Chapter 2
Part B – Descriptive Statistics: Tabular and Graphical Displays

- Summarizing Data for Two Variables Using Tables
- Summarizing Data for Two Variables Using Graphical Displays
- Data Visualization: Best Practices in Creating Effective Graphical Displays

Summarizing Data for Two Variables Using Tables

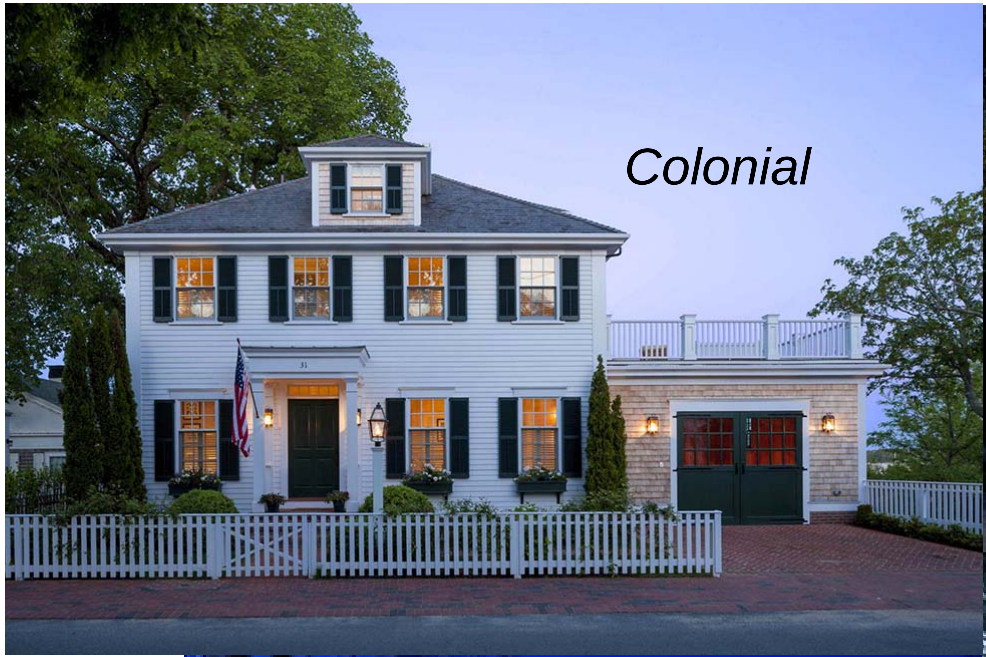
- Thus far we have focused on methods that are used to summarize the data for one variable at a time.
- Often a manager is interested in tabular and graphical methods that will help understand the relationship between two variables.
- Crosstabulation (交叉表) is a method for summarizing the data for two variables.

Crosstabulation

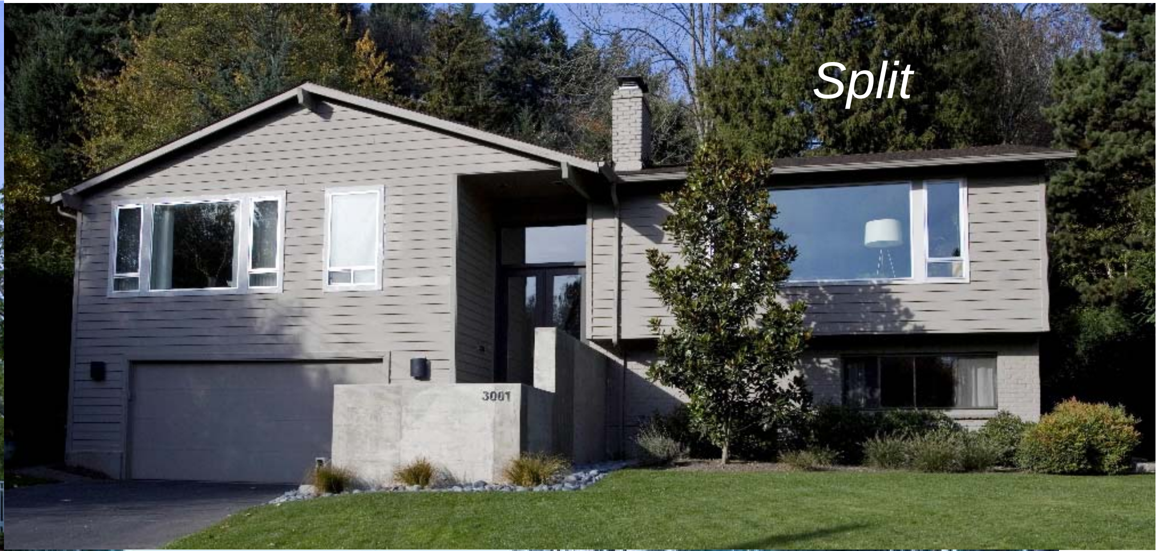
- A crosstabulation is a tabular summary of data for two variables.

- Crosstabulation can be used when:
  - one variable is categorical and the other is quantitative,
  - both variables are categorical, or
  - both variables are quantitative.

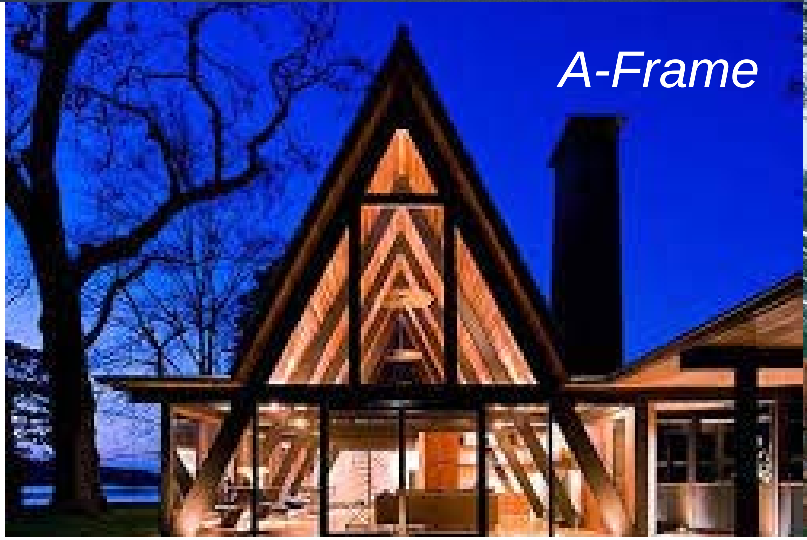
- The left and top margin labels define the classes for the two variables.



