

Data Visualization

Exploring and Explaining with Data



Data Visualization, 1e

Chapter 2: Selecting a Chart Type

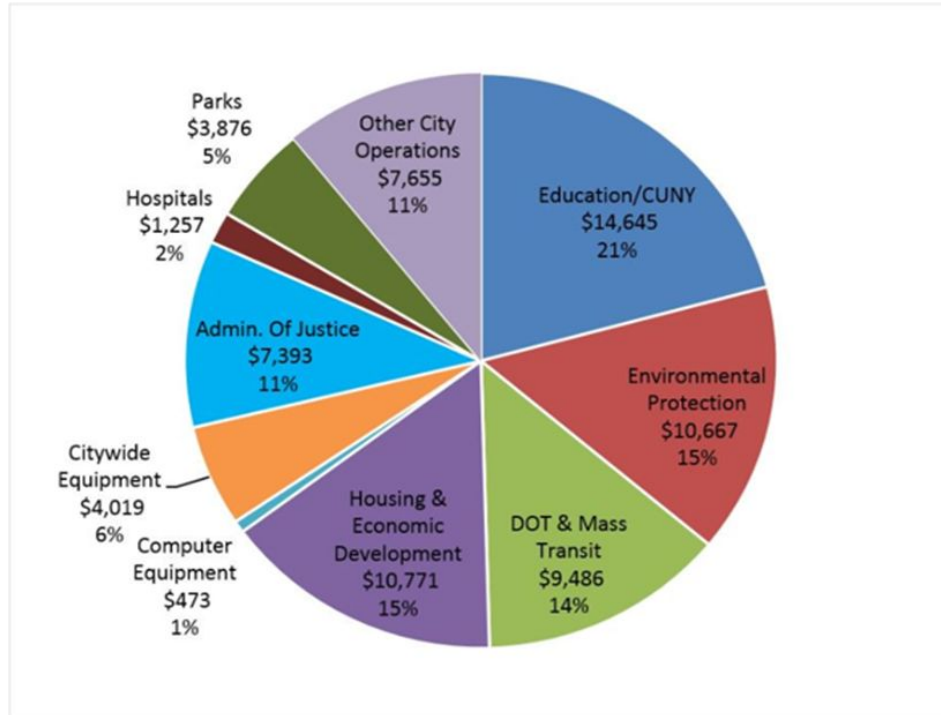
Chapter Objectives

After completing this chapter, you will be able to:

- LO 2.1 Create charts and graphs using Excel
- LO 2.2 Edit charts and graphs created using Excel
- LO 2.3 Select an appropriate chart type for a given goal and data type
- LO 2.4 Glean insights from charts and graphs
- LO 2.5 Describe which chart types should be avoided and why

Data Visualization Makeover

A pie chart showing the NYC budget allocation



Source: NYC Office of Management and Budget, FY 2020 Adopted Capital Commitment Plan, October 2019 Pie Chart figures may not tie due to rounding.

A bar chart showing the NYC budget allocation



*See additional comments in the notes.

2.1 Selecting an Appropriate Chart

Some of the common goals for charts are to show the following:

- The **composition** – the breakdown of a whole entity into its components.
- The **ranking** – is the ordering of items in order of contextual importance.
- The **relationship** – is the identification of patterns existing between pairs of variables.
- The **distribution** – is the dispersion of components of a whole entity.

2.2 Creating a Chart in Excel

	A	B
1	Month	Attendance
2	Jan	5422
3	Feb	4878
4	Mar	6586
5	Apr	6943
6	May	7876
7	Jun	17843
8	July	21967
9	Aug	14542
10	Sept	8751
11	Oct	6454
12	Nov	5677
13	Dec	11422

The following steps show how to create a column chart in Excel using the data in the file *Zoo*.

- **Step 1.** Select cells A1:B13
- **Step 2.** Click the **Insert** tab on the Ribbon
- **Step 3.** Click the **Insert Column or Bar Chart** button in the **Charts group**
- **Step 4.** When the list of column and bar charts subtypes appears, click the **Clustered Column** button

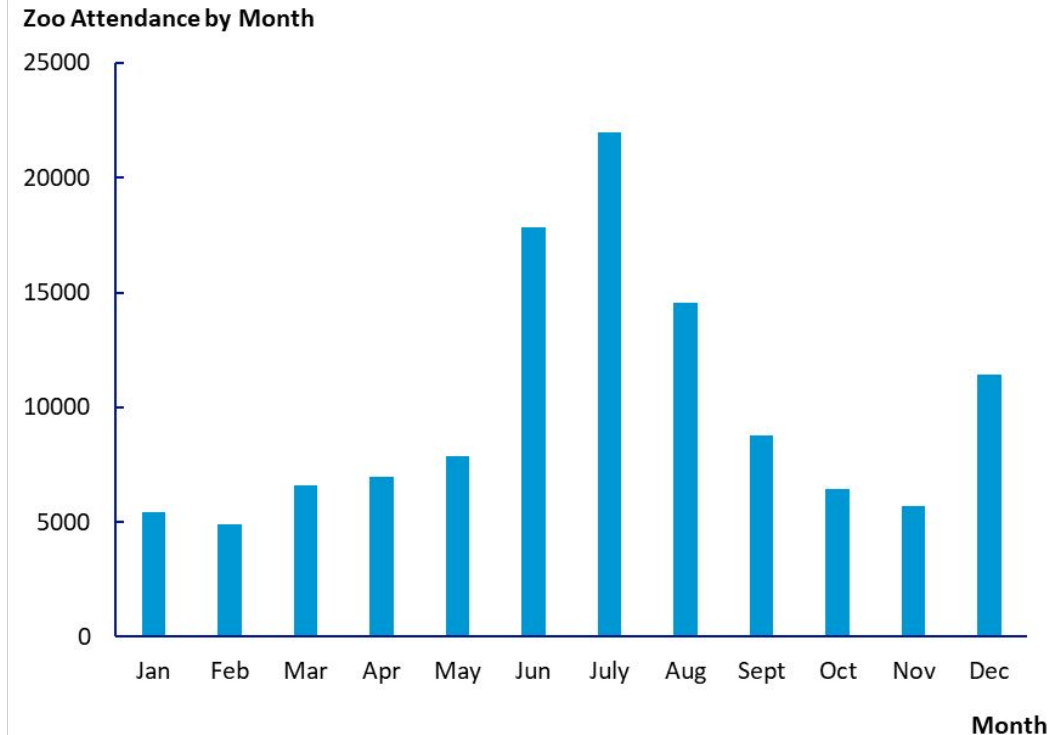
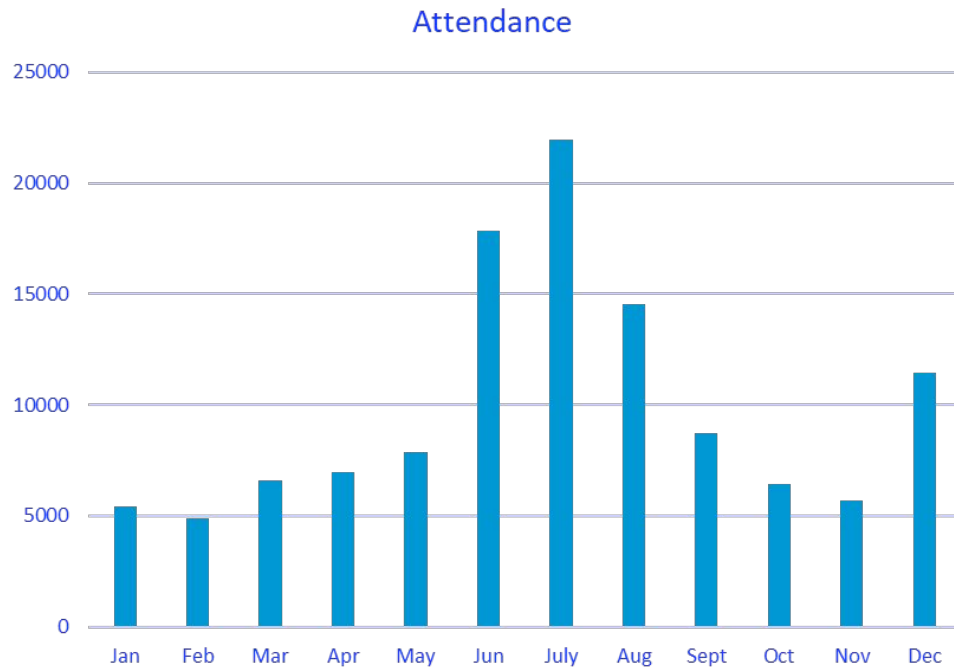
In Excel, the horizontal axis on the chart generally represents the leftmost column of data.

