Section 4.1, Frequency Distributions and Their Graphs

Section 4.1 Part I Objectives

- Can you construct and interpret frequency distributions?
- Can you calculate midpoints and frequencies?
- Can you explain in words the difference between a relative frequency and a cumulative frequency?
- Can you construct and interpret frequency histograms?
- Can you construct and interpret frequency polygons?

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.

frequency: "how often" did that response show?

Class Width 6-1=5

Class	Frequency, *f*
1-5	5
6-10	8
11-15	6
16-20	8
21-25	5
26-30	4

6 classes

Upper Class Limits

Lower Class Limits

To find class width:
- Range = 30-1 = 29
- 29 ÷ 6 = 4.83 ≈ 5 → class width

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. (max-min, high # − low #)
   - Divide the range by the number of classes.
   - Round up to the next convenient number. ALWAYS!

table above: 30-1=29; 29/6=4.83 → 5

3. →

4. →

5. →

3. Find the class limits. → 7

- You can use the Minimum data entry as the lower limit of the first class. 7 - □
- Find the remaining lower limits (add the class width to the lower limit of the next class). +12 → 19 - □ +12 → 31 - □ +12 → 43 - □
- Find the upper limit of the first class. Remember that classes cannot overlap (look at the lower limit of the next class to help.)
- Find the remaining upper class limits.

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

The following sample data set lists the number of minutes 50 Internet subscribers spent on the Internet during their most recent session. Construct a frequency distribution that has seven classes.

Range: 86 − 7 = 79

50 40 41 7 11 (7) 22 44 28 21 19 23 37 51 54 42 (86)

41

78 56 72 56 17 7 69 30 80 56 29 33 46 31 39 20

18 29 34 59 73 77 36 39 30 62 54 67 39 31 53 44

1. **Number of classes =** 7
2. **Find the class width** $\frac{79}{7}$ (Range / #classes) = 11.3 **Round up to** 12

3. Use 7 (minimum value) as first lower limit. Add the class width of 12 to get the lower limit of the next class.

   7 + 12 = 19

   Find the remaining lower limits by adding 12 each time.

Class Width = 12

Lower Limit	Upper Limit
7	18
+12 → 19	30
+12 → 31	42
+12 → 43	54
55	66
67	78
79	90

Why so much higher than max, 86?

Rounded up

The upper limit of the first class is 18 (one less than the lower limit of the second class). Add the class width of 12 to get the upper limit of the next class.

Find the remaining upper limits by adding 12.

Lower Limit	Upper Limit
7	18
19	30
31	42
43	54
55	66
67	78
79	90

+12 +12 +12

Class Width = 12

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*
7 – 18	𝍸 I	6
19 – 30	𝍸 𝍸	10
31 – 42	𝍸 𝍸 III	13
43 – 54	𝍸 III	8
55 – 66	𝍸	5
67 – 78	𝍸 I	6
79 – 90	II	2

$\Sigma f = 50$

What is that symbol at the bottom of the chart?

Σ is the Greek letter sigma, and it means the "sum".

So Σ*f* means the "sum of the frequencies." → matches # of data entries → we didn't miss any ☺

Determining the Midpoint:

Midpoint of a class: the point in the middle of a class.

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

AVERAGE!

Midpt. is useful on graphs, data displays

Class		Midpoint	Frequency, f
7-18	(7+18)/2 =	12.5	6
19-30	(19+30)/2 =	24.5	10
31-42	(31+42)/2	36.5	13

Class Width=12

Rel. freq.: how often, compared to the rest?

Determining the Relative Frequency

Relative Frequency of a class

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

$$\text{relative frequency} = \frac{\text{Class freq.}}{\text{Sample size}} = \frac{f}{n} \quad (n = \text{total})$$

Class	Frequency, f	Relative frequency
7-18	6	6/50 = 0.12
19-30	10	10/50 = 0.20
31-42	13	13/50 = 0.26

Cumulative frequency of a class

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

Class	Frequency, f	Cumulative frequency
7-18	6	6
19-30	+ 10	16
31-42	+ 13	29
43-54	8	37

Expanded Frequency Distribution

Class	Frequency, f (Internet use in min.)	Midpoint	Relative Frequency	Cumulative Frequency
7 – 18	6	12.5	0.12	6
19 – 30	10	24.5	0.20	16
31 – 42	13 (most common)	36.5	0.26	29
43 – 54	8	48.5	0.16	37
55 – 66	5	60.5	0.10	42
67 – 78	6	72.5	0.12	48
79 – 90	2	84.5	0.04	50

Put it all together!

