STATISTICS

INFORMED DECISIONS USING DATA

Fifth Edition

Chapter 2

Organizing and Summarizing Data

Pearson

Wed 2/26

2.2 Organizing Quantitative Data: The Popular Displays

Learning Objectives

1. Organize discrete data in tables
2. Construct histograms of discrete data
3. Organize continuous data in tables
4. Construct histograms of continuous data
5. Draw stem-and-leaf plots
6. Draw dot plots
7. Identify the shape of a distribution

2.2 Organizing Quantitative Data: The Popular Displays

2.2.1 Organize Discrete Data in Tables (1 of 2)

EXAMPLE Constructing Frequency and Relative Frequency Distribution from Discrete Data

The following data represent the number of available cars in a household based on a random sample of 50 households.

Construct a frequency and relative frequency distribution.

3	0	1	2	1	1	1	2	0	2
4	2	2	2	1	2	2	0	2	4
1	1	3	2	4	1	2	1	2	2
3	3	2	1	2	2	0	3	2	2
2	3	2	1	2	2	1	1	3	5

Data based on results reported by the United States Bureau of the Census.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.1 Organize Discrete Data in Tables (2 of 2)

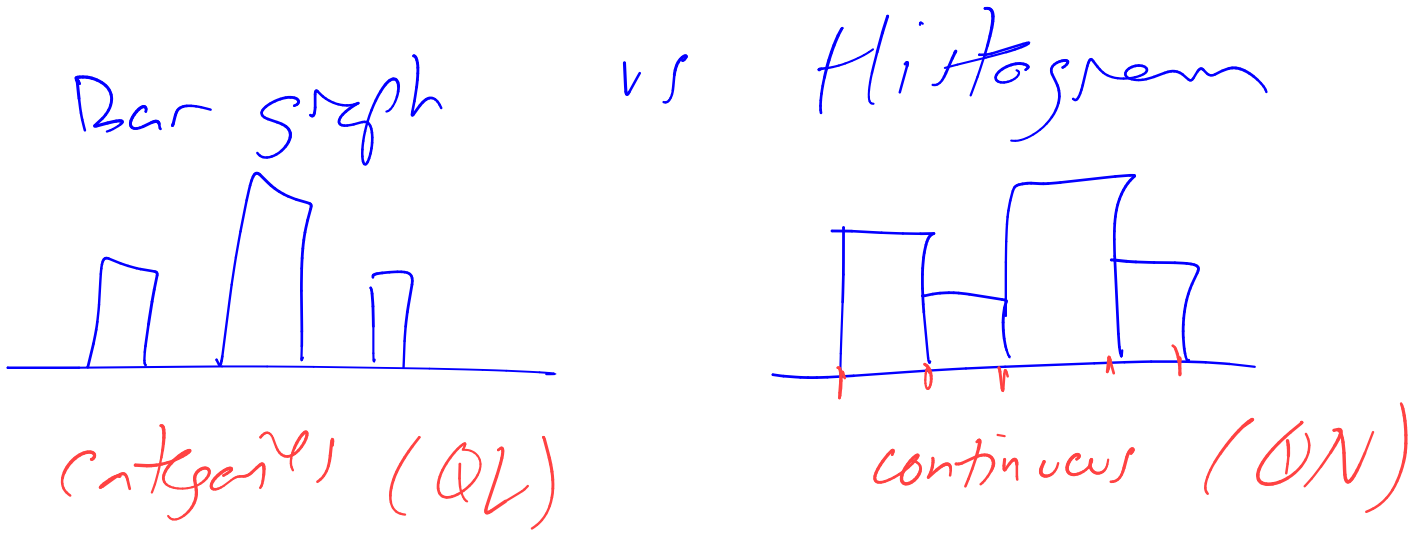
# of Cars	Tally	Frequency	Relative Frequency
0	||||	4	$\frac{4}{50} = 0.08$
1	|||| |||| |||	13	$\frac{13}{50} = 0.26$
2	|||| |||| |||| |||| ||	22	0.44
3	|||| ||	7	0.14
4	|||	3	0.06
5	|	1	0.02

total = 50

2.2 Organizing Quantitative Data: The Popular Displays

2.2.2 Construct Histograms of Discrete Data (1 of 4)

A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same and the rectangles touch each other.



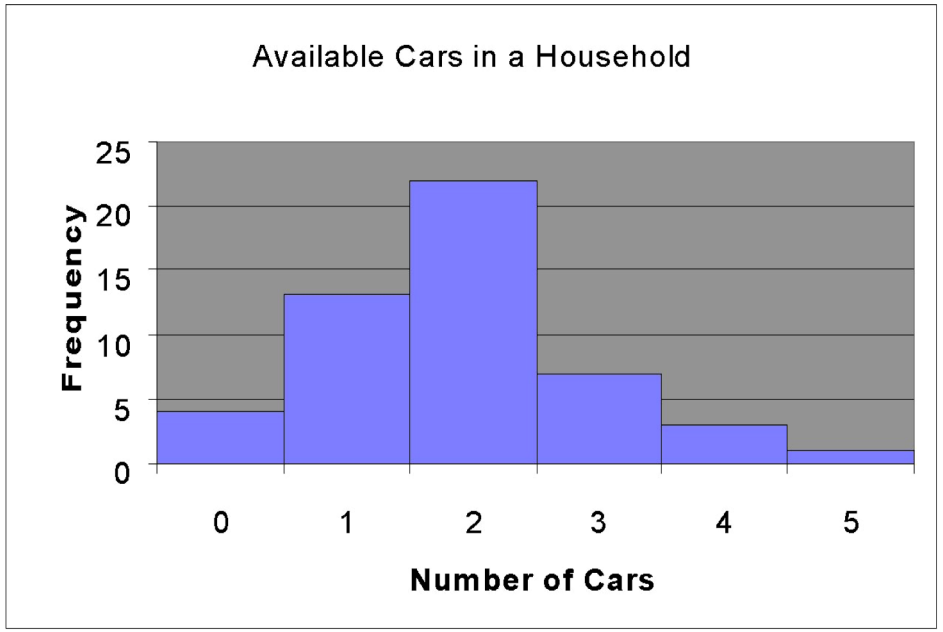
2.2 Organizing Quantitative Data: The Popular Displays

2.2.2 Construct Histograms of Discrete Data (2 of 4)

EXAMPLE Drawing a Histogram for Discrete Data

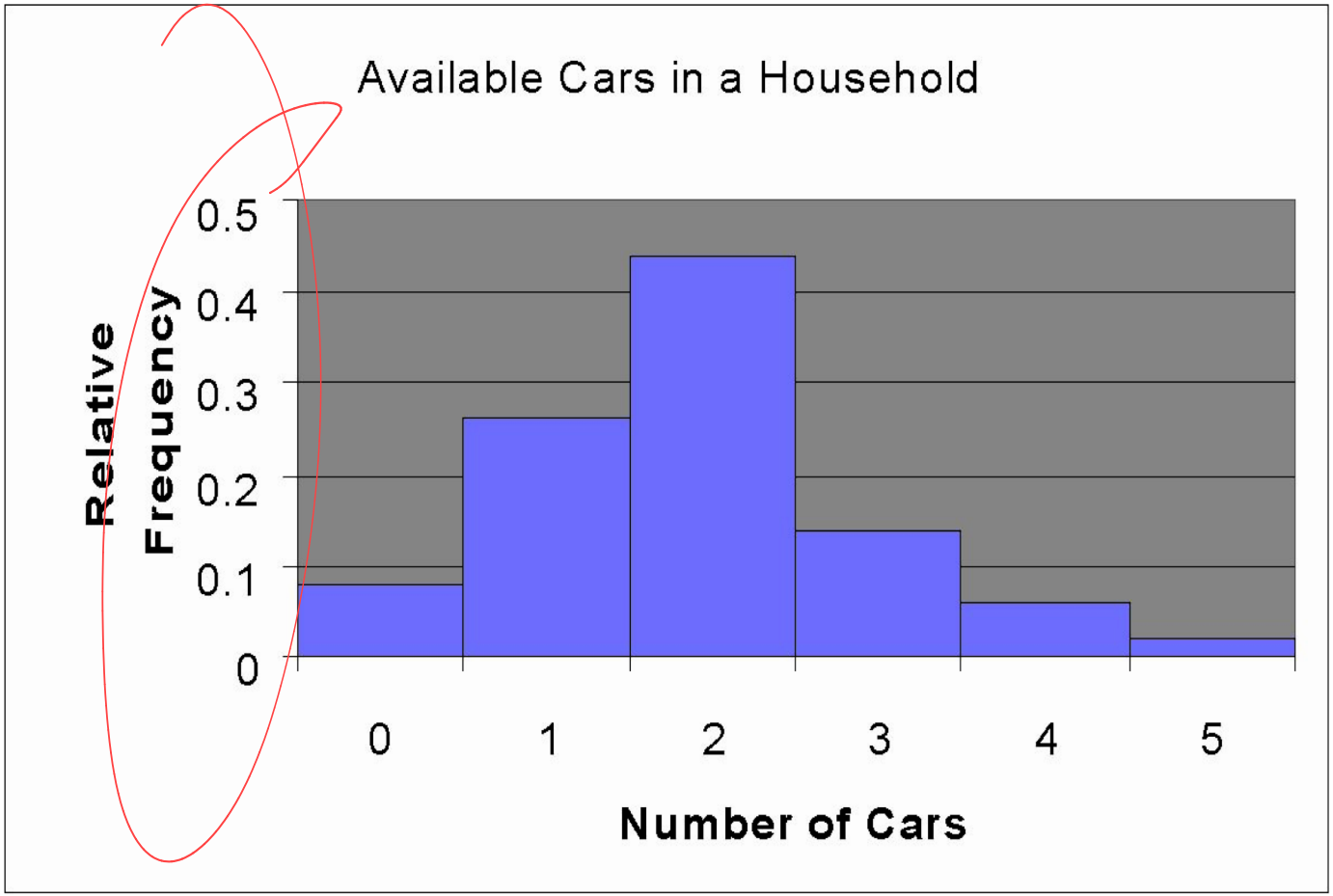
Draw a frequency and relative frequency histogram for the "number of cars per household" data.