The assignment of data to the horizontal and vertical axes can be switched by right-clicking on the chart, choosing **Select Data Source**, and on the **Select Data Source** dialog box, click on **Switch Row/Column**.

To create a chart using nonadjacent columns of data in Excel, select the leftmost column you wish to include in the chart, press and hold down the control key (**Ctrl**), select the other columns of data you want to include.

2.2 Editing a Chart in Excel

Step-by-step instructions on how to improve the chart shown to the left so that it appears as shown to the right are shown in the notes.



2.3 Build a Scatter Chart

	A	B	C	D
1	City	State	Average Low temp	Average Snowfall
2	Atlanta	Georgia	53	2.9
3	Austin	Texas	59	0.6
4	Baltimore	Maryland	45	20.2
5	Birmingham	Alabama	53	1.6
6	Boston	Massachusetts	44	43.8
7	Buffalo	New York	40	94.7
8	Charlotte	North Carolina	49	4.3
9	Chicago	Illinois	41	36.7
10	Cincinnati	Ohio	43	11.2
11	Cleveland	Ohio	43	68.1
12	Columbus	Ohio	44	27.5
13	Dallas	Texas	57	1.5
14	Denver	Colorado	36	53.8
15	Detroit	Michigan	42	42.7
16	Hartford	Connecticut	40	40.5

The following steps show how to create a scatter chart in Excel using the data in the file *Snow*.

Step 1. Select cells C1:D52

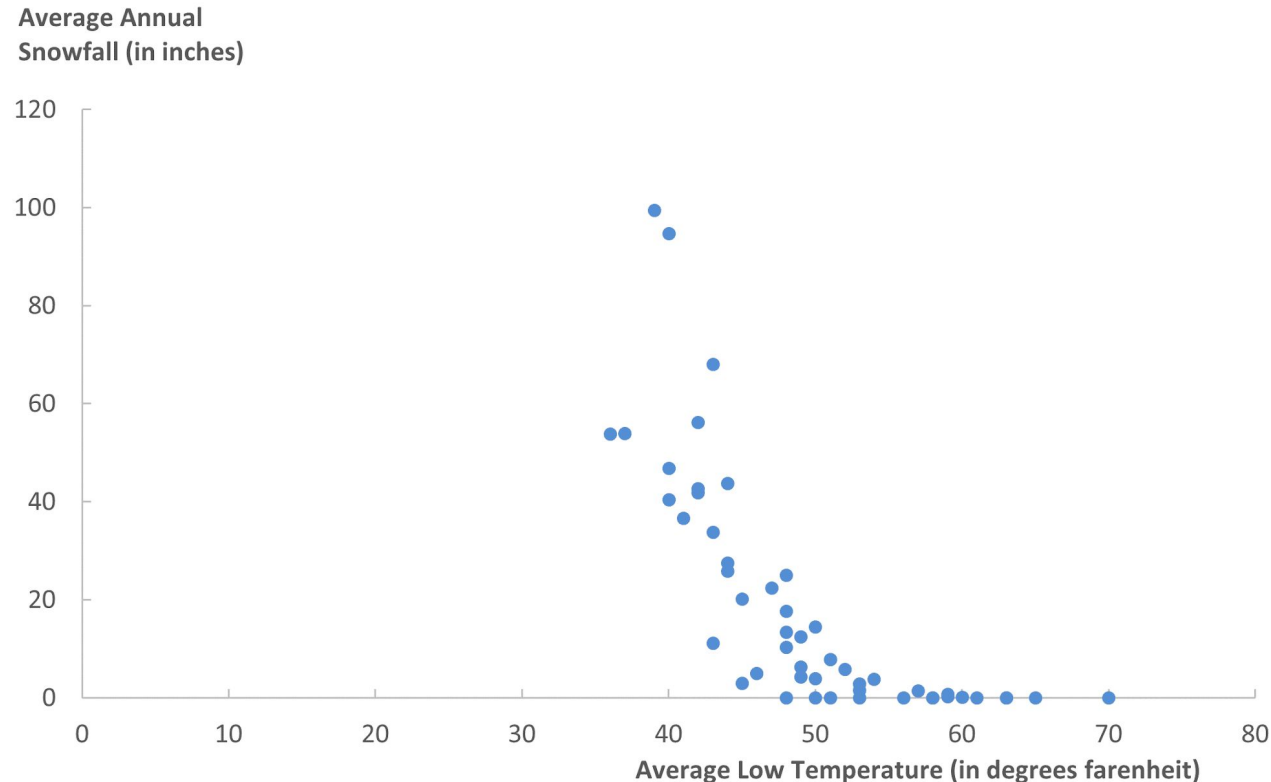
Step 2. Click the **Insert** tab on the Ribbon

Step 3. Click the **Insert Scatter (X,Y) or Bubble Chart** button in the **Charts** group

Step 4. When the list chart subtypes appear, click the **Scatter** button. Then edit the chart as outlined in Section 2.2.

2.3 Interpret a Scatter Chart

Average Snowfall versus Average Low Temperature for 51 U.S. Cities



- Each point on the chart represents a pair of measurements for 51 U.S. cities.
- The measurements are average low temperature (in degrees Fahrenheit) and the average annual amount of snowfall (inches).
- The chart shows how the average annual amount of snowfall intuitively levels off at zero for warm-weather cities.
- Scatter charts are among the most useful charts for exploring pairs of quantitative data.