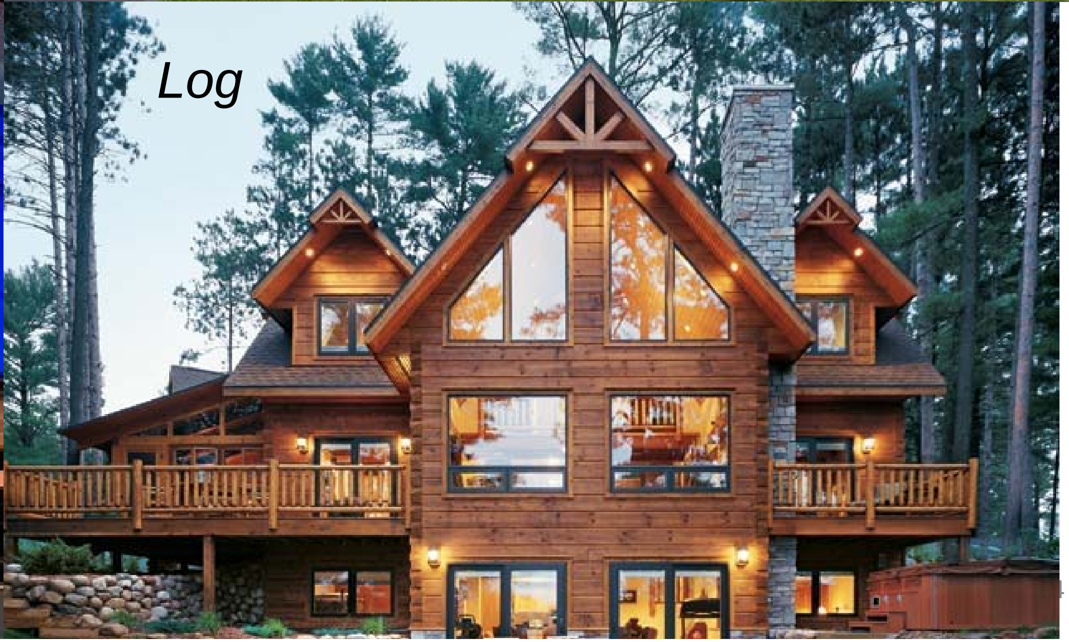
Colonial



Split



A-Frame



Log

Crosstabulation

- Example: Finger Lakes Homes

  The number of Finger Lakes homes sold for each style and price for the past two years is shown below.

Price	Home Style				
Range	Colonial	Log	Split	A-Frame	Total
< $250,000	18	6	19	12	55
≥ $250,000	12	14	16	3	45
Total	30	20	35	15	100

- The greatest number of homes (19) in the sample are a split-level style and priced at less than $250,000.
- Only three homes in the sample are an A-Frame style and priced at $250,000 or more.

Crosstabulation: Row Percentages

- Example: Finger Lakes Homes

Price	Home Style				
Range	Colonial	Log	Split	A-Frame	Total
< $250,000	32.73	10.91	34.55	21.82	100
≥ $250,000	26.67	31.11	35.56	6.67	100

Note: row totals are actually 100.01 due to rounding.

(Colonial and ≥ $250K)/(All ≥ $250K) x 100 = (12/45) x 100

Crosstabulation: Column Percentages

- Example: Finger Lakes Homes

Price	Home Style			
Range	Colonial	Log	Split	A-Frame
< $250,000	60.00	30.00	54.29	80.00
≥ $250,000	40.00	70.00	45.71	20.00
Total	100	100	100	100

(Colonial and ≥ $250K)/(All Colonial) x 100 = (12/30) x 100

Summary Crosstabulation

Verdict (判決)	Judge Luckett	Judge Kendall	Total
Upheld (維持原判)	129 (86%)	110 (88%)	239
Reversed (駁回重審)	21 (14%)	15 (12%)	36
Total	150 (100%)	125 (100%)	275

Verdict (判決)	Common Pleas (民事訴訟)	Municipal court (市政法院)	Total
Judge Luckett			
Upheld (維持原判)	29 (91%)	100 (85%)	129
Reversed (駁回重審)	3 (9%)	18 (15%)	21
Total	32 (100%)	118 (100%)	150
Judge Kendall			
Upheld (維持原判)	90 (90%)	20 (80%)	110
Reversed (駁回重審)	10 (10%)	5 (20%)	15
Total	100 (100%)	25 (100%)	125