$\Sigma f = 50$ sum of freq.

$$\Sigma \frac{f}{n} = 1$$

Observ.:

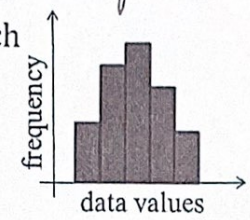
- Most common time span online for internet users surveyed: 31-42 min.
- Almost half (46%) spent between 19 & 42 minutes.

GRAPHS →

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



Class Boundaries

- The numbers that separate classes without forming gaps between them
- The distance from the upper limit of the first class to the lower limit of the second class is 19 – 18 = 1.
- Half this distance is 0.5.
- First class lower boundary = 7 – 0.5 = 6.5
- First class upper boundary = 18 + 0.5 = 18.5

Put 0.5 on either end of the class.

Class	Class Boundaries	Frequency, *f*
7 – 18	6.5-18.5	6
19 – 30	18.5-30.5	10
31 – 42	30.5-42.5	13

Class Boundaries

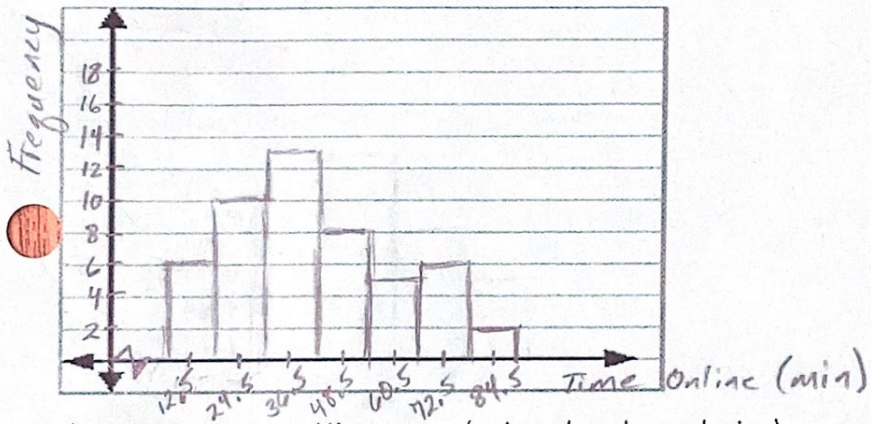
Class	Class Boundaries	Frequency, *f*
7 – 18	6.5 – 18.5	6
19 – 30	18.5 - 30.5	10
31 – 42	30.5 - 42.5	13
43 – 54	42.5 - 54.5	8
55 – 66	54.5 - 66.5	5
67 – 78	66.5 - 78.5	6
79 – 90	78.5 - 90.5	2

Example: Frequency Histogram

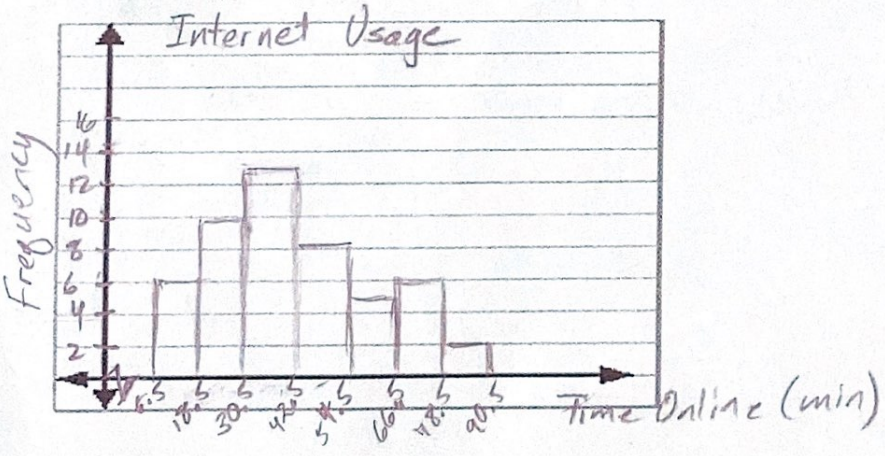
Construct a frequency histogram for the Internet usage frequency distribution.

Class	Class Boundaries	Midpoint	Frequency, f
7 – 18	6.5 – 18.5	12.5	6
19 – 30	18.5 – 30.5	24.5	10
31 – 42	30.5 – 42.5	36.5	13
43 – 54	42.5 – 54.5	48.5	8
55 – 66	54.5 – 66.5	60.5	5
67 – 78	66.5 – 78.5	72.5	6
79 – 90	78.5 – 90.5	84.5	2

Solution: Frequency Histogram (using Midpoints)

