



## Statistics & Data Analysis Concepts for Data Science and ML **2**

---

# Descriptive Statistics : Graphical Methods – *Data Visualization*

# Learning Objectives

- Learn the techniques of summarizing and describing data using charts and graphs.
- Learn to construct a frequency distribution from a set of data
- Calculate relative frequency, cumulative frequency, and relative cumulative frequency from a frequency table and interpret their meaning
- Construct different types of graphs using quantitative data including histograms, frequency polygons, ogives, stem-and-leaf plots, dot plots and interpret these plots
- Construct bar charts and pie charts using qualitative (categorical) data and learn their applications
- Construct other types of charts and graphs including time series plots and scatter plots
- Use computer packages to construct the above charts and graphs

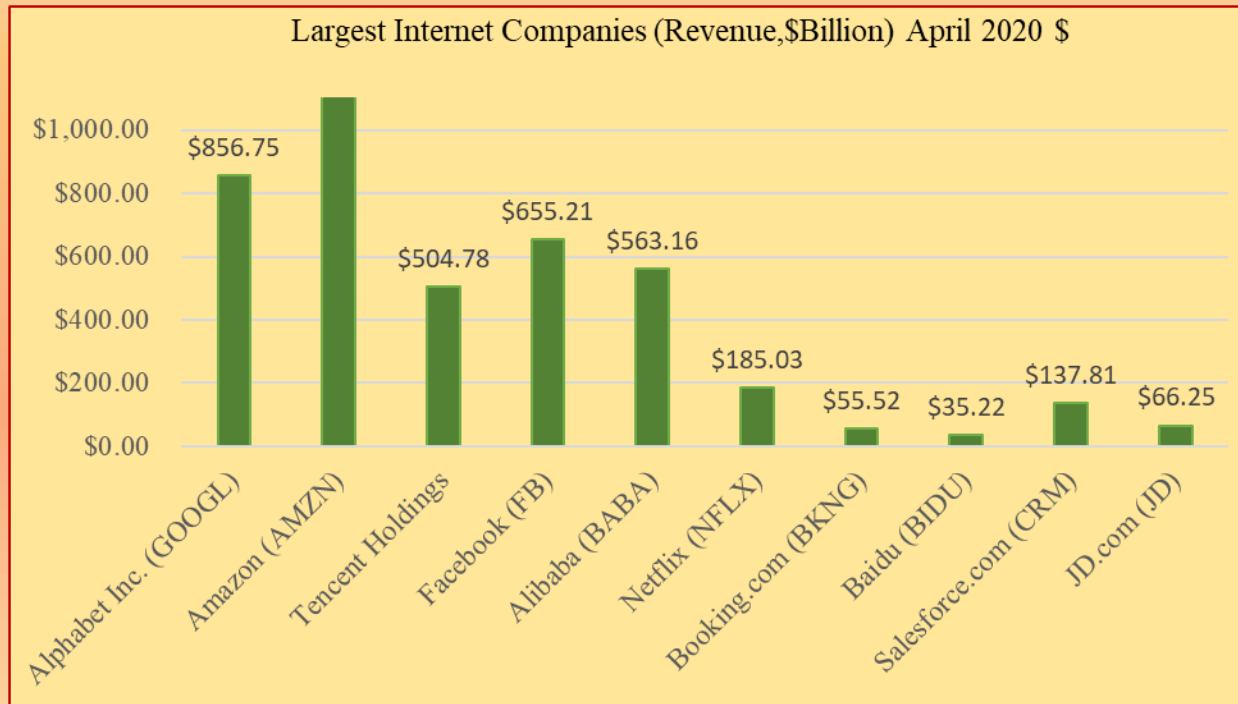
# Graphical/Visual Techniques

---

- It is said that a picture is worth a thousand words; this is particularly true when a large set of data is effectively presented using charts and graphs that quickly reveal important features.
- Visual displays of the data are easily recognizable and found ubiquitously in business periodicals, financial magazines, on the internet, and televisions.

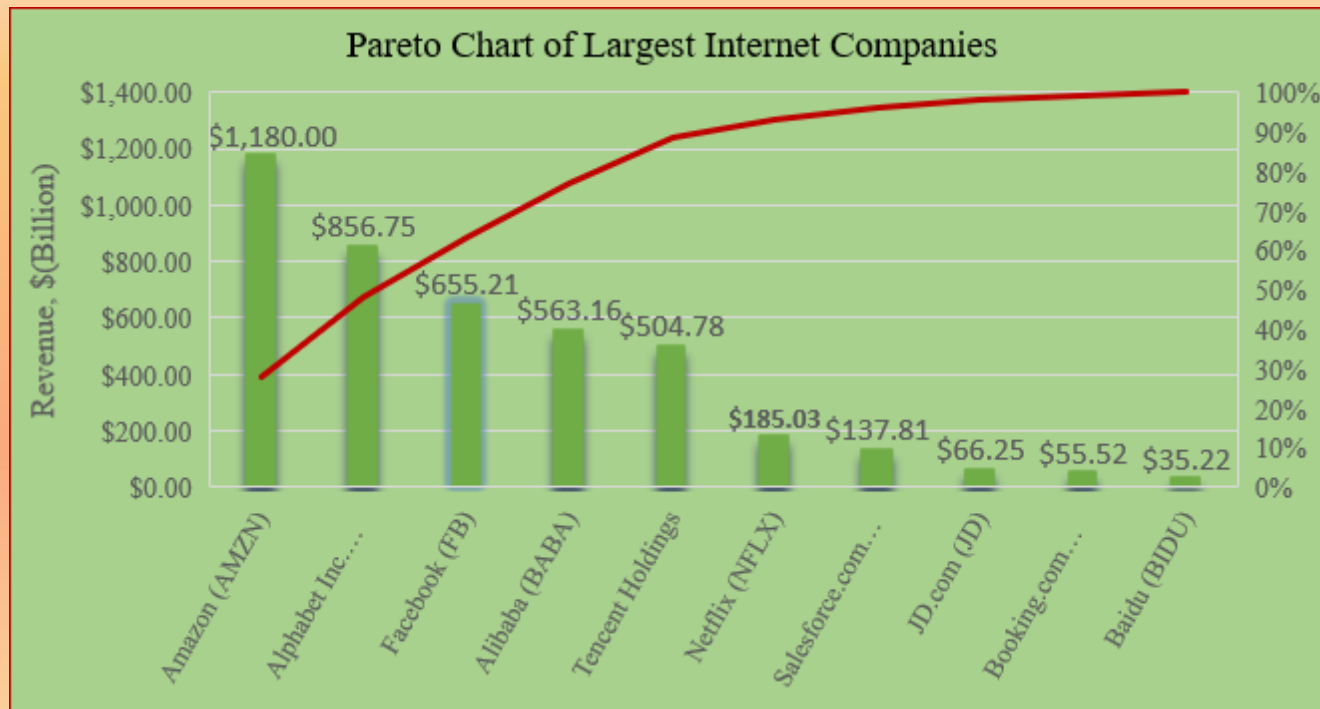
# Real World Examples of Graphs

Graphs summarizing the market value of largest Internet Companies April 2020



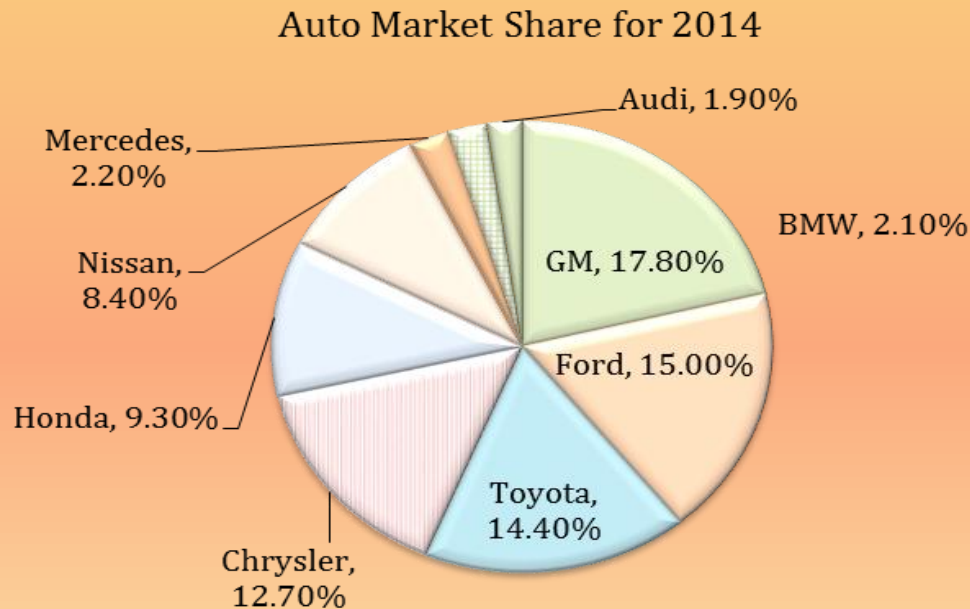
# Real World Examples of Graphs

Graphs summarizing the market value of Internet Companies  
( Largest to Smallest) April 2020

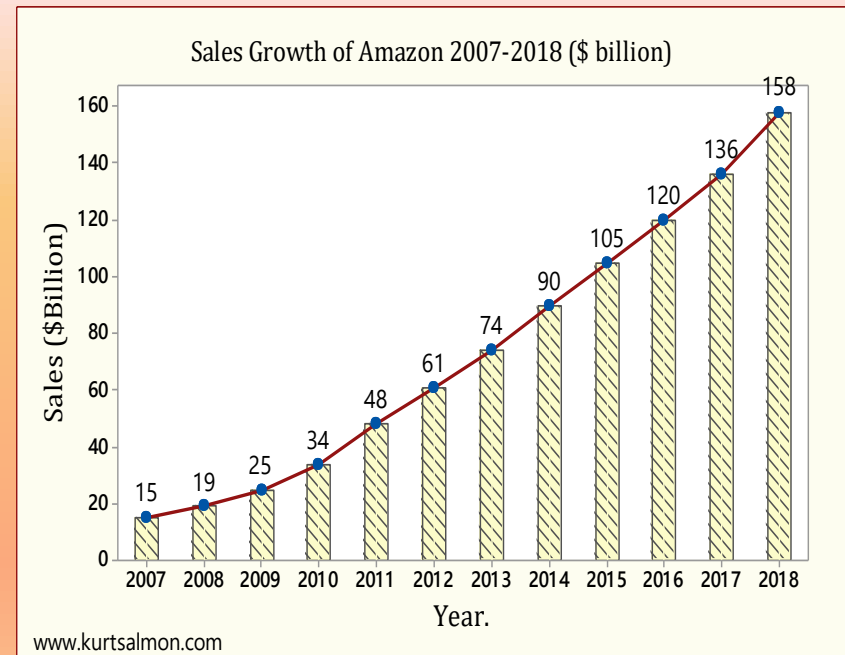
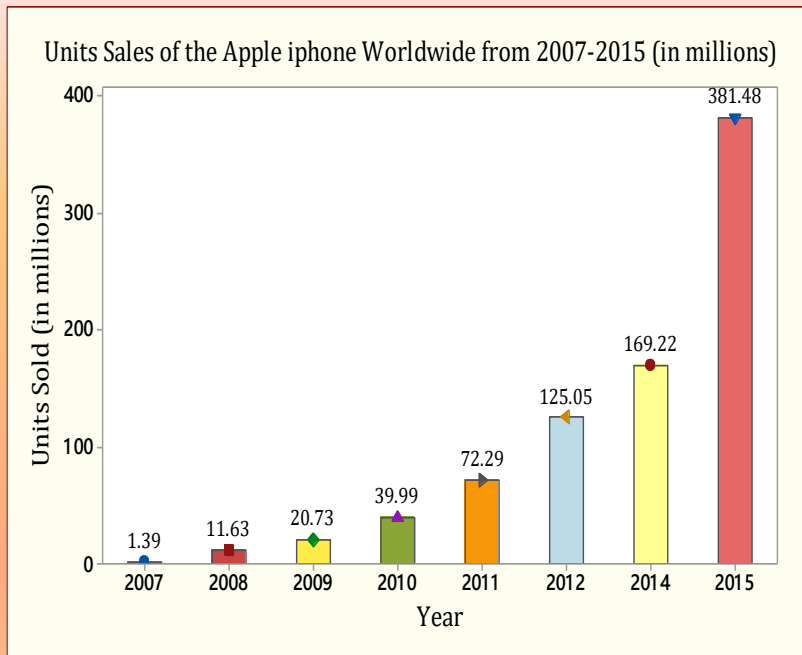