Crosstabulation: Simpson's Paradox

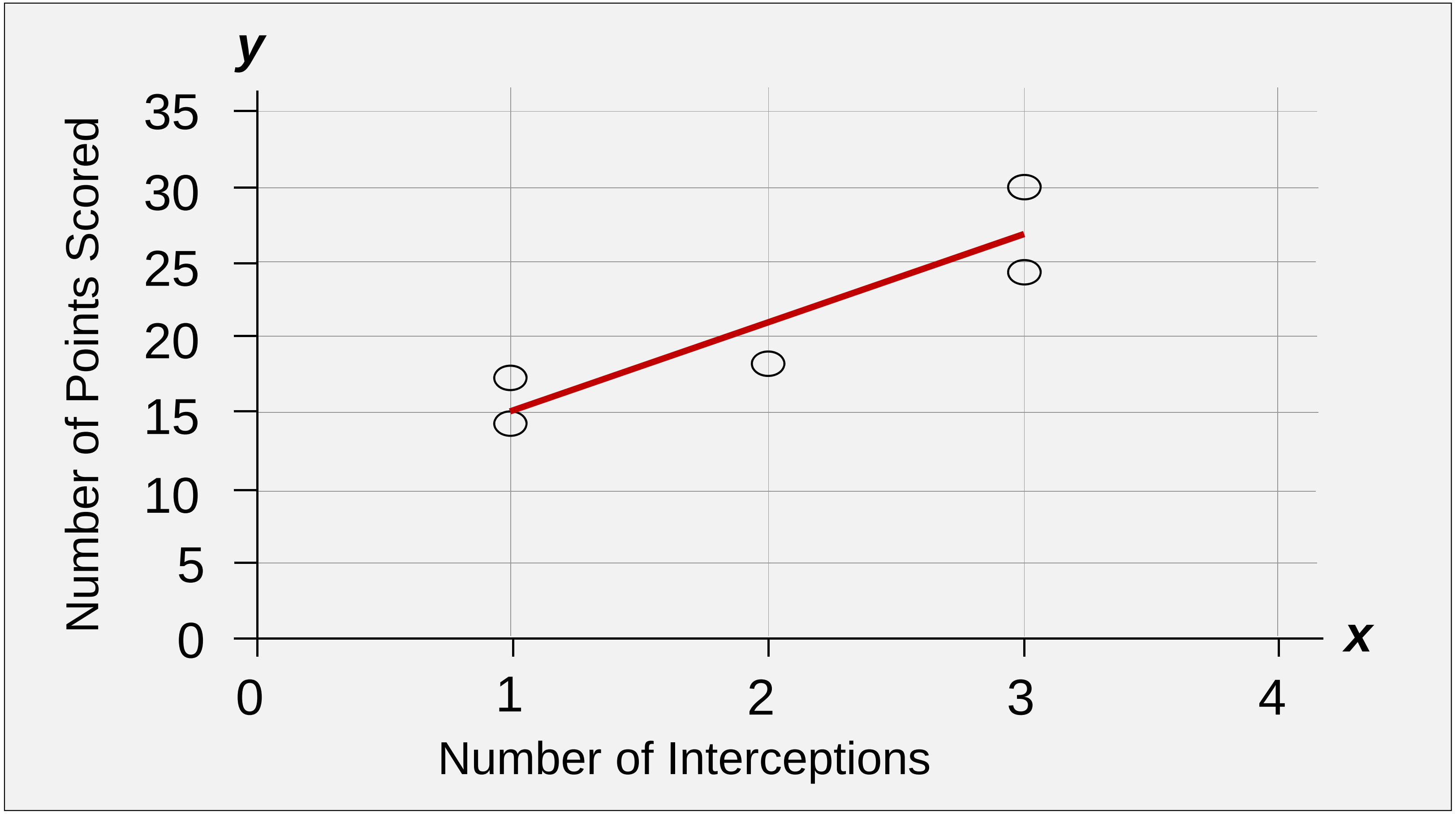
- Data in two or more crosstabulations are often aggregated to produce a summary crosstabulation.
- We must be careful in drawing conclusions about the relationship between the two variables in the aggregated crosstabulation.
- In some cases the conclusions based upon an aggregated crosstabulation can be completely reversed if we look at the unaggregated data. The reversal of conclusions based on aggregate and unaggregated data is called Simpson's paradox.

- Example: Panthers Football Team

  The Panthers football team is interested in investigating the relationship, if any, between interceptions made and points scored.

x = Number of Interceptions	*y* = Number of Points Scored
1	14
3	24
2	18
1	17
3	30