# of Cars	Frequency	Relative Frequency
0	4	$\frac{4}{50} = 0.08$
1	13	$\frac{13}{50} = 0.26$
2	22	0.44
3	7	0.14
4	3	0.06
5	1	0.02



2.2 Organizing Quantitative Data: The Popular Displays

2.2.2 Construct Histograms of Discrete Data (4 of 4)



2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (1 of 10)

Classes are categories into which data are grouped. When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, we must create classes by using intervals of numbers.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (2 of 10)

The following data represents the number of persons aged 25 - 64 who are currently work-disabled.

Age	Number (in thousands)
25-34	2,132
35-44	3,928
45-54	4,532
55-64	5,108

The **lower class limit** of a class is the smallest value within the class while the **upper class limit** of a class is the largest value within the class. The lower class limit of first class is 25. The lower class limit of the second class is 35. The upper class limit of the first class is 34. The **class width** is the difference between consecutive lower class limits. The class width of the data given above is 35 − 25 = 10.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (3 of 10)

The following data represent the time between eruptions (in seconds) for a random sample of 45 eruptions at the Old Faithful Geyser in Wyoming. Construct a frequency and relative frequency distribution of the data.

728	678	723	735	703
730	722	708	714	713
726	716	736	719	672
698	702	738	725	711
721	703	735	699	695
722	718	695	702	731
700	703	706	733	726
720	723	711	696	695
729	699	714	700	718

670-679
//

Source: Ladonna Hansen, Park Curator

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (4 of 10)

The smallest data value is 672 and the largest data value is 738. We will create the classes so that the lower class limit of the first class is 670 and the class width is 10 and obtain the following classes:

LCL

670 – 679

680 – 689

690 – 699

700 – 709

710 – 719

720 – 729

730 – 739

CW = 680 − 670 = 10

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (5 of 10)

Time between Eruptions (seconds)	Tally	Frequency	Relative Frequency
670-679	||	2	$\frac{2}{45} = 0.044$
680 - 689		0	0
690 - 699	||||| ||	7	0.1556
700 - 709	||||| ||||	9	0.2
710 - 719	||||| ||||	9	0.2
720 - 729	||||| ||||| |	11	0.2444
730 - 739	||||| ||	7	0.1556

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (6 of 10)

The choices of the lower class limit of the first class and the class width were rather arbitrary.

There is not one correct frequency distribution for a particular set of data.

However, some frequency distributions can better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form.

Use the distribution that seems to provide the best overall summary of the data.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (7 of 10)

A different table with class width of 5 seconds:

Time between Eruptions (seconds)	Tally	Frequency	Relative Frequency
670 - 674	|	1	$\frac{1}{45} = 0.0222$
675 - 679	|	1	0.0222
680 - 684		0	0
685 - 689		0	0
690 - 694		0	0
695 - 699	||||| ||	7	0.1556
700 - 704	||||| ||	7	0.1556
705 - 709	||	2	0.0444
710 - 714	|||||	5	0.1111
715 - 719	||||	4	0.0889
720 - 724	||||| |	6	0.1333
725 - 729	|||||	5	0.1114
730 - 734	|||	3	0.0667
735 - 739	||||	4	0.0889

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (8 of 10)

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Choosing the Lower Class Limit of the First Class

- Choose the smallest observation in the data set or a convenient number slightly lower than the smallest observation in the data set.

dont need.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (9 of 10)

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Determining the Class Width

- Decide on the **number of classes**. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer classes you should have.
- Determine the **class width** by computing

$$\text{Class width} = \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$$

optional

2.2 Organizing Quantitative Data: The Popular Displays

2.2.3 Organize Continuous Data in Tables (10 of 10)

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

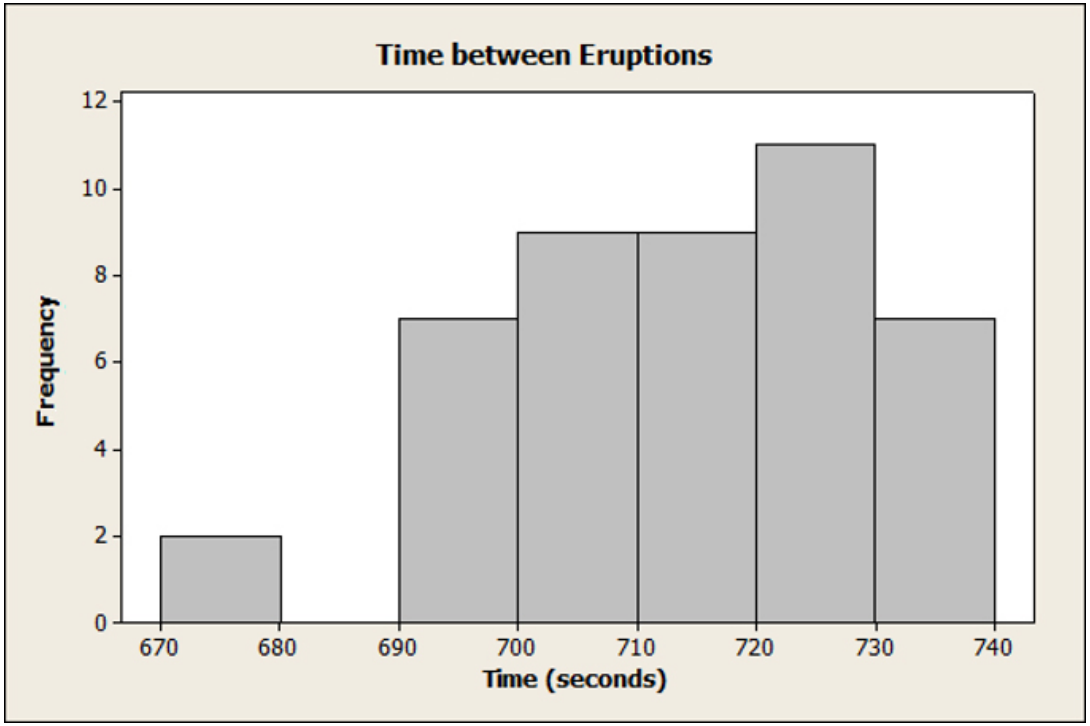
Round this value **up** to a convenient number.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.4 Construct Histograms of Continuous Data (1 of 3)

EXAMPLE Constructing a Frequency Histogram for Continuous Data

Using class width of 10:

