Chapter 1

Introduction to Statistics

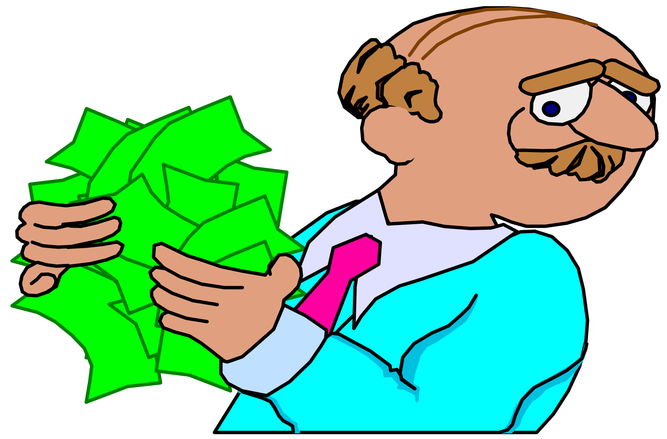
1.1

The Science of Statistics

What is Statistics?

Statistics

The science of collecting, organizing, analyzing, and interpreting data in order to make decisions.



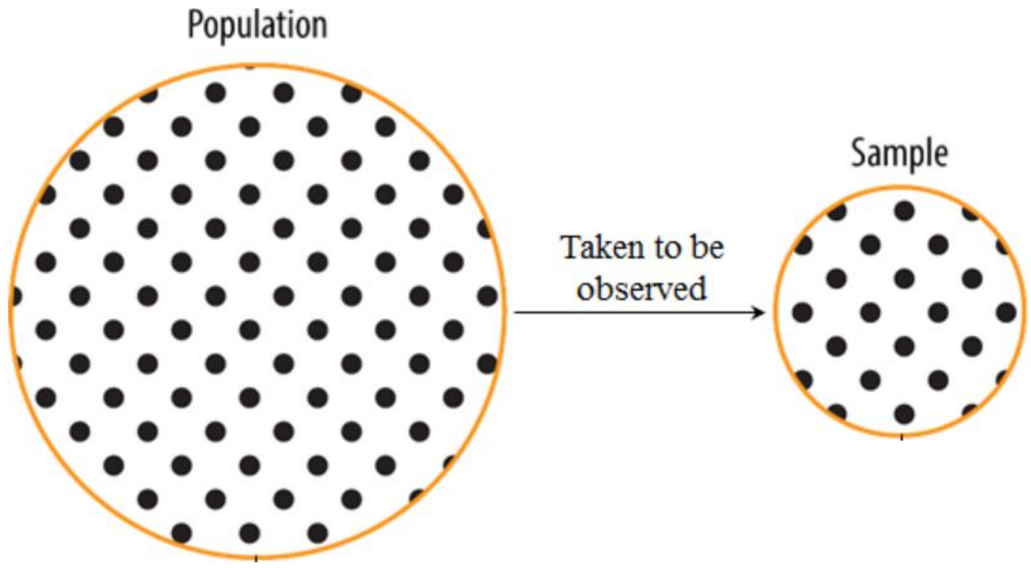
Data Sets

Population

The collection of ***all*** outcomes, responses, measurements, or counts that are of interest.

Sample

A subset of the population .



Parameter and Statistic

Parameter (measure of population)

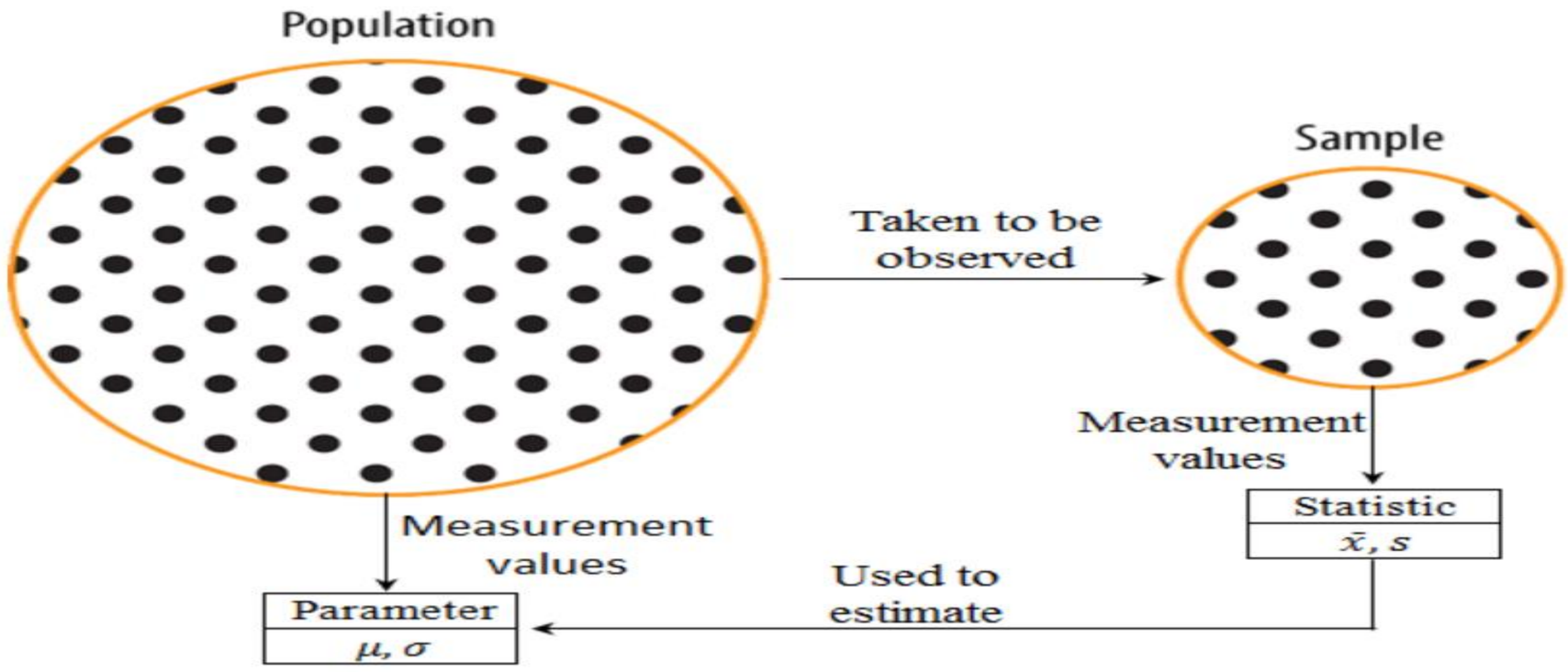
A number that describes a population characteristic.

Average age of all people in the United States

Statistic → (measure of sample)

A number that describes a sample characteristic.