2.3 Build a Bubble Chart

	A	B	C	D
1	Airport Code	TSA wait time in minutes	Cheapest on-site parking cost in \$ per day	Annual Enplanements in millions of passengers
2	ATL	10.30	\$14.0	49.1
3	CLT	9.45	\$7.0	22.2
4	DEN	8.35	\$8.0	27.0
5	AUS	5.30	\$8.0	5.8
6	STL	4.80	\$7.0	6.2
7	SMF	7.30	\$10.0	4.6
8	RDU	6.75	\$6.5	4.8
9	EWR	9.90	\$18.0	18.8
10	SFO	7.25	\$18.0	24.0
11	LAX	6.40	\$12.0	36.1
12	SLC	4.50	\$10.0	10.8
13	SAN	8.80	\$15.0	9.7
14	MSY	4.60	\$16.0	5.3
15	IAD	3.0	\$10	10.7
16	BNA	2.50	\$12.0	5.6

The following steps show how to create a bubble chart in Excel using the data in the file *AirportData*.

Step 1. Select cells B1:D16

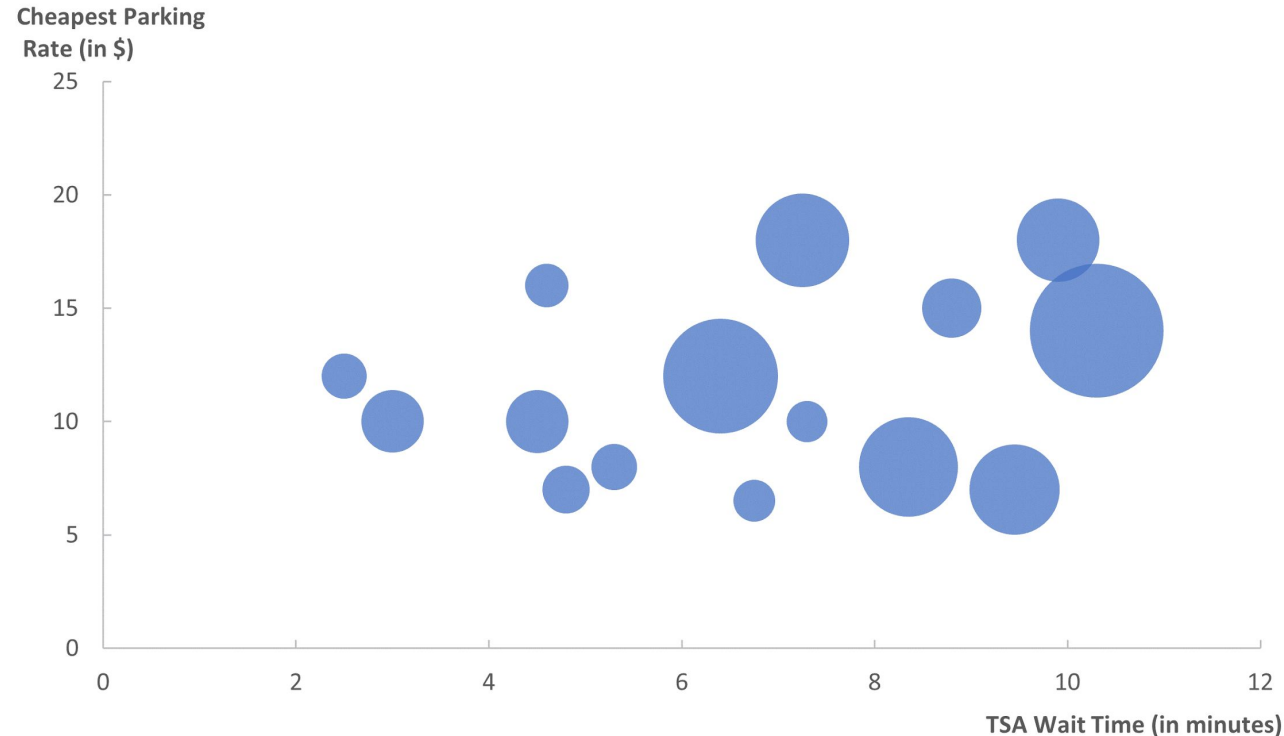
Step 2. Click the **Insert** tab on the Ribbon

Step 3. Click the **Insert Scatter (X,Y) or Bubble Chart** button in the **Charts** group

Step 4. When the list chart subtypes appear, click the **Bubble** button. Then, edit the chart as outlined in Section 2.2.