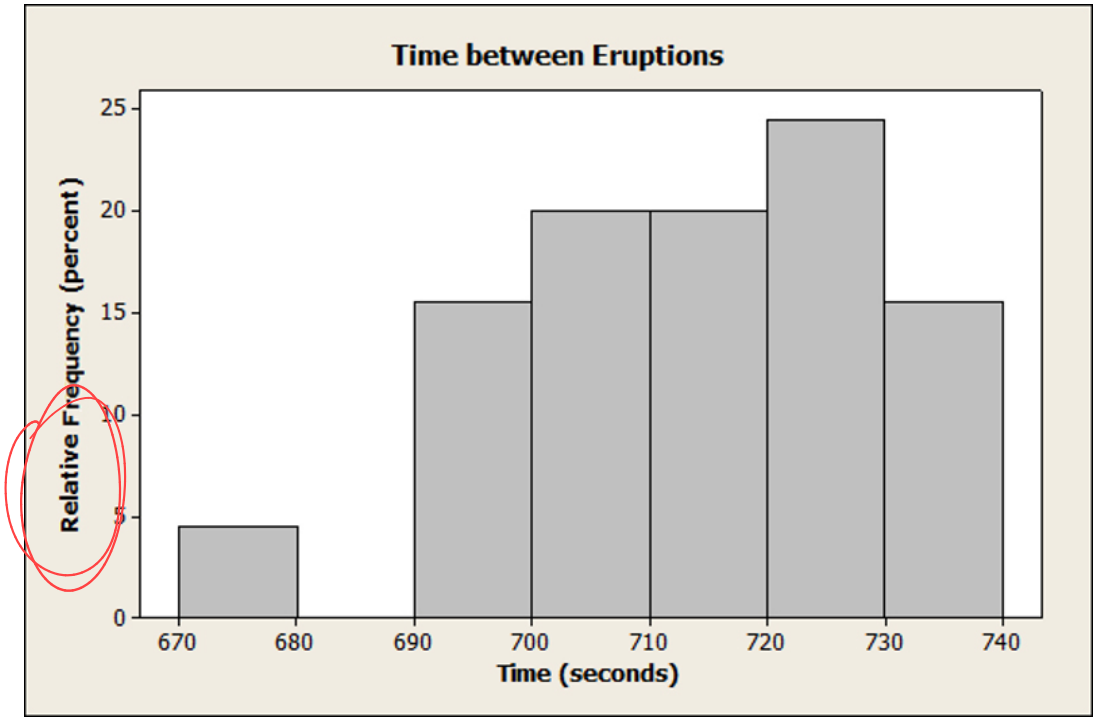


2.2 Organizing Quantitative Data: The Popular Displays

2.2.4 Construct Histograms of Continuous Data (2 of 3)

EXAMPLE Constructing a Relative Frequency Histogram for Continuous Data

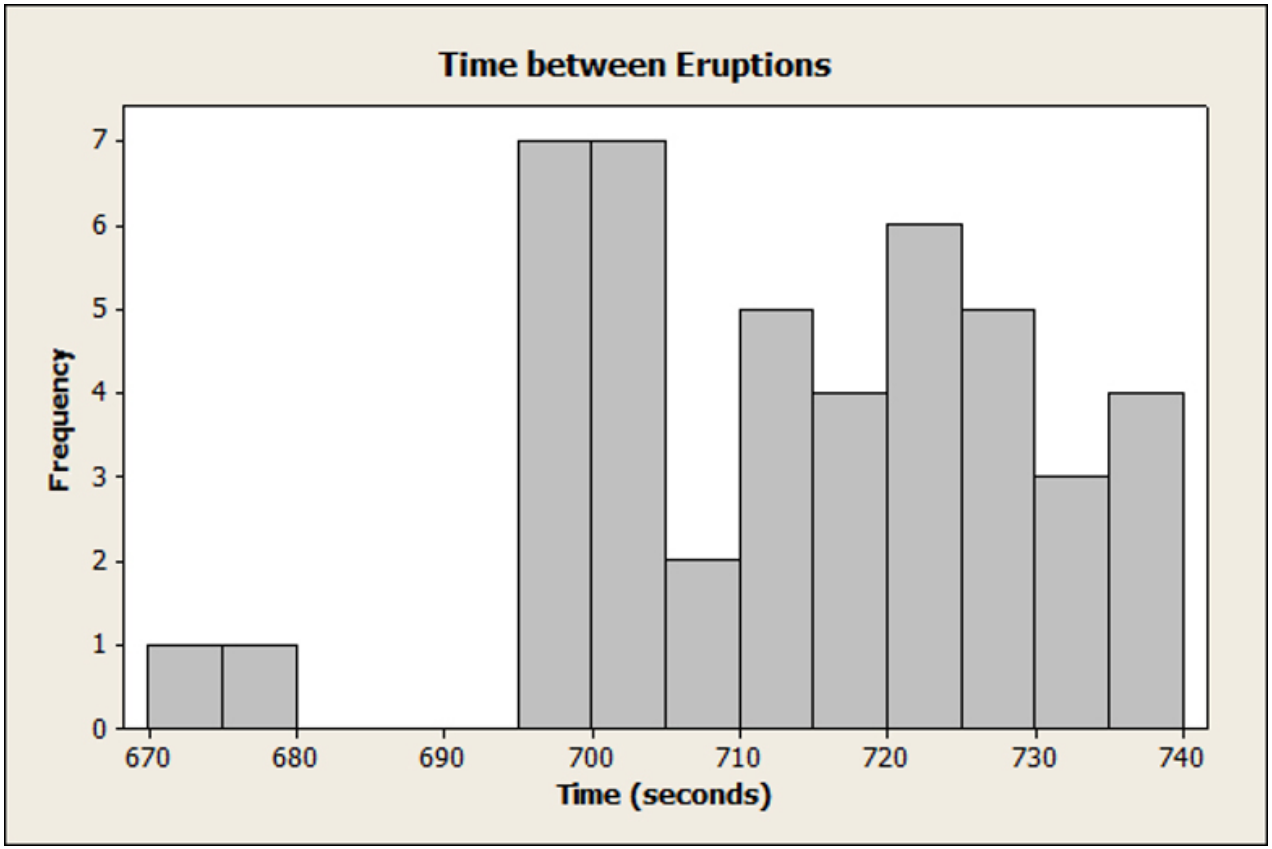
Using class width of 10:



2.2 Organizing Quantitative Data: The Popular Displays

2.2.4 Construct Histograms of Continuous Data (3 of 3)

Using class width of 5:

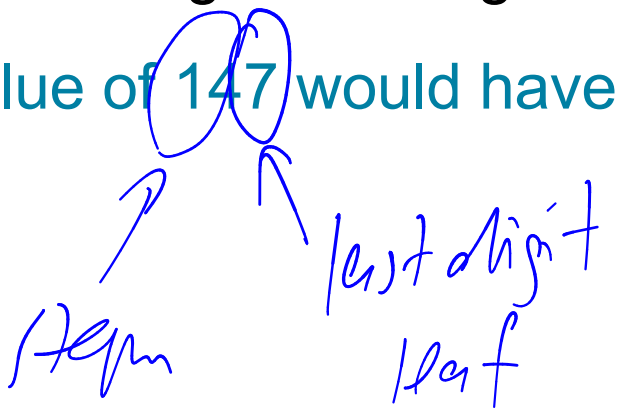


2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (1 of 10)

A **stem-and-leaf plot** uses digits to the left of the rightmost digit to form the **stem**. Each rightmost digit forms a **leaf**.

For example, a data value of 147 would have 14 as the stem and 7 as the leaf.



2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (2 of 10)

EXAMPLE Constructing a Stem-and-Leaf Plot

An individual is considered to be unemployed if they do not have a job, but are actively seeking employment. The following data represent the unemployment rate in each of the fifty United States plus the District of Columbia in June, 2008.

State	Unemployment Rate
Alabama	4.7
Alaska	6.8
Arizona	4.8
Arkansas	5.0
California	6.9
Colorado	5.1

We let the stem represent the integer portion of the number and the leaf will be the decimal portion. For example, the stem of Alabama (4.7) will be 4 and the leaf will be 7.

stem	leaves
4	7 8 2
5	0 1 4
6	8 9

etc...

2.2 Organizing Quantitative Data: The Popular Displays

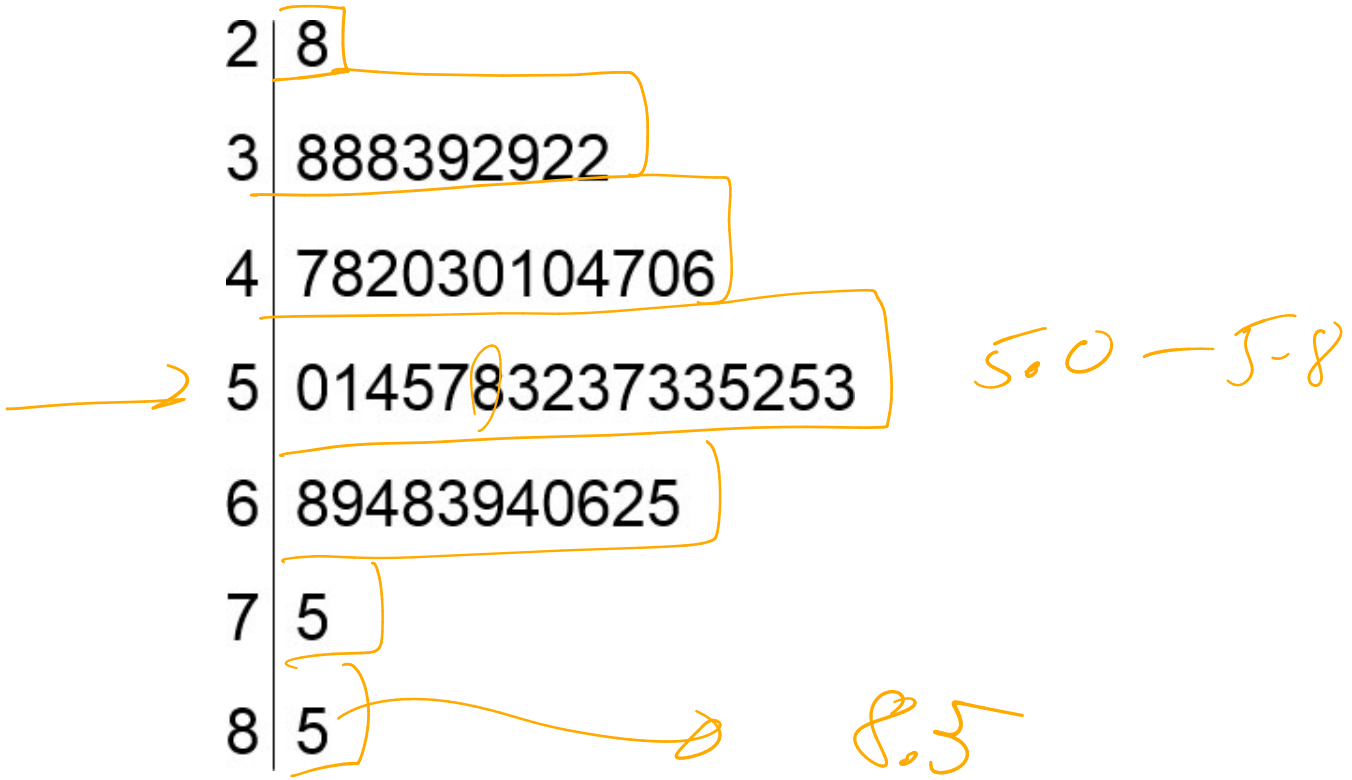
2.2.5 Draw Stem-and-Leaf Plots (3 of 10)

