Chapter 4 Summary
Displaying Quantitative Data

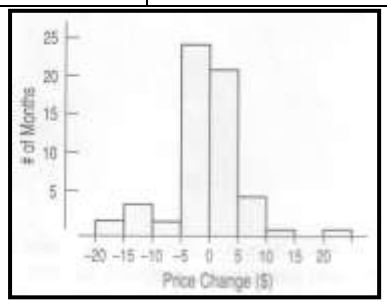
What did you learn?
Make a picture for quantitative data to help see the story the data has to tell.

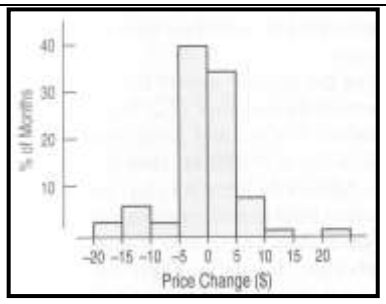
- Distribution of quantitative data can be shown using a histogram, a stem-and-leaf plot, or a dotplot.
- Examine the shape, center, spread, and unusual features of the data.
- We can compare two different groups using displays. If we use the same scale, with can compare them using shape, center, spread, or unusual features of the groups.
- Trends can be viewed on a timeplot of the data that is collected over time.

Enron story – Stock price change:

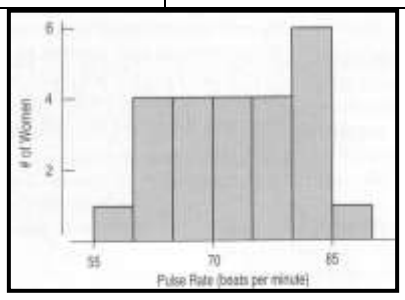
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
1997	-$1.44	-0.75	-0.69	-0.88	0.12	0.75	0.81	-1.75	0.69	-0.22	-0.16	0.34
1998	0.78	0.62	2.44	-0.28	2.22	-0.50	2.06	-0.88	-4.50	4.12	1.16	-0.50
1999	3.28	3.34	-1.22	0.47	5.62	-1.59	4.31	1.47	-0.72	-0.38	-3.25	0.03
2000	5.72	21.06	4.50	4.56	-1.25	-1.19	-3.12	8.00	9.31	1.12	-3.19	-17.75
2001	14.38	-1.08	-10.11	-12.11	5.84	-9.37	-4.74	-2.69	-10.61	-5.85	-17.16	-11.59

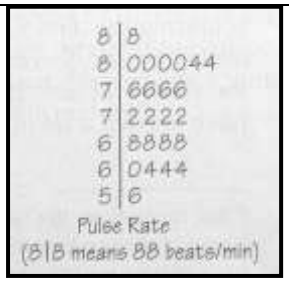
Distribution	The distribution of a variable gives the possible values of the variable and the relative frequency of each value.
Histogram	A histogram uses adjacent bars to show the distribution of values in a quantitative variable. Each bar represents the frequency of values falling in an interval of values.
Relative frequency histogram	A histogram uses adjacent bars to show the distribution of values in a quantitative variable. Each bar represents the relative frequency of values falling in an interval of values.





Stem-and-leaf display	A stem-and-leaf display shows quantitative data values in a way that sketches the distribution of the data.



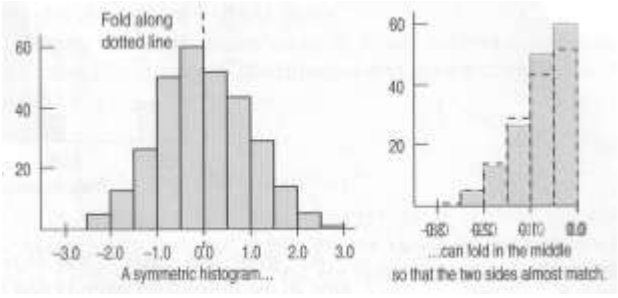


Dotplot	A dotplot graphs a dot for each case against a single axis.
Quantitative data condition	The data are values of a quantitative variable whose units are known.