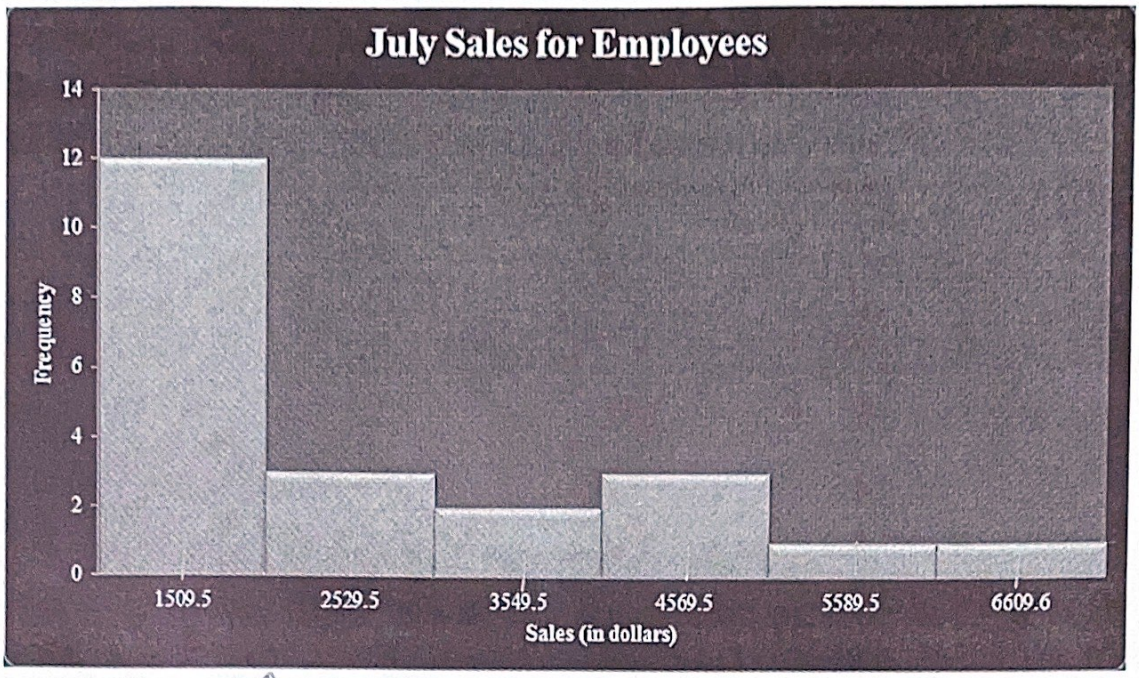


Solution: Frequency Histogram (using class boundaries)



Interpret: Most reported internet usage between 19 & 54 min.

Example: Interpreting frequency histograms



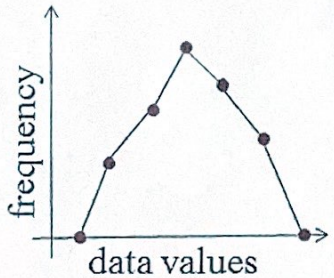
↑ 2019

1. How many classes? 6 2. What is the class width? 2529.5 − 1509.5 = $1020
3. What is the frequency of the class with the greatest frequency? 12 sales
4. Find Σf. (Sum of freq) = 12+3+2+3+1+1 = 22 5. Describe any patterns. Most employees had about $1000 - 2000 in sales for July

Graphs of Frequency Distributions

Frequency Polygon

- A line graph that emphasizes the continuous change in frequencies.



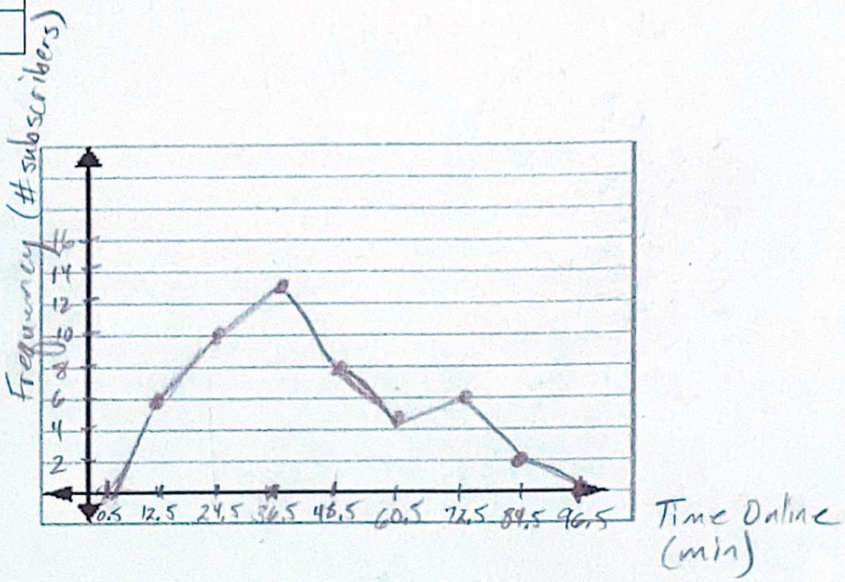
Example: Frequency Polygon

Construct a frequency polygon for the Internet usage frequency distribution. Plot the midpoint on the horiz. axis and the frequency on the vertical axis.