Average age of people from a sample of three states



Branches of Statistics

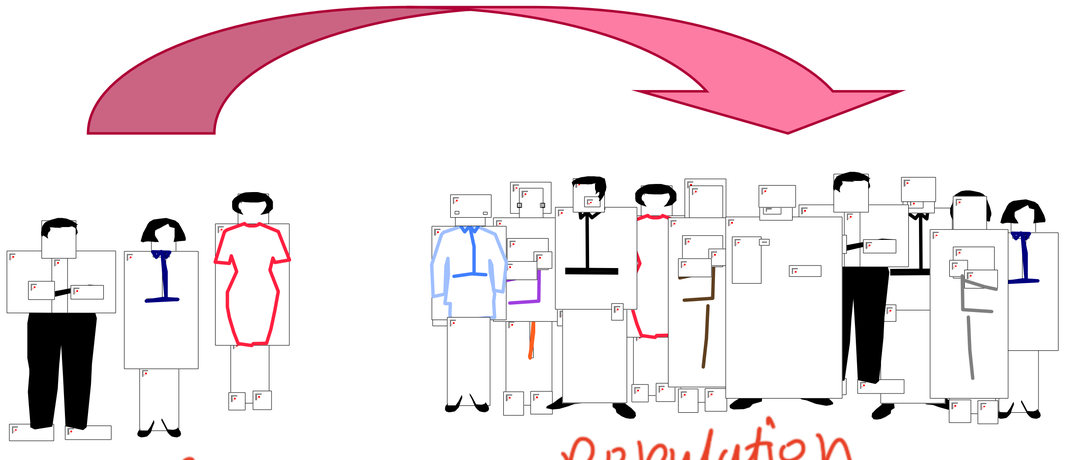
(describe)

* **Descriptive Statistics**
Involves organizing, summarizing, and displaying data.

e.g. Tables, charts, averages

(الاستنتاج)

Inferential Statistics
Involves using ***sample data*** to draw conclusions about a ***population.***



Types of Data

Data

Quantitative

Numerical
(Age, weight, temperature)

Qualitative

Categorical

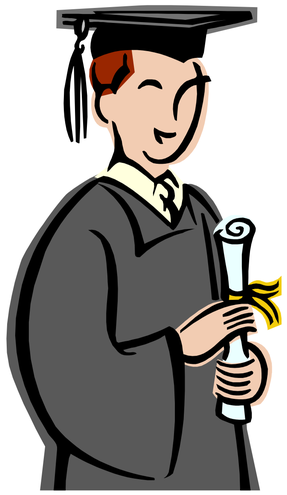
(colours, size, Nationalities)

Types of Data

Qualitative Data

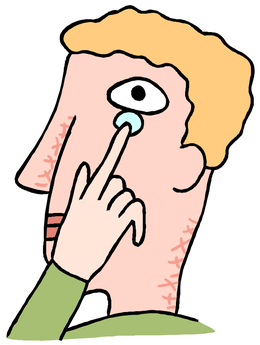
Consists of attributes, labels, or nonnumerical entries.

Major



Place of birth

Eye color



Types of Data

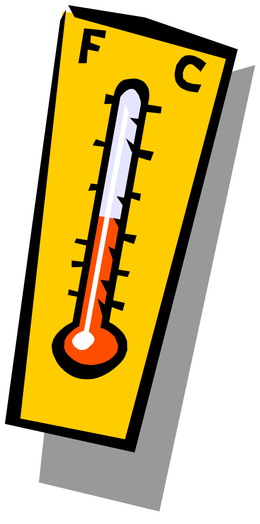
Quantitative data

Numerical measurements or counts.

Age

Weight of a letter

Temperature



Example: Classifying Data by Type

The suggested retail prices of several vehicles are shown in the table. Which data are qualitative data and which are quantitative data? *(Source Ford Motor Company)*

Qualitative → Quantitative

Model	Suggested retail price
Focus Sedan	$15,995
Fusion	$19,270
Mustang	$20,995
Edge	$26,920
Flex	$28,495
Escape Hybrid	$32,260
Expedition	$35,085
F-450	$44,145

Solution: Classifying Data by Type

Model	Suggested retail price
Focus Sedan	$15,995
Fusion	$19,270
Mustang	$20,995
Edge	$26,920
Flex	$28,495
Escape Hybrid	$32,260
Expedition	$35,085
F-450	$44,145

Qualitative Data
(Names of vehicle models are nonnumerical entries)

Quantitative Data
(Suggested retail prices of vehicle models are numerical entries)

Section 1.2

Frequency Distributions and Their Graphs

Frequency Distribution

Frequency Distribution

- A table that shows **classes** or **intervals** of data with a count of the number of entries in each class.
- The **frequency, *f*,** of a class is the number of data entries in the class.

Class width
6 – 1 = 5

Class	Frequency, *f*
1–6	5
6–11	8
11–16	6
16–21	8
21–26	5
26–31	4

Lower class limits

Upper class limits

Constructing a Frequency Distribution

1. Decide on the number of classes.
   - Usually between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width.
   - Determine the range of the data.
   - Divide the range by the number of classes.
   - *Round up to the next convenient number.*

Constructing a Frequency Distribution

3. Find the class limits.
   - You can use the minimum data entry as the lower limit of the first class.
   - Find the remaining lower limits (add the class width to the lower limit of the preceding class).
   - Find the upper limit of the first class. Remember that classes cannot overlap.
   - Find the remaining upper class limits.

Constructing a Frequency Distribution

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency *f* for each class.

Example: Constructing a Frequency Distribution

Quantitative

The following sample data set lists the prices (in dollars) of 30 portable global positioning system (GPS) navigators. Construct a frequency distribution that has seven classes. → 7

90	130	400	200	350	70	325	250	150	250
275	270	150	130	59	200	160	450	300	130
220	100	200	400	200	250	95	180	170	150

steps

① No. of classes = 7 given.

② Class width = $\frac{\text{Max} - \text{min}}{\text{No. of classes}} = \frac{150 - 59}{7} = 55.86$ Round up