# Real World Examples of Graphs

Graphs summarizing the market value of largest Internet Companies and Auto Market Share for 2014



# More Examples of Graphs



A number of charts and graphs can be found in reports of financial periodicals like *The Economist*, *Business Week*, *Fortune* and many others. Almost every issue of *USA Today* and *The Wall Street Journal* contain a number of visual displays in their articles.

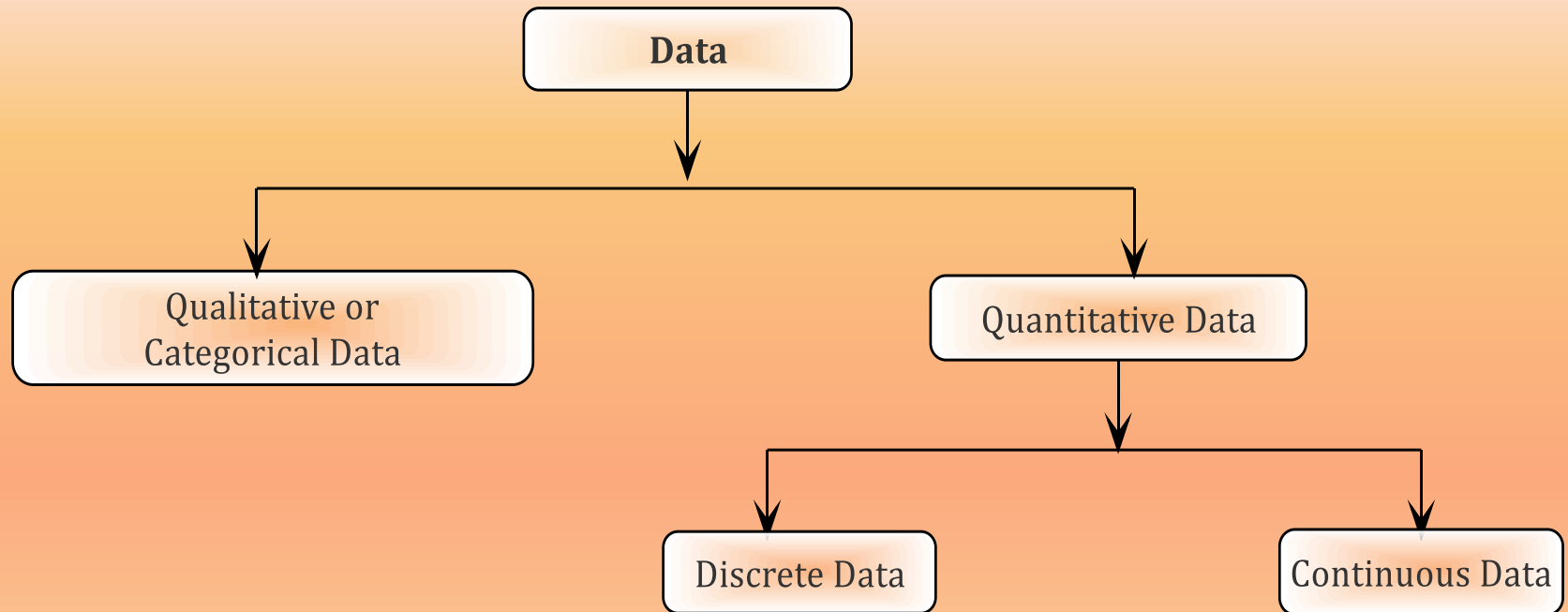
# What you will learn

---

In this chapter, you will learn:

- A number of charts and graphs that are used to describe the key features of data.
- How these graphs enable one to summarize a large sets of data.
- To create Graphs of both the quantitative and categorical data
- You will also learn how to use the computer software including Excel and MINITAB to construct and interpret the charts and graphs.

# Classification of Data: Review



# Different Forms of Data

Lifetime of 200 Television Components (in Hours)  
Ungrouped and Raw Data

314	330	371	365	267	307	371	297	291	398
276	253	286	344	385	349	269	304	319	283
430	253	378	306	376	308	339	368	289	344
340	298	330	311	318	358	354	406	369	254
322	242	331	236	344	418	328	393	267	305
325	282	315	328	319	353	336	384	298	398
343	203	373	297	276	333	257	367	296	349
322	325	252	345	373	317	307	289	363	340
309	246	302	260	292	231	338	372	226	365
271	302	331	374	355	336	312	354	329	345
276	329	379	288	356	302	263	364	337	361
416	360	337	273	298	390	215	382	329	306
306	279	414	262	372	303	346	331	362	366
387	304	302	280	287	368	281	329	309	310
375	346	413	309	283	299	335	330	376	260
277	366	345	409	312	266	383	289	294	370
359	363	243	339	323	297	333	299	302	384
370	357	314	348	257	291	358	409	337	347
215	277	313	300	322	304	282	410	390	332
373	280	339	349	363	297	274	334	359	330

Data Array: Data Arranged in Increasing Order  
(Read row-wise)

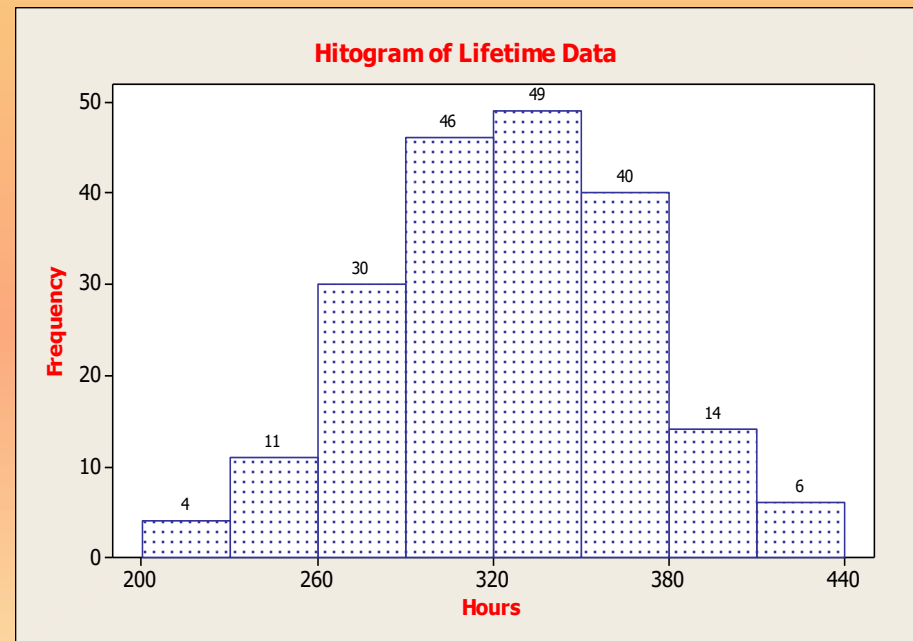
203	215	215	226	231	236	242	243	246	252
253	253	254	257	257	260	260	262	263	266
267	267	269	271	273	274	276	276	276	277
277	279	280	280	281	282	282	283	283	286
287	288	289	289	289	291	291	292	294	296
297	297	297	297	298	298	298	299	299	300
302	302	302	302	302	303	304	304	304	305
306	306	306	307	307	308	309	309	309	310
311	312	312	313	314	314	315	317	318	319
319	322	322	322	323	325	325	328	328	329
329	329	329	330	330	330	330	331	331	331
332	333	333	334	335	336	336	337	337	337
338	339	339	339	340	340	343	344	344	344
345	345	345	346	346	347	348	349	349	349
353	354	354	355	356	357	358	358	359	359
360	361	362	363	363	363	364	365	365	366
366	367	368	368	369	370	370	371	371	372
372	373	373	373	374	375	376	376	378	379
382	383	384	384	385	387	390	390	393	398
398	406	409	409	410	413	414	416	418	430

# Frequency Distribution – Grouping and Graphical Presentation - Example

## Frequency Distribution (Grouped Data)

Class- interval	Frequency (f)
200 - 230	4
230 - 260	11
260 - 290	30
290 - 320	46
320 - 350	49
350 - 380	40
380 - 410	14
410 - 440	6
	$\sum f = 200$

## Graph of frequency Distribution or Histogram



# Frequency Distribution

Frequency distribution provides a compact representation of data.

This is another name for grouping

- Compact representation is obtained by **arranging the data into groups** or **class intervals** usually of **equal width**, and
- recording or counting the number of observations in each interval. Counting the number of observations in each group is called the **class frequency**

**A grouped data or frequency distribution would look like:**

<b>Class- interval</b>	<b>Frequency</b>
200 – 230	4
230 - 260	11
260 – 290	30
:	
and so forth	

The class interval 200 – 230 means  $200 \leq x < 230$  (includes the value 200 but not 230). The value 200 is known as the **lower class boundary or lower class limit** and the value 230 is known as the **upper class boundary or upper class limit**.

# Frequency Distribution...cont.

---

- There is no unique frequency distribution for a given set of data
- Several frequency distributions are possible for the same set of data
- When dividing the data into class intervals, 5 to 15 class intervals are recommended
- If there are *too many class intervals*, the class frequency (count) is low and the saving in computational effort is small
- If there are too few class intervals, the true characteristic of the distribution may be obscured and some information may be lost.

# Frequency Distribution...cont.

---

- How many class-intervals or groups?

The *number of class intervals* should be governed by the *amount* and *scatter* of data present.