Class	Midpoint	Frequency, f
7 – 18	12.5	6
19 – 30	24.5	10
31 – 42	36.5	13
43 – 54	48.5	8
55 – 66	60.5	5
67 – 78	72.5	6
79 – 90	84.5	2

Solution: Frequency Polygon

The graph should begin and end on the horizontal axis, so extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.



You can see that the frequency of subscribers increases up to 36.5 minutes and then decreases

4.1 Part II Notes

More Frequencies Graphs

Section 4.1 Part II Objectives

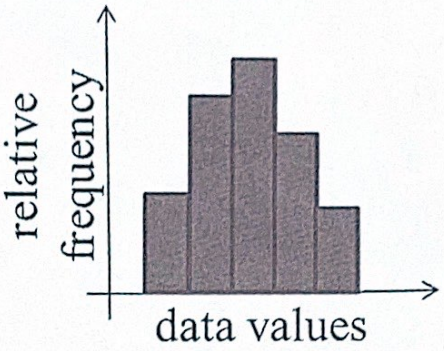
- Can you construct and interpret relative frequency histograms?
- Can you explain the difference between relative frequencies and cumulative frequencies?
- Can you construct and interpret ogives? Ogive
- Can you use technology to create displays

Graphs of Frequency Distributions

Relative Frequency Histogram

- Has the same shape and the same horizontal scale as the corresponding frequency histogram.
- The vertical scale measures the relative **frequencies**, not frequencies.

Remember... rel. freq. are decimal values (or %) from freq ÷ sample size $\frac{f}{n}$



Example: Relative Frequency Histogram

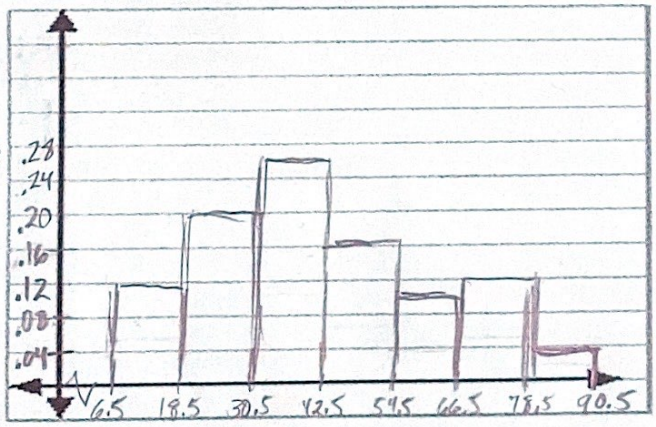
Construct a relative frequency histogram for the Internet usage frequency distribution.

$n = 50$

Class	Class Boundaries	Frequency, f	Relative Frequency
7 – 18	6.5 – 18.5	6	0.12 ← 6/50
19 – 30	18.5 – 30.5	10	0.20 ← 10/50 etc.
31 – 42	30.5 – 42.5	13	0.26
43 – 54	42.5 – 54.5	8	0.16
55 – 66	54.5 – 66.5	5	0.10
67 – 78	66.5 – 78.5	6	0.12
79 – 90	78.5 – 90.5	2	0.04

Solution: Relative Frequency Histogram (using class boundaries)

Rel. Freq. (portion of subscribers)



Vert. axis: 0 to .26

count by .04

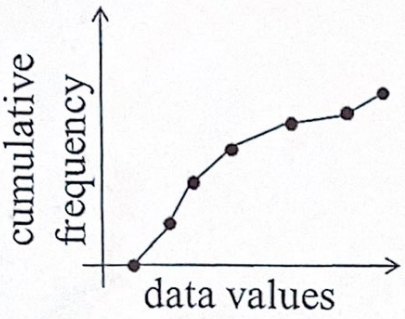
From this graph you can see that 20 % of Internet subscribers spent between 18.5 minutes and 30.5 minutes online.

Graphs of Frequency Distributions

Cumulative: add up everything so far

Cumulative Frequency Graph or Ogive

- A line graph that displays the cumulative frequency of each class at its upper class boundary.
- The upper boundaries are marked on the horizontal axis.
- The cumulative frequencies are marked on the vertical axis.



Constructing an Ogive

1. Construct a frequency distribution that includes cumulative frequencies as one of the columns.
2. Specify the horizontal and vertical scales.
   - The horizontal scale consists of the upper class boundaries.
   - The vertical scale measures cumulative frequencies.
3. Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size *n*.).