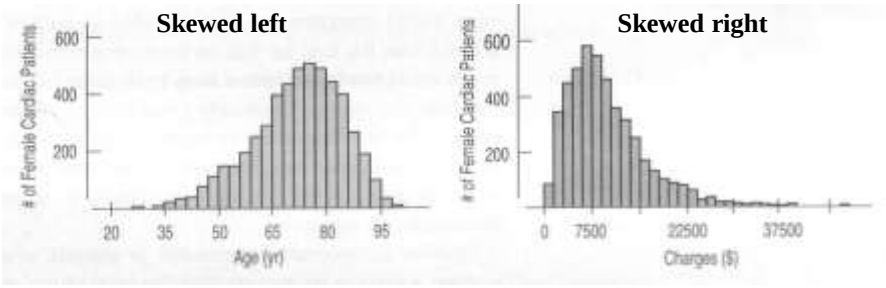
Shape	To describe the shapes of a distribution, look for single versus multiple modes and symmetry versus skewness.
Center	A value that attempts the impossible by summarizing the entire distribution with a single number, a "typical" value.
Spread	A numerical summary of how tightly the values are clustered around the "center."

Modes	A hump or local high point in the shape of the distribution of a variable is called a "mode." The apparent location of modes can change as the scale of a histogram is changed.
Unimodal	Having one mode. This a useful term for describing the shape of a histogram when it's generally mound-shaped.
Bimodal	Distributions with two modes.
Multimodal	Distributions with more than two modes.

Uniform	A distribution that's roughly flat.
Symmetric	A distribution is symmetric if the two halves on either side of the center look approximately like mirror images of each other.



Tails	The tails of a distribution are the parts that typically trail off on either side. Distributions can be characterized as having long tails (if they straggle off for some distance) or short tails (if they don't).
Skewed	A distribution is skewed if it's not symmetric and one tail stretches out farther than the other. Distributions are said to be **skewed left** when the longer tail stretches to the left, and **skewed right** when it goes to the right.

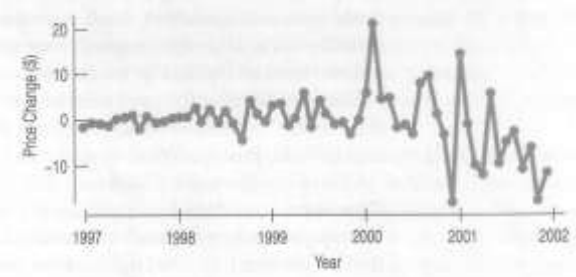


Chapter 4 Summary Continued

Outliers	Outliers are extreme values that don't appear to belong with the rest of the data. They may be unusual values that deserve further investigation, or just mistakes; there's no obvious way to tell. Don't delete outliers automatically - you have to think about them. Outliers can affect many statistical analyses, so you should always be alert for them.

Gaps	Regions of a histogram that have no values for a given data set.

Timeplot	A timeplot displays data that change over time. Often, successive values are connected with lines to show trends more clearly.



Monthly Enron stock price changes stretched out over time in a time-plot. Now it's easy to see that after being relatively stable, the stock price became somewhat volatile in 1998 and then even more so starting in 2000. **Figure 4.11**

Re-express / Transform	Applying a simple function to a set of data to make a skewed distribution more symmetric.

What can go wrong?

- Don't make a histogram of a categorical variable.
- Don't look for shape, center, and spread of a bar chart.
- Don't use bars in every display – save them for histograms and bar charts.
- Choose a bin width appropriate to the data.
- Avoid inconsistent scales.
- Label clearly.

A plot gone wrong:

- Horizontal scales are inconsistent – one starts in 1965 and the other in 1989.
- Vertical axis isn't labeled, not consistent.
- Vertical scales don't point in the same direction and ranking going lower (from 15th to 6th) should be viewed as an improvement.