2.3 Interpret a Bubble Chart

Airport Passengers, TSA Wait Time and Cost Of Parking

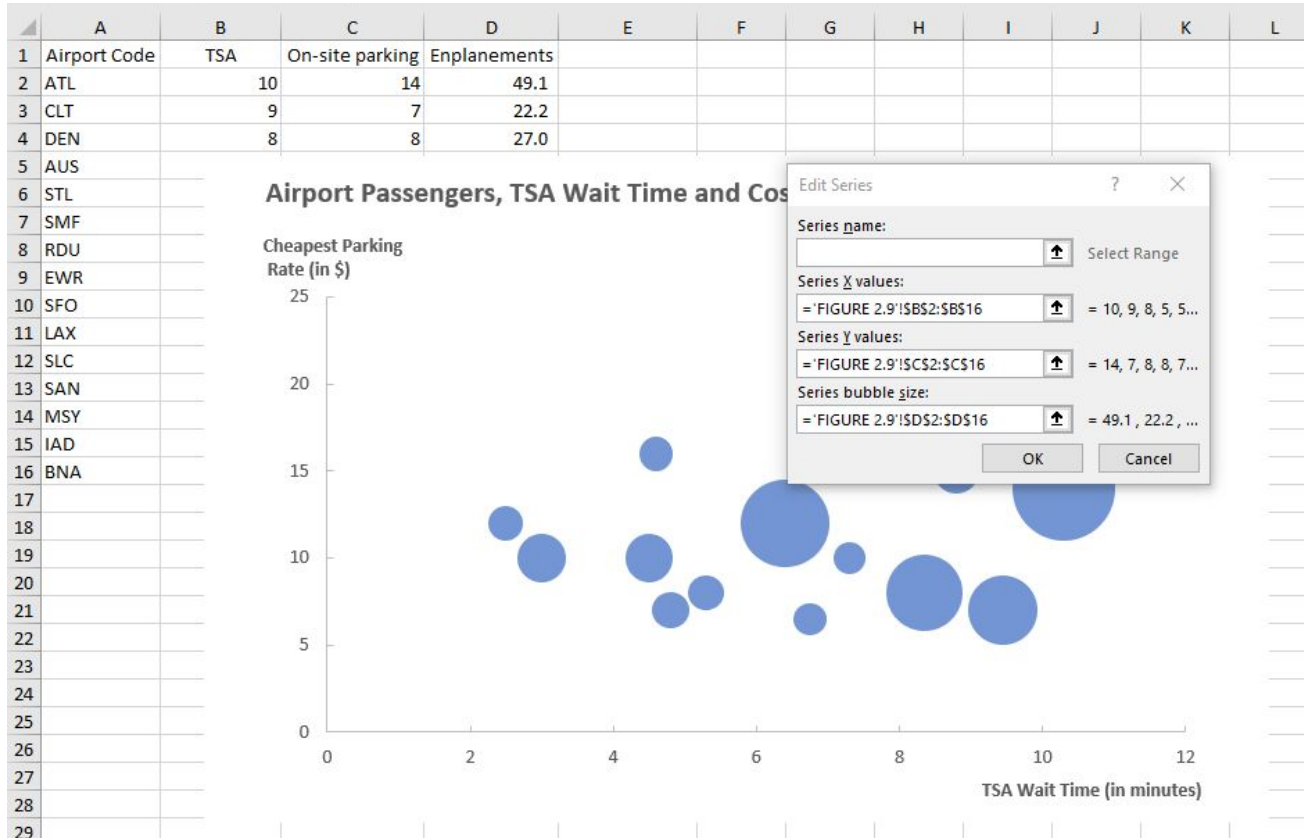


A bubble chart allows us to introduce a third quantitative variable, displayed by using different sized dots, which we refer to as bubbles.

For each of the 15 airports, we have the following quantitative variables:

- Horizontal axis: Average wait time [in minutes] in the non-priority TSA queue.
- Vertical axis: The cheapest on-site daily rate [in \$] for parking at the airport.
- Bubble size: The number of enplanements [in millions of passengers] in a year, including transfers.

2.3 Edit a Bubble Chart



You can use the following steps to change the variable assignments to the x-axis, y-axis, and bubbles:

Step 1. Right-click on any bubble

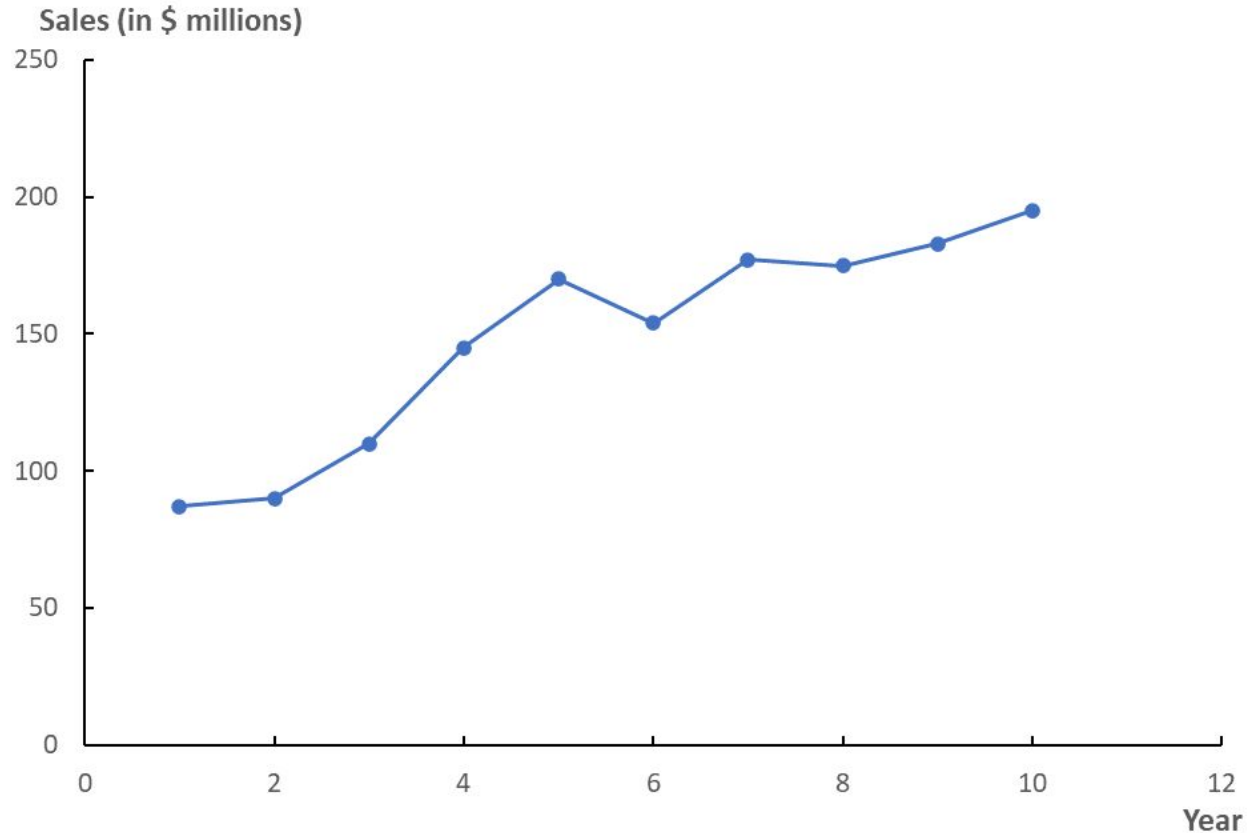
Step 2. Click on **Select Data...**

Step 3. When the **Select Data Source** dialog box appears, click the **Edit** button under **Legend Entries (Series)**

Step 4. Enter the location of the data you want to correspond to the horizontal values in the **Series X values:** box.
Note: do not include column headers.

Step 5. Repeat Step 4 for the **Series Y Values:** box and **Series bubble size:** box and click **OK**

2.4 Line Charts



A **line chart** is like scatter chart with a line connecting the points.

Line charts are handy for time series data because the addition of lines connecting the points suggests continuity and makes it easier for the viewer to see and interpret changes over time.

2.4 Build a line Chart

	A	B
1	Year	Sales (in \$ millions)
2	1	87
3	2	90
4	3	110
5	4	145
6	5	170
7	6	154
8	7	177
9	8	175
10	9	183
11	10	195

The following steps show how to create a line chart in Excel using the data in the file *Cheetah*.