Example: Ogive

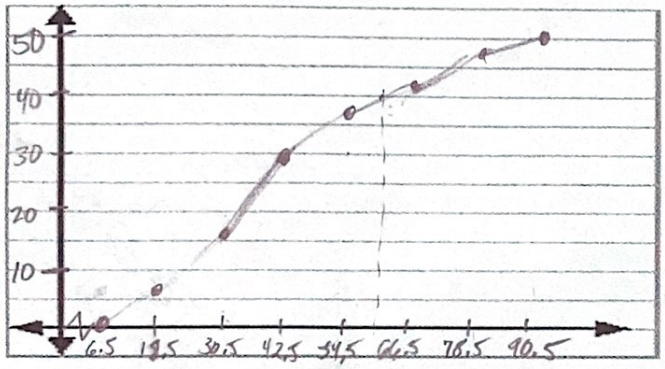
Construct an ogive for the Internet usage frequency distribution.

Class	Class Boundaries	Frequency, f	Cumulative Frequency
7 – 18	6.5 – 18.5	6	6
19 – 30	18.5 – 30.5	10	16
31 – 42	30.5 – 42.5	13	29
43 – 54	42.5 – 54.5	8	37
55 – 66	54.5 – 66.5	5	42
67 – 78	66.5 – 78.5	6	48
79 – 90	78.5 – 90.5	2	50

Solution: Ogive

Similar to polygon - also line graph

How to recognize: always incr.



From the ogive, you can see that about 40 subscribers spent 60 minutes or less online during their last session. The greatest increase in usage occurs between 30.5 minutes and 42.5 minutes

Technology Directions: Graphing Calculator directions (entering a list)

Button	Comments
STAT EDIT	(or just hit ENTER)
Arrow up to L1 CLEAR ENTER	to erase previous entries
Type in 1st entry ENTER	

Also: Y= Clear

Enter all other items in the same fashion.

Don't forget to erase L1 if a list has already been entered.
There could be "hidden" entries that will impact your calculations.

Technology Directions: Graphing Calculator directions (plotting a list)

Button	Comments
STATPLOT 2nd y=	Note: you must have entered L1 (and possibly L2) before you can display any data.
ENTER	to choose Plot 1 (Use arrow and ENTER to turn the plot on, if it already isn't on.)
Arrow down and then right to choose appropriate graph. ENTER	Some displays require L1 **and** L2.
ZOOM 9	This chooses the window to fit your data automatically.

Q: How does it choose boundaries?

Section 4.2 Part I
More Graphs and Displays

Section 4.2 Part I Objectives

- Can you graph quantitative data using stem-and-leaf plots and dot plots?
- Can you graph qualitative data using pie charts and Pareto charts?

Graphing Quantitative Data Sets

Stem-and-leaf plot

- Each number is separated into a stem and a leaf.
- Similar to a histogram. (shape, just sideways)
- Still contains original data values.
- No horiz axis !

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

26 → see patterns

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

Key: 4|5 = 45

Example: Constructing a Stem-and-Leaf Plot

The following are the numbers of text messages sent last month 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

138 139 122 78 133 126 123 145 121 134 124 119 132 133 124

129 112 126 148 147

70's ✓
80's X
90's X
100-109 - 5 #s
110-119 - 13 #s

Solution: Constructing a Stem-and-Leaf Plot

- Start by ordering the data! Min 78 Max 159
- The smallest stem should be the 10s digit from the smallest number in the list.
- Include a key.

64 22 88 93
78 99 2
sort

70's → 7 | 8 ← outlier

8	
9	
10	5 8 9 9 9
11	2 2 2 3 4 6 7 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

Include a key to identify the values of the data.

15|5 = 155

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

Example: A stem-and-leaf plot with decimals...

Make a stem-and-leaf plot of the shoe sizes for family members.

1.5, 7, 6.5, 10, 7.5, 3, 10.5, 11, 6.5
7.0 10.0 3.0 11.0 ✓9 entries

1	5
2	
3	0
4	
5	
6	55
7	05
8	
9	
10	05
11	0

Key 1|5 = 1.5

Interp: spread out

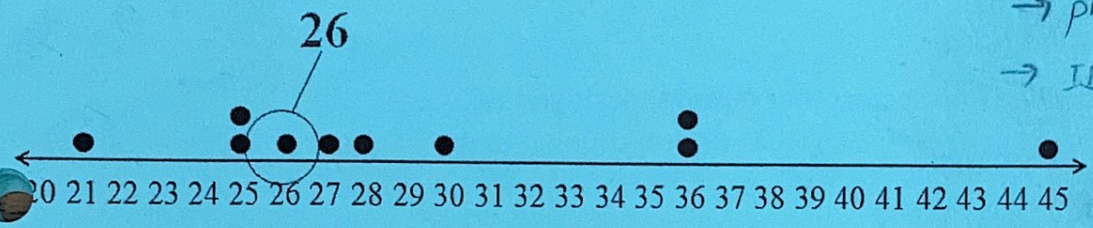
Graphing Quantitative Data Sets

Dot plot

- Each data entry is plotted, using a point, above a horizontal axis (a ___horiz___ line.)