An *estimate* for the number of class intervals can be calculated using the following formula:

where, K = number of class intervals

$$K = 1 + 3.33 \log_{10} n$$

- If there are n=200 observations then using the above formula,

$$\begin{aligned} K &= 1 + 3.33 \log_{10} 200 \\ &= 1 + 3.33 (2.3010) \\ &= 8.66 \text{ or approximately } 9 \text{ classes} \end{aligned}$$

# Frequency Distribution...cont.

---

- Class intervals should be chosen so that no result falls on a class boundary.
- Class intervals may be written two different ways

## *Class Intervals 1 (upper boundary inclusive)*

200 - 229

230 - 259

260 - 289

:

and so on (In this case, the upper boundary is inclusive)

## *Class Intervals 2 (upper boundary exclusive)*

200 - 230

230 - 260

260 - 290

:

and so forth (In this case, the upper boundary is exclusive)

# Frequency Distribution...cont.

- What should be the width of each class?

The class width is of equal size, the number of classes determines the width of each class. Use the following formula to determine the class width:

*(Largest value in the data - Smallest value)*

$$\text{Width of class interval} = \frac{\text{Largest value in the data} - \text{Smallest value}}{\text{Number of Class Intervals}}$$

If the number of observations,  $n = 200$  and  $K =$  number of classes is 9. Using these values, the minimum and maximum values in the data are 203 and 430 then class width,

$$= \frac{430 - 203}{8} = 28.375$$

**Note: The number of class intervals and the class width using the above formulas are approximate and are not exact. They can be adjusted to get desired class intervals and width.**

# Summarizing Quantitative Data: Frequency Distribution – another example

Table 2.1: Driving Speed (mph), n=100

51	46	62	70	54	59	59	57	61	66	49	57	57	65	61	62	51	63	62	65	55	55	65
64	60	55	70	61	63	55	70	65	51	53	49	62	56	61	64	54	60	63	69	72	69	60
57	63	60	56	60	61	57	57	61	54	58	55	69	63	55	58	58	62	59	59	62	53	69
56	59	57	60	63	60	56	52	65	58	60	62	54	57	60	53	56	60	71	59	64	58	71
68	62	61	61	67	59	58	49															

Table 2.2: Driving Speed (mph) - (Sorted Data)

46	49	49	49	51	51	51	52	53	53	53	54	54	54	54	55	55	55	55	55	55	56	56
56	56	56	57	57	57	57	57	57	57	57	58	58	58	58	58	58	59	59	59	59	59	59
59	60	60	60	60	60	60	60	60	60	60	61	61	61	61	61	61	61	61	62	62	62	62
62	62	62	62	63	63	63	63	63	63	63	64	64	64	65	65	65	65	65	66	67	68	69
69	69	70	70	70	71	71	72															

# Forming Frequency Distribution

1. Approximate number of classes can be found using the formula:

$$K = 1 + 3.33 \log_{10} n$$

$$K = 1 + 3.33 \log_{10} 100$$

$$\begin{aligned} K &= 1 + 3.33(2.0) \\ &= 7.66 \text{ or approximately 8 classes} \end{aligned}$$

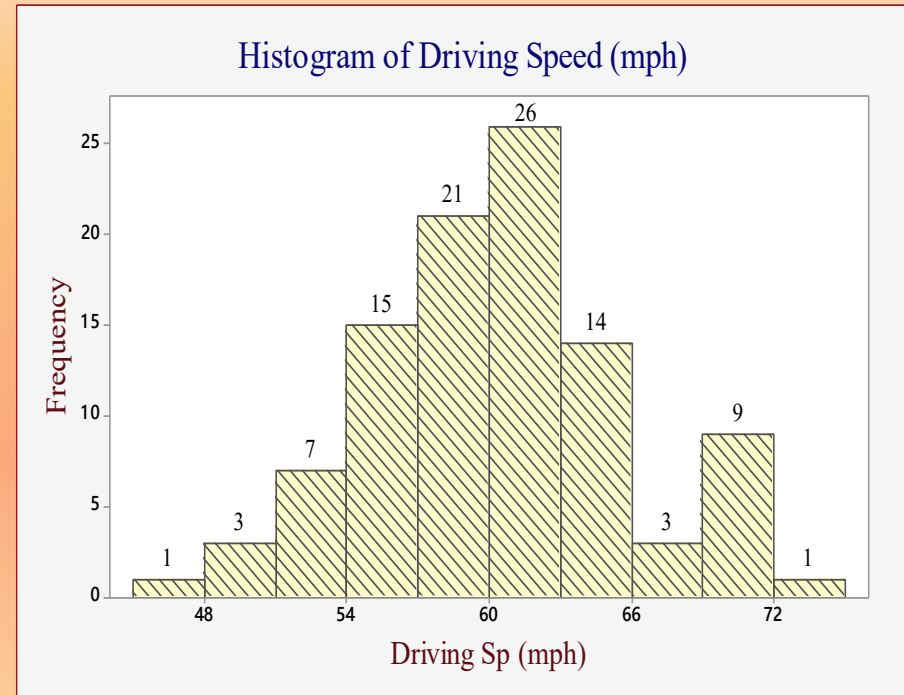
2. Class- width using 10 class intervals:

$$\text{Class width} = \frac{72 - 46}{10} = 2.6$$

# Frequency Distribution and Histogram

Frequency distribution of 100 drivers with 60 miles per hour (mph) speed limit

Class-interval (mph)	Frequency (f)
45- 48	1
48 - 51	3
51 - 54	7
54 - 57	15
57 - 60	21
60 - 63	26
63 - 66	14
66 - 69	3
69 - 72	9
72 - 75	1
Total	$\sum f_i = 100$



Histogram of Driving Speed (mph)  
(10 class-intervals)

# Different Frequency Distribution for the Same Data Set

Table 2.4: Utility Charges of 50 Customers (rounded to nearest dollar)

Utility Charge (\$)												
98	173	204	180	149	104	155	199	129	84	159	187	92
118	174	113	150	215	132	167	143	151	208	177	125	130
146	170	111	169	97	165	152	156	132	145	189	168	141
151	110	121	185	153	116	137	193	139	131	160		

Table 2.5: Sorted Data

Utility Charge (Sorted) – Read Row wise												
84	92	97	98	104	110	111	113	116	118	121	125	129
130	131	132	132	137	139	141	143	145	146	149	150	151
151	152	153	155	156	159	160	165	167	168	169	170	173
174	177	180	185	187	189	193	199	204	208	215		



Form a frequency distribution having: [a] Five class intervals, [b] Six class intervals, and [c] Seven class intervals.

# Different Frequency Distribution for the Same Data Set ...cont.

---

- Frequency Distribution with Five Class-intervals

Determine the class-width using

$$\text{Width} = \text{Range} / \text{Number of classes} = (215-84)/5 = 26.2$$

Use a width of 30

Table 2.6: Frequency Distribution with Five Class Intervals

Class Intervals	Frequency( $f_i$ )
80 - 110	5
110 - 140	14
140 - 170	18
170 - 200	10
200 - 230	3
	$\sum f_i = 50$



# Different Frequency Distribution for the Same Data Set ...cont.

## Frequency Distribution with Six and Seven Class-intervals

$$\text{Width} = (\text{Maximum value} - \text{Minimum value}) / \text{Number of classes} = (215-84)/6 = 21.8$$

Use a class-width of 25 in this case.

Table 2.8: Frequency Distribution with Seven Class Intervals

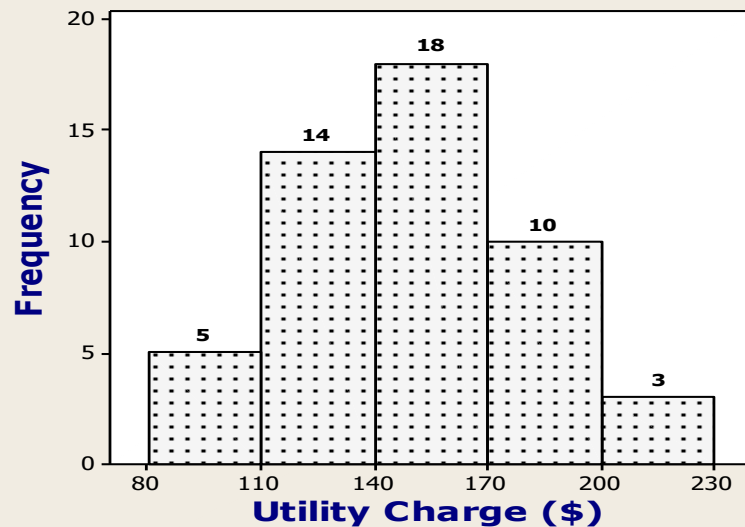
Table 2.7: Frequency Distribution with Six Class Interval

Class Intervals	Frequency ( $f_i$ )
75-100	4
100-125	7
125-150	13
150-175	16
175-200	7
200-225	3
	$\sum f_i = 50$

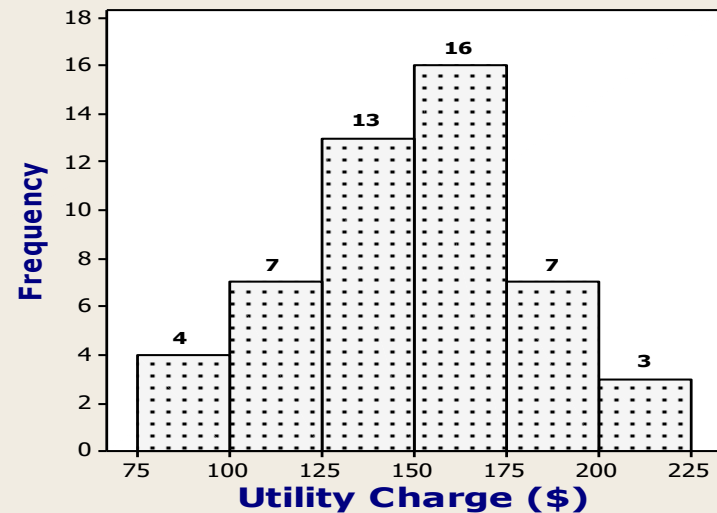
Class Intervals	Frequency ( $f_i$ )
80-100	4
100-120	6
120-140	9
140-160	13
160-180	9
180-200	6
200-220	3
	$\sum f_i = 50$

# Histograms or the Plot of frequency Distribution

**Histogram with Five Class Intervals**



**Histogram with Six Class Intervals**



# Types of Histogram

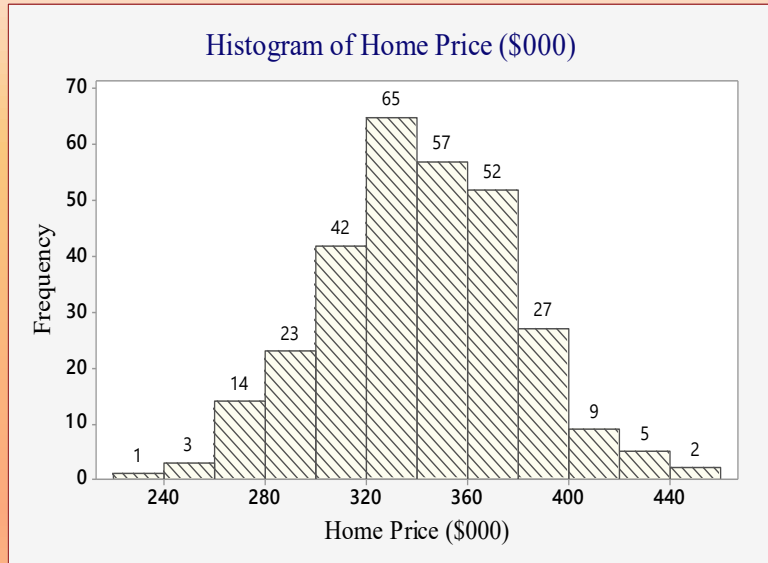
---

- Frequency Histogram
- Relative Frequency Histogram
- Percent Frequency Histogram

All of the above histograms have the same shape. The difference is in plotting the values on the y-axis.