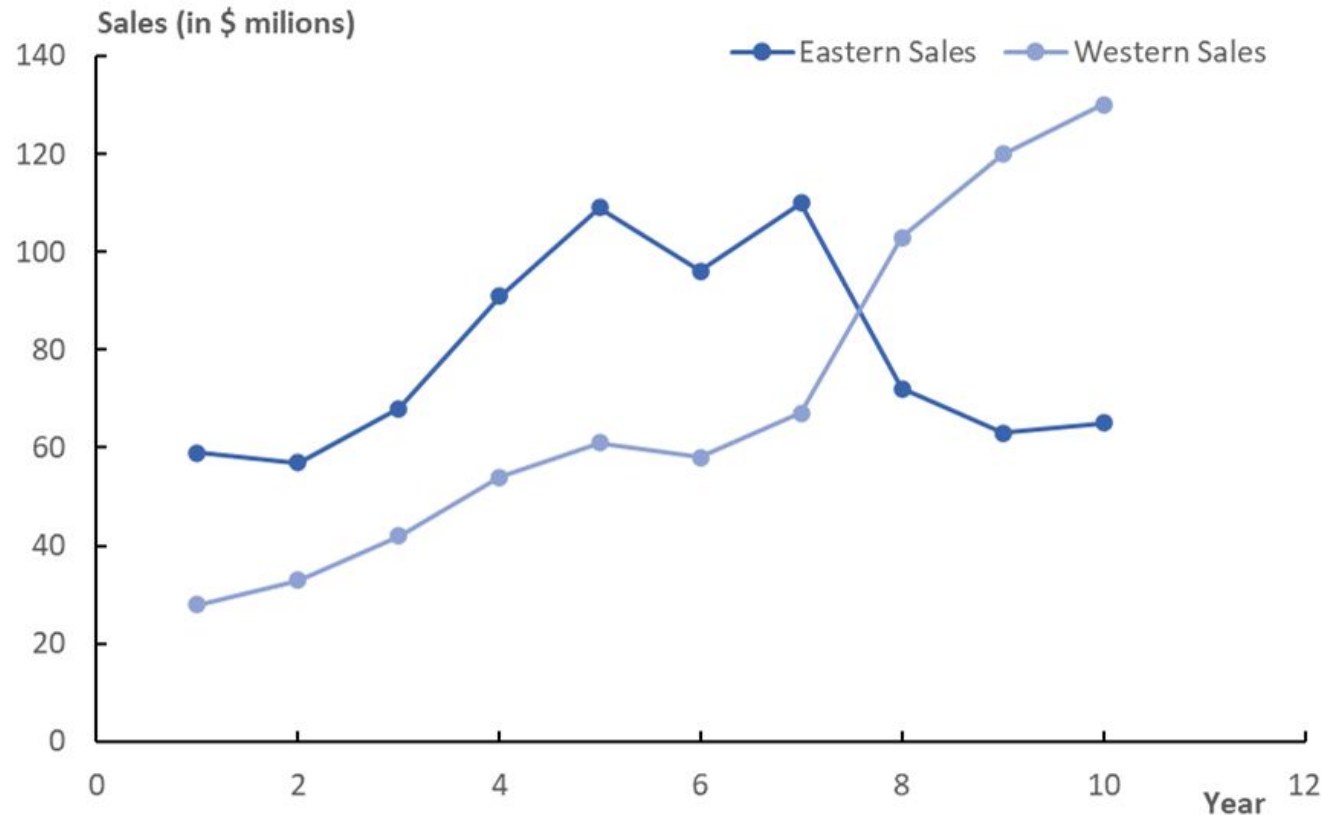
Step 1. Select cells A1:B11

Step 2. Click the **Insert** tab on the Ribbon

Step 3. Click the **Insert Scatter (X,Y) or Bubble Chart** button in the **Charts** group

Step 4. Select **Scatter with Straight Lines and Markers**. Then edit the chart as described in Section 2.2

2.4 Multiple-Line Charts



Sales in the western region have increased over the last three of the given ten-year period.

In contrast, sales in the eastern region have dropped significantly since year seven.

2.4 Build a Multiple-Line Chart

	A	B	C	D
1	Year	Eastern Sales	Western Sales	Total Sales (in \$ millions)
2	1	59	28	87
3	2	57	33	90
4	3	68	42	110
5	4	91	54	145
6	5	109	61	170
7	6	96	58	154
8	7	110	67	177
9	8	72	103	175
10	9	63	120	183
11	10	65	130	195

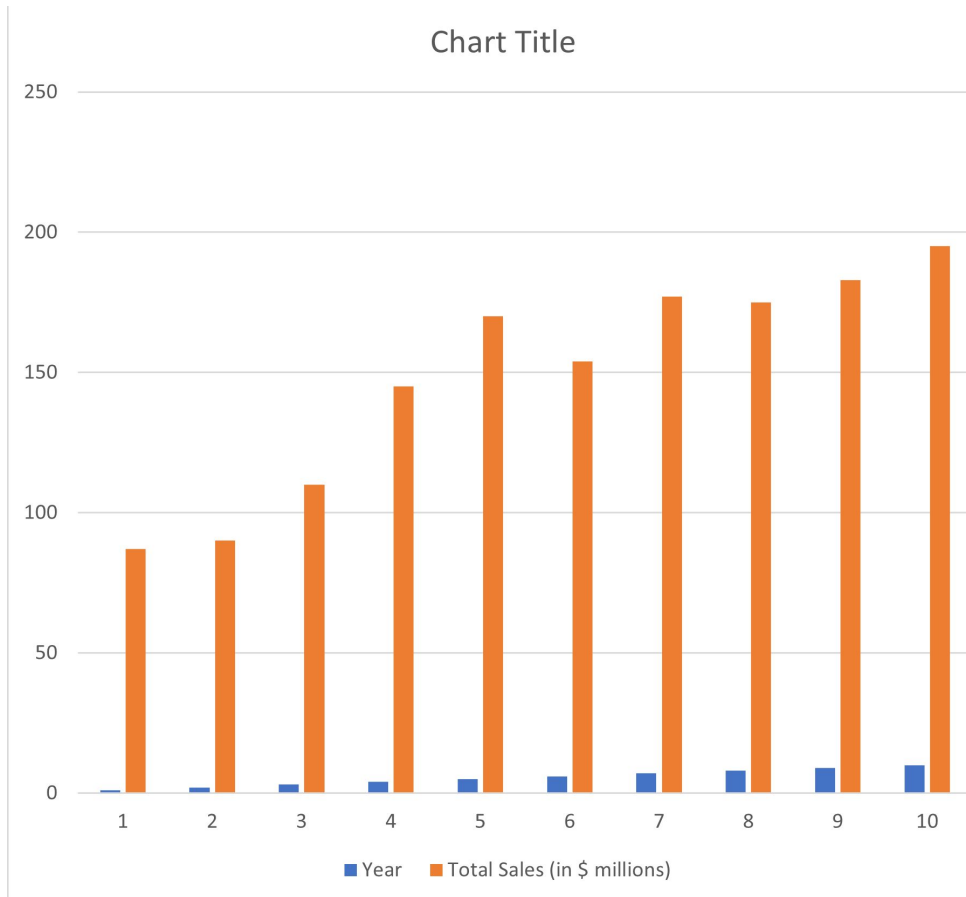
The following steps show how to create a multiple-line chart in Excel using the data in the file *CheetahRegion*.

Step 1. Select cells A1:C11 (do not select cells D1:D11)

Step 2. Repeat Steps 2-4 shown before to create a line chart.

Step 3. Select **Chart Style**. Click the **Color** tab and select the **Monochromatic Palette 1** option.

2.4 Column Charts



The following steps create the column chart of the *Cheetah* Sports sales data used before building a column chart.

Step 1. Select cells A1:B11

Step 2. Click the **Insert** tab on the Ribbon.