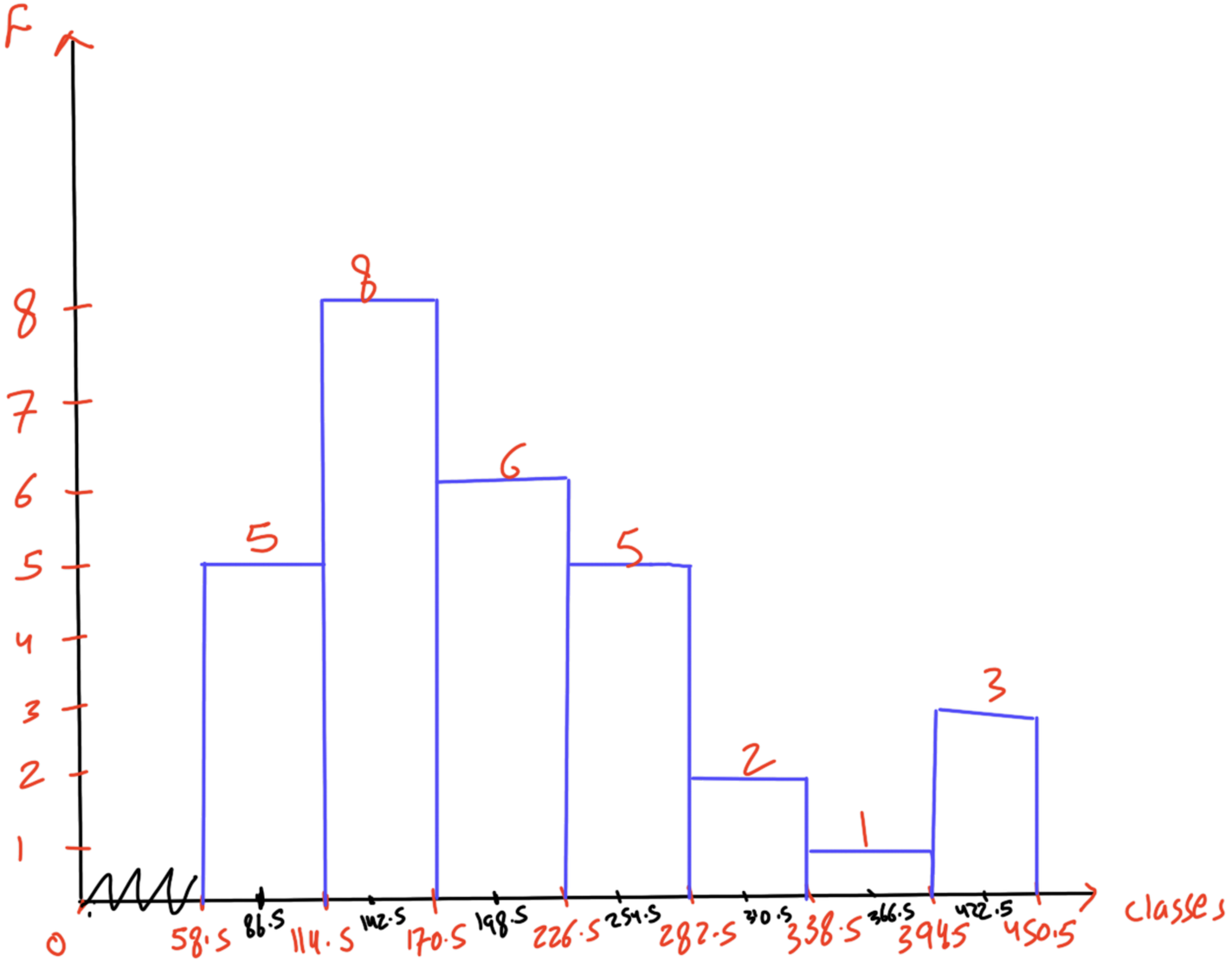
$\simeq 56$

59 + 56 = 115

③ Classes	④ Tally	⑤ F Frequency	Midpoint = $\frac{\text{lower limit} + \text{upper limit}}{2}$	RF Relative Frequency = $\frac{F}{\text{Total}}$	CF Cumulative Frequency
lower limit + class width = upper limit 59 – 115	𝍸	5	$\frac{59+114}{2} = 86.5$	$\frac{5}{30} = 0.17$	5 ↓+8
115 – 171	𝍸 |||	8	142.5	$\frac{8}{30} = 0.27$	13 ↓+6
171 – 227	𝍸 |	6	198.5	$\frac{6}{30} = 0.2$	19 ↓+5
227 – 283	𝍸	5	254.5	$\frac{5}{30} = 0.17$	24 ↓+2
283 – 339	||	2	310.5	$\frac{2}{30} = 0.07$	26 ↓+1
339 – 395	|	1	366.5	$\frac{1}{30} = 0.03$	27 ↓+3
395 – 451	|||	3	422.5	$\frac{3}{30} = 0.1$	30 → Total
	Total	30		Total 1 pi

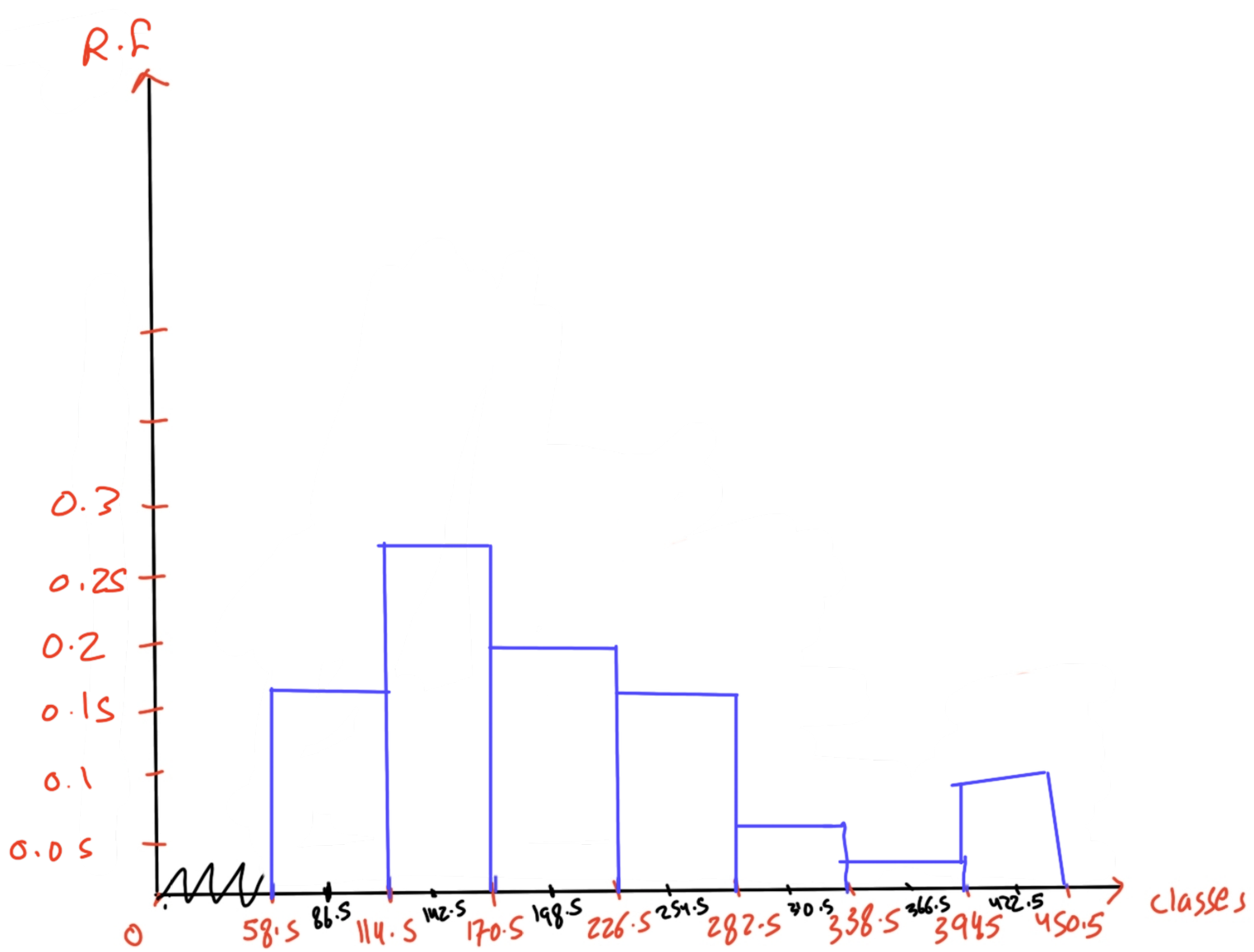
+ class width 56

histogram using boundaries & frequency



using midpoint

Histogram using Relative frequency



Solution: Constructing a Frequency Distribution

90 130 400 200 350 70 325 250 150 250

275 270 150 130 59 200 160 450 300 130

220 100 200 400 200 250 95 180 170 150

1. Number of classes = 7 (given)
2. Find the class width

$$\frac{\max - \min}{\#\text{classes}} = \frac{450 - 59}{7} = \frac{391}{7} \approx 55.86$$

Round up to 56

Solution: Constructing a Frequency Distribution

3. Use 59 (minimum value) as first lower limit. Add the class width of 56 to get the lower limit of the next class.

   59 + 56 = 115

   Find the remaining lower limits.