benefits:
→ see how data is distr. (patterns)
→ preserve spec. data entries
→ ID unusual values (outliers)

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



Example: Constructing a Dot Plot

Use a dot plot to 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.
- There is no vert. axis.
- To represent a data entry, plot a point above the entry's position on the axis.
- There is no need to SORT the data prior to creating the plot.
- If an entry is repeated, plot another point above the previous point.

p. 57

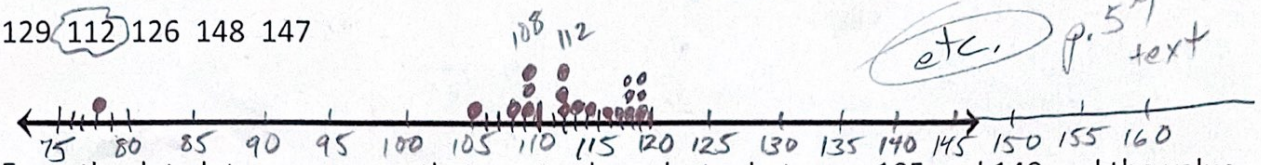
Solution: Constructing a Dot Plot

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



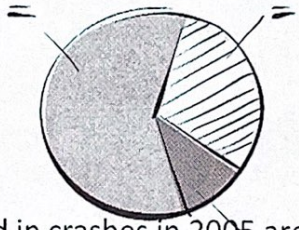
etc. p. 57 text

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 (an outlier.)

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 motor vehicle occupants killed in crashes in 2005 are shown in the table. Use a pie chart to organize the data. *(Source: U.S. Department of Transportation, National Highway Traffic Safety Administration)*

Vehicle Type	Killed:
Cars	18,440
Trucks	13,778
Motorcycles	4,553
Other	823

Need calc.

Solution: Constructing a Pie Chart

Find the relative frequency (percent) of each category

Vehicle Type	Frequency, f	Relative Frequency
Cars	18,440	.49
Trucks	13,778	.37
Motorcycles	4,553	.12
Other	823	.02

f/n

Give time to calc.

Total sum $\Sigma f = 37{,}594 = n$ (sample size)

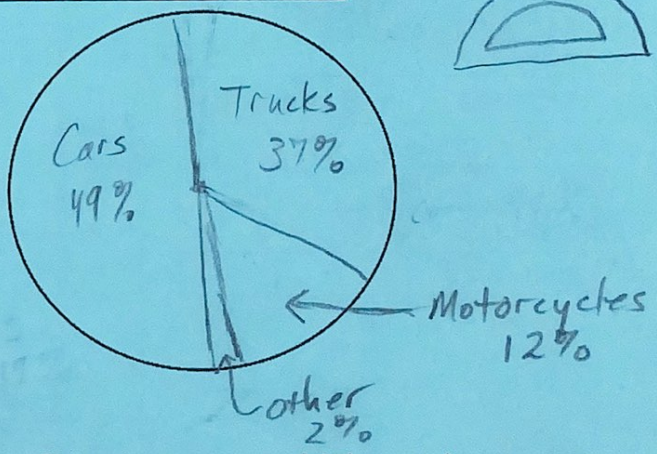
- Construct the pie chart using the central angle that corresponds to each category.
- To find the central angle, multiply 360 ° by the category's relative frequency.
- For example, the central angle for cars is

$360(0.49) \approx$ 176 °

→ "49% of 360" = 176.4

Vehicle Type	Frequency, f	Relative Frequency	Central Angle
Cars	18,440	0.49	176
Trucks	13,778	0.37	133
Motorcycles	4,553	0.12	43
Other	823	0.02	7

Vehicle type	Relative frequency	Central angle
Cars	0.49	176°
Trucks	0.37	133°
Motorcycles	0.12	43°
Other	0.02	7°



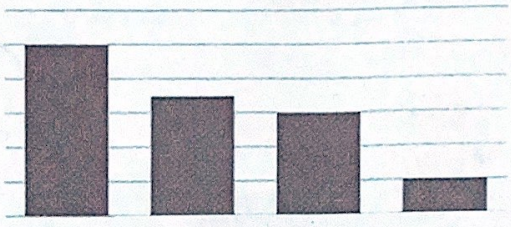
From the pie chart, you can see that most fatalities in motor vehicle crashes were those involving cars.

