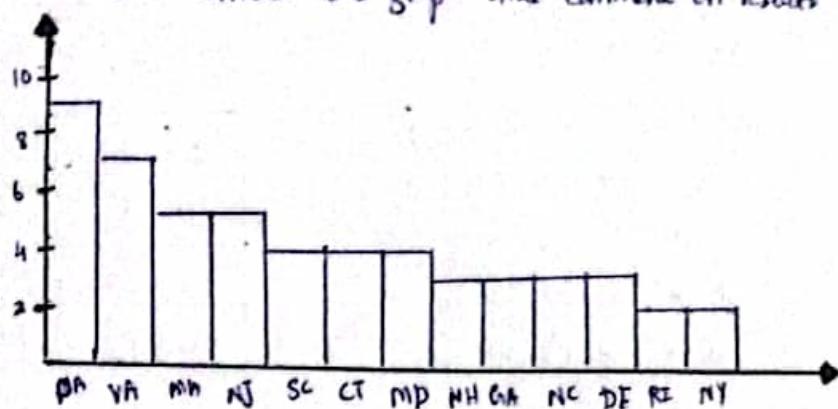


Bar Graph



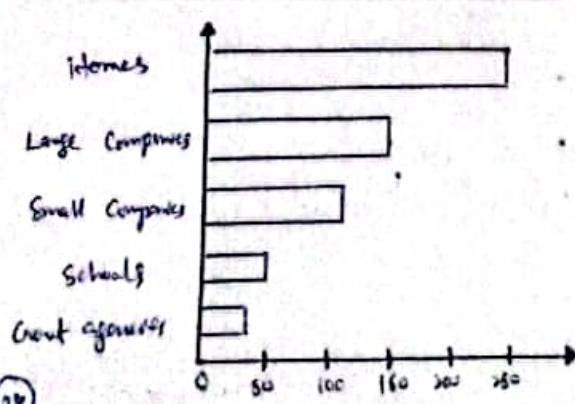
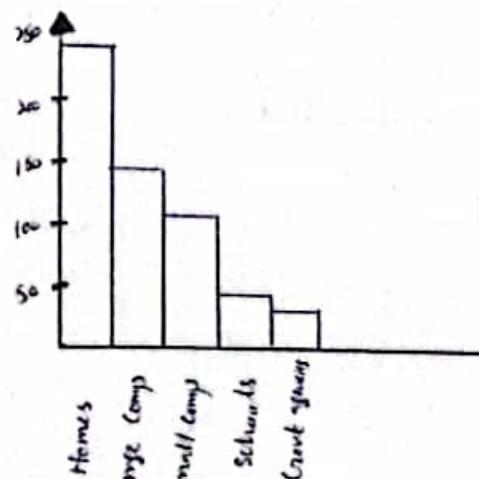
2. Delegates Who Signed - the Declaration of Independence The state represented by each delegate who signed the declaration of independence is indicated. Organize the data in a pareto chart and a vertical bar graph and comment on results

MA 5	PA 8	SC 4
NH 3	RI 2	CT 4
VA 7	NY 2	DE 3
MD 4	GA 3	
NJ 5	NC 3	



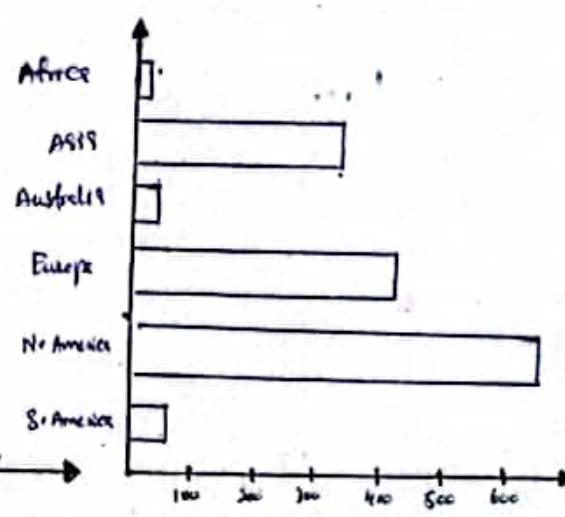
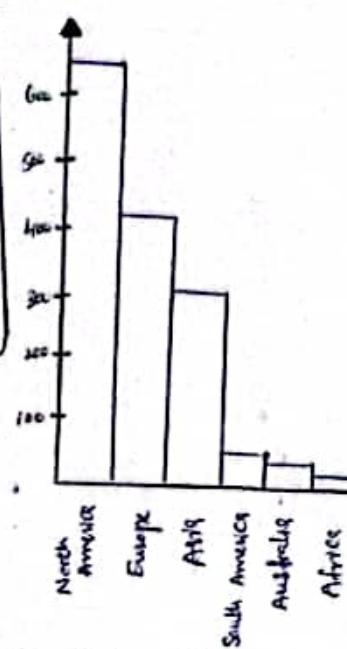
3. Internet Connections The following data represent the estimated number (in millions) of computers connected to the Internet worldwide. Construct a Pareto Chart and a horizontal bar graph for the data. Based on the data, suggest the best place to market appropriate Internet products.