State13	Unemployment Rate	State	Unemployment Rate	State	Unemployment Rate
Alabama	4.7	Kentucky	6.3	North Dakota	3.2
Alaska	6.8	Louisiana	3.8	Ohio	6.6
Arizona	4.8	Maine	5.3	Oklahoma	3.9
Arkansas	5.0	Maryland	4.0	Oregon	5.5
California	6.9	Mass	5.2	Penn	5.2
Colorado	5.1	Michigan	8.5	Rhode Island	7.5
Conn	5.4	Minnesota	5.3	South Carolina	6.2
Delaware	4.2	Mississippi	6.9	South Dakota	2.8
Dist Col	6.4	Missouri	5.7	Tenn	6.5
Florida	5.5	Montana	4.1	Texas	4.4
Georgia	5.7	Nebraska	3.3	Utah	3.2
Hawaii	3.8	Nevada	6.4	Vermont	4.7
Idaho	3.8	New Hamp	4.0	Virginia	4.0
Illinois	6.8	New Jersey	5.3	Washington	5.5
Indiana	5.8	New Mexico	3.9	W. Virginia	5.3
Iowa	4.0	New York	5.3	Wisconsin	4.6
Kansas	4.3	North Carolina	6.0	Wyoming	3.2

2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (5 of 10)

Unordered leaves:



2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (6 of 10)

Ordered leaves:

2	8
3	222388899
4	000012346778
5	0122333334555778
6	02344568899
7	5
8	5

2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (7 of 10)

Construction of a Stem-and-leaf Plot

Step 1 The stem of a data value will consist of the digits to the left of the right-most digit. The leaf of a data value will be the rightmost digit.

Step 2 Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.

Step 3 Write each leaf corresponding to the stems to the right of the vertical line.

Step 4 Within each stem, rearrange the leaves in ascending order, title the plot, and include a legend to indicate what the values represent.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.5 Draw Stem-and-Leaf Plots (10 of 10)

Advantage of Stem-and-Leaf Diagrams over Histograms

Once a frequency distribution or histogram of continuous data is created, the raw data is lost (unless reported with the frequency distribution), however, the raw data can be retrieved from the stem-and-leaf plot.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.6 Draw Dot Plots (1 of 3)

A **dot plot** is drawn by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed.

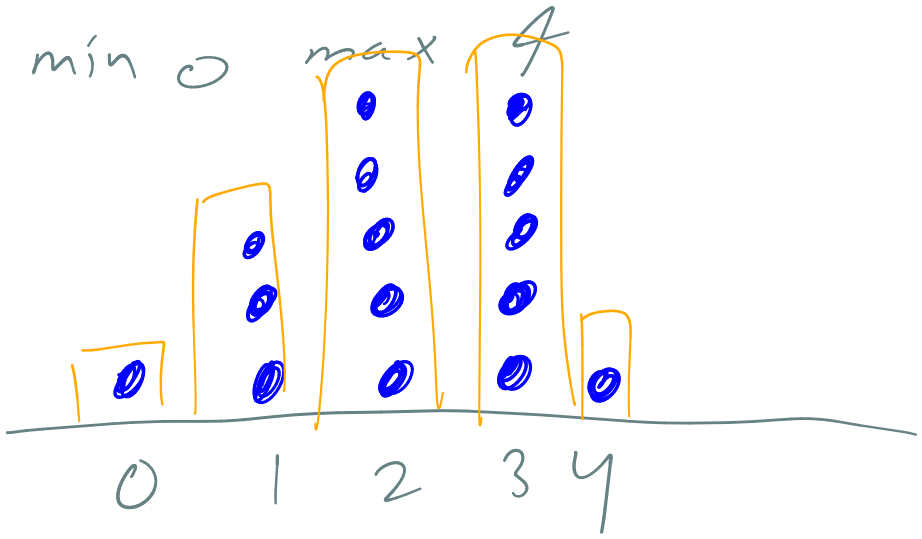
2.2 Organizing Quantitative Data: The Popular Displays

2.2.6 Draw Dot Plots (2 of 3)

EXAMPLE Drawing a Dot Plot

The following data represent the number of available cars in a household based on a random sample of 15 households. Draw a dot plot of the data.

3	0	1
4	2	2
1	1	3
3	3	2
2	3	2



Data based on results reported by the United States Bureau of the Census.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.6 Draw Dot Plots (2 of 3)

EXAMPLE Drawing a Dot Plot

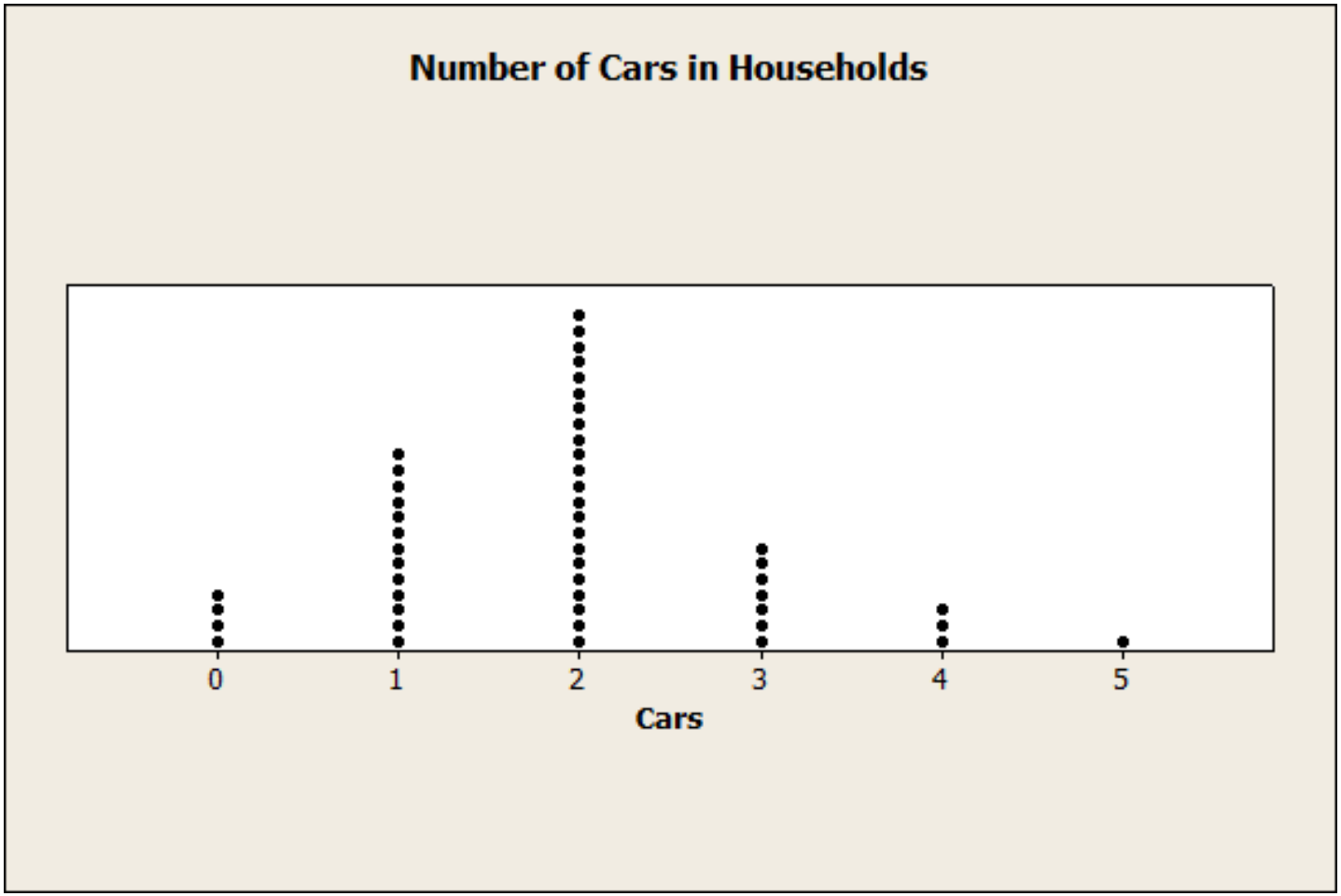
The following data represent the number of available cars in a household based on a random sample of 50 households. Draw a dot plot of the data.

3	0	1	2	1	1	1	2	0	2
4	2	2	2	1	2	2	0	2	4
1	1	3	2	4	1	2	1	2	2
3	3	2	1	2	2	0	3	2	2
2	3	2	1	2	2	1	1	3	5

Data based on results reported by the United States Bureau of the Census.