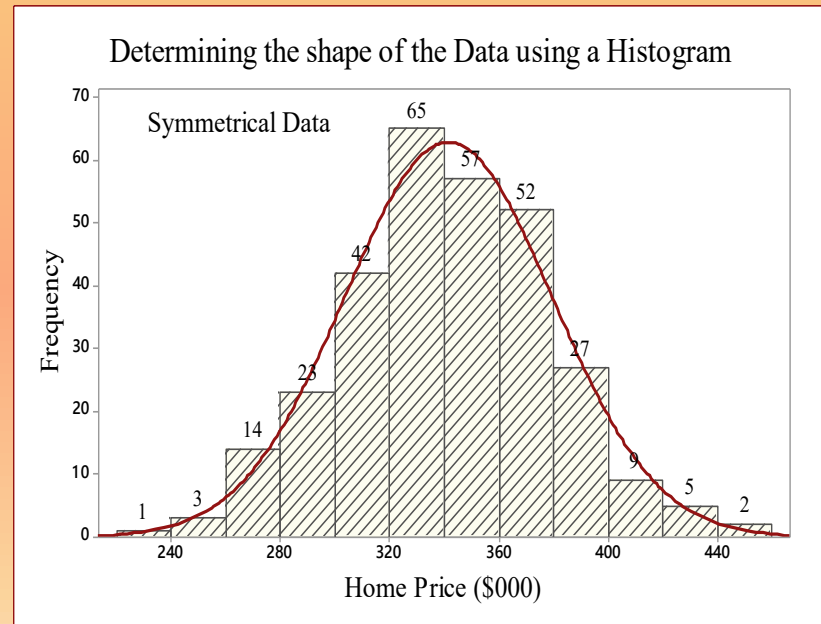
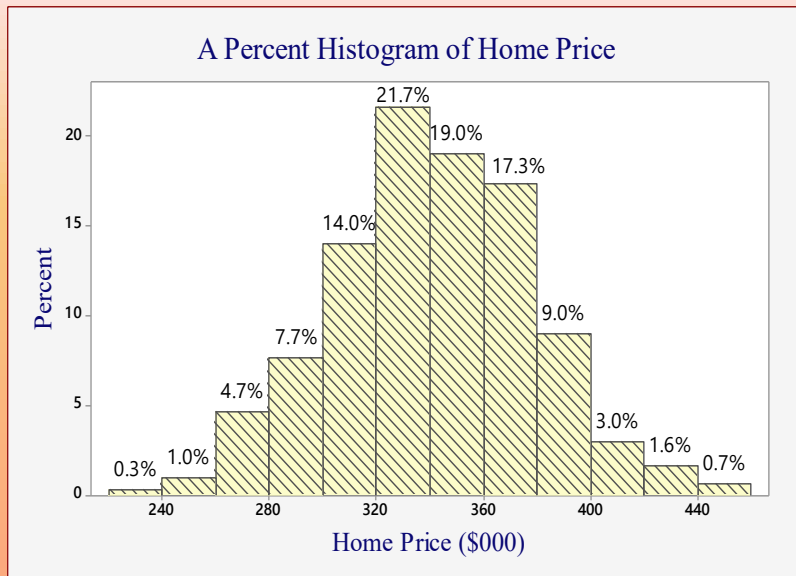
# Histogram: Summarizing the data and examining the distribution (shape)

Class-interval Selling Price(\$000)	Frequency ( $f$ ) No. of Houses	Relative Frequency
220 – 240	1	0.003
240 – 260	3	0.010
260 – 280	14	0.047
280 – 300	23	0.077
300 – 320	42	0.140
320 – 340	65	0.217
340 – 360	57	0.190
360 – 380	52	0.173
380 – 400	27	0.090
400 – 420	9	0.030
420 – 440	5	0.017
440 – 460	2	0.007
	$\sum f = 300$	Sum=1.00

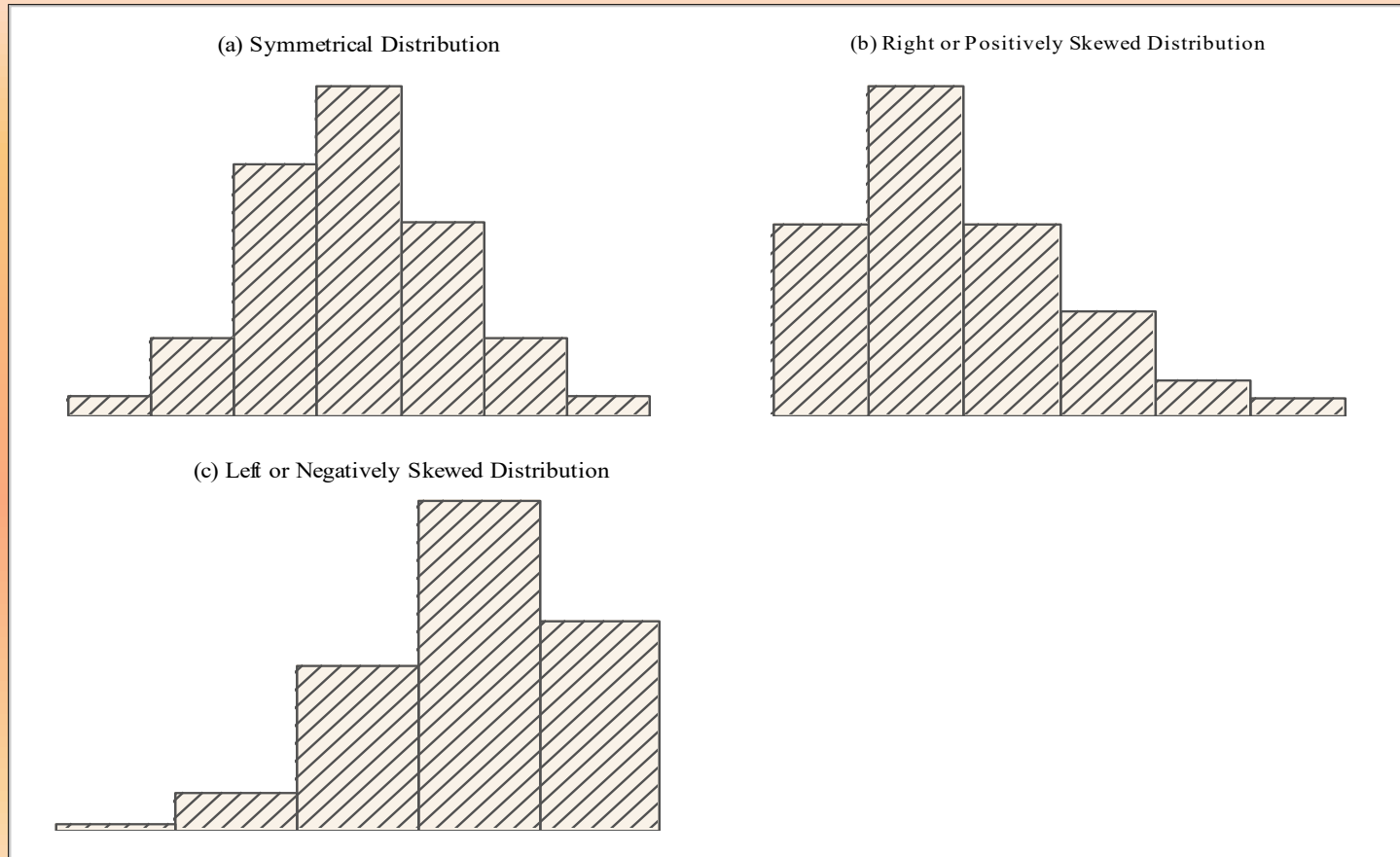


Frequency Histogram of Home Price

# Determining the shape or distribution of the data using a histogram



# (a) Symmetrical Distribution (b) Right or Positively Skewed Distribution (c) Left or Negatively Skewed Distribution



## Example 1:

A data set contains 250 observations. The minimum value is 60 and the maximum value is 510. What class-width would you recommend for the data?

Solution:

First, determine the possible number of class-intervals using

$$K = 1 + 3.33 \log n = 1 + 3.33 \log 250 = 8.99 \text{ or } 9 \text{ classes}$$

The class width is given by

$$\begin{aligned} \text{Width} &= \text{Max. Value} - \text{Min. Value} / \text{No. of classes} \\ &= 510 - 60 / 9 = 50 \end{aligned}$$

A class-width of 50 should be used.

## Example 2

A Frequency Distribution of 100 Drivers with 60 Miles per Hour Speed Limit is given below. Calculate the relative frequency, cumulative frequency, and relative cumulative frequency.

Class-interval (mph)	Frequency (f)
45 - 48	1
48 - 51	4
51 - 54	9
54 - 57	14
57 - 60	25
60 - 63	24
63 - 66	10
66 - 69	5
69 - 72	7
72 - 75	1
Total	$\sum f_i = 100$



# Solution

Class-interval, Frequency, Relative Frequency, Cumulative and Relative Cumulative Frequency

Class-interval (mph)	Frequency (f)	Relative Frequency	Cumulative Frequency	Relative Cumulative Frequency
45- 48	1	0.01	1	0.01
48 - 51	4	0.04	5	0.05
51 - 54	9	0.09	14	0.14
54 - 57	14	0.14	28	0.28
57 - 60	25	0.25	53	0.53
60 - 63	24	0.24	77	0.77
63 - 66	10	0.10	87	0.87
66 - 69	5	0.05	92	0.92
69 - 72	7	0.07	99	0.99
72 - 75	1	0.01	100	1.00
Total	$\sum f_i = 100$	1.00		

(a) Find the percentage of drivers who were speeding (driving 66 mph or above).

**13%**

(c) Find the percentage of the drivers whose speed does not exceed 60 mph.

**53%**

## Example 3

The data in Table 2.16 show the sales (in thousands of dollars) for the past 50 days for a particular store. Table 2.17 shows the sorted data.

Table 2.16: Sales (in thousands of dollars)

Sales in thousands of dollars for the past 50 days

63	74	77	83	85	86	88	90	94	96	97	99	100
102	103	103	105	106	108	109	110	111	112	112	113	113
114	115	115	116	117	120	121	121	121	122	125	126	126
128	135	135	135	137	138	139	140	140	155	161		

Table 2.17:(Sorted Data from Table 2.16: Read row wise)

63	74	77	83	85	86	88	90	94	96	97	99	100
102	103	103	105	106	108	109	110	111	112	112	113	113
114	115	115	116	117	120	121	121	121	122	125	126	126
128	135	135	135	137	138	139	140	140	155	161		

- A frequency distribution is to be formed of the sales data in Table 2.16. Suggest the number of class intervals for the frequency distribution.
- Based on the number of class intervals in part (a), determine the class-width for the classes.
- Use the number of class-intervals in (a) and class width in (b) to form a frequency distribution. Also, calculate the relative frequency, cumulative frequency, and relative cumulative frequency.