Location	No of Computers
Homes	240
Small Companies	102
Large Companies	148
Government agencies	33
Schools	47



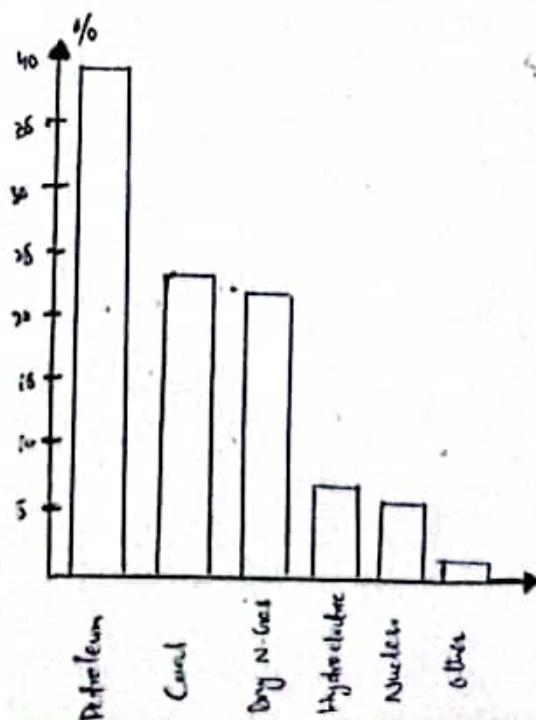
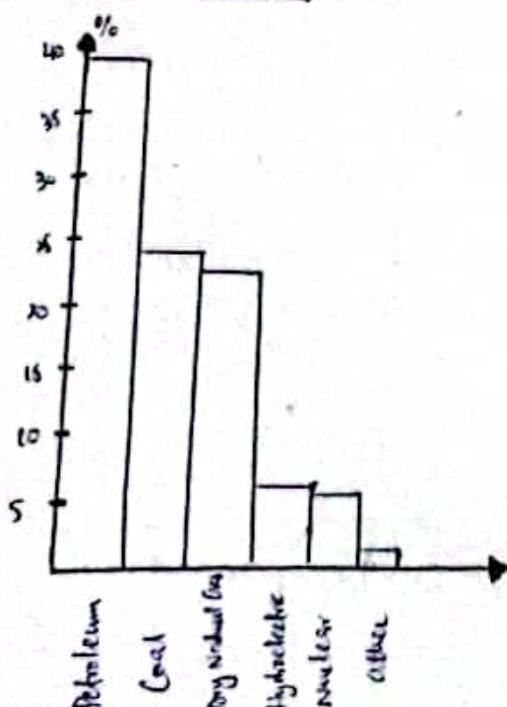
4. Roller Coaster Mania The World Roller Coaster Census Report lists the following number of roller coasters on each continent. Represent the data graphically, using a pareto chart and a horizontal bar graph.

Africa	17
Asia	315
Australia	22
Europe	413
North America	643
South America	45



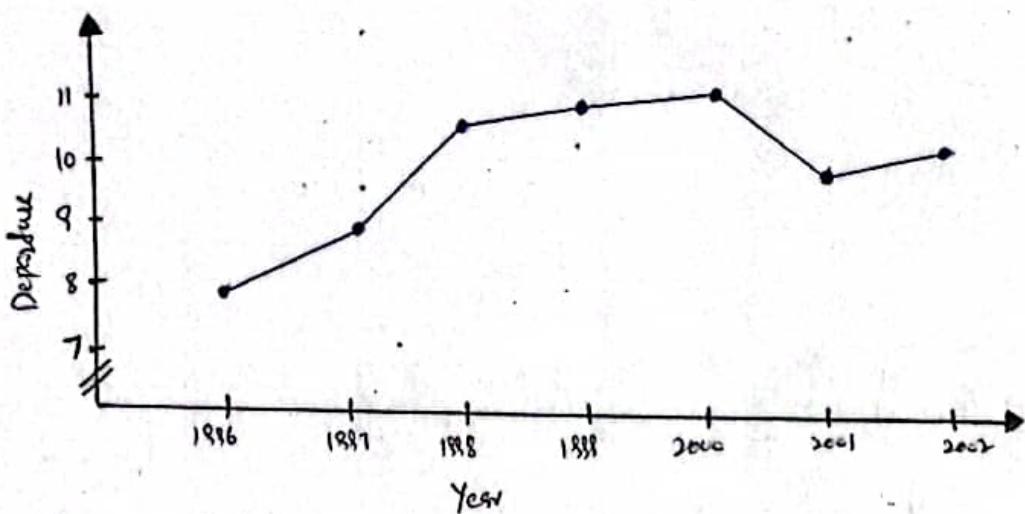
5. World Energy Use The following percentages indicate the source of energy used worldwide. Construct a pareto chart and a vertical bar graph for the energy used