Class width = 56

Lower limit	Upper limit
59	
115	
171	
227	
283	
339	
395	

Solution: Constructing a Frequency Distribution

The upper limit of the first class is 114 (one less than the lower limit of the second class).

Add the class width of 56 to get the upper limit of the next class.

$$114 + 56 = 170$$

Find the remaining upper limits.

Lower limit	Upper limit
59	115
115	171
171	227
227	283
283	339
339	395
395	451

Class width = 56

Solution: Constructing a Frequency Distribution

4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

Class	Tally	Frequency, f
59–115	𝍸	5
115–171	𝍸 III	8
171–227	𝍸 I	6
227–283	𝍸	5
283–339	II	2
339–395	I	1
395–451	III	3

Determining the Midpoint

Midpoint of a class

$$\frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

Class	Midpoint	Frequency, f
59–115	$\frac{59+114}{2} = 86.5$	5
115–171	$\frac{115+170}{2} = 142.5$	8
171–227	$\frac{171+226}{2} = 198.5$	6

Class width = 56

Determining the Relative Frequency

Relative Frequency of a class

- Portion or percentage of the data that falls in a particular class.

- $$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}$$

Class	Frequency, f	Relative Frequency
59–115	5	$\frac{5}{30} \approx 0.17$
115–171	8	$\frac{8}{30} \approx 0.27$
171–227	6	

Determining the Cumulative Frequency

Cumulative frequency of a class

- The sum of the frequencies for that class and all previous classes.

Class	Frequency, *f*	Cumulative frequency
59–115	5	5
115–171	+ 8	13
171–227	+ 6	19

Expanded Frequency Distribution

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
59–115	5	86.5	0.17	5
115–171	8	142.5	0.27	13
171–227	6	198.5	0.2	19
227–283	5	254.5	0.17	24
283–339	2	310.5	0.07	26
339–395	1	366.5	0.03	27
395–451	3	422.5	0.1	30