# Solution

(a)  $K = 1 + 3.33 \log n = 1 + 3.33 \log 50 = 6.66$  or seven class intervals

(b) Width = (Maximum - Minimum) / No. of classes =  $(161 - 63) / 7 = 14$ ;  
use a width of 15.

(c) The calculations are shown in the Table

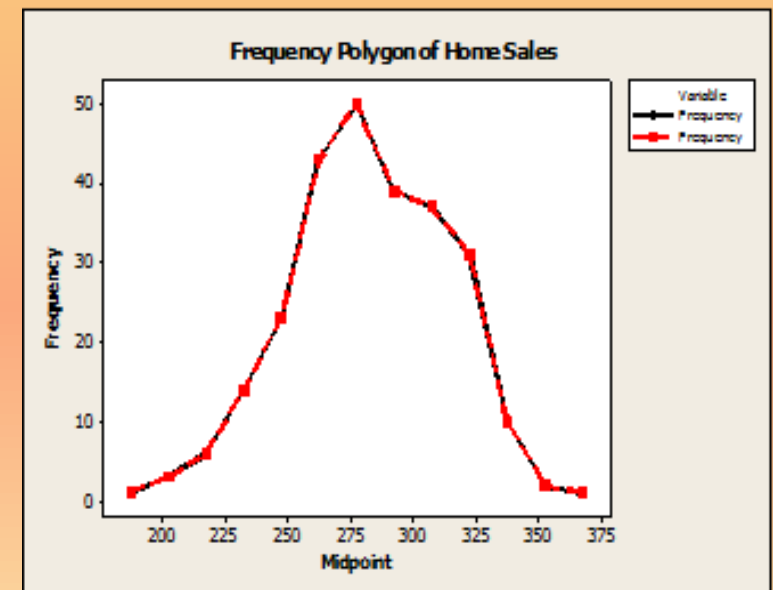
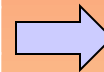
Class-interval (\$)	Frequency (f) (days)	Relative Frequency	Cumulative Frequency	Relative Cumulative Frequency
60 - 75	2	0.04	2	0.04
75 - 90	5	0.10	7	0.14
90 - 105	9	0.18	16	0.32
105 - 120	15	0.30	31	0.62
120 - 135	9	0.18	40	0.80
135 - 150	8	0.16	48	0.96
150 - 165	2	0.04	50	1.00
	$\sum f_i = 50$	1.00		

# Frequency Polygon

A frequency polygon is the plot of class frequencies and the midpoint – this plot is helpful in visualizing the pattern from the data (that is, if the data are symmetrical or skewed). See the example below

Midpoint for the Frequency Distribution and the Midpoints

Class-interval Selling Price(\$000)	Midpoint	Frequency No. of Houses
180 – 195	187.5	1
195 – 210	202.5	3
210 – 225	217.5	6
225 – 240	232.5	14
240 – 255	247.5	23
255 - 270	262.5	43
270 - 285	277.5	50
285 - 300	292.5	39
300 - 315	307.5	37
315 - 330	322.5	31
330 - 345	337.5	10
345 - 360	352.5	2
360 - 375	367.5	1
		$\sum f = 260$



# Calculating Relative Frequency, Cumulative frequency, and Relative Cumulative Frequency

Table 2.9: Frequency Distribution with Seven Class Intervals

(1) Class Intervals (\$)	(2) Frequency (f <sub>i</sub> ) No. of Customers	(3) Relative Frequency	(4) Cumulative Frequency	(5) Relative Cumulative Frequency
80-100	4	4/50=0.08	4	4/50=0.08
100-120	6	6/50=0.12	6+4= 10	10/50=0.20
120-140	9	9/50=0.18	9+6+4=19	19/50=0.38
140-160	13	13/50=0.26	13+9+6+4=32	32/50=0.64
160-180	9	9/50=0.18	9+13+9+6+4=41	41/50=0.82
180-200	6	6/50=0.12	6+9+13+9+6+4=47	47/50=0.94
200-220	3	3/50=0.06	3+6+9+13+9+6+4=50	50/50=1.00
	$\sum f_i = 50$	1.00		

$$\text{Relative Frequency of a Class} = \frac{\text{Frequency of that class}}{\text{Total Number of Observations}}$$

$$\text{Cumulative Relative Frequency of a Class} = \frac{\text{Cumulative Frequency of that class}}{\text{Total Number of Observations}}$$

# Ogive: The Plot of Cumulative or Relative Cumulative Frequency

- An ogive is a very useful plot in estimating the data values between the class intervals when the data are grouped into a frequency distribution. It also can be used in estimating different percentiles from the data

- Example:**

Table 2.24: Less than, Cumulative and Relative Cumulative Freq. of Data in Table 2.23

Table 2.23: Spot Speed for 1200 Cars

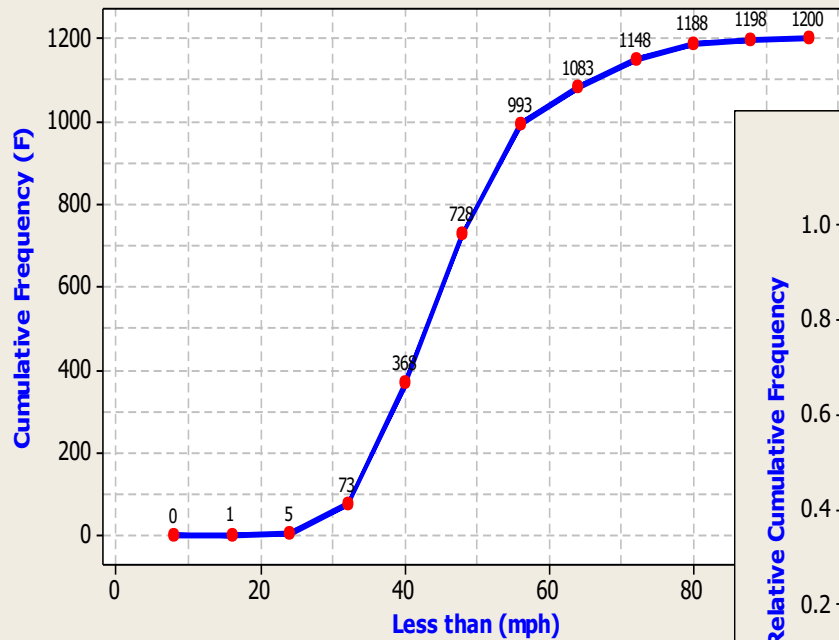
Class Interval (speed in mph)	Frequency (f) (No. of cars)
8 - 16	1
16 - 24	4
24 - 32	68
32 - 40	295
40 - 48	360
48 - 56	265
56 - 64	90
64 - 72	65
72 - 80	40
80 - 88	10
88 - 96	2
	$\sum f = 1200$



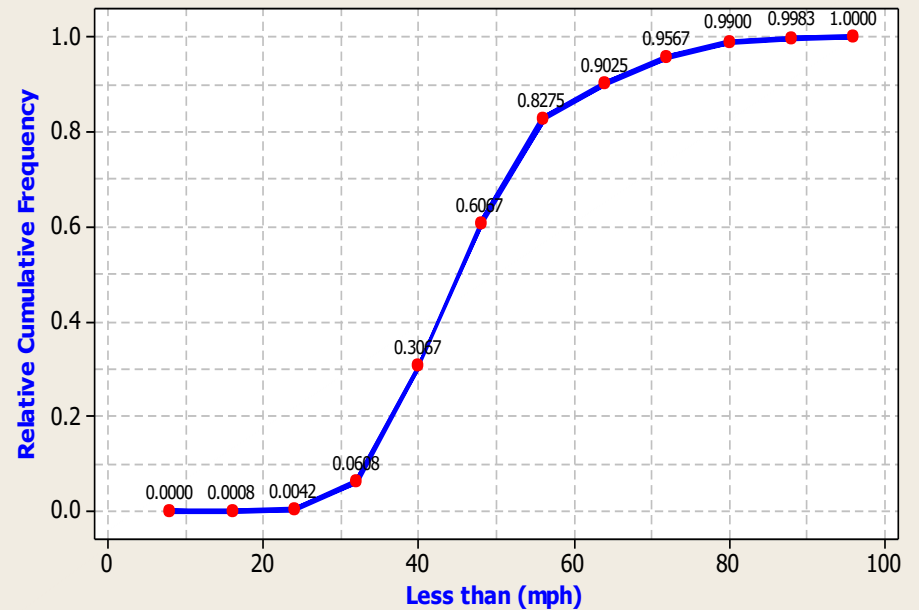
Less than (mph)	Cumulative Frequency (F)	Relative Cumulative Frequency
8	0	0.0000
16	1	0.0008
24	5	0.0042
32	73	0.0608
40	368	0.3067
48	728	0.6067
56	993	0.8275
64	1083	0.9025
72	1148	0.9567
80	1188	0.9900
88	1198	0.9983
96	1200	1.0000

# Ogives

A Less-than Ogive (Plot of Cumulative Frequencies)



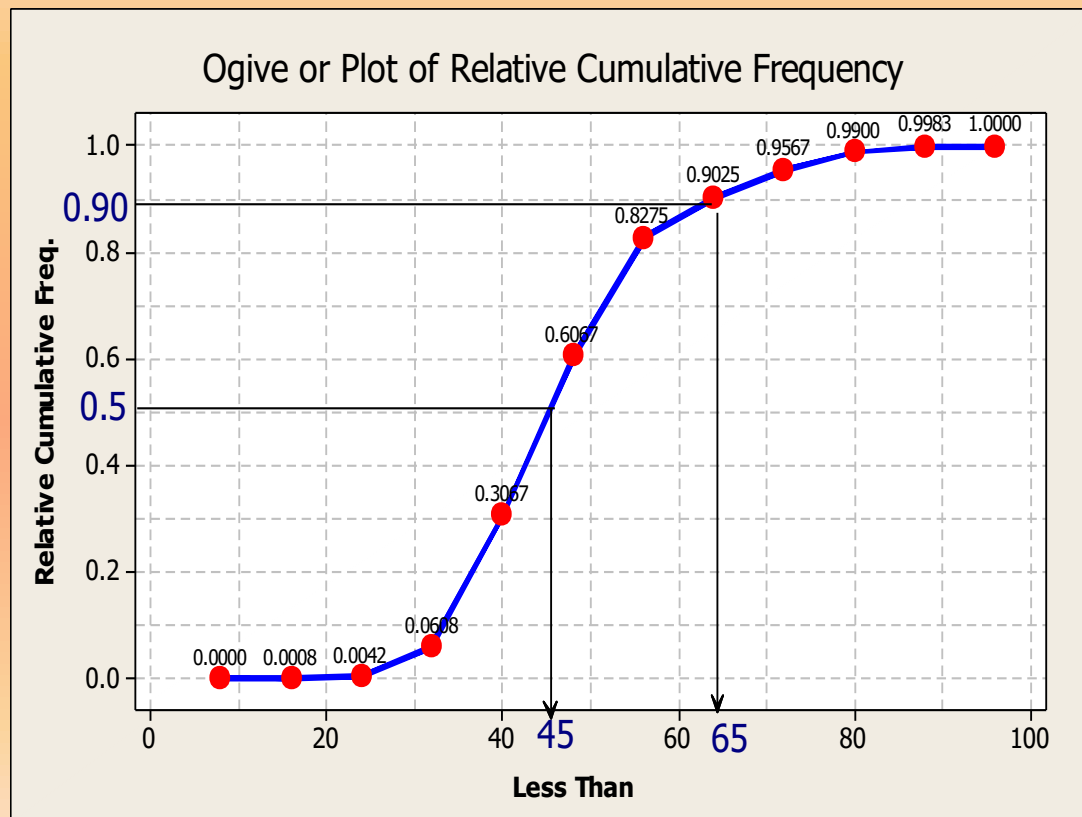
A Less-than Ogive (Plot of Relative Cumulative Frequency)