Petroleum	38.8%
Coal	23.2
Dry Natural Gas	22.4
Hydroelectric	7.0
Nuclear	6.1
Other	1.2



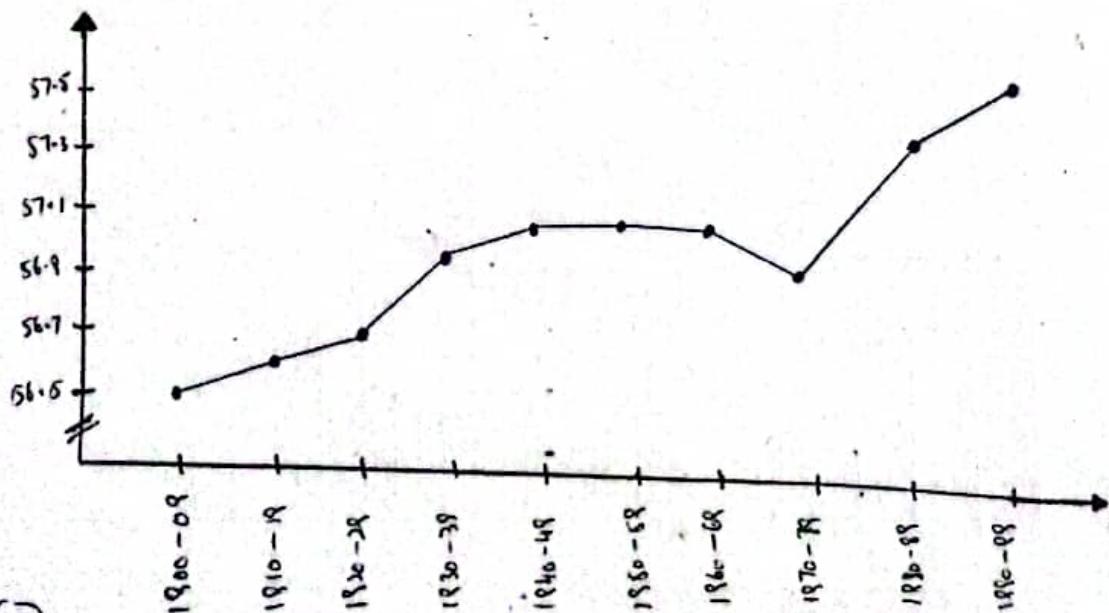
6. Airline Departures Draw a time series graph to represent the data for the number of airline departures (in millions) for the given years. Over the years, is the number of departures increasing, decreasing, or about the same?

Year	1996	1997	1998	1999	2000	2001	2002
Number of Departures	7.8	8.9	10.5	10.9	11.0	8.8	10.1



7. Average Global Temperatures Represent these average global temperatures in a time series graph.

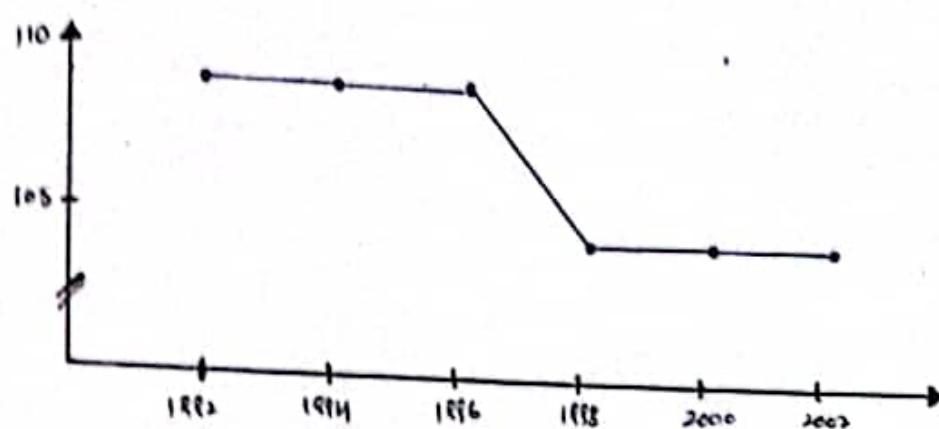
Year	1900-09	1910-19	1920-29	1930-39	1940-49	1950-59	1960-69	1970-79	1980-89	1990-99
Temperature	56.5	56.6	56.7	57.0	57.2	57.1	57.1	57.0	57.1	57.6



(30)

8. Nuclear Power Reactors Draw a time series graph for the data shown and comment on the trend. The data represent the number of active nuclear reactors.