2.2 Organizing Quantitative Data: The Popular Displays

2.2.6 Draw Dot Plots (3 of 3)



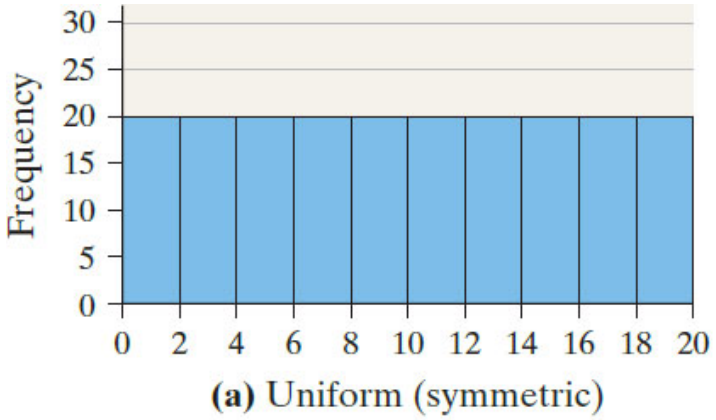
2.2 Organizing Quantitative Data: The Popular Displays

2.2.7 Identify the **Shape of a Distribution** (1 of 4)

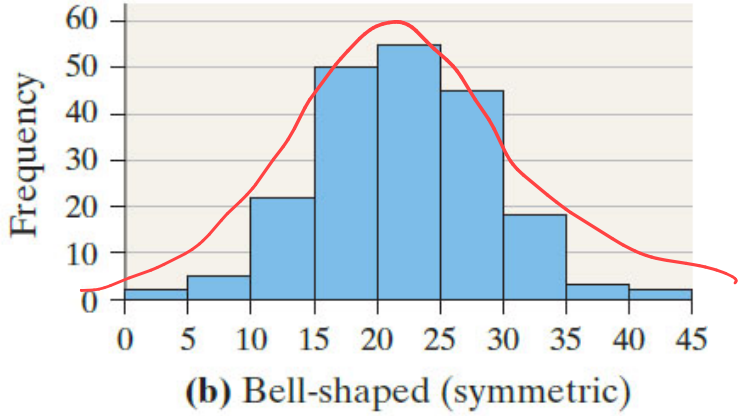
- **Uniform distribution** - the frequency of each value of the variable is evenly spread out across the values of the variable
- **Bell-shaped distribution** - the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle (Other names: **Symmetric** or **Normal**)
- **Skewed right** - the tail to the right of the peak is longer than the tail to the left of the peak
- **Skewed left** - the tail to the left of the peak is longer than the tail to the right of the peak.

2.2 Organizing Quantitative Data: The Popular Displays

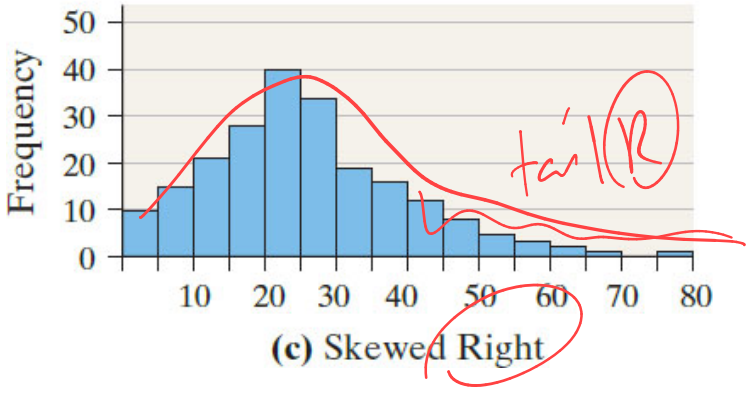
2.2.7 Identify the Shape of a Distribution (2 of 4)



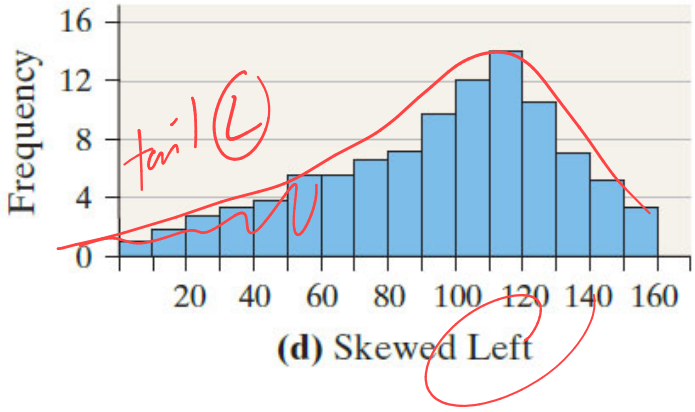
(a) Uniform (symmetric)



(b) Bell-shaped (symmetric)



(c) Skewed Right



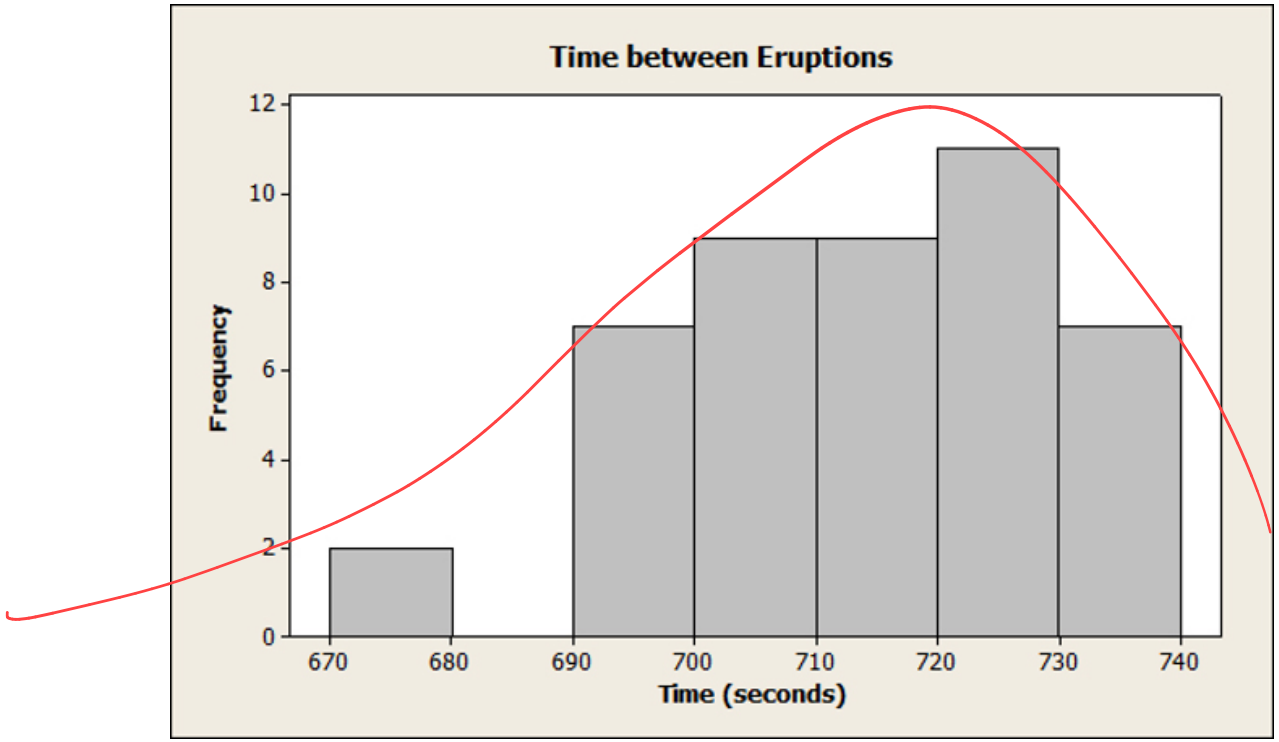
(d) Skewed Left

2.2 Organizing Quantitative Data: The Popular Displays

2.2.7 Identify the Shape of a Distribution (3 of 4)

EXAMPLE Identifying the Shape of the Distribution

Identify the shape of the following histogram which represents the time between eruptions at Old Faithful.



2.3 Additional Displays of Quantitative Data

Learning Objectives

1. Construct frequency polygons
2. Create cumulative frequency and relative frequency tables
3. Construct frequency and relative frequency ogives
4. Draw time-series graphs