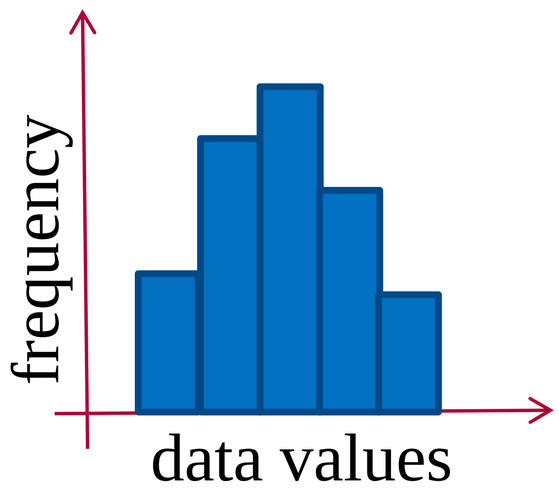
$\Sigma f = 30$

$\sum \frac{f}{n} \approx 1$

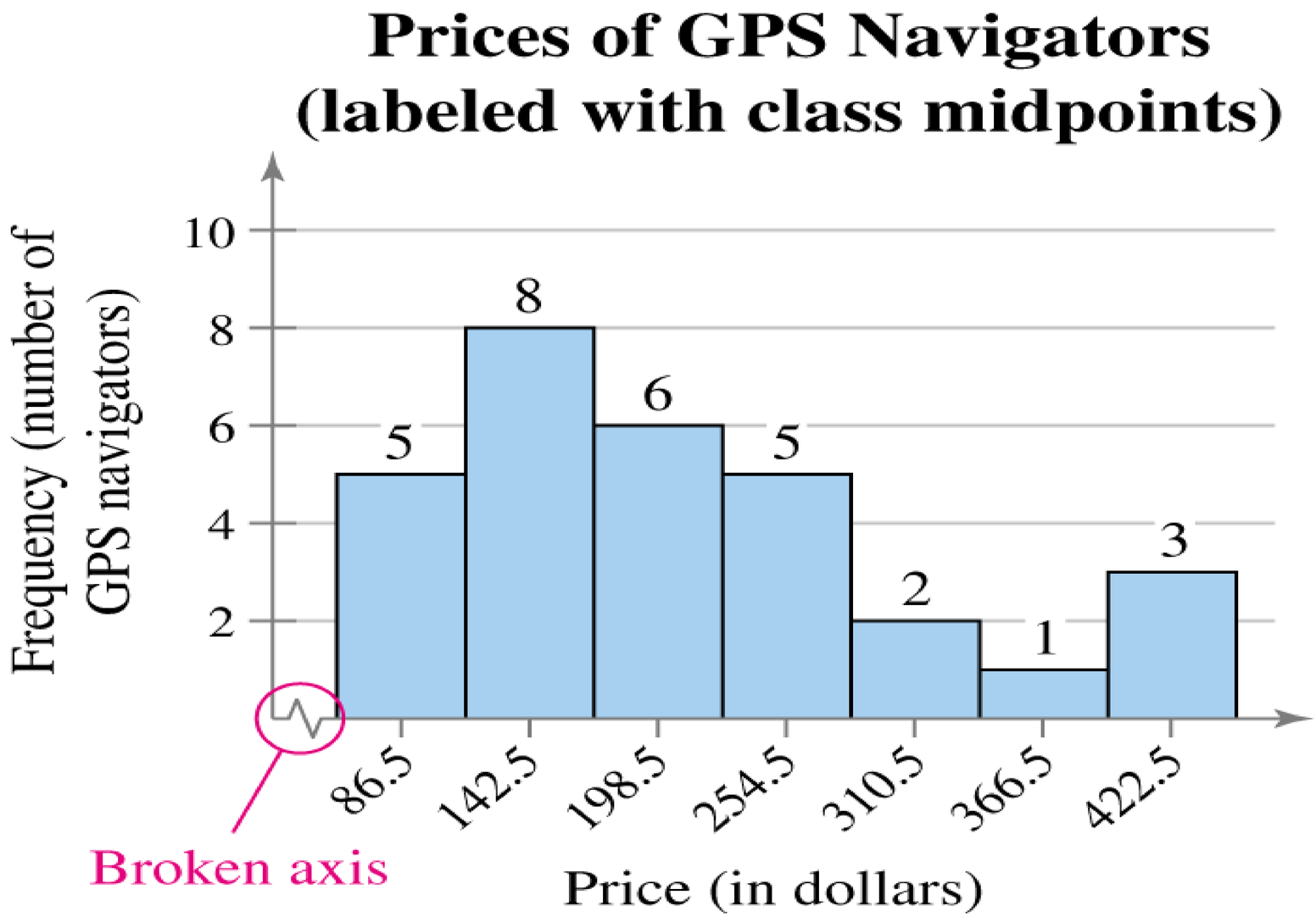
Graphs of Frequency Distributions

Frequency Histogram

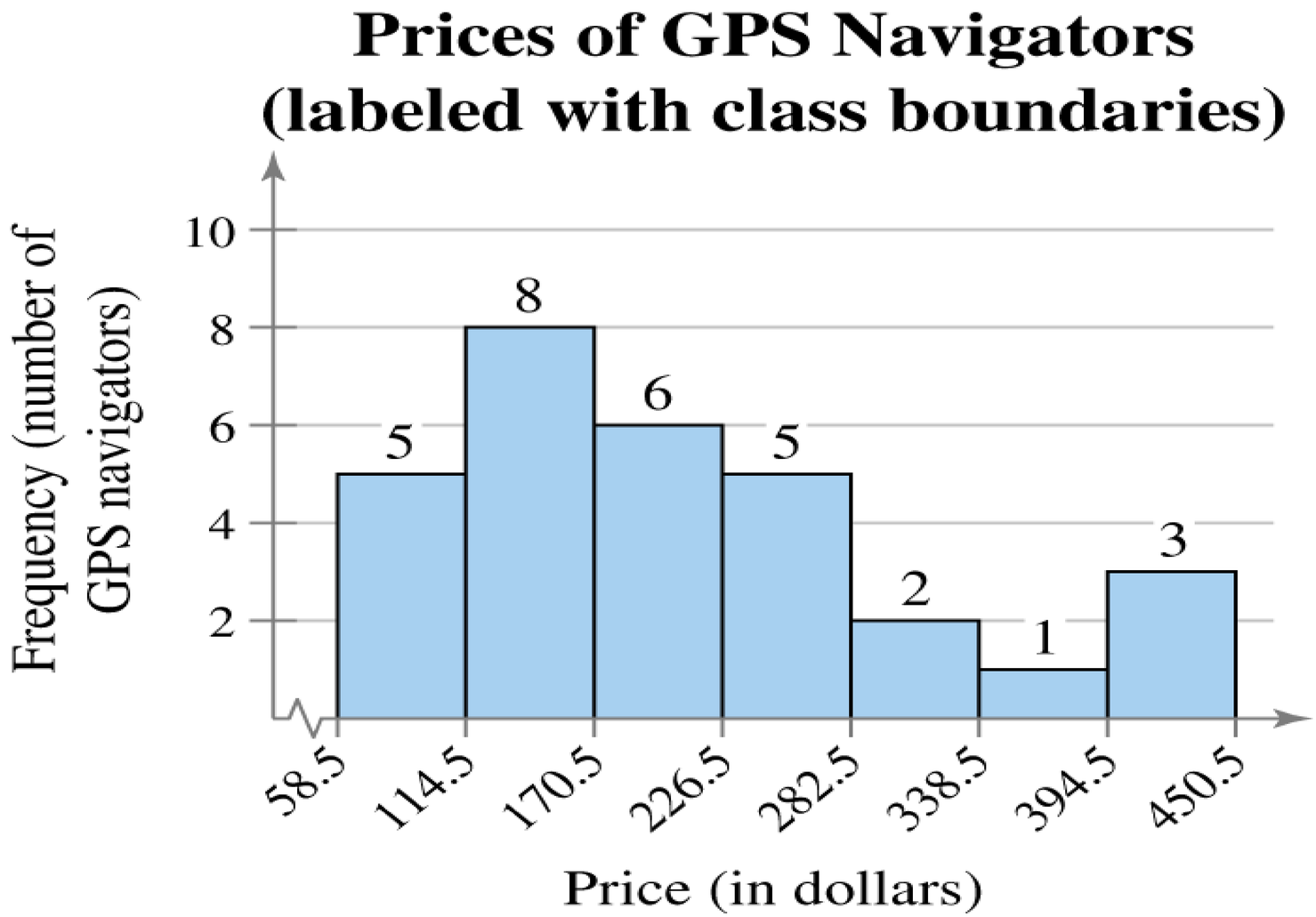
- A bar graph that represents the frequency distribution.
- The horizontal scale is quantitative and measures the data values.
- The vertical scale measures the frequencies of the classes.
- Consecutive bars must touch.



Solution: Frequency Histogram (using Midpoints)

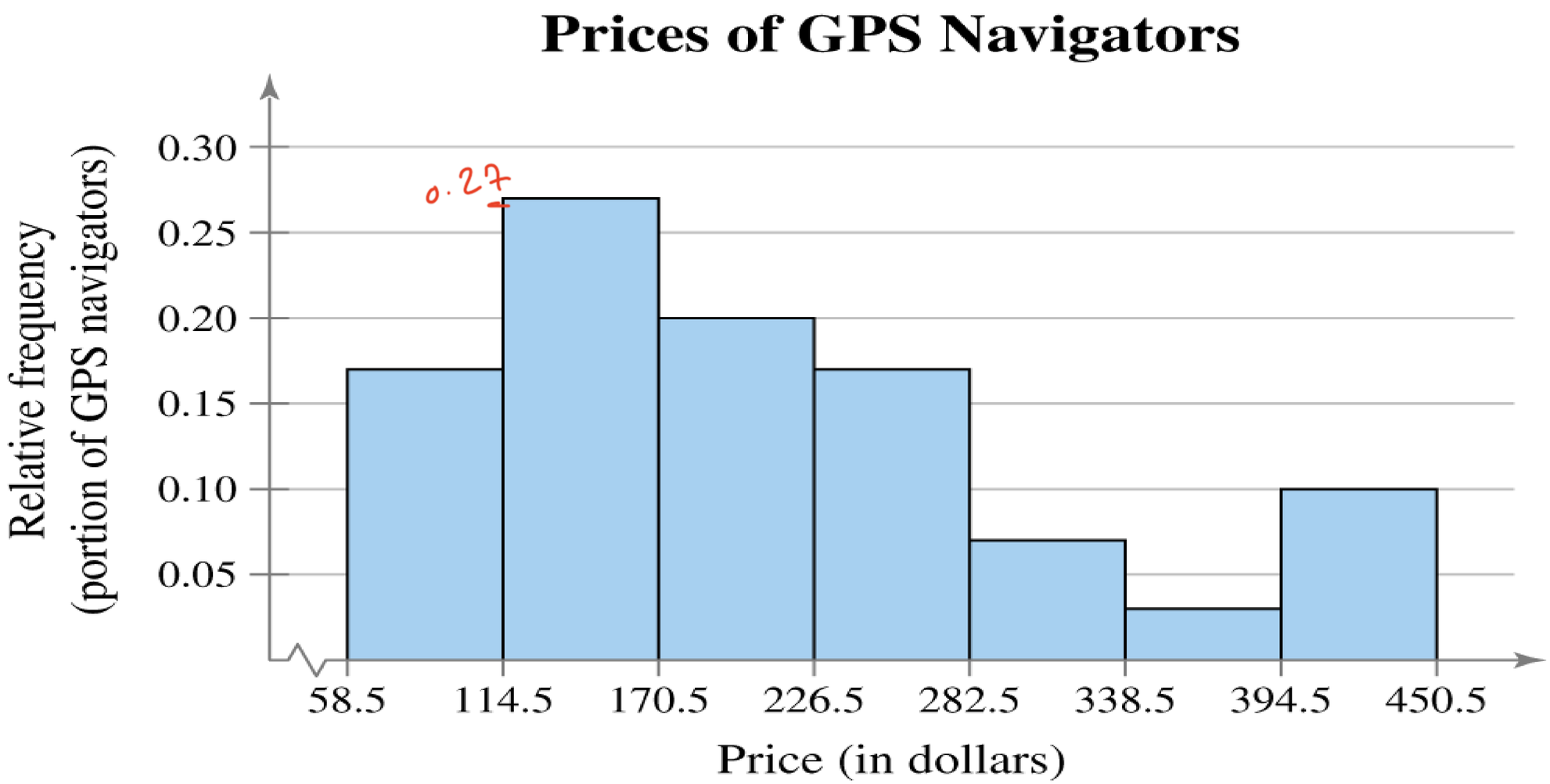


Solution: Frequency Histogram (using class boundaries)



You can see that more than half of the GPS navigators are priced below $226.50.

Solution: Relative Frequency Histogram



From this graph you can see that 27% of GPS navigators are priced between $114.50 and $170.50.