Graphing Qualitative Data Sets

Pareto Chart

- A vertical bar graph in which the height of each bar represents frequency or relative frequency.
- The vertical axis should start at 0.
- The bars are positioned in order of decreasing height, with the tallest bar positioned at the left.
- horiz axis: qualitative

Frequency



Categories

Example: Constructing a Pareto Chart

In a recent year, the retail industry lost $41.0 million in inventory shrinkage. Inventory shrinkage is the loss of inventory through breakage, pilferage, shoplifting, and so on. The causes of the inventory shrinkage are administrative error ($7.8 million), employee theft ($15.6 million), shoplifting ($14.7 million), and vendor fraud ($2.9 million). Use a Pareto chart to organize this data.

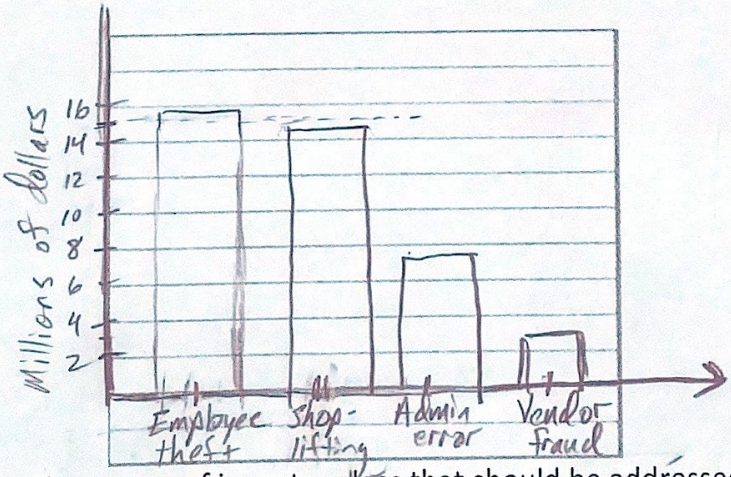
(Source: National Retail Federation and Center for Retailing Education, University of Florida)

Solution: Constructing a Pareto Chart

Cause	$ (million)
Admin. error	7.8
Employee theft	15.6
Shoplifting	14.7
Vendor fraud	2.9

most

smallest



From the graph, it is easy to see that the causes of inventory loss that should be addressed first are employee theft and shoplifting.

Section 4.2 Part II and Section 4.2 Part II Objectives

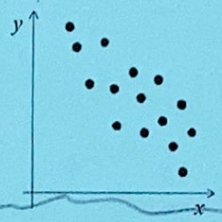
More Graphs and Displays

Can you graph paired data sets using scatter plots and time series charts?

Graphing Paired Data Sets

Paired Data Sets

- Each entry in one data set corresponds to one entry in a second data set.
- Graph using a scatter plot.
- The ordered pairs are graphed as points in a coordinate plane.
- Used to show the relationship between two quantitative variables.



A marketing manager conducts a study to see if there is a relationship between a person's age and the number of magazines to which that person subscribes. The results are shown below.

Solution: Create a scatter plot.