2.3 Additional Displays of Quantitative Data

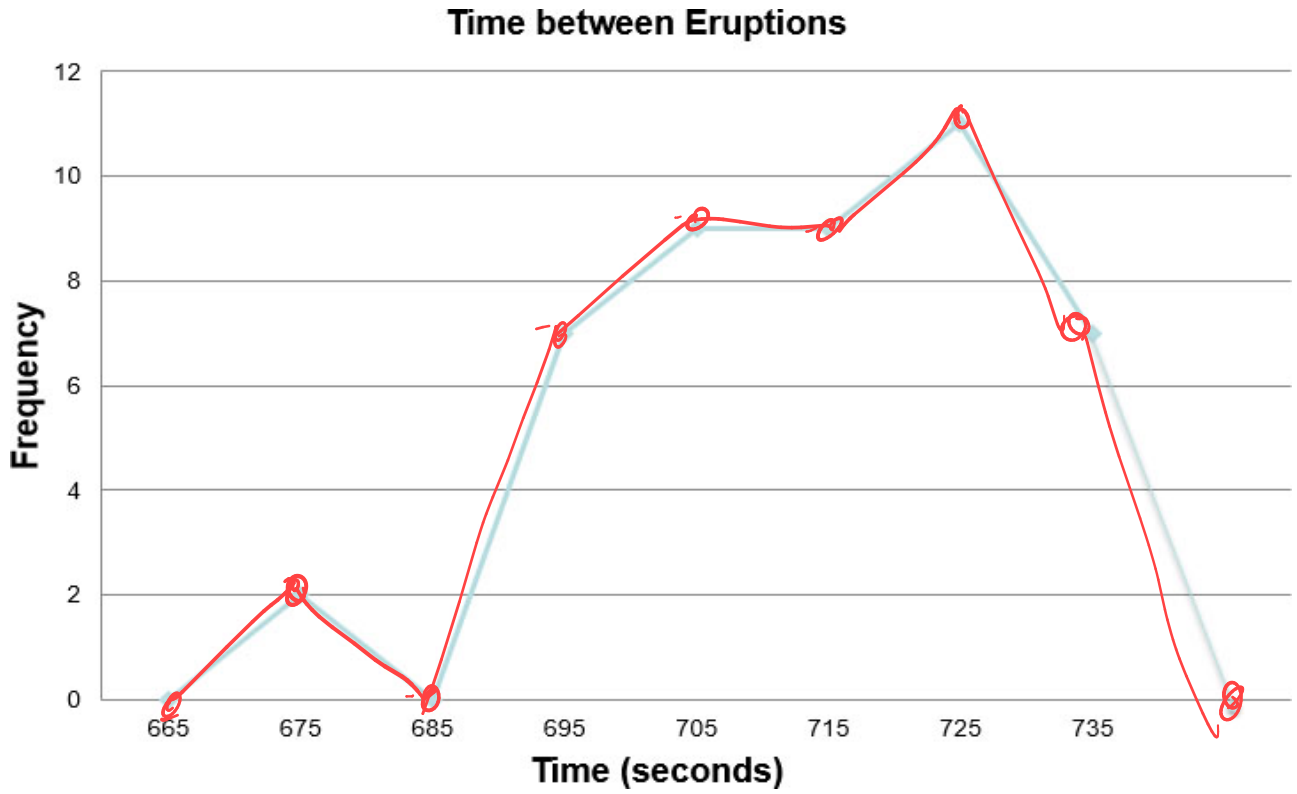
2.3.1 Construct Frequency Polygons (1 of 4)

A **class midpoint** is the sum of consecutive lower class limits divided by 2.

A **frequency polygon** is a graph that uses points, connected by line segments, to represent the frequencies for the classes. It is constructed by plotting a point above each class midpoint on a horizontal axis at a height equal to the frequency of the class. Next, line segments are drawn connecting consecutive points. Two additional line segments are drawn connecting each end of the graph with the horizontal axis.

2.3 Additional Displays of Quantitative Data
2.3.1 Construct Frequency Polygons (3 of 4)

Frequency Polygon



2.3 Additional Displays of Quantitative Data

2.3.2 Create Cumulative Frequency and Relative Frequency Tables (1 of 2)

A **cumulative frequency distribution** displays the aggregate frequency of the category. In other words, for discrete data, **it displays the total number of observations less than or equal to the category**. For continuous data, it displays the total number of observations less than or equal to the upper class limit of a class.

A **cumulative relative frequency distribution** displays the proportion (or percentage) of observations less than or equal to the category for discrete data and the proportion (or percentage) of observations less than or equal to the upper class limit for continuous data.

2.3 Additional Displays of Quantitative Data

2.3.2 Create Cumulative Frequency and Relative Frequency Tables (2 of 2)

Time between Eruptions (seconds)	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
670 – 679	2	0.0444	2	2/45
680 – 689	0	0	2+0 = 2	2/45
690 – 699	7	0.1556	2+0+7 = 9	9/45
700 – 709	9	0.2	18	18/45
710 – 719	9	0.2	27	27/45
720 – 729	11	0.2444	38	38/45
730 – 739	7	0.1556	45	45/45 = 1.00

total

2.3 Additional Displays of Quantitative Data

2.3.2 Construct Frequency and Relative Frequency Ogives (1 of 3)

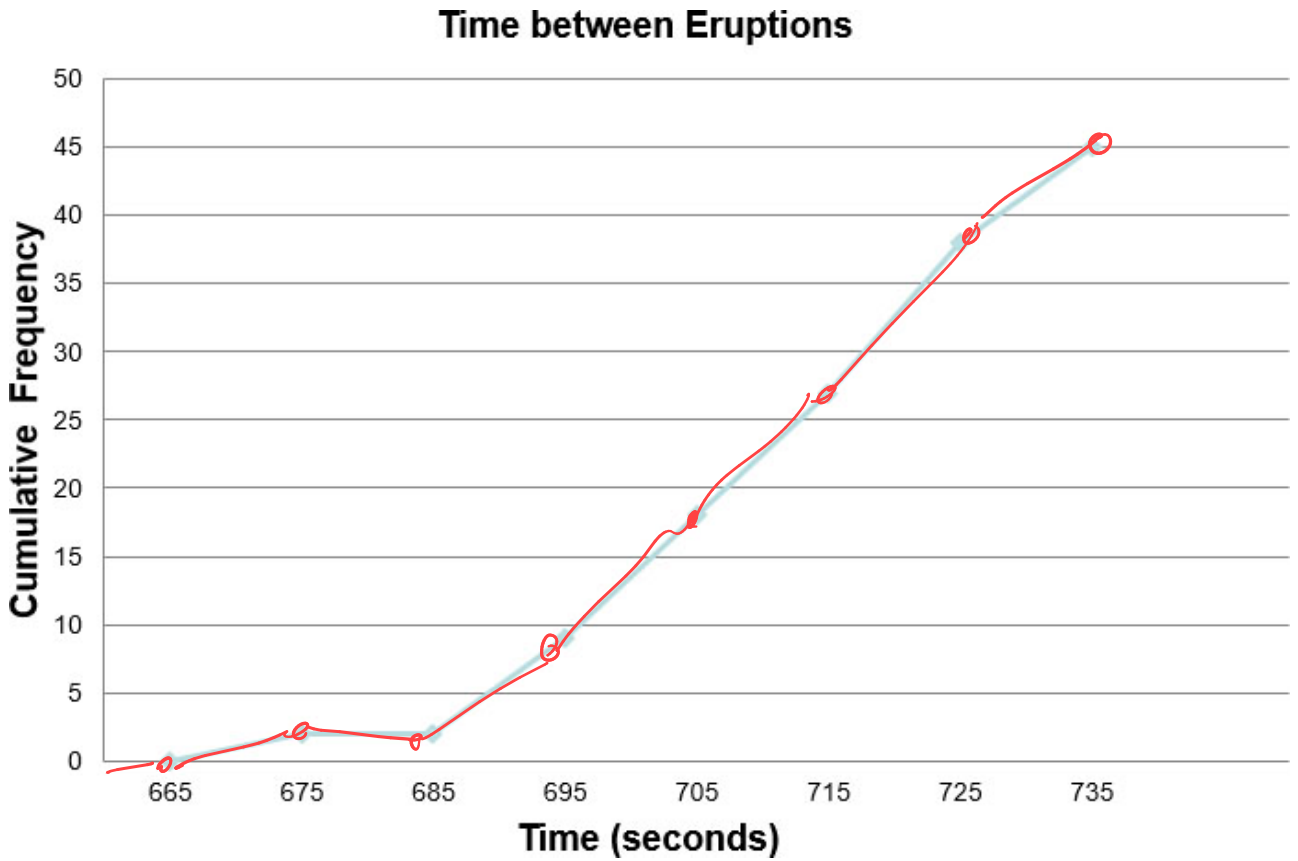
An **ogive** (read as “oh jive”) is a graph that represents the cumulative frequency or cumulative relative frequency for the class. It is constructed by plotting points whose ***x*-coordinates are the upper class limits** and whose ***y*-coordinates are the cumulative frequencies or cumulative relative frequencies** of the class.

Then line segments are drawn connecting consecutive points. An additional line segment is drawn connecting the first point to the horizontal axis at a location representing the upper limit of the class that would precede the first class (if it existed).