0.27 X 100 = 27%

Section 1.3

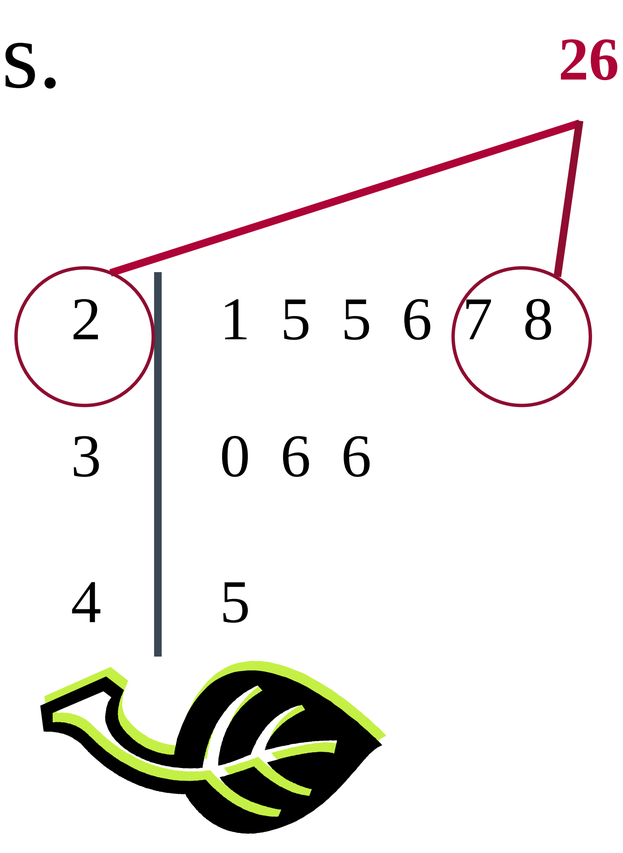
More Graphs and Displays

Graphing Quantitative Data Sets

Stem-and-leaf plot

- Each number is separated into a **stem** and a **leaf**.
- Similar to a histogram.
- Still contains original data values.

Data: 21, 25, 25, **26**, 27, 28,
30, 36, 36, 45



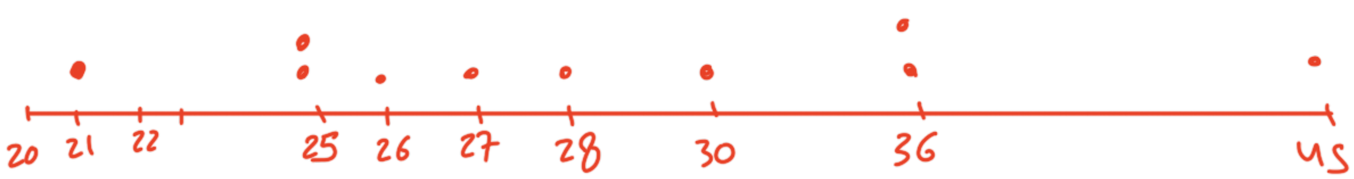
21, 25, 25, 26, 27, 28, 30, 36, 36, 45

Quantitative data

stem & leaf

stem	leaf
2	1 5 5 6 7 8
3	0 6 6
4	
5	

dot plot



Example: Constructing a Stem-and-Leaf Plot

The following are the numbers of text messages sent last week by the cellular phone users on one floor of a college dormitory. Display the data in a stem-and-leaf plot.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

Solution: Constructing a Stem-and-Leaf Plot

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

- The data entries go from a low of 78 to a high of 159.
- Use the rightmost digit as the leaf.
  - For instance,

    78 = 7 | 8 and 159 = 15 | 9
- List the stems, 7 to 15, to the left of a vertical line.
- For each data entry, list a leaf to the right of its stem.

Solution: Constructing a Stem-and-Leaf Plot

Number of Text Messages Sent

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

Key: 15|5 = 155

Unordered Stem-and-Leaf Plot

Number of Text Messages Sent

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

Key: 15|5 = 155

Include a key to identify the values of the data.

Ordered Stem-and-Leaf Plot

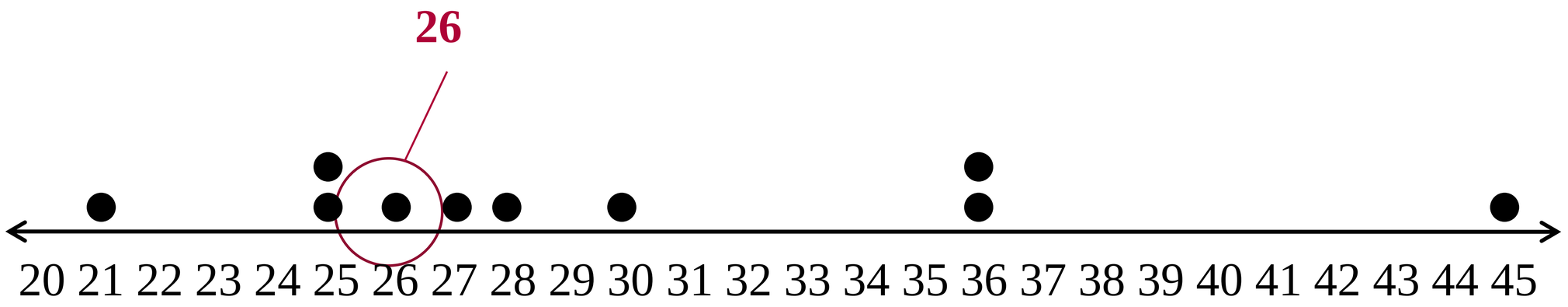
From the display, you can conclude that more than 50% of the cellular phone users sent between 110 and 130 text messages.

Graphing Quantitative Data Sets

Dot plot

- Each data entry is plotted, using a point, above a horizontal axis.

Data: 21, 25, 25, **26**, 27, 28, 30, 36, 36, 45



Example: Constructing a Dot Plot

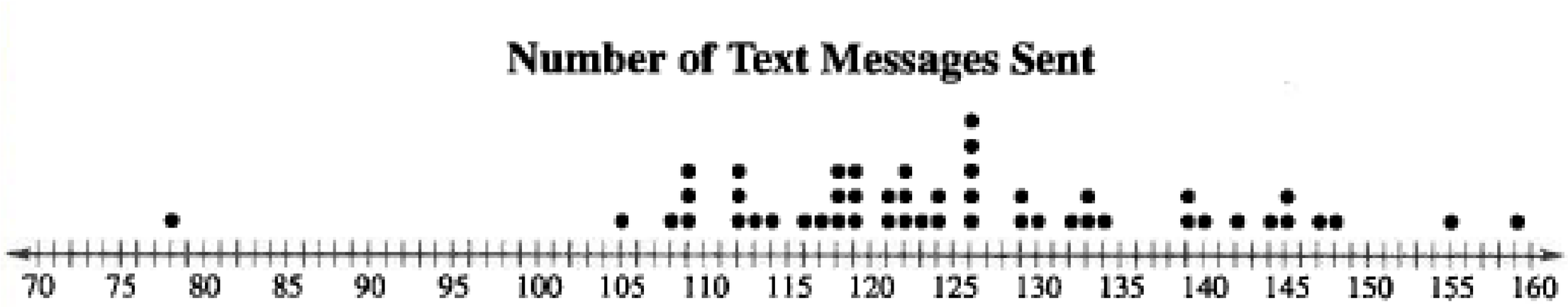
Use a dot plot organize the text messaging data.