Scatter Diagram and Trendline

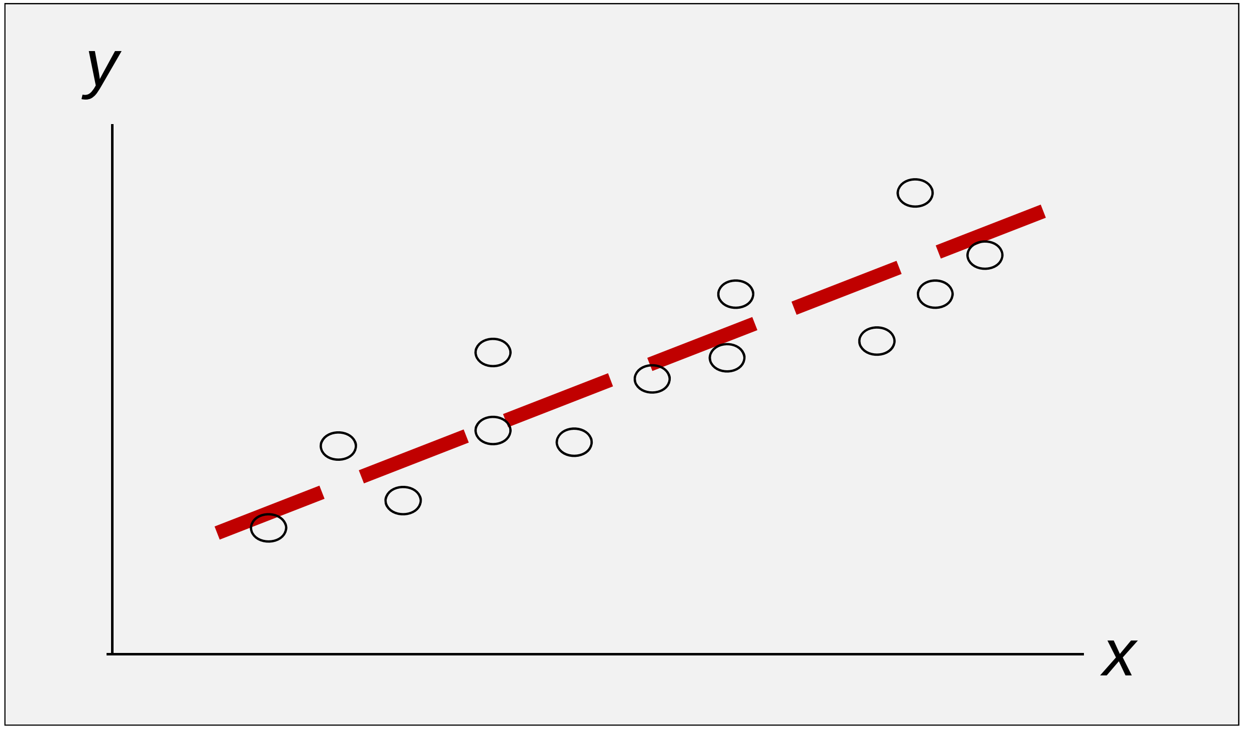


Summarizing Data for Two Variables
Using Graphical Displays: Scatter Diagram and Trendline

- Scatter diagrams and trendlines are useful in exploring the relationship between two variables.
- A scatter diagram is a graphical presentation of the relationship between two **quantitative variables**.
- One variable is shown on the horizontal axis and the other variable is shown on the vertical axis.
- The general pattern of the plotted points suggests the overall relationship between the variables.
- A trendline provides an approximation of the relationship.