2.3 Additional Displays of Quantitative Data

2.3.2 Construct Frequency and Relative Frequency Ogives (2 of 3)

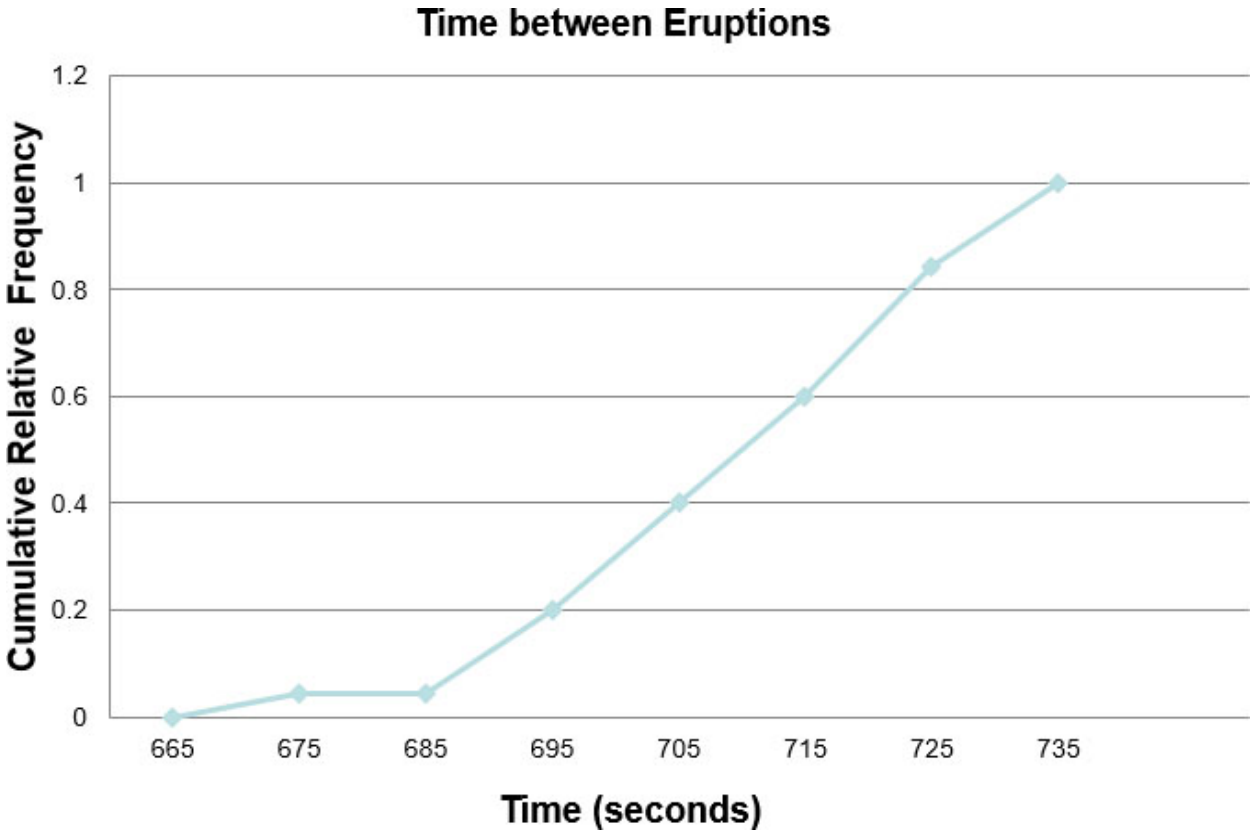
Frequency Ogive



2.3 Additional Displays of Quantitative Data

2.3.2 Construct Frequency and Relative Frequency Ogives (3 of 3)

Relative Frequency Ogive



2.3 Additional Displays of Quantitative Data

2.3.4 Draw Time Series Graphs (1 of 3)

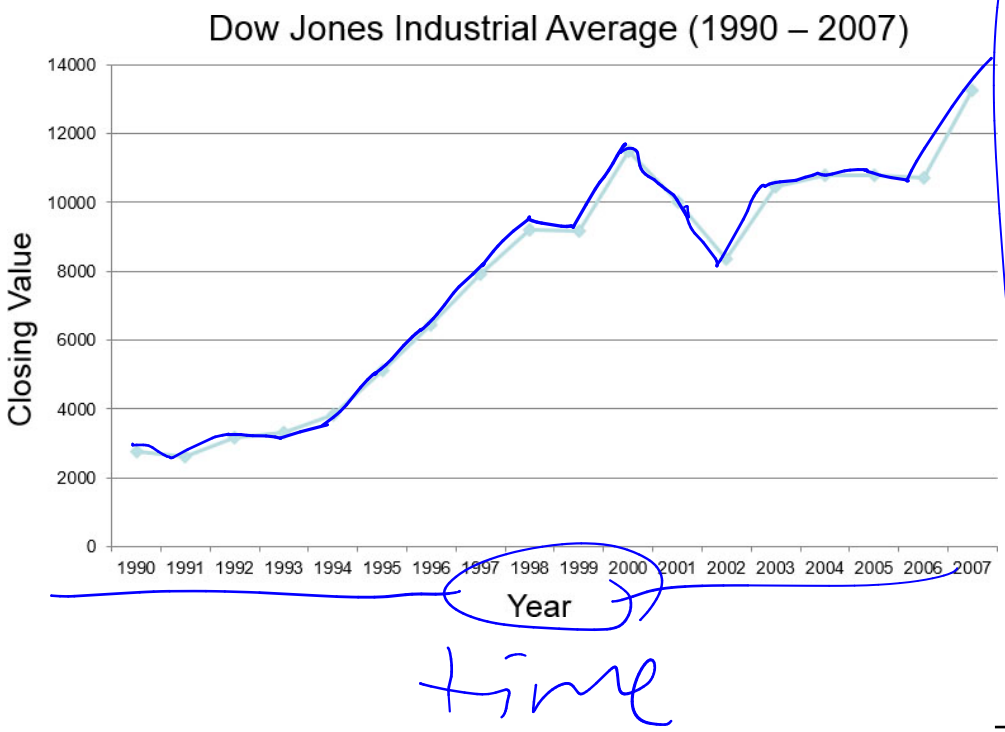
If the value of a variable is measured at different points in time, the data are referred to as **time series data**.

A **time-series plot** is obtained by plotting the time in which a variable is measured on the horizontal axis and the corresponding value of the variable on the vertical axis. Line segments are then drawn connecting the points.

2.3 Additional Displays of Quantitative Data

2.3.4 Draw Time Series Graphs (2 of 3)

The data to the right shows the closing prices of the Dow Jones Industrial Average for the years 1990 - 2007.



time

Year	Closing Value
1990	2753.2
1991	2633.66
1992	3168.83
1993	3301.11
1994	3834.44
1995	5117.12
1996	6448.27
1997	7908.25
1998	9212.84
1999	9,181.43
2000	11,497.12
2001	10021.71
2002	8342.38
2003	10452.74
2004	10783.75
2005	10,783.01
2006	10,717.50
2007	13264.82

Pearson

2.4 Graphical Misrepresentations of Data

Learning Objectives

1. Describe what can make a graph misleading or deceptive

2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (1 of 10)

Statistics: The only science that enables different experts using the same figures to draw different conclusions. – **Evan Esar**

Example: Non-zero vertical axis

common way misrepresent data

2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (2 of 10)

EXAMPLE Misrepresentation of Data

The data in the table represent the historical life expectancies (in years) of residents of the United States.

a) Construct a misleading time series graph that implies that life expectancies have risen sharply.

b) Construct a time series graph that is not **misleading**.

Year, *x*	Life Expectancy, *y*
1950	68.2
1960	69.7
1970	70.8
1980	73.7
1990	75.4
2000	77.0

Source: National Center for Health Statistics

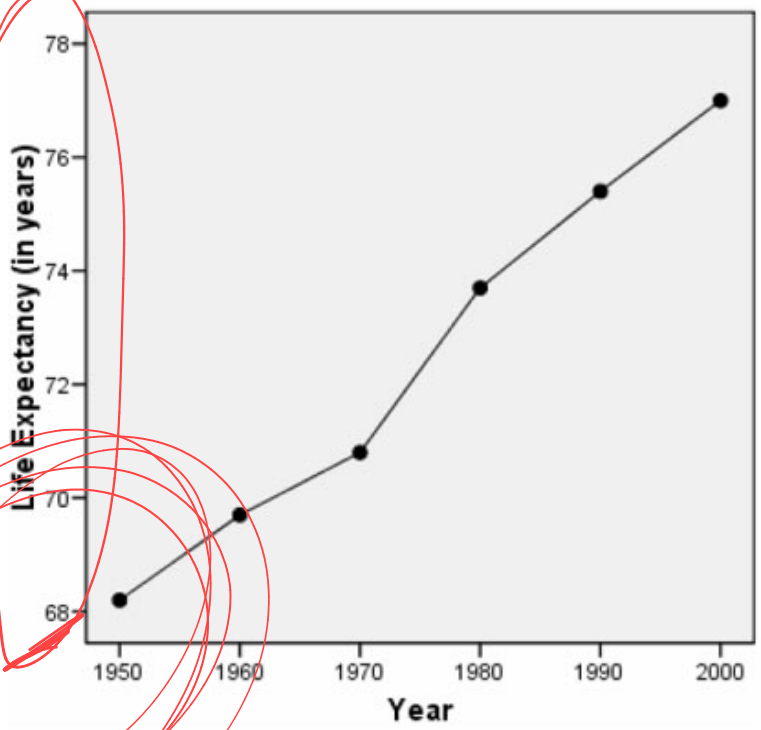
~10 apart

2.4 Graphical Misrepresentations of Data

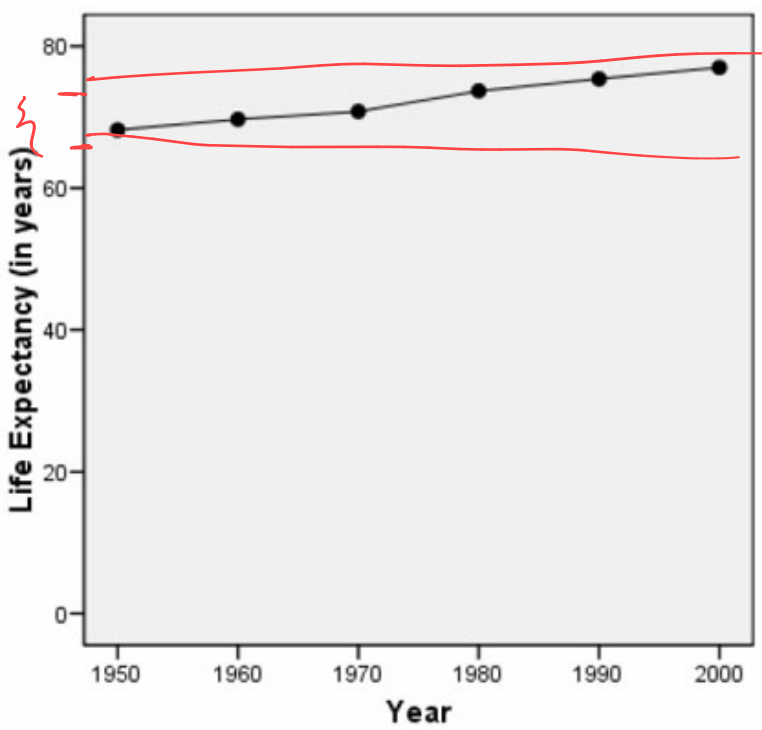
2.4.1 Describe What Can Make a Graph Misleading or Deceptive (3 of 10)

EXAMPLE Misrepresentation of Data

(a) Non-zero Vertical Axis



(b)



Pearson

2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (4 of 10)

EXAMPLE Misrepresentation of Data

The National Survey of Student Engagement is a survey that (among other things) asked first year students at liberal arts colleges how much time they spend preparing for class each week. The results from the 2007 survey are summarized on the next slide.

a) Construct a pie chart that exaggerates the percentage of students who spend between 6 and 10 hours preparing for class each week.

b) Construct a pie chart that is not misleading.