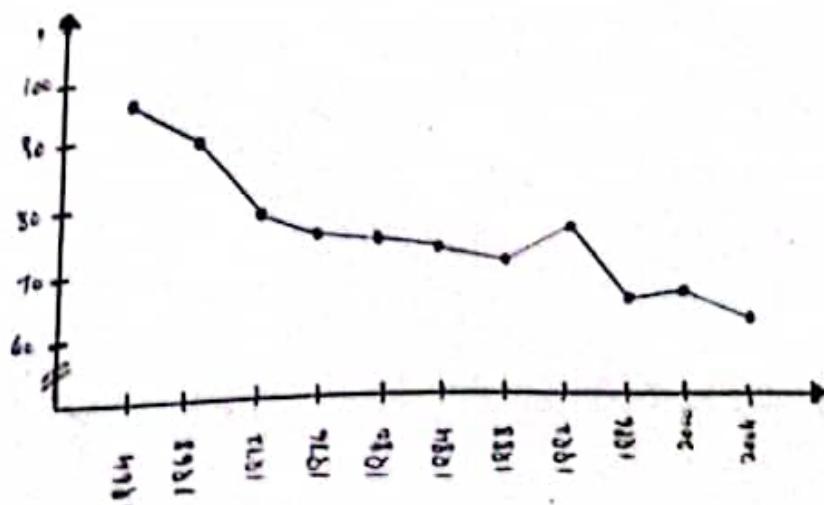
Year	1992	1994	1996	1998	2000	2002
Number	103	109	108	104	101	104



8. Percentage of Voters in Presidential Elections

Listed are the percentages of voters who voted in past presidential elections since 1964. Illustrate the data with a time series graph. The day before the 2006 election a website published a survey when 80% of the respondents said they voted in 2004 election. Give possible reasons for discrepancy.

1964	95.83	1980	76.53	1996	65.97
1968	88.65	1984	74.63	2000	61.50
1972	79.85	1988	72.48	2004	64.0
1976	77.64	1992	78.04		



(31)

10. Reasons We Travel The following data are based on a Survey from American Travel Survey on why people travel. Construct a pie graph for the data and analyze the results.

Purpose	Number
Personal business	146
Visit friends/relative	330
Work-related	225
Leisure	289
$n = 3000$	

$$\text{Step 1: } \frac{f}{n} \cdot 360^\circ$$

Personal business $\frac{146}{3000} \times 360^\circ = 52.56^\circ = 52.56^\circ$

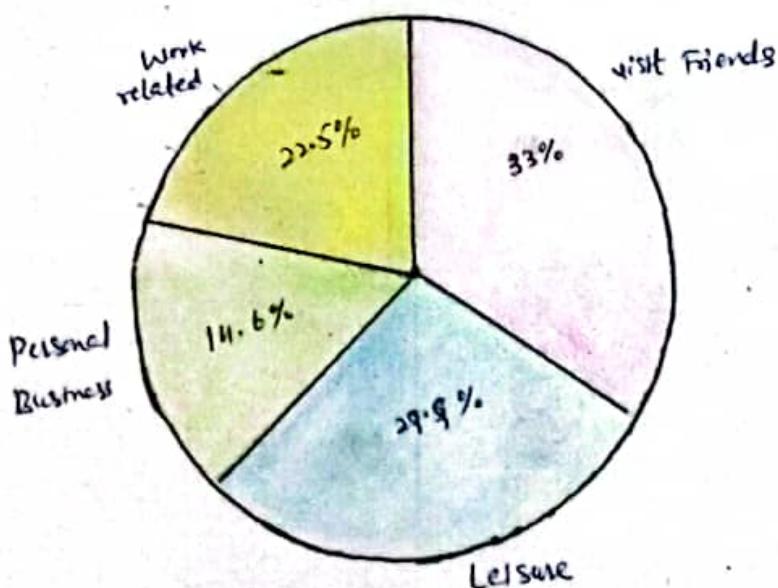
Visit friends $\frac{330}{3000} \times 360^\circ = 128^\circ = 118.8^\circ$