Step 3. In the **Charts** group, click the **Insert Column or Bar Chart** button

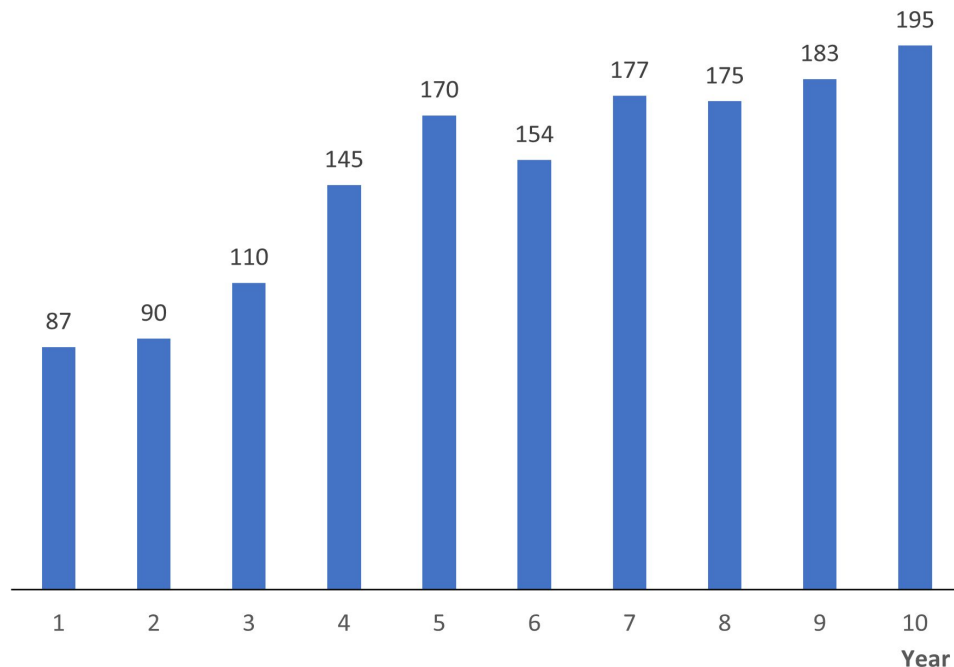
Step 4. Select **Clustered Column**

As you can see, the chart is incorrectly rendered because Excel displays the year as if it were a quantitative variable.

We need to correct the chart as shown in the following slide.

2.4 Edit a Column Chart

Total Sales (in \$ millions)



To fix the chart so that it displays the correct number of quantitative variables – in this case, one – we need to follow these additional steps:

Step 5. Right-click on the chart and select **Change Chart Type...**

Step 6. When the **Change Chart Type** task pane opens, select the **Cluster Column** type that plots the single variable Sales with ten monochromatic columns, and click **OK**

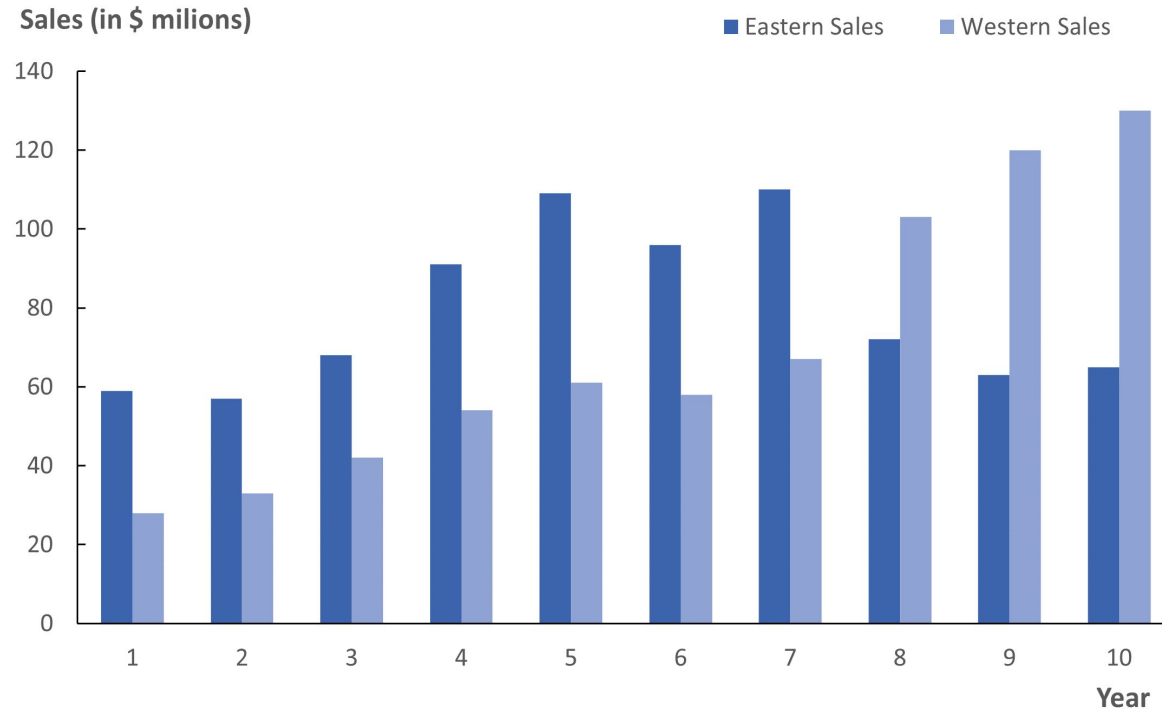
Then edit the chart as outlined in Section 2.2.

The next step adds data labels to the bars:

Step 7. Click on the **Chart Elements** button and select **Data Labels**.

*See additional comments in the notes.

2.4 Clustered Column Charts



The following steps create the clustered column chart of the *CheetahRegion* sports sales data previously used for a multiple line chart.