155 159 144 129 105 145 126 116 130 114 122 112 112 142 126
118 118 108 122 121 109 140 126 119 113 117 118 109 109 119
139 139 122 78 133 126 123 145 121 134 124 119 132 133 124
129 112 126 148 147

- So that each data entry is included in the dot plot, the horizontal axis should include numbers between 70 and 160.
- To represent a data entry, plot a point above the entry's position on the axis.
- If an entry is repeated, plot another point above the previous point.

Solution: Constructing a Dot Plot

155	159	144	129	105	145	126	116	130	114	122	112	112	142	126
118	118	108	122	121	109	140	126	119	113	117	118	109	109	119
139	139	122	78	133	126	123	145	121	134	124	119	132	133	124
129	112	126	148	147										

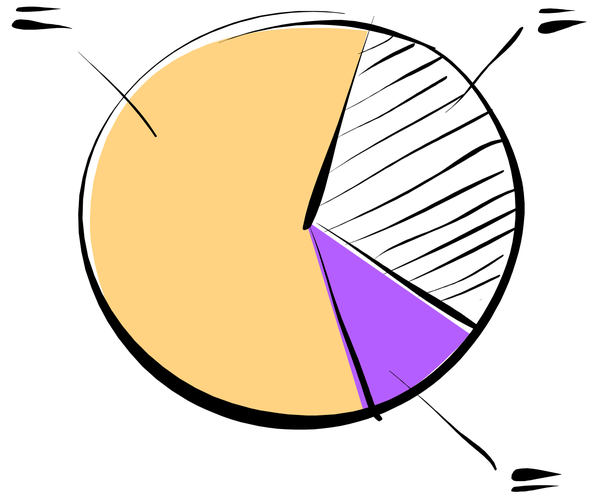


From the dot plot, you can see that most values cluster between 105 and 148 and the value that occurs the most is 126. You can also see that 78 is an unusual data value.

Graphing Qualitative Data Sets

Pie Chart

- A circle is divided into sectors that represent categories.
- The area of each sector is proportional to the frequency of each category.



Example: Constructing a Pie Chart

The numbers of earned degrees conferred (in thousands) in 2007 are shown in the table. Use a pie chart to organize the data. ***(Source: U.S. National Center for Educational Statistics)***

Type of degree	Number (thousands)
Associate's	728
Bachelor's	1525
Master's	604
First professional	90
Doctoral	60

Qualitative data

Pie chart

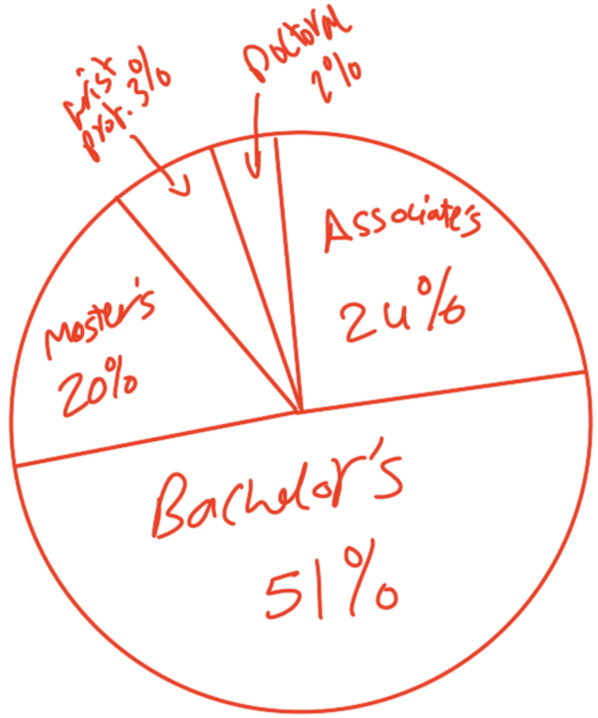
Type of degree	Number (thousands)
Associate's	728
Bachelor's	1525
Master's	604
First professional	90
Doctoral	60
	3007

f

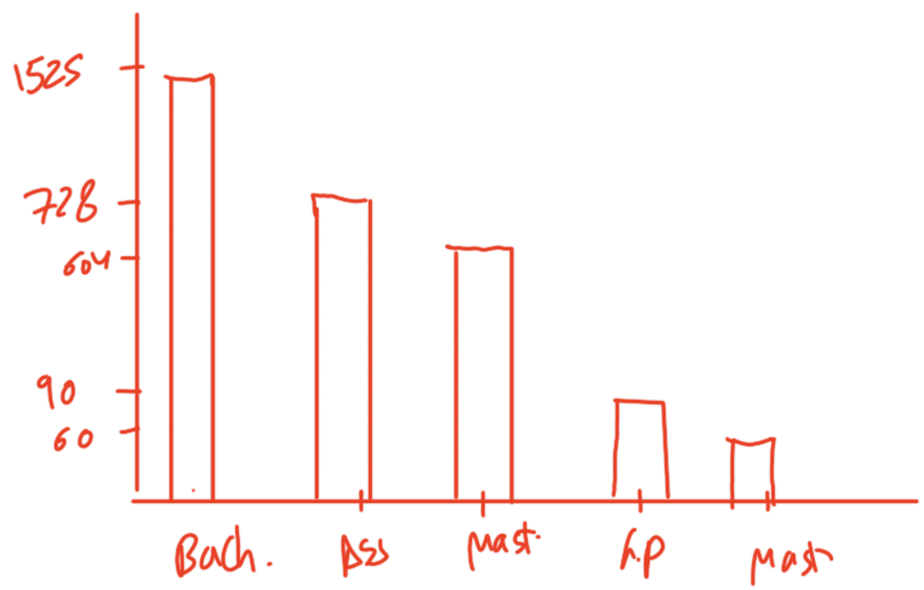
R.f x100

0.24 x100 = 24%
0.51 x100 = 51%
0.2 x100 = 20%
0.03 x100 = 3%
0.02 x100 = 2%

1 100%



pareto chart



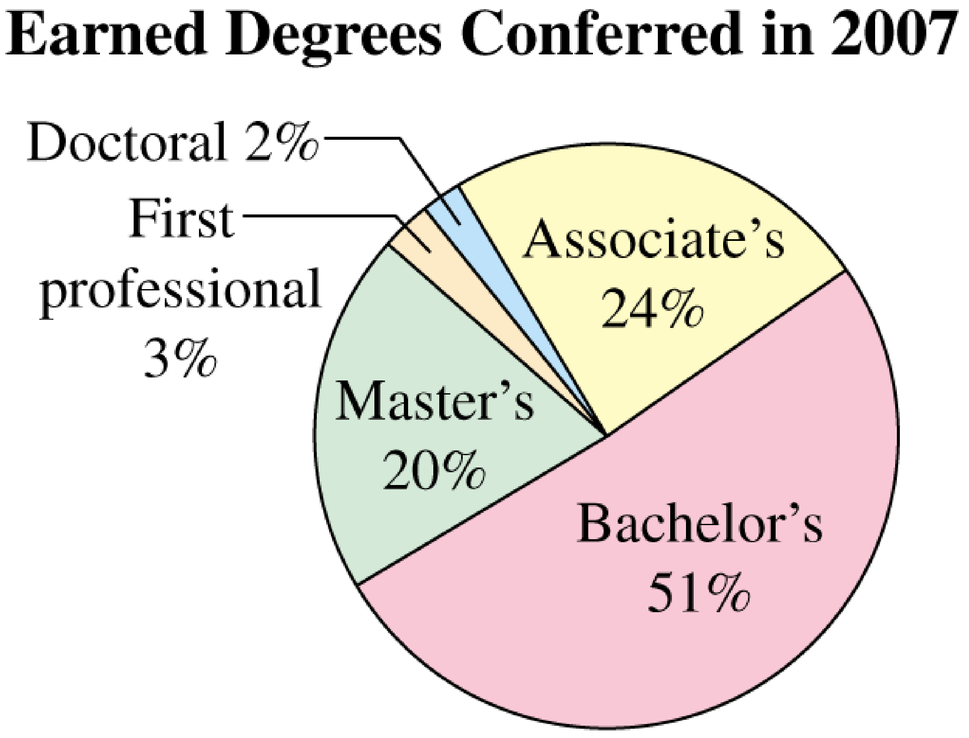
Solution: Constructing a Pie Chart

- Find the relative frequency (percent) of each category.