Work related $\frac{225}{3000} \times 360^\circ = 81^\circ$

Leisure $\frac{289}{100} \times 360^\circ = 107.64^\circ$

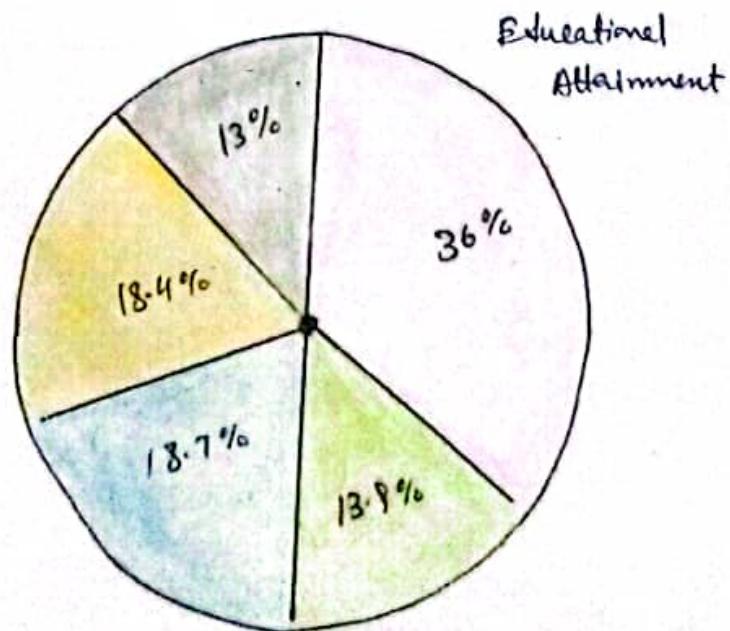
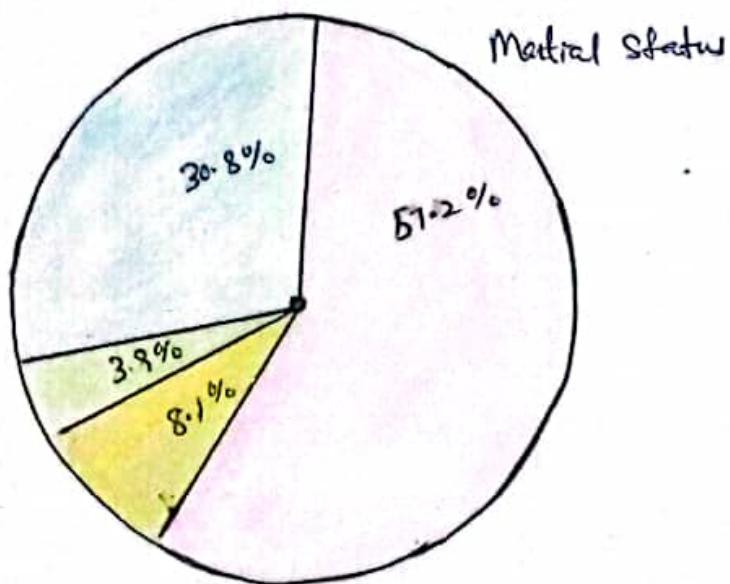
Purpose	Number	Percentage	Degree
Personal business	146	14.6%	52.56
visit friends	330	33.0%	118.8
work-related	225	22.5%	81
Leisure	289	28.9%	107.64

Reasons we Travel

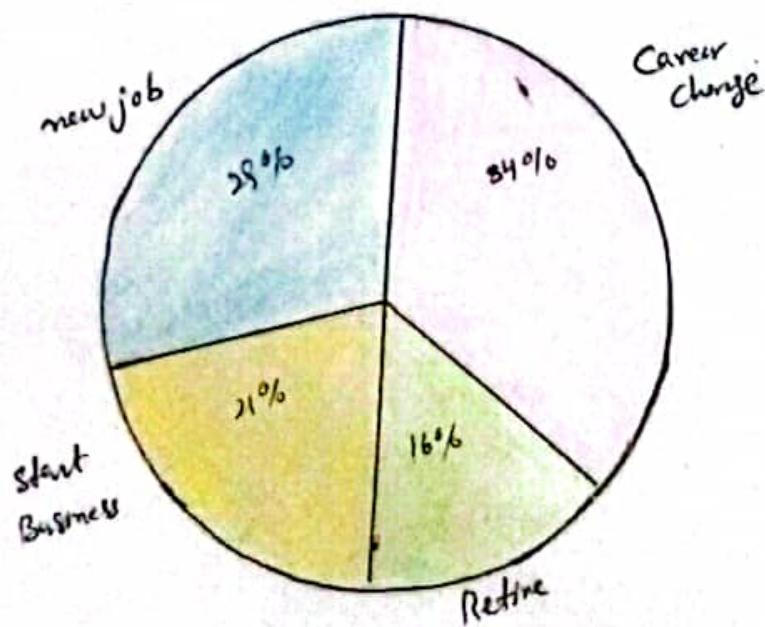
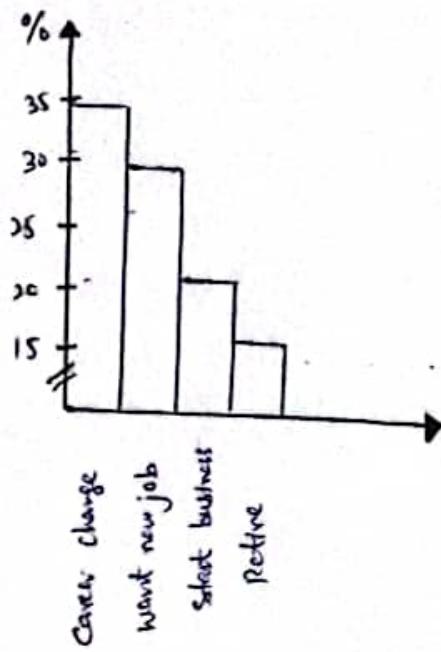


1.1. Characteristic of the Population 65 and over Two characteristics of the population aged 65 and over are shown below for 2004. Illustrate each characteristic with a pie graph.