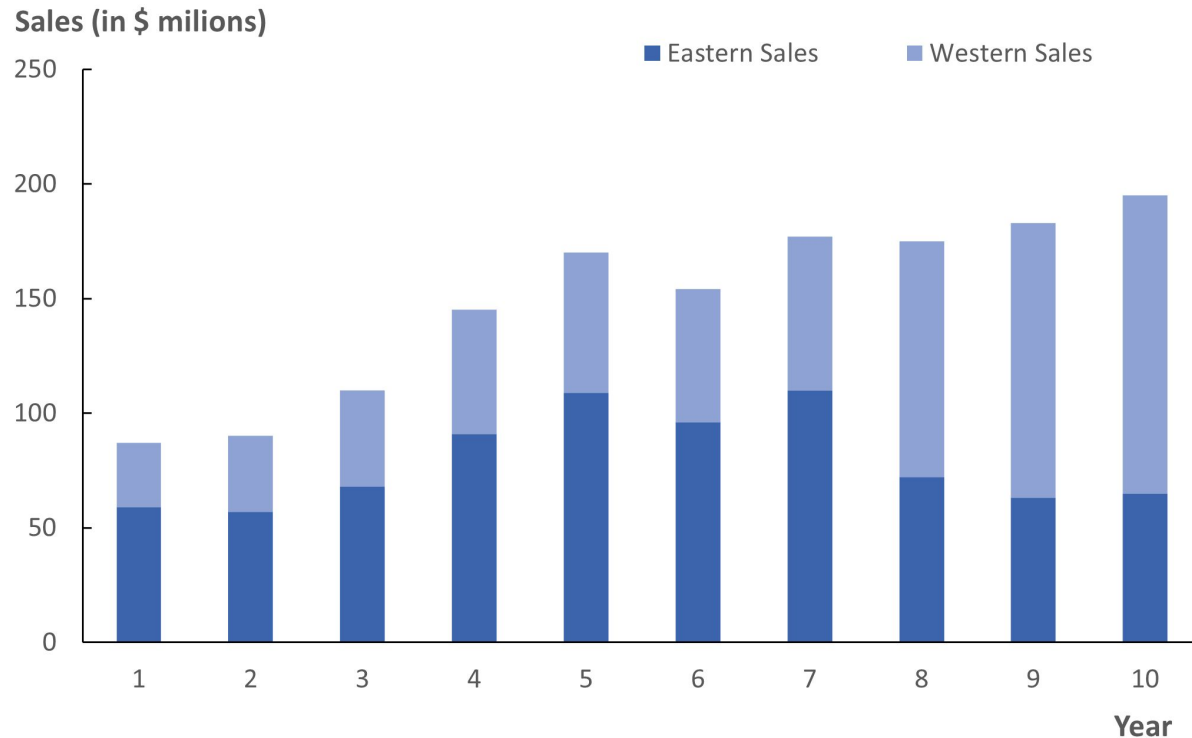
Step 1. Select cells A1:C11 (do not select cells D1:D11)

Step 2. Follow Steps 2-6 previously outlined for a column chart, including the change of color scheme to **Monochromatic Palette 1** (using the **Chart Styles** option)

The multiple line chart option is less cluttered than the clustered column chart shown here.

The best option to convey how total sales are changing is the stacked column chart.

2.4 Stacked Column Charts



The following steps create the stacked column chart of the *CheetahRegion* sports sales data previously used to build a clustered column chart.

Step 1. Select cells A1:C11 (do not select cells D1:D11)

Step 2. Follow Steps 2-6 previously outlined for a column chart, except that in Step 4 we click on the **Insert Column or Bar Chart** in the **Charts** group and select **Stacked Column**

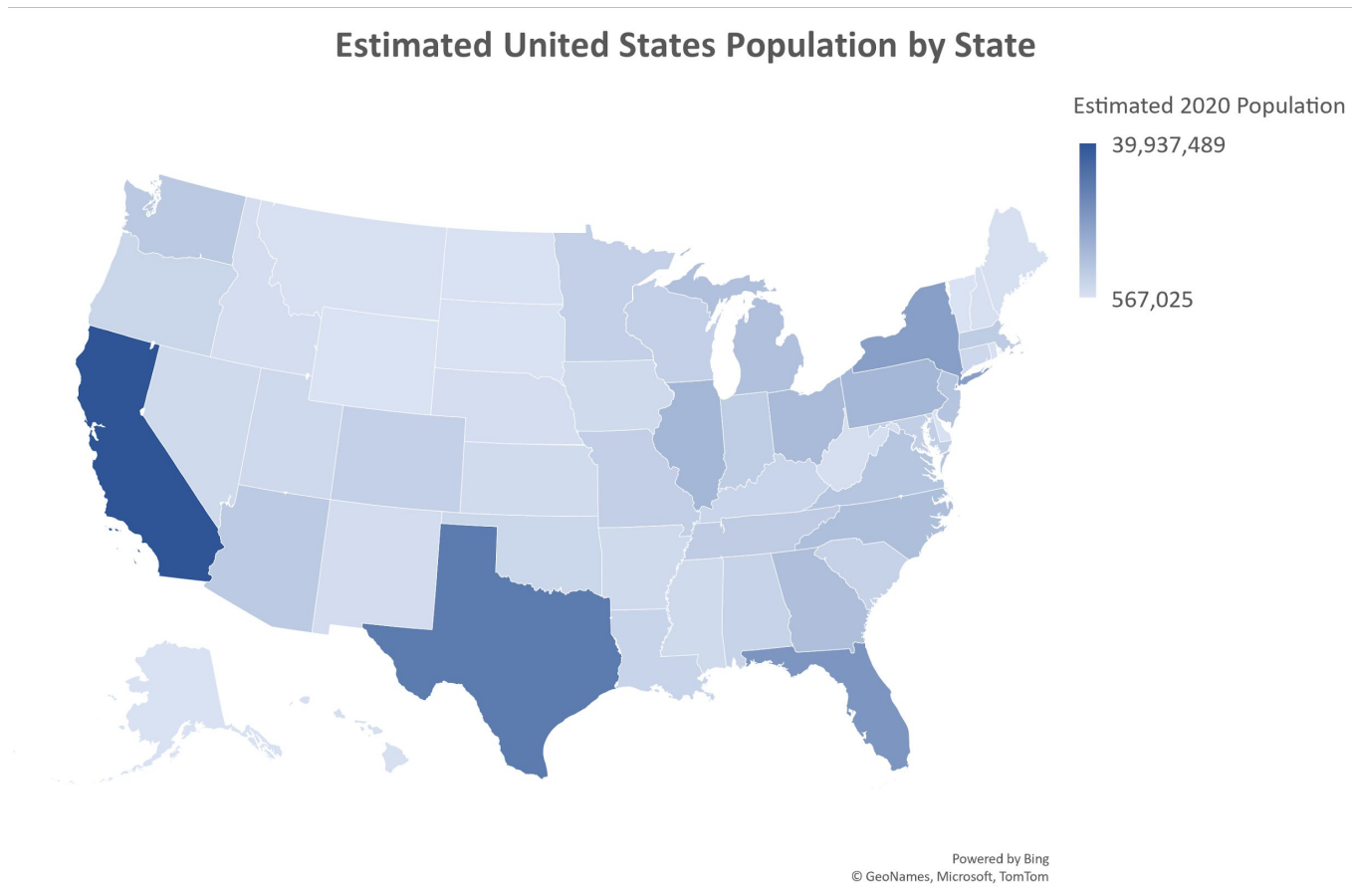
This chart shows the combination of eastern and western sales by year and the total height of the column indicates total sales.

*See additional comments in the notes.

2.4 Bar Charts

- All column charts, clustered and stacked alike, can also be represented horizontally as bar charts.
- A bar chart may be preferable over the corresponding column chart when there are lengthy category names that are easier to display horizontally.
- For time series data, a column chart provides for a more natural display of the passage of time horizontally, from left to right.
- To build a bar chart, follow the same steps used for a column chart earlier in this section, but on step 4 select **Clustered Bar** or **Stacked Bar** instead.