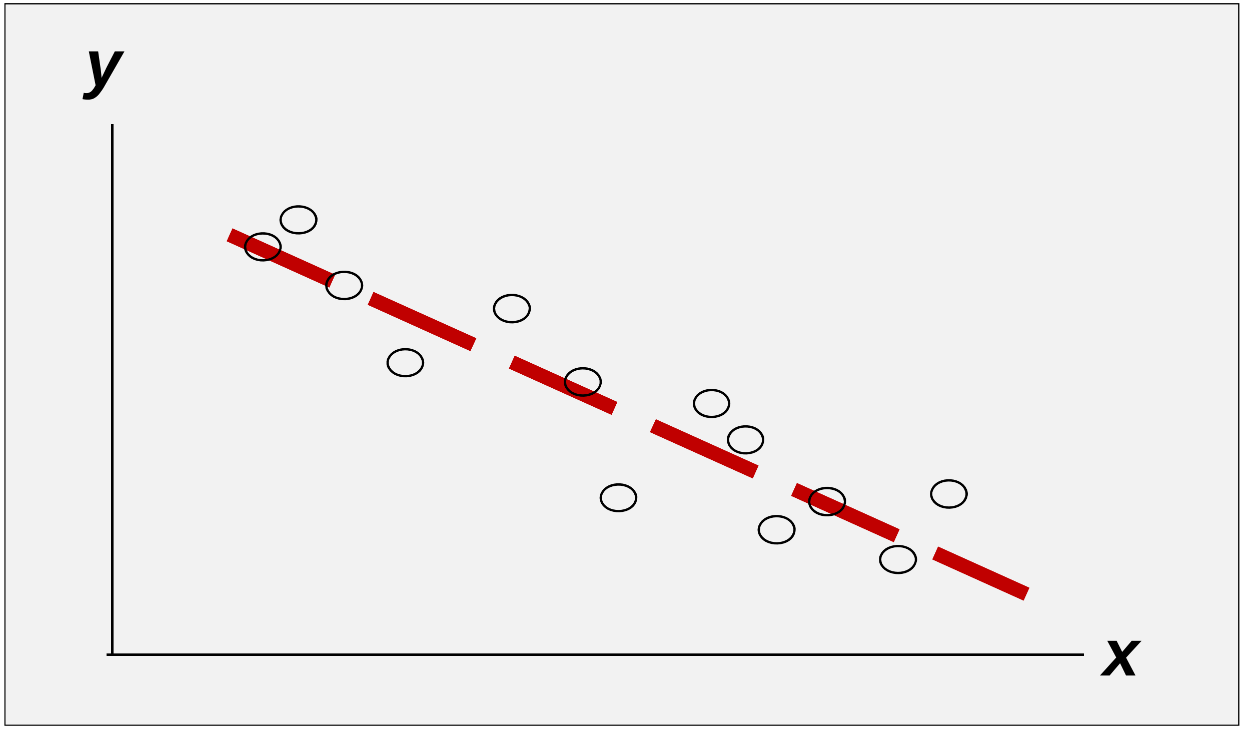
Scatter Diagram

- A Positive Relationship



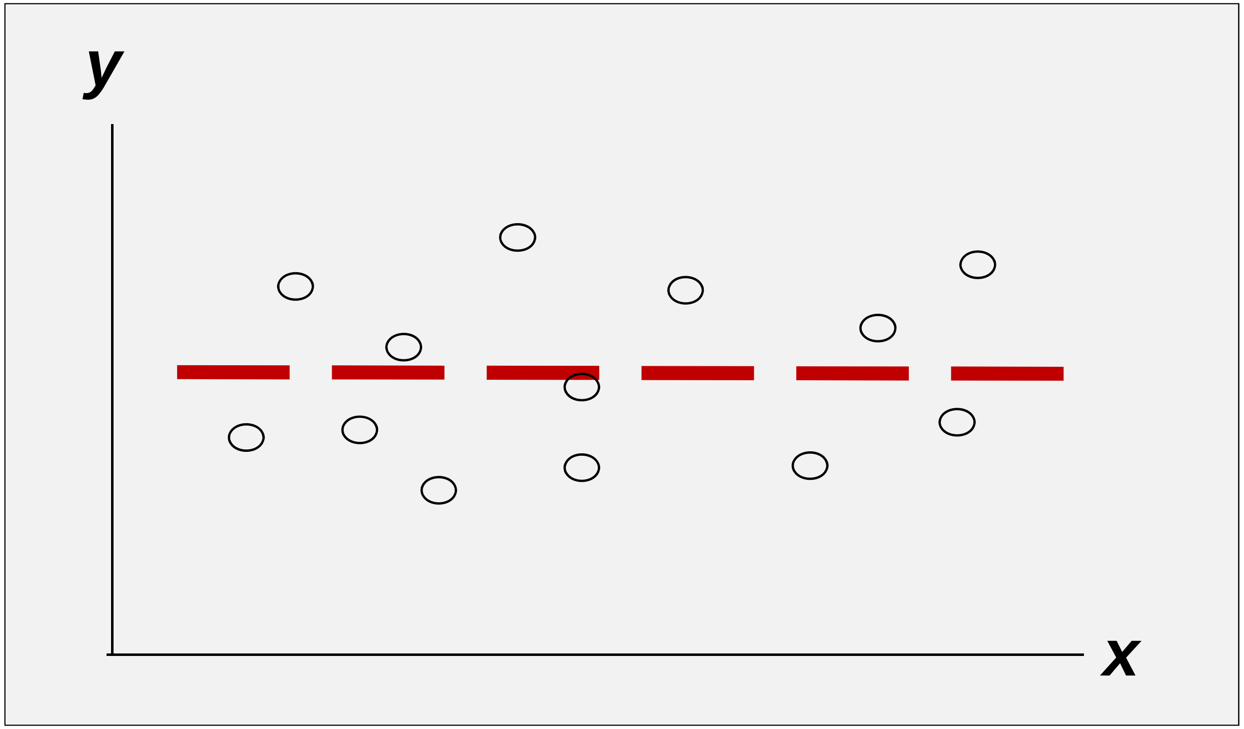
Scatter Diagram

- A Negative Relationship



Scatter Diagram

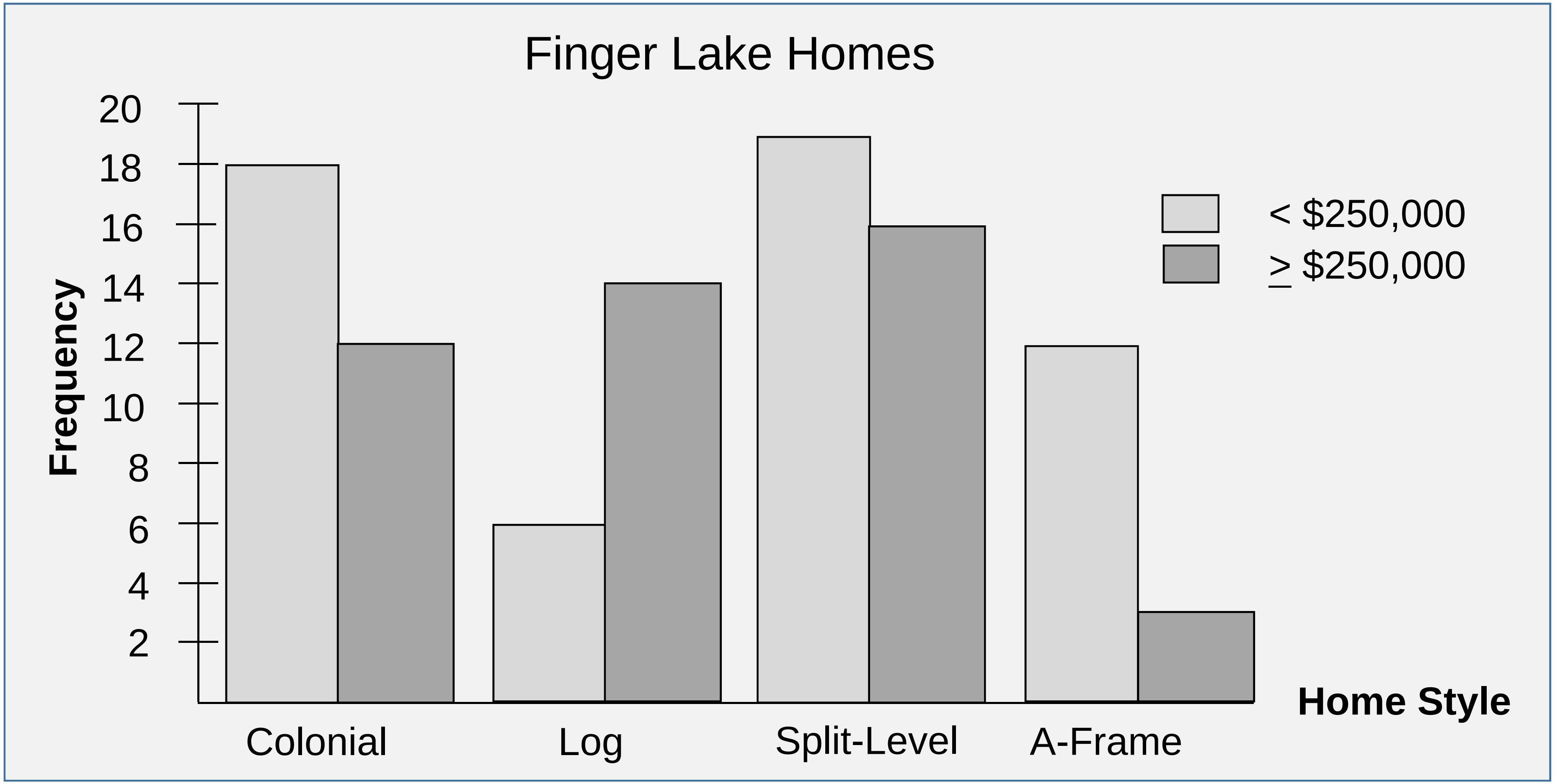
- No Apparent Relationship



Side-by-Side Bar Chart

- A side-by-side bar chart is a graphical display for depicting multiple bar charts on the same display.
- Each cluster of bars represents one value of the first variable.
- Each bar within a cluster represents one value of the second variable.