<u>Marital Status</u>	<u>Educational attainment</u>
Never Married	3.9%
Married	57.2
Widowed	30.8
Divorced	8.1
	Less than ninth grade 13.9%
	Completed 9-12 but no diploma 13.0
	1+G graduate 36.0
	Some college/associates degree 18.4
	Bachelor's/Advanced degree 18.7



13. Workers Switch Jobs In a recent survey, 3 in 10 people indicated that they are likely to leave their jobs when the economy improves. Of those surveyed, 34% indicated that they would make a career change, 28% want a new job in the same industry, 21% are going to start a new business, and 16% are going to retire. Make a pie chart and a pareto chart for the data. Which chart do you think better represents the data?

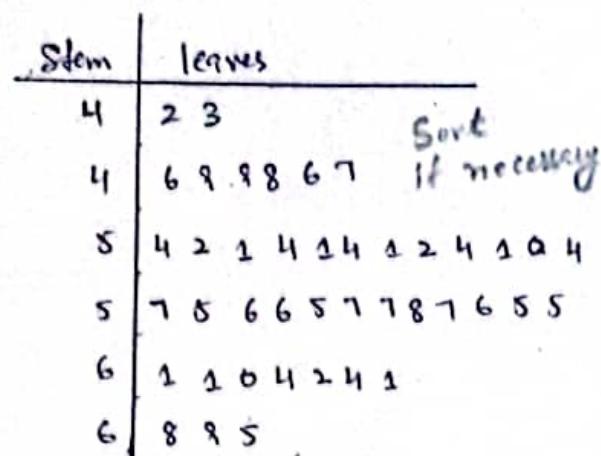


The pie chart better represent the data.

18. President's Ages at Inauguration

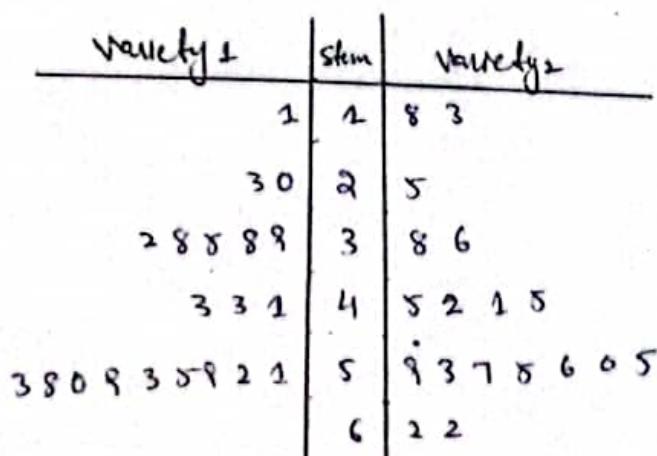
Construct a stem and leaf plot and analyze the data.

57	54	52	55	51	56
61	63	56	55	54	61
57	51	46	54	51	52
57	49	54	42	60	69
58	64	48	51	62	64
57	48	50	56	43	46
61	65	47	55	55	54



17. Twenty Days of plant Growth The growth (in centimeters) of two varieties of plant after 20 days is shown in this table. Construct back to back stem and leaf plot for the data, and compare the distributions.

Variety 1	Variety 2
20	18
12	45
41	62
43	59
51	53
59	53
55	59
50	42
58	55
35	56
38	38
23	41
32	50
43	62
53	45