Type of degree	Frequency, f	Relative frequency
Associate's	728	$\frac{728}{3007} \approx 0.24$
Bachelor's	1525	$\frac{1525}{3007} \approx 0.51$
Master's	604	$\frac{604}{3007} \approx 0.20$
First professional	90	$\frac{90}{3007} \approx 0.03$
Doctoral	60	$\frac{60}{3007} \approx 0.02$
	$\Sigma f = 3007$	

Solution: Constructing a Pie Chart

Type of degree	Relative frequency
Associate’s	0.24
Bachelor’s	0.51
Master’s	0.20
First professional	0.03
Doctoral	0.02

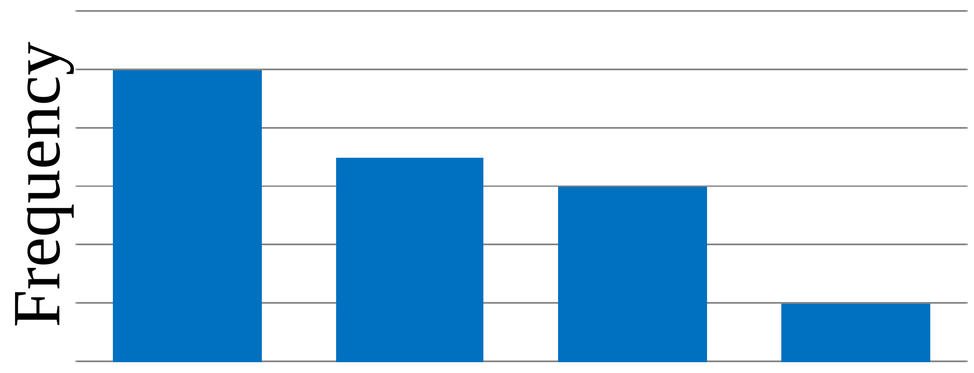


From the pie chart, you can see that over one half of the degrees conferred in 2007 were bachelor’s degrees.

Graphing Qualitative Data Sets

Pareto Chart

- A vertical bar graph in which the height of each bar represents frequency or relative frequency.
- The bars are positioned in order of decreasing height, with the tallest bar positioned at the left.



Pareto Diagram

- **Pareto Diagram:** A bar graph with the bars arranged from the most numerous category to the least numerous category. It includes a line graph displaying the cumulative percentages and counts for the bars.

Notes:

- The Pareto diagram is often used in quality control applications

- Used to identify the number and type of defects that happen within a product or service

Example

✓ **Example:** The final daily inspection defect report for a cabinet manufacturer is given in the table below:

Defect	Number
Dent	5
Stain	12
Blemish	43
Chip	25
Scratch	40
Others	10
	135

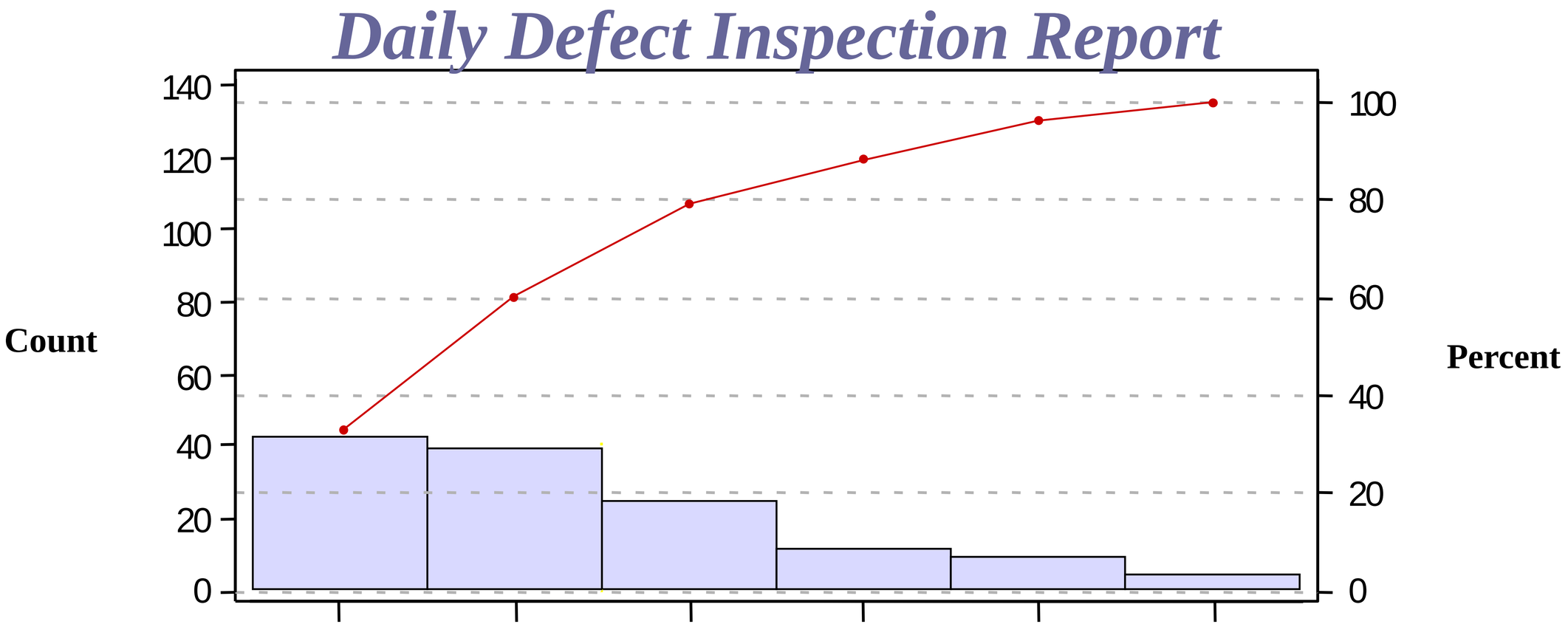
R.F X 100
3.7%
8.89%

1) Construct a Pareto diagram for this defect report. Management has given the cabinet production line the goal of reducing their defects by 50%.

2) What two defects should they give special attention to in working toward this goal?

Solutions

1)



Count	43	40	25	12	10	5
Percent	31.9	29.6	18.5	8.9	7.4	3.7
Cum%	31.9	61.5	80.0	88.9	96.3	100.0

R.f X100

2) The production line should try to eliminate blemishes and scratches. This would cut defects by more than 50%.