# WHAT INFORMATION CAN BE OBTAINED FROM THE OGIVE?

- What is the median speed of the cars sampled? 45
- Determine the speed of 90% of the cars crossing the intersection. 65 mph

See the figure



# Stem-and-Leaf Plot

- **Stem-and-leaf** plot is a very efficient way of displaying data, checking the variation and shape of the distribution.
- Stem-and-leaf plots are obtained by dividing each data value into two parts; stem and leaf.
- For example, if the data are two-digit numbers, e.g., 34, 56, 67, etc., then the first number (the tens digit) is considered the stem value, and the second number (the ones digit) is considered the leaf value. Thus, in data value 56, 5 is the stem and 6 is the leaf. In a three digit data value, the first two digits are considered as the stem and the third digit as the leaf.
- A data set with values 80, 82, 85, 87, and 89 have the common stem of 8
- A stem-and-leaf can be constructed easily using sorted data

# Stem-and-Leaf Plot

Table 2.27: Number of Defective Products

No. of Defects (out of 1000)

42	14	55	54	30	58	76	44	28	20	60	68	26
93	35	81	82	38	27	83	52	82	71	54	62	73
68	64	66	66	75	63	56	34	88	39	49	79	76
54	46	44	86	58	45	10	41	34	65	53	54	37
57	46	30	55	40	46	38	25	12	105	59	95	103

Table 2.28: Sorted Data from Table 2.27

No. of Defects (Sorted— read row-wise)

10	12	14	20	25	26	27	28	30	30	34	34	35
37	38	38	39	40	41	42	44	44	45	46	46	46
49	52	53	54	54	54	54	55	55	56	57	58	58
59	60	62	63	64	65	66	66	68	68	71	73	75
76	76	79	81	82	82	83	86	88	93	95	103	105



Stem-and-leaf of No. of Defects (out of 1000)

N = 65

Leaf Unit = 1.0

1

2

3

3 1 024

8 2 05678

17 3 004457889

27 4 0124456669

(13) 5 2344445567889

25 6 023456688

16 7 135669

10 8 122368

4 9 35

2 10 35

# Stem-and-Leaf Plot : Example


The stem-and leaf plot shows the number of orders received per day by a company.

1	9	2
2	10	3
5	11	245
7	12	78
8	13	2
11	14	137
15	15	1229
22	16	2266778
27	17	01599
(11)	18	00013346799
17	19	03346
12	20	4679
8	21	0177
4	22	45
2	23	18



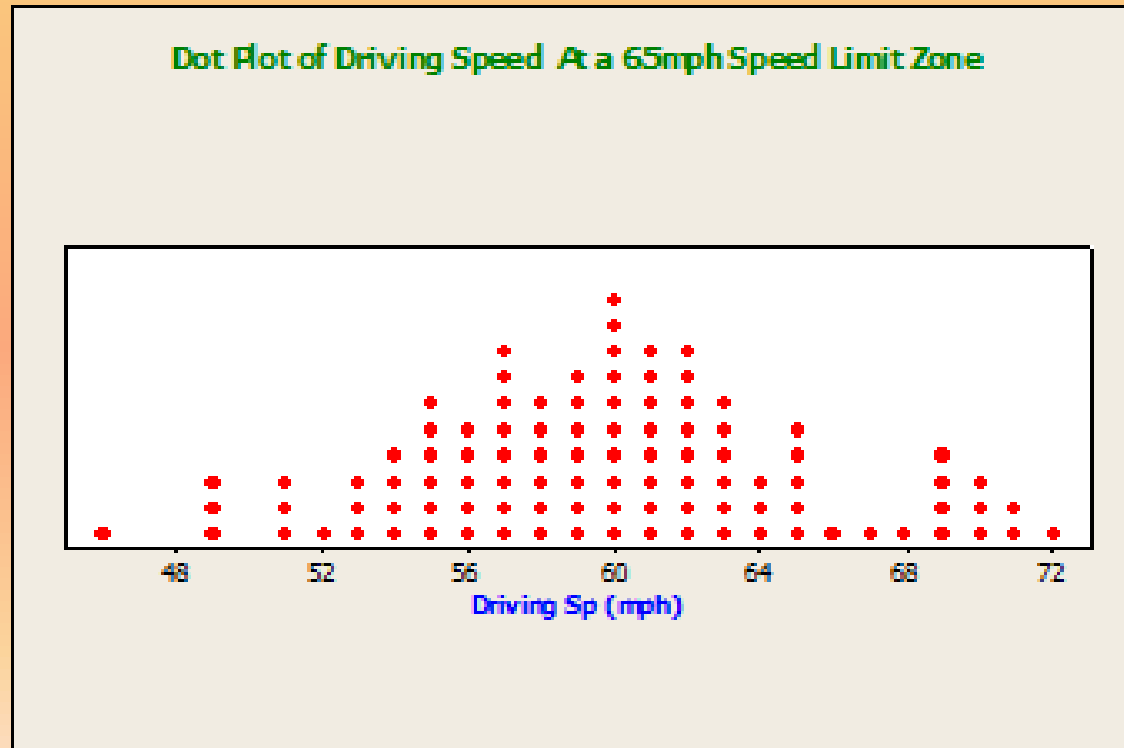
- [a] How many days were studied? **55**
- [b] How many observations are in the fourth class? **2**
- [c] What are the smallest and largest orders? **92, 238**
- [d] List the actual values in the sixth class? **141, 143, 147**
- [e] How many days did the firm receive less than 140 orders? **8**
- [f] How many days did the firm receive 200 or more orders? **12**
- [g] How many days did the firm receive 180 orders? **3**
- [h] What is the middle value? **180**
- [i] What can you say about the shape of the data? **Left or negatively skewed**

# Dot Plot

- A ***dot plot*** may be used to study the shape of the distribution or to compare two or more than two sets of data.
- In a dot plot, the horizontal axis shows the range of values in the data. Each observation is represented by a dot placed above the axis.
- If the data value repeats, the dots are piled up at that location, showing a dot for each repeated value.
- A dot in this plot may also represent multiple observations. 
- These plots are particularly useful when the data sets are small

# Dot Plot

Plot below shows the spot speed of 100 cars at a 65 mph speed limit zone. The plot shows that most of the cars were at or below the speed limit. There were 13 cars over the speed limit of 65 mph. The shape of the data is approximately symmetrical.



# Describing Categorical Data

## Tallies

A tally is a count of numbers in each category. It provides a table like a frequency distribution displaying the counts and percentages for each category.

A product was rated by 200 customers using a scale of 1 to 10 where 1 is unacceptable, 10 is outstanding, and 5 is average

Product Rating														
5	5	10	5	5	2	4	7	3	1	10	8	8	2	1
5	8	4	5	1	8	3	5	2	2	3	8	2	5	2
3	3	4	8	6	3	2	7	6	6	5	3	4	6	6
6	2	5	2	4	1	2	6	1	9	5	5	6	1	7
5	6	4	7	6	6	3	7	3	6	10	4	5	1	5
6	4	2	5	6	3	4	6	4	6	4	4	6	7	6
1	3	4	3	4	4	7	3	7	4	10	3	1	7	5
6	8	4	4	2	5	5	1	3	5	4	4	9	5	7
6	3	2	7	2	7	1	5	4	5	10	6	7	10	1
2	5	5	2	6	2	1	6	9	8	5	3	8	2	1
10	3	2	10	5	5	1	6	5	2	2	2	2	6	6
6	9	1	8	2	7	2	3	2	2	5	4	6	6	1
10	2	6	3	3	2	10	5	5	10	4	2	10	5	4
1	7	8	2	6										



# Tallies

## Tally of Product Rating

### Tally for Discrete Variables: Product Rating

Product

Rating	Count	Percent
1	18	9.00
2	31	15.50
3	21	10.50
4	24	12.00
5	33	16.50
6	31	15.50
7	15	7.50
8	11	5.50
9	4	2.00
10	12	6.00
N=	200	

# Cross Tabulation

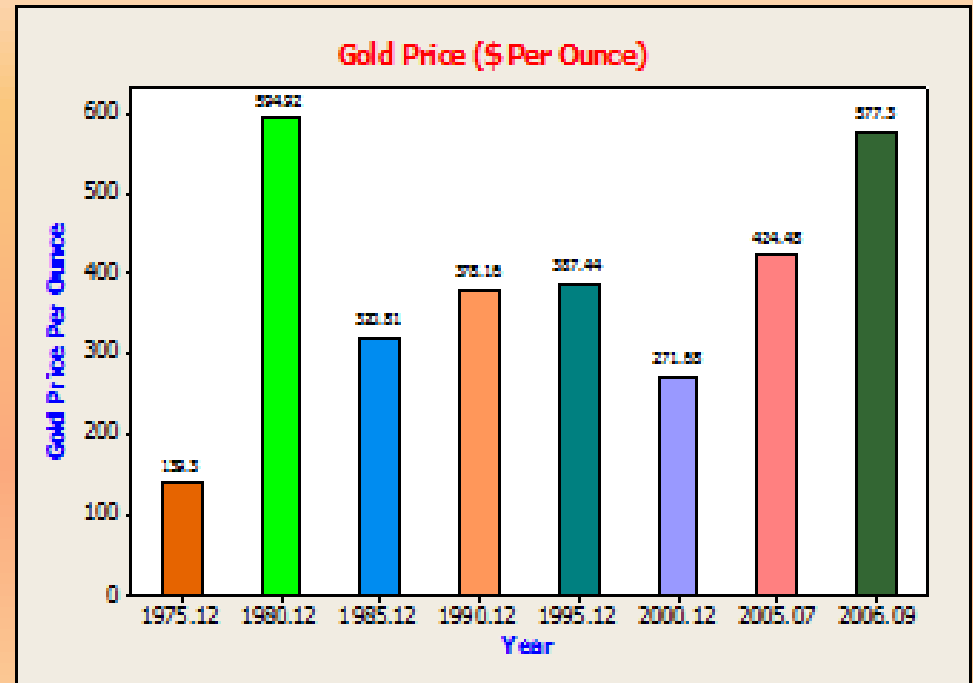
Cross tabulation is used to summarize the data for two variables. Suppose that a manufacturing firm is in the process of implementing a Lean Six Sigma Quality program. The company hired a consulting firm to provide training to 100 executives at different levels. At the conclusion of the training, the management asked the executives at different levels to rate the training program as effective, somewhat effective, useless, or very effective. The response can be shown as a cross tabulation.