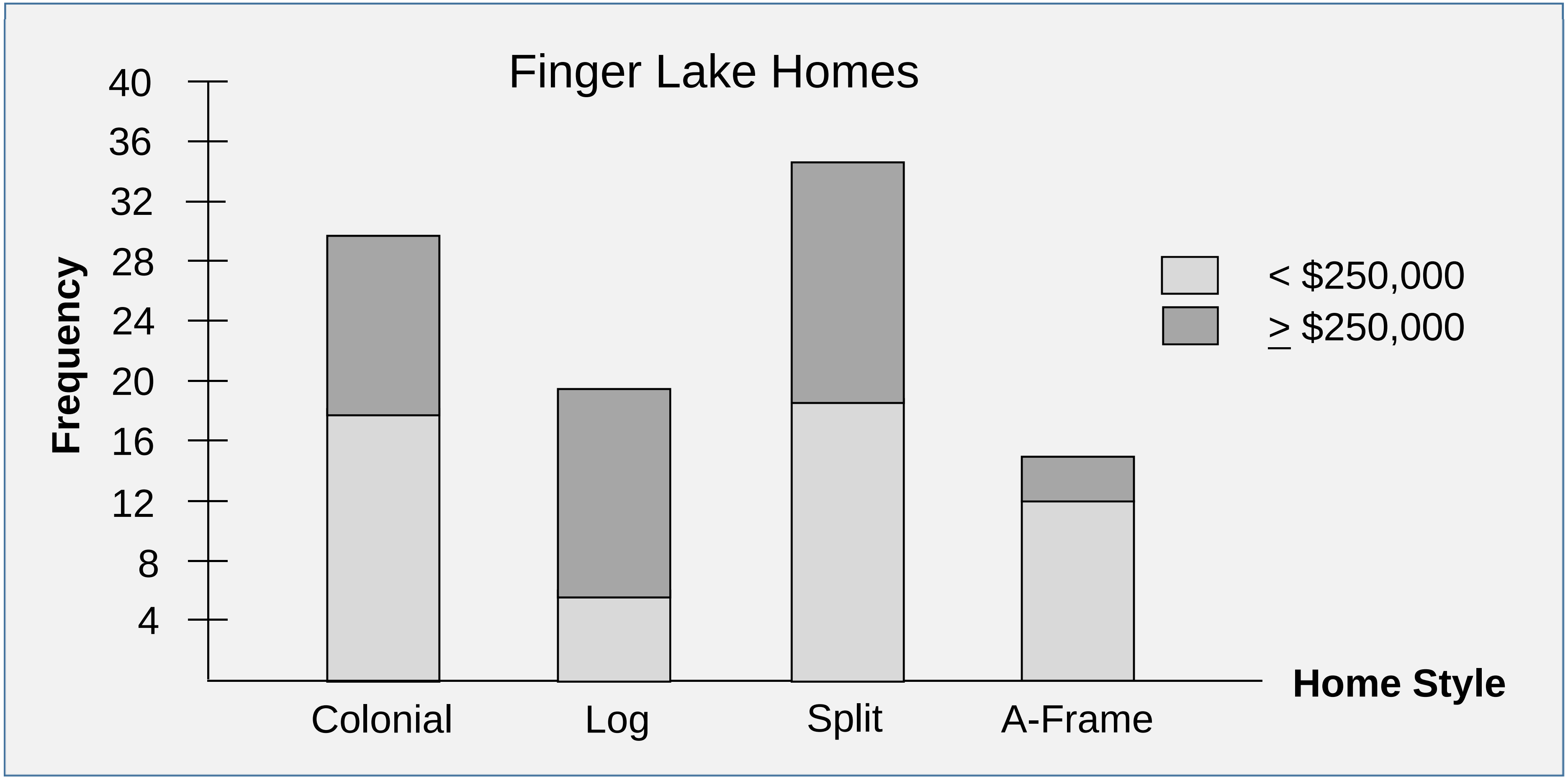
Side-by-Side Bar Chart



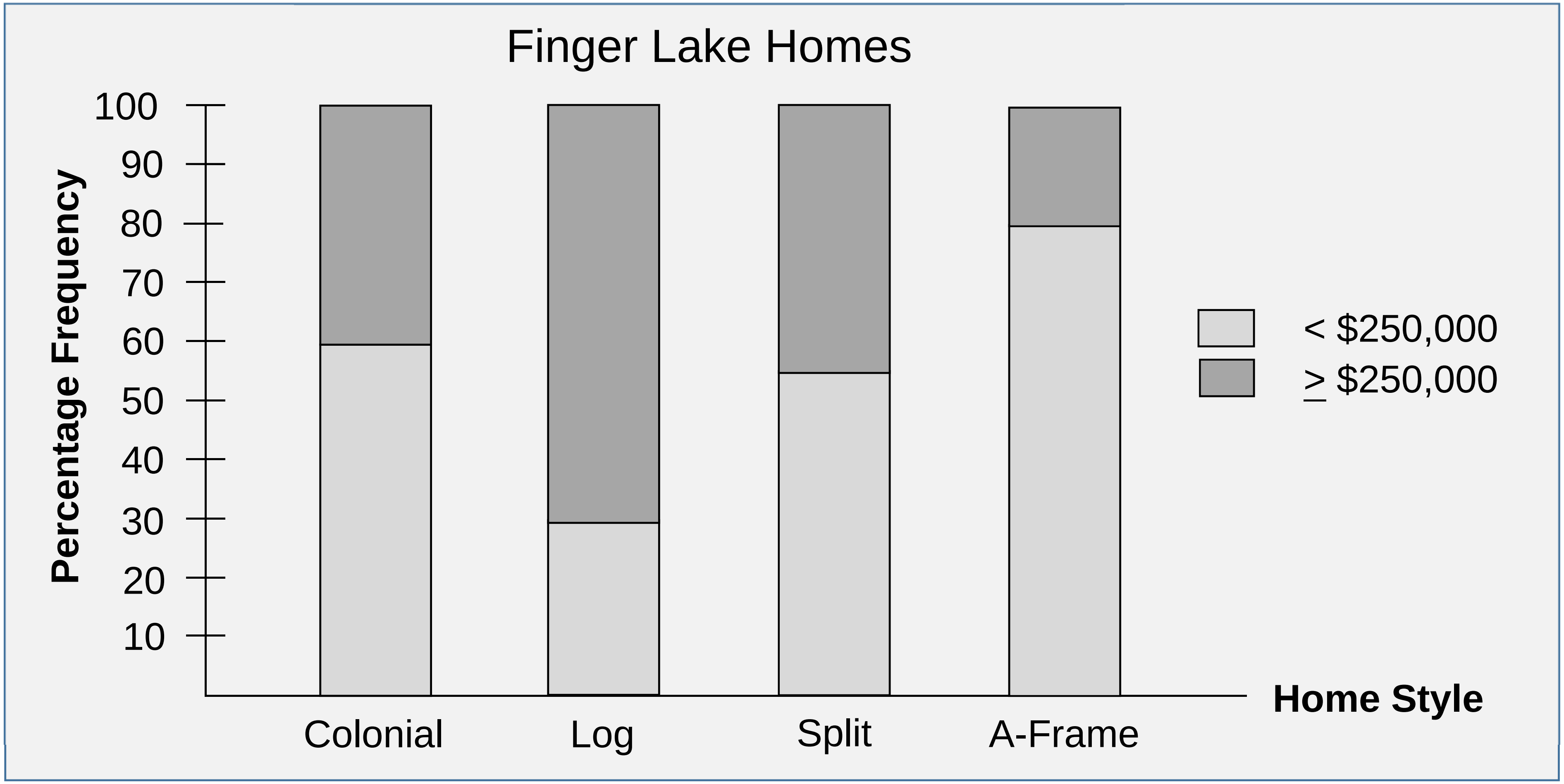
Stacked Bar Chart

- A stacked bar chart is another way to display and compare two variables on the same display.
- It is a bar chart in which each bar is broken into rectangular segments of a different color.
- If percentage frequencies are displayed, all bars will be of the same height (or length), extending to the 100% mark.

Stacked Bar Chart



Stacked Bar Chart



Data Visualization: Best Practices in Creating Effective Graphical Displays

- Data visualization is the use of graphical displays to summarize and present information about a data set.
- The goal is to communicate as effectively and clearly as possible, the key information about the data.

Creating Effective Graphical Displays

- Creating effective graphical displays is as much art as it is science.
- Here are some guidelines . . .
    - Give the display a clear and concise title.
    - Keep the display simple.
    - Clearly label each axis and provide the units of measure.