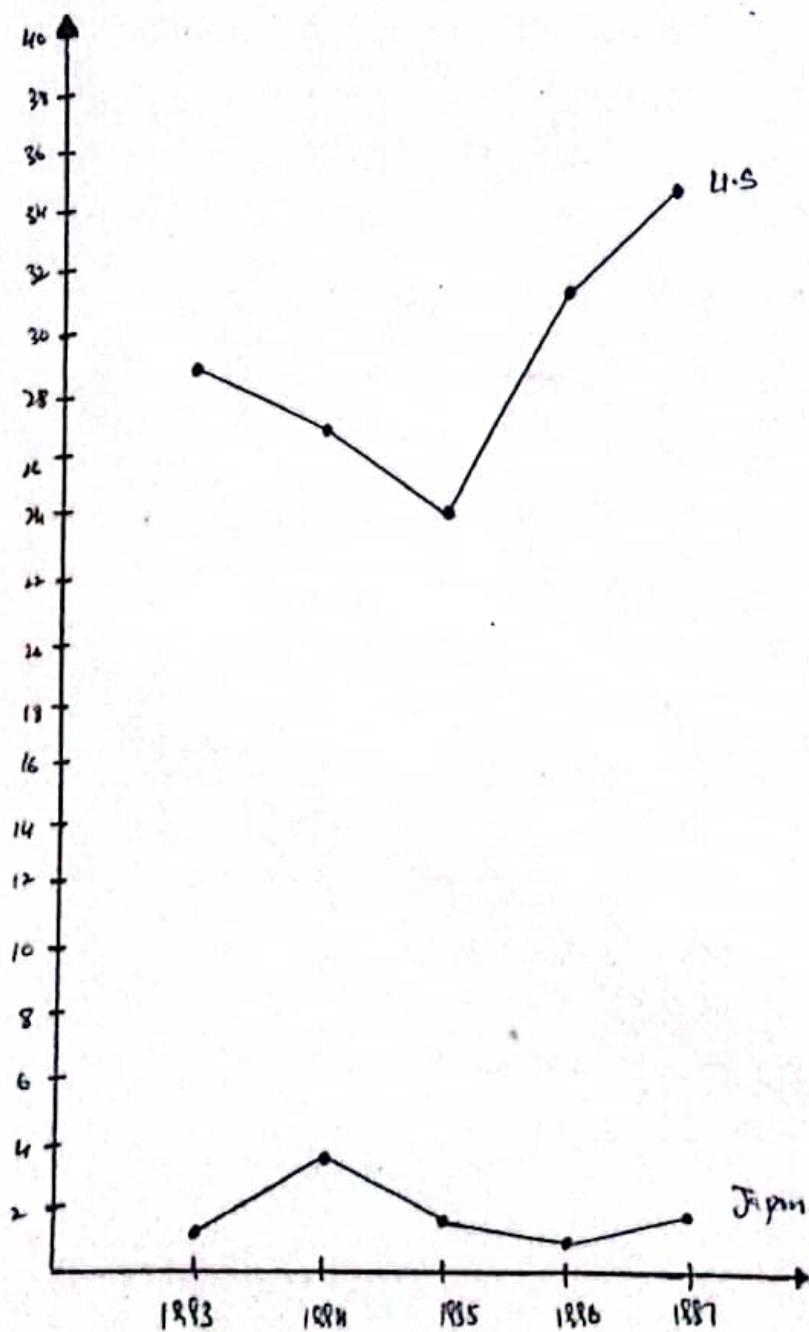


(35)

Extending the Concept

20. Successful Space Launches The number of successful space launches by the U.S and Japan for years 1983-1987 is shown here. Construct a compound time series graph for data. What comparison can be made regarding the launches?

Year	1983	1984	1985	1986	1987
U.S	28	27	24	32	37
Japan	1	4	2	1	2



Review Exercises

1. How People Get Their News

Construct a categorical frequency distribution for data and interpret results.

N	N	T	T	T	I	R	R	I	T
I	N	R	R	I	N	N	I	T	N
I	R	T	T	T	T	N	R	R	I
R	R	I	N	T	R	T	I	I	T
T	I	N	T	T	I	R	N	R	T

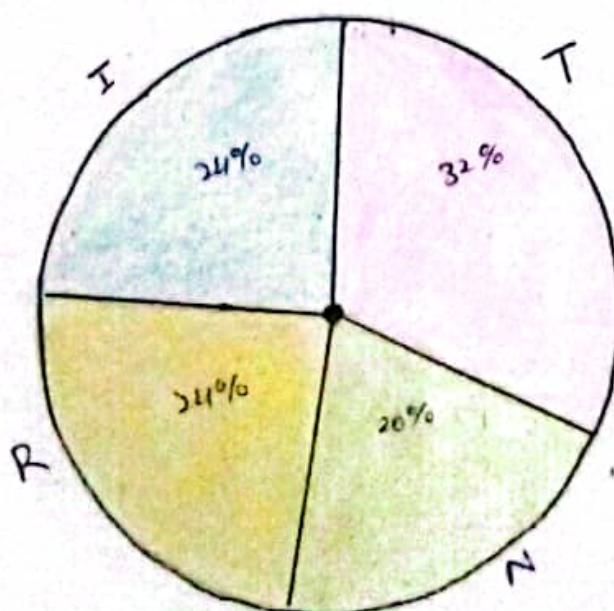
Category	Tally	F
N		10
T		16
I		12
R		12

2. Construct a pie graph for the data in exercise 1, and analyze results

Formula for the percentage of values in each class: $\% = \frac{f}{n} \cdot 100\%$

$$\% N = \frac{10}{50} \times 100\% = 20\% ; \% T = \frac{16}{50} \times 100\% = 32\%$$