**Two-way Table of Product Rating**

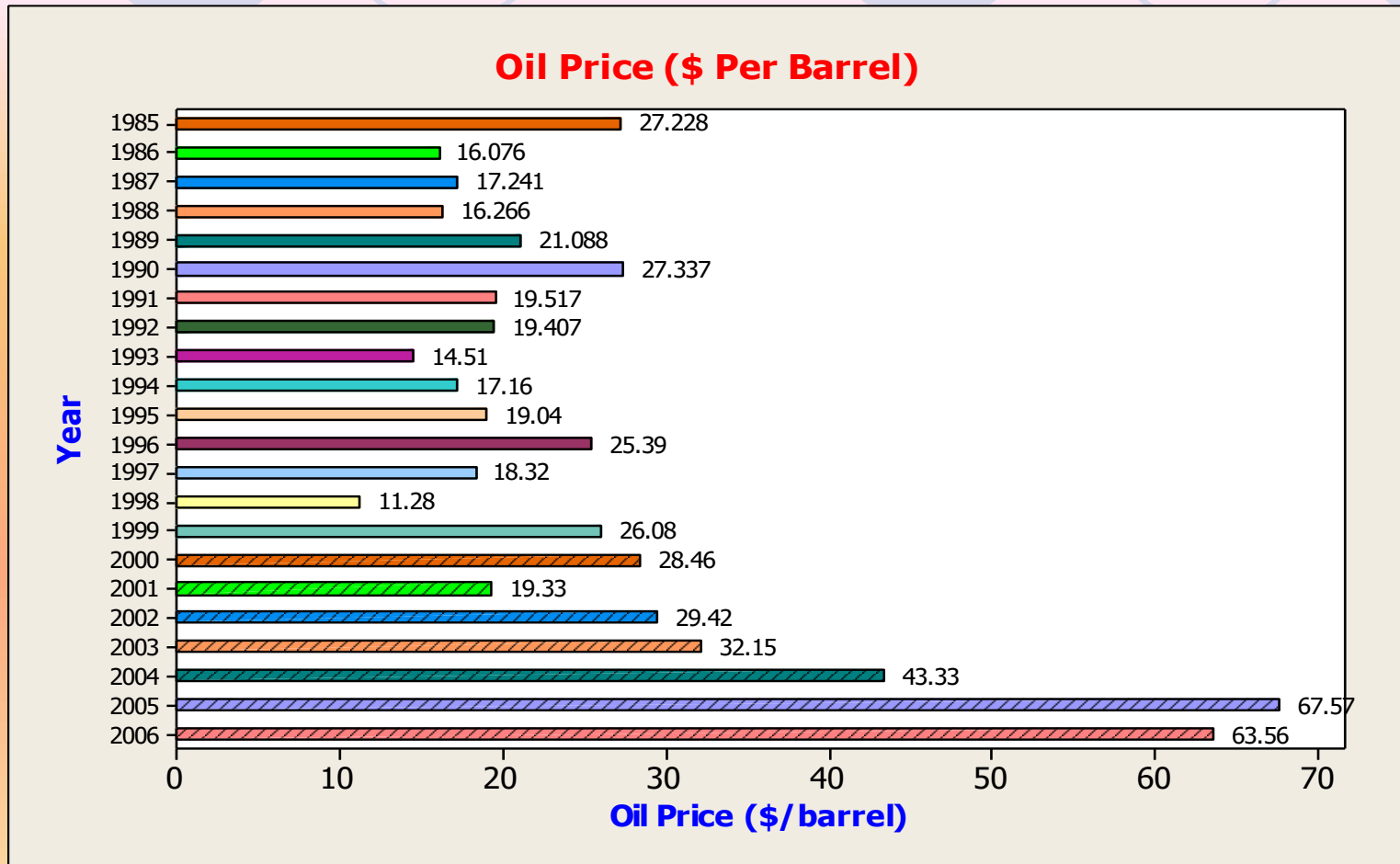
Tabulated Statistics: Executive Level, Rating					
Rows: Executive Level	Columns: Rating				
		Somewhat		Very	
	Effective	Effective	Useless	Effective	All
Manager	11	3	11	13	38
Senior Manager	7	4	8	7	26
Supervisor	7	9	4	16	36
All	25	16	23	36	100
Cell Contents:	Count				

# The Bar Chart: Summarizing Qualitative Or Categorical Data

Year	Gold Price/Ounce(\$)
1975.12	139.30
1980.12	594.92
1985.12	320.81
1990.12	378.16
1995.12	387.44
2000.12	271.68
2005.07	424.48
2006.09	577.30



# Bar Chart...cont.



# The Pie Chart

---

- A pie chart is used to show the relative magnitudes of parts to a whole.
- In this chart relative frequencies of each group of data are plotted.
- A circle is constructed and is divided into distinct sections. Each section represents one category of data.
- The area of each section is determined by multiplying the relative frequency of each section by the angle of a circle. Since there are 360 degrees in a circle, each section is multiplied by 360 degrees to obtain the correct number of degrees to represent each section.

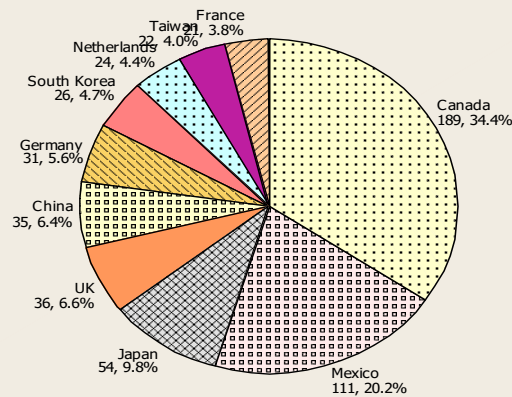


# The Pie Chart...cont.

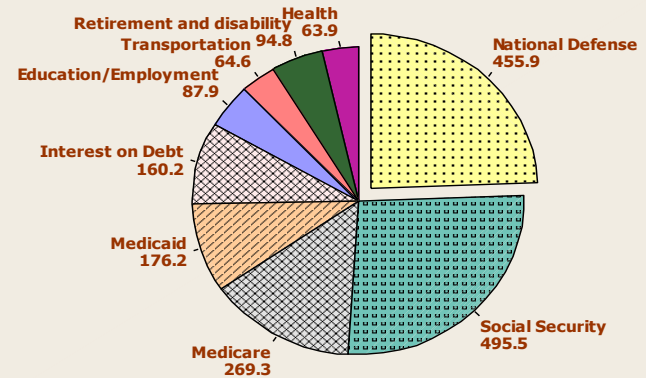
Country	Export (\$ Billions)	Percentage of Export (%)
1 Canada	189	34.4262
2 Mexico	111	20.2186
3 Japan	54	9.8361
4 UK	36	6.5574
5 China	35	6.3752
6 Germany	31	5.6466
7 South Korea	26	4.7359
8 Netherlands	24	4.3716
9 Taiwan	22	4.0073
10 France	21	3.8251

Row	Federal Expenditure 2004	Billions of \$
1	National Defense	455.9
2	Social Security	495.5
3	Medicare	269.3
4	Medicaid	176.2
5	Interest on Debt	160.2
6	Education/Employment	87.9
7	Transportation	64.6
8	Retirement and disability	94.8
9	Health	63.9