2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (5 of 10)

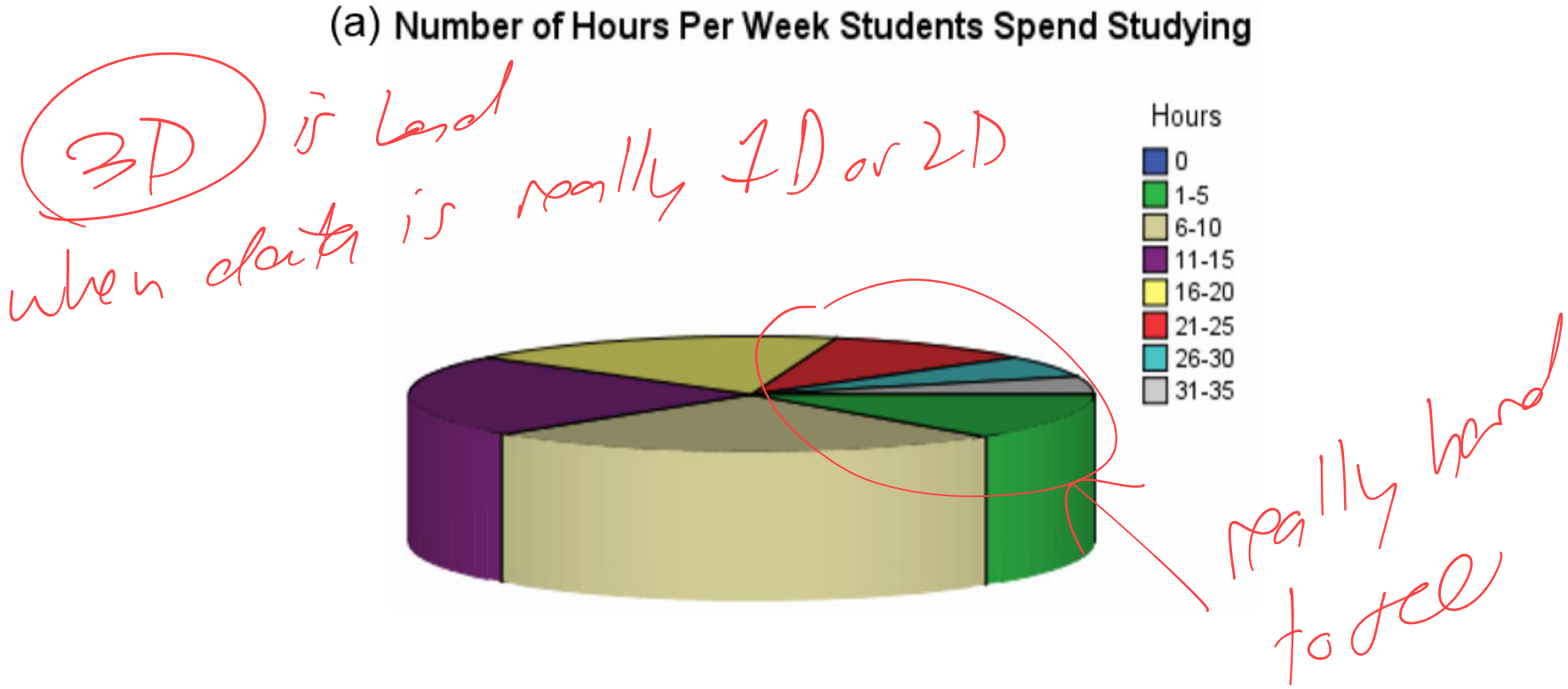
EXAMPLE Misrepresentation of Data

Hours	Relative Frequency
0	0
1 – 5	0.13
6 – 10	0.25
11 – 15	0.23
16 – 20	0.18
21 – 25	0.10
26 – 30	0.06
31 – 35	0.05

Source:
http://nsse.iub.edu/NSSE_2007_Annual_Report/docs/withhold/NSSE_2007_Annual_Report.pdf

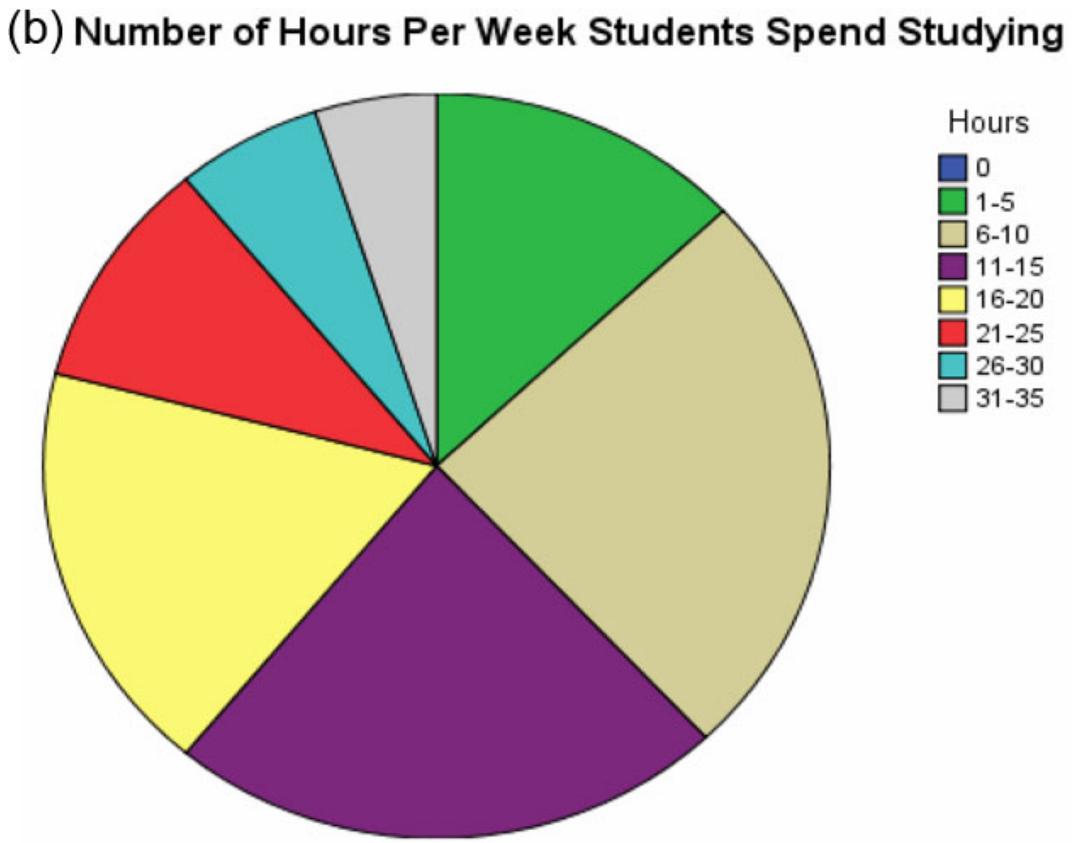
2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (6 of 10)



2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (7 of 10)



2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (8 of 10)

Guidelines for Constructing Good Graphics

- Title and label the graphic axes clearly, providing explanations, if needed. Include units of measurement and a data source when appropriate.
- Avoid distortion. Never lie about the data.
- Minimize the amount of white space in the graph. Use the available space to let the data stand out. If scales are truncated, be sure to clearly indicate this to the reader.
- Avoid clutter, such as excessive gridlines and unnecessary backgrounds or pictures. Don't distract the reader.
- Avoid three dimensions. Three-dimensional charts may look nice, but they distract the reader and often lead to misinterpretation of the graphic.
- Do not use more than one design in the same graphic. Sometimes graphs use a different design in one portion of the graph to draw attention to that area. Don't try to force the reader to any specific part of the graph. Let the data speak for themselves.
- Avoid relative graphs that are devoid of data or scales.

2.4 Graphical Misrepresentations of Data

2.4.1 Describe What Can Make a Graph Misleading or Deceptive (8 of 10)

Misleading or Deceptive Graphs