$$I = 24\% ; R = 24\%$$

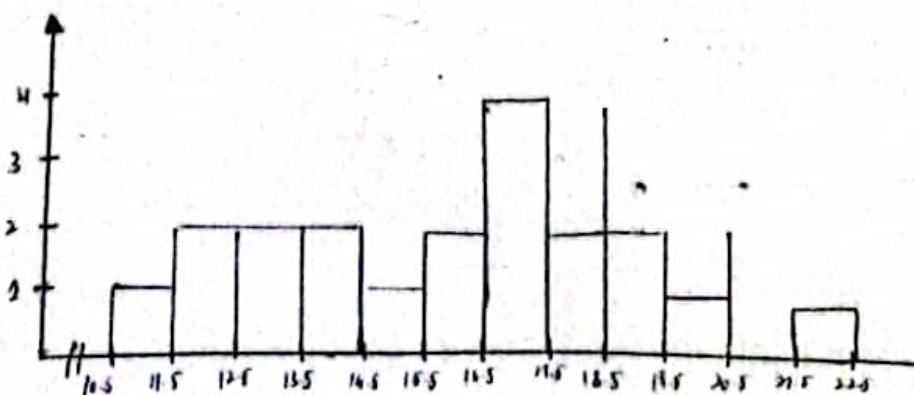
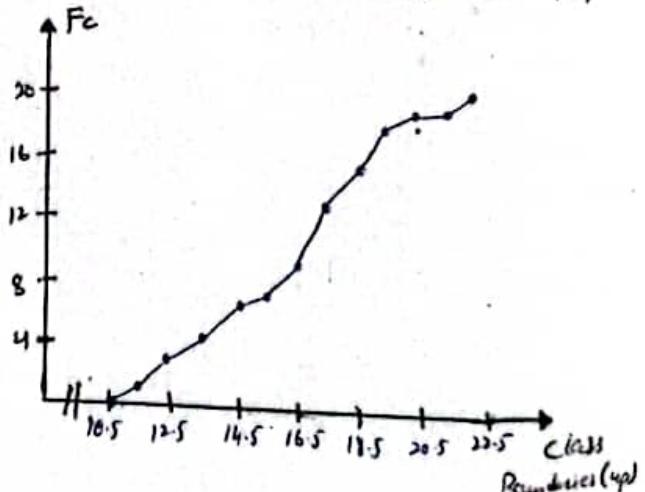
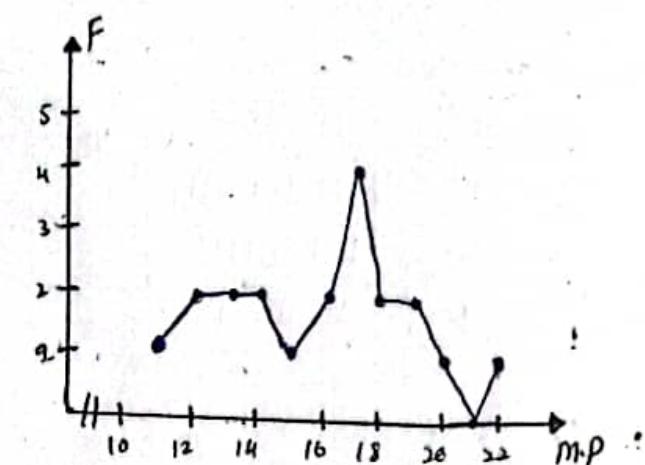


5. BUN Count The Blood urea nitrogen (BUN) count of 20 randomly selected patients is given here in (mg/dl). Construct an ungrouped frequency distribution for the data. (data will be used for exercise 6)

17	18	13	14
12	17	11	20
13	18	18	17
14	16	17	12
16	15	18	22

6. Construct a histogram, a frequency polygon, and an ogive for data in Exercise 5 in this section, and analyze the results.

X	F	F _c	Boundaries	M.P
11	1	1	10.5 - 11.5	11
12	2	3	11.5 - 12.5	12
13	2	5	12.5 - 13.5	13
14	2	7	13.5 - 14.5	14
15	1	8	14.5 - 15.5	15
16	2	10	15.5 - 16.5	16
17	4	14	16.5 - 17.5	17
18	2	16	17.5 - 18.5	18
19	2	18	18.5 - 19.5	19
20	1	19	19.5 - 20.5	20
21	0	19	20.5 - 21.5	21
22	1	20	21.5 - 22.5	22



B. NBA Champions The NBA champions from 1985 on are listed. Use the data to construct a pareto chart and a vertical bar graph.

1985	Los Angeles	1986	Chicago
1986	Boston	1987	Chicago
1987	Los Angeles	1988	Chicago
1988	Detroit	1989	San Antonio
1989	Detroit	1990	Los Angeles
1990	Detroit	1991	Los Angeles
1991	Chicago	1992	Chicago
1992	Chicago	1993	Chicago
1993	Chicago	1994	Detroit
1994	Houston	1995	Houston
		1996	Los Angeles
		1997	Los Angeles
		1998	Chicago
		1999	San Antonio
		2000	Los Angeles
		2001	Los Angeles
		2002	Chicago
		2003	San Antonio
		2004	Detroit
		2005	San Antonio

Champions	F
Chicago	6
Los Angeles	5
Detroit	4
San Antonio	3
Houston	2
Boston	1