**Pie Chart of U.S. Export for 2004**

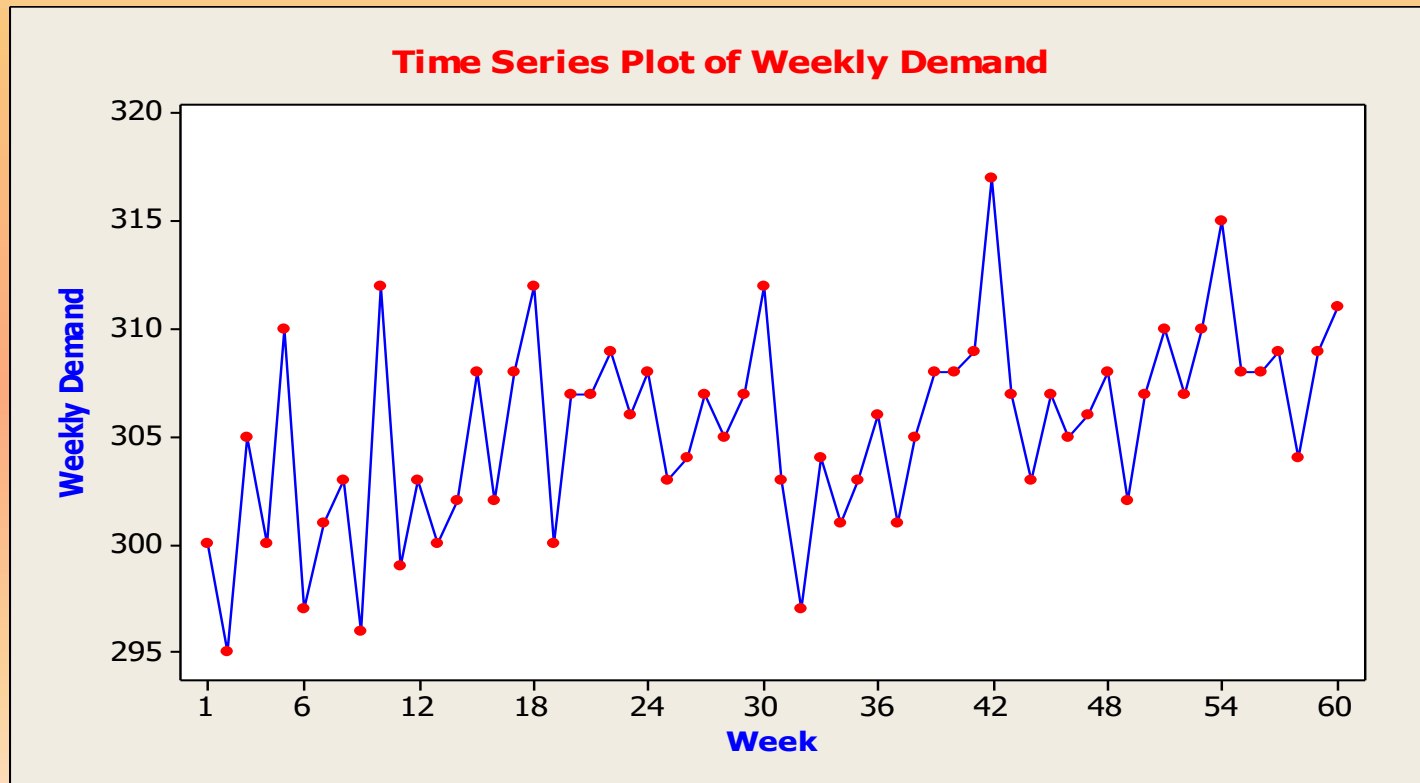


**Pie Chart Federal Expenditure (2004)**

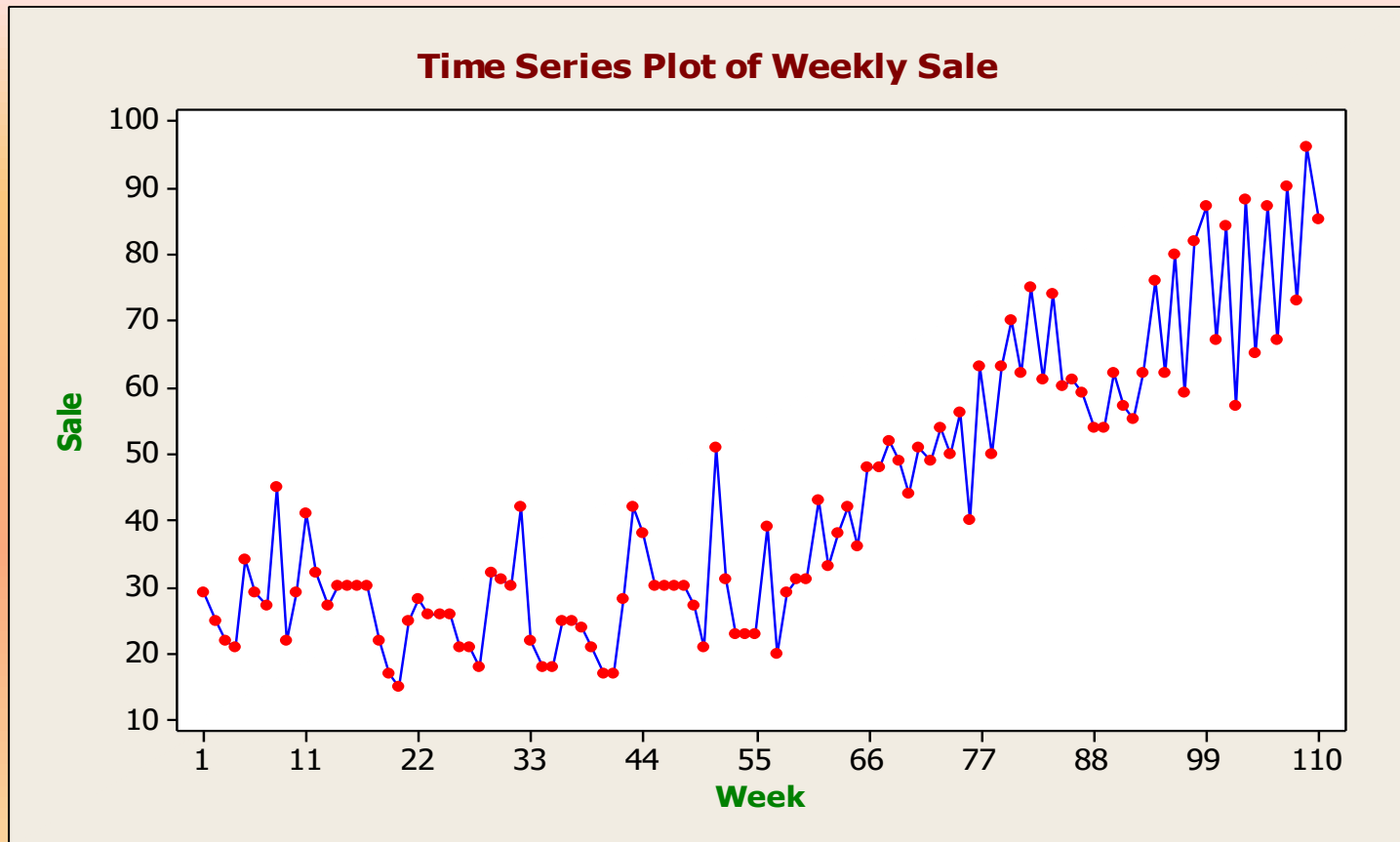


# Time Series Plots

A time series plots the data over time. The graph plots the pairs of points and connects these points using a straight line. In a time series plot, the x values are time. The plot is helpful in visualizing a trend or a pattern in a data set over time.

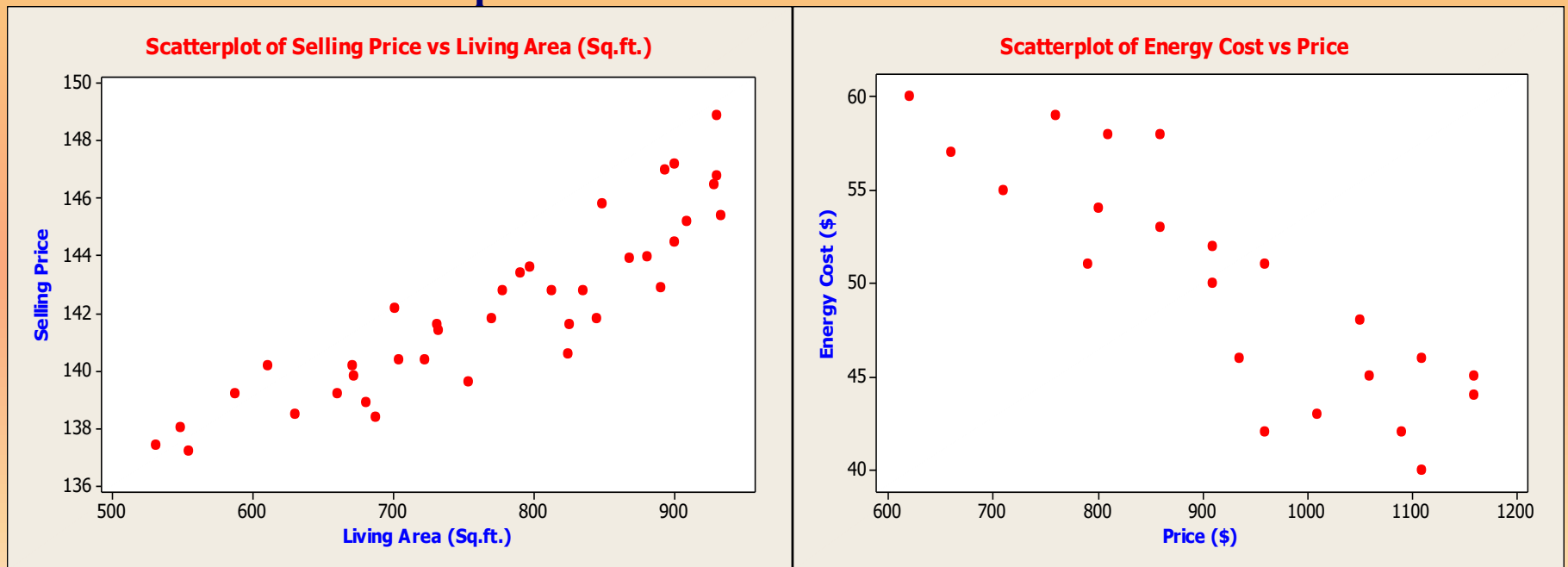


# A Time Series Plot Showing Trend



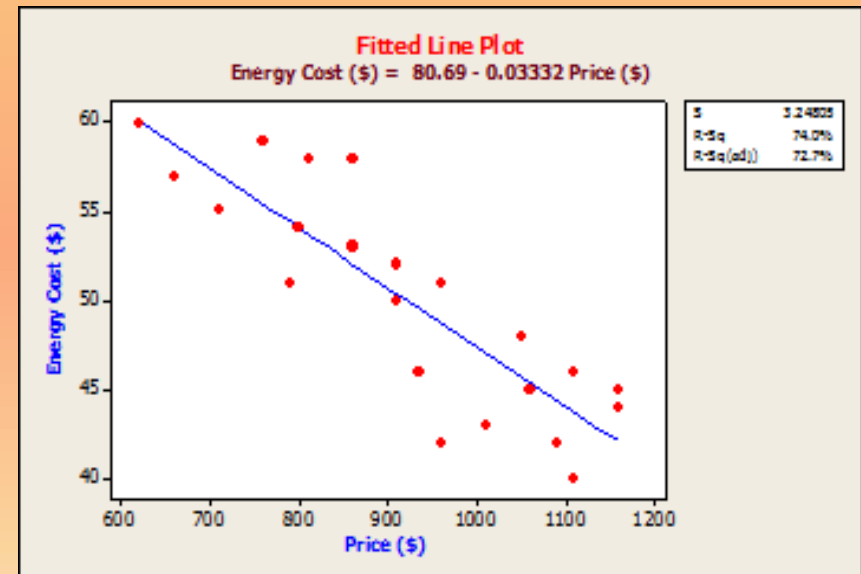
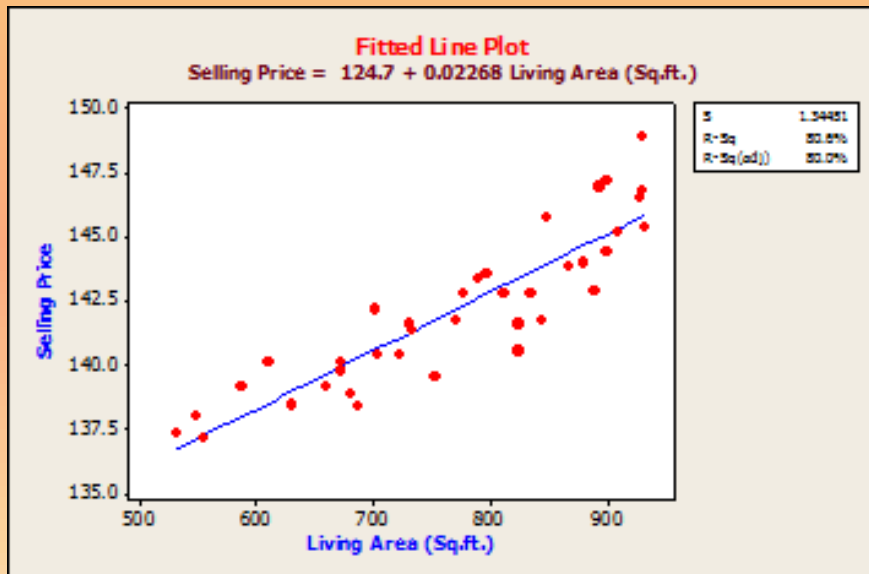
# Scatter Plots

Scatter plots are helpful in investigating the relationship between two variables. One of these variables is considered a dependent variable and the other an independent variable. The data value is thought of as having a  $x$  value and a  $y$  value. Thus, we have  $(x_i, y_i)$ ,  $i=1,2,3,\dots,n$  pairs. The relationships between the variables can be seen from scatter plots.



# Fitted Line Plots

The scatter plot shows a linear or non-linear relationship or no relationship between the two variables. If the relationship between the two variables is linear a best fitting line over the scatter plot can be drawn using a computer package. Some packages also provide the equation of the best fitting line that can be used to predict one variable using the other.



# Widely used Charts and Graphs: Examples