    - If colors are used, make sure they are distinct.
    - If multiple colors or line types are used, provide a legend.

Choosing the Type of Graphical Display

- Displays used to show the distribution of data:

  Bar Chart to show the frequency distribution and relative frequency distribution for categorical data

  Pie Chart to show the relative frequency and percent frequency for categorical data

  Dot Plot to show the distribution for quantitative data over the entire range of the data

  Histogram to show the frequency distribution for quantitative data over a set of class intervals

  Stem-and-Leaf Display to show both the rank order and shape of the distribution for quantitative data

Choosing the Type of Graphical Display

- Displays used to make comparisons:

  Side-by-Side Bar Chart to compare two variables

  Stacked Bar Chart to compare the relative frequency or percent frequency of two categorical variables

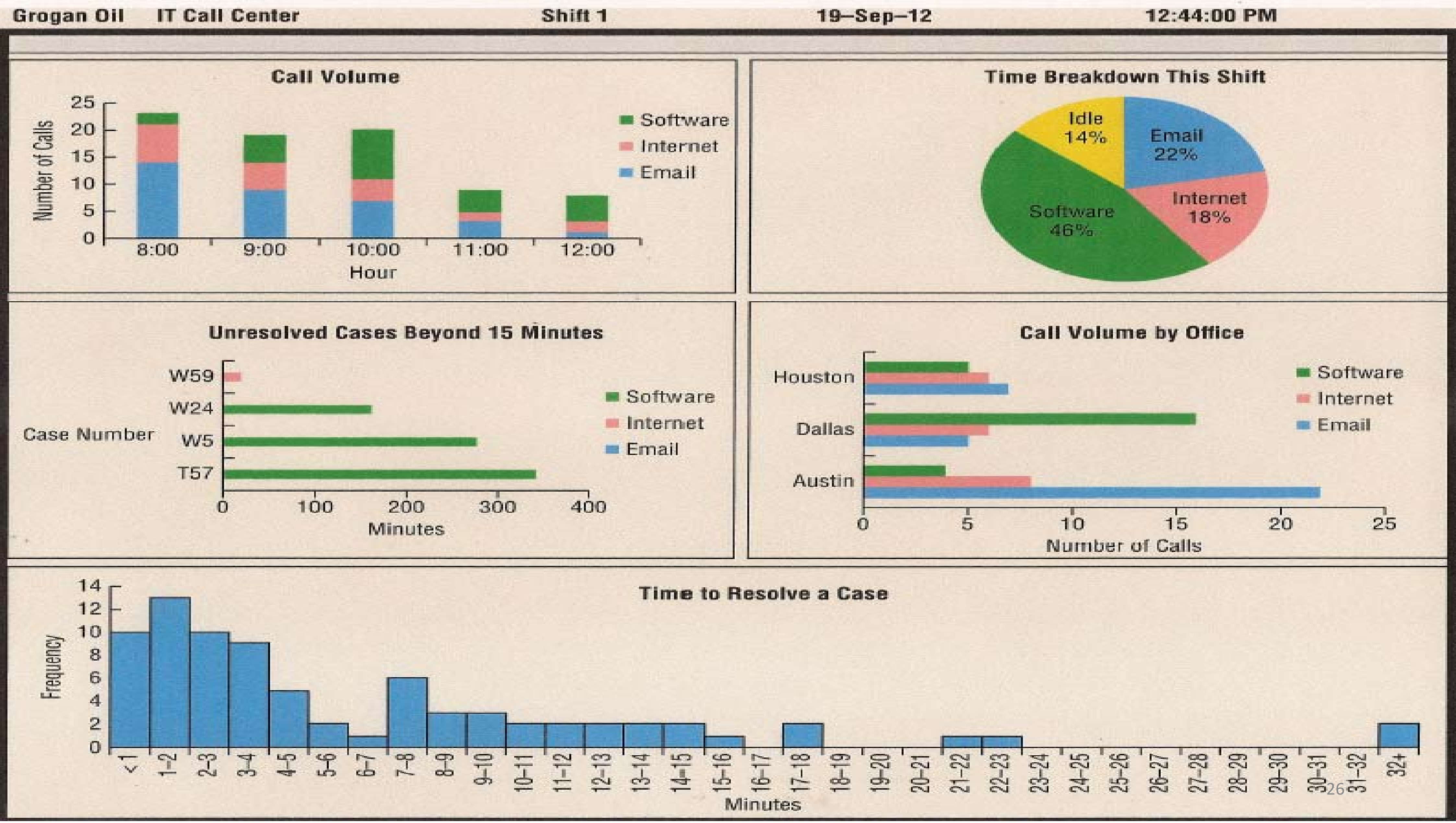
- Displays used to show relationships:

  Scatter Diagram to show the relationship between two quantitative variables

  Trendline to approximate the relationship of data in a scatter diagram

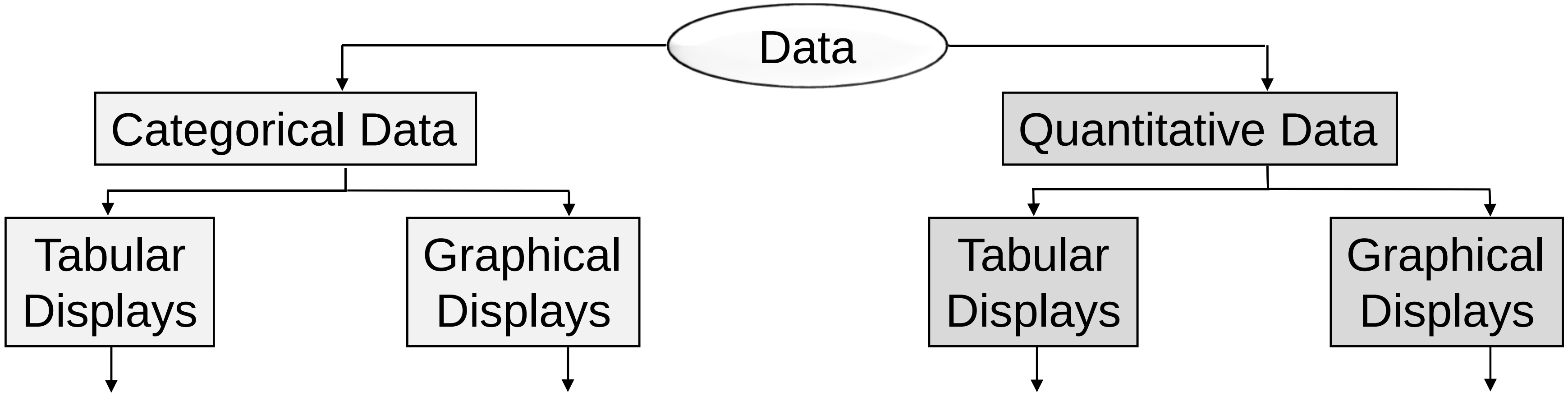
Data Dashboards

- A data dashboard is a widely used data visualization tool.
- It organizes and presents key performance indicators (KPIs) used to monitor an organization or process.
- It provides timely summary information that is easy to read, understand, and interpret.
- Some additional guidelines include . . .
  - Minimize the need for screen scrolling.
  - Avoid unnecessary use of color or 3D displays.
  - Use borders between charts to improve readability.



Grogan Oil IT Call Center Shift 1 19–Sep–12 12:44:00 PM

Tabular and Graphical Displays



Data

Categorical Data

Tabular Displays

- Frequency Distribution
- Rel. Freq. Dist.
- Percent Freq. Distribution
- Crosstabulation

Graphical Displays

- Bar Chart
- Pie Chart
- Side-by-Side Bar Chart
- Stacked Bar Chart

Quantitative Data

Tabular Displays

- Frequency Dist.
- Rel. Freq. Dist.
- % Freq. Dist.
- Cum. Freq. Dist.
- Cum. Rel. Freq. Dist.
- Cum. % Freq. Dist.
- Crosstabulation

Graphical Displays

- Dot Plot
- Histogram
- Stem-and-Leaf Display
- Scatter Diagram

FuelData2012

- Size: Compact, Midsize, and Large
- Displacement: Engine size in liters
- Cylinders (汽缸): Number of cylinders in the engine
- Drive: All wheel (A), front wheel (F), and rear wheel (R)
- Fuel Type: Premium (P) or regular (R) fuel
- City MPG: Fuel efficiency rating for city driving in terms of miles per gallon
- Hwy MPG: Fuel efficiency rating for highway driving in terms of miles per gallon

1. *Size vs. Hwy MPG*
2. *Drive vs. City MPG*
3. *Fuel Type vs. City MPG*

Car	Size	Displacement	Cylinders	Drive	Fuel Type	City MPG	Hwy MPG
1	Compact	2.0	4	F	P	22	30
2	Compact	2.0	4	A	P	21	29
			…				
57	Midsize	3.5	6	F	P	20	29
58	Midsize	3.7	6	A	P	18	26
			…				
126	Large	4.2	8	A	P	18	28