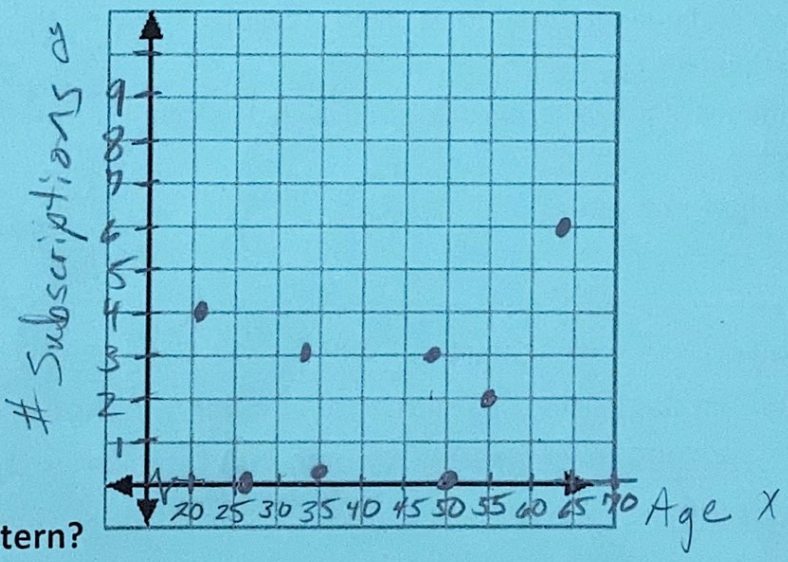
x

Age	55	48	26	21	33	50	64	35
# of Subscriptions	2	3	0	4	3	0	6	1

y

Age → low: 21, high: 64 → x-axis

# of Subscriptions → low: 0, high: 6 → y-axis

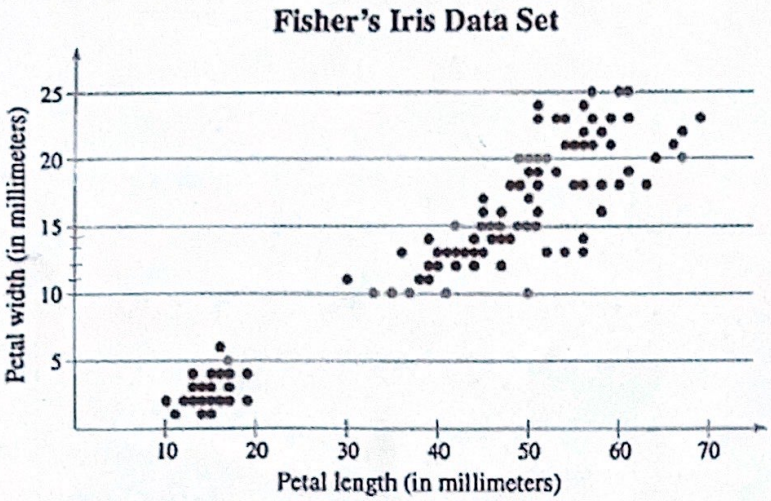


Is there a pattern?

No

Example: Interpreting a Scatter Plot

As the petal length increases, what tends to happen to the petal width?



Each point in the scatter plot represents the petal length and petal width of one flower.

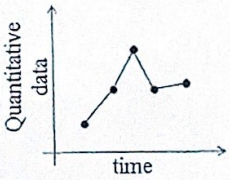
As petal length incr., petal width incr. also.

The British statistician Ronald Fisher introduced a famous data set called Fisher's Iris data set. This data set describes various physical characteristics, such as petal length and petal width (in millimeters), for three species of iris. The petal lengths form the first data set and the petal widths form the second data set.

Graphing Paired Data Sets

Time Series

- Data set is composed of quantitative entries taken at regular intervals over a period of time.
  - e.g., The amount of precipitation measured each day for one month.
- Use a **time series** chart to graph.



trend: incr, decr, incr. slightly

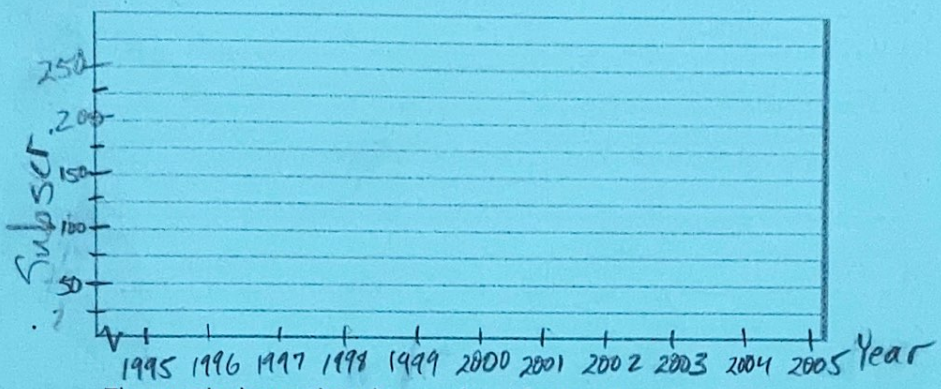
Example: Constructing a Time Series Chart

The table on the following page lists the number of cellular telephone subscribers (in millions) for the years 1995 through 2005. Construct a time series chart for the number of cellular subscribers. *(Source: Cellular Telecommunication & Internet Association)*

Solution: Constructing a Time Series Chart

- Let the horizontal axis represent the time (in years).
- Let the vertical axis represent the number of subscribers (in millions).
- Plot the paired data and connect them with line segments.

Year	Subscribers (in millions)
1995	33.8
1996	44.0
1997	55.3
1998	69.2
1999	86.0
2000	109.5
2001	128.4
2002	140.8
2003	158.7
2004	182.1
2005	207.9



The graph shows that the number of subscribers has been increasing since 1995, with greater increases recently

Using the graphing calculator to create a scatter plot

Age	55	48	26	21	33	50	64	35
# of Subscriptions	2	3	0	4	3	0	6	1

Button	Comments
STAT	
EDIT	(or just hit ENTER)
Clear L1 and L2	
Input data for L1 and L2	Scatterplots are used with 2 quantitative variables
STATPLOT	2^{nd}, Y=
ENTER	to choose Plot 1
Use arrow and ENTER to turn on the plot, if it isn't already on.	
Use arrow to choose the scatterplot.	L1 and L2 are automatically chosen. If you want two different lists, use 2ND and the appropriate lists, separated by a common (above 7)
ZOOM	
9	