2.5 Choropleth Maps



- Choropleth maps can be built for quantitative or categorical data.
- Color shading is used to denote the population of each state, a quantitative variable.
- A darker shade will indicate a higher population and a lighter shade a lower population.
- The map shows that the states with the highest population are California, Texas, Florida, and New York.
- We can also build choropleth map for categorical data.

2.5 Build a Choropleth Map

A choropleth map is a geographic map that uses colors or symbols to indicate aggregate measures or statistics by geographic region.

The following steps show how to create a choropleth map in Excel using the data in the file *StatePopulation*.

Step 1. Select cells A1:B51

Step 2. Click the **Insert** tab on the Ribbon

Step 3. In the **Charts** group, click the **Maps** button

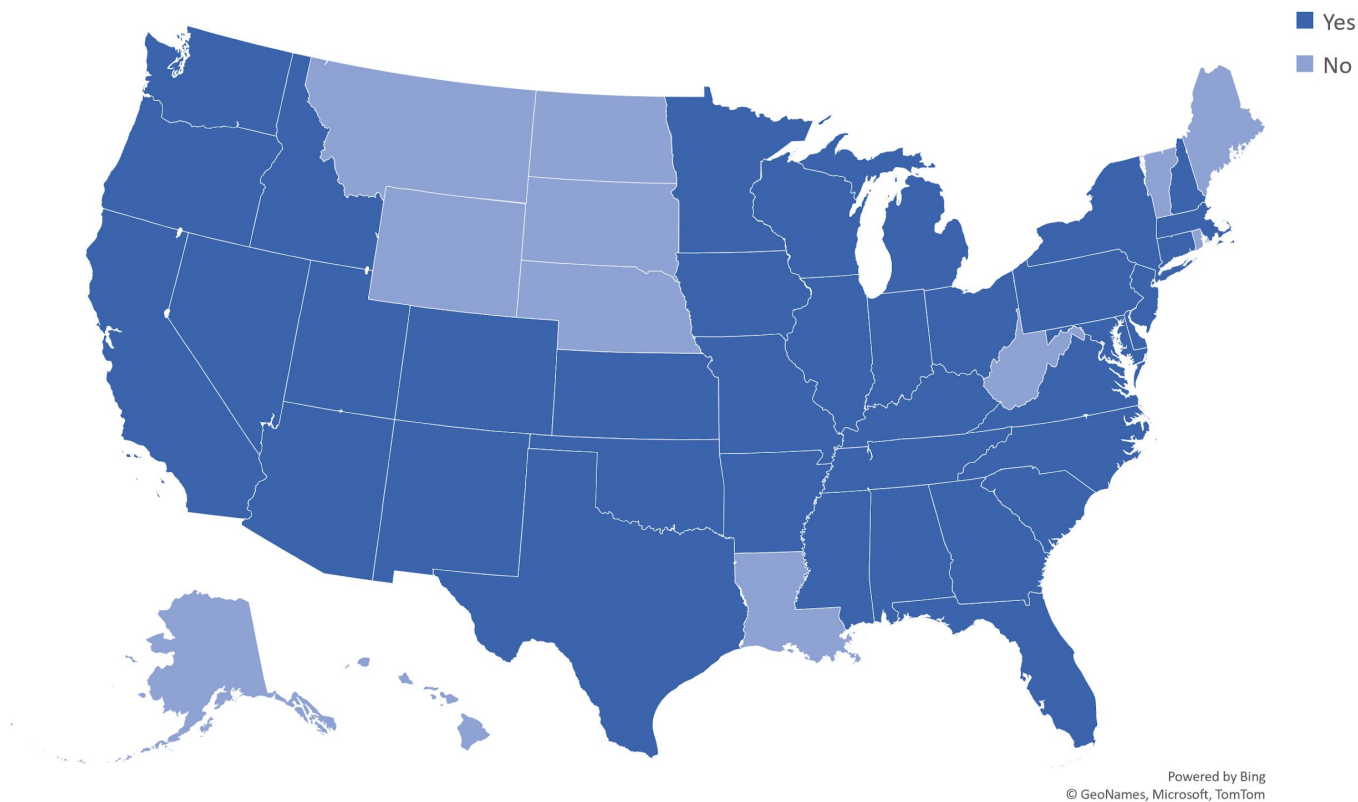
Step 4. Select **Filled Map**

Then edit the chart as outlined in Section 2.2.

	A	B
1	State	Estimated 2020 Population
2	California	39,937,489
3	Texas	29,472,295
4	Florida	21,992,985
5	New York	19,440,469
6	Pennsylvania	12,820,878
7	Illinois	12,659,682
8	Ohio	11,747,694
9	Georgia	10,736,059
10	North Carolina	10,611,862

2.5 Choropleth Maps for Categorical Data

Which States Have At Least One Amazon Fulfillment Center?



- This choropleth map was created in Excel using categorical data from the file *AmazonFulfill* and following the steps used in the previous example.
- The map shows that Amazon has at least one fulfillment center in 38 states.
- States without a fulfillment center tend to be either relatively sparsely populated or a geographic outlier.
- Amazon has many fulfillment centers to ensure quick customer delivery times for many of the products it sells.

2.5 Heat Maps

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2	St. Louis	-2%	-1%	-1%	0%	2%	4%	3%	5%	6%	7%	8%	8%
3	Phoenix	5%	4%	4%	2%	2%	-2%	-5%	-8%	-6%	-5%	-7%	-8%
4	Albany	-5%	-6%	-4%	-5%	-2%	-5%	-5%	-3%	-1%	-2%	-1%	-2%
5	Austin	16%	15%	15%	16%	18%	17%	14%	15%	16%	19%	18%	16%
6	Cincinnati	-9%	-6%	-7%	-3%	3%	6%	8%	11%	10%	11%	13%	11%
7	San Francisco	2%	4%	5%	8%	4%	2%	4%	3%	1%	-1%	1%	2%
8	Seattle	7%	7%	8%	7%	5%	4%	2%	0%	-2%	-4%	-6%	-5%
9	Chicago	5%	3%	2%	6%	8%	7%	8%	5%	8%	10%	9%	8%
10	Atlanta	12%	14%	13%	17%	12%	11%	8%	7%	7%	8%	5%	3%
11	Miami	2%	3%	0%	1%	-1%	-4%	-6%	-8%	-11%	-13%	-11%	-10%
12	Minneapolis	-6%	-6%	-8%	-5%	-6%	-5%	-5%	-7%	-5%	-2%	-1%	-2%
13	Denver	5%	4%	1%	1%	2%	3%	1%	-1%	0%	1%	2%	3%
14	Salt Lake City	7%	7%	7%	13%	12%	8%	5%	9%	10%	9%	7%	6%
15	Raleigh	4%	2%	0%	5%	4%	3%	5%	5%	9%	11%	8%	6%
16	Boston	-5%	-5%	-3%	4%	-5%	-4%	-3%	-1%	1%	2%	3%	5%
17	Pittsburgh	-6%	-6%	-4%	-5%	-3%	-3%	-1%	-2%	-2%	-1%	-2%	-1%

A heat map is a 2D graphical representation of a data table that uses shades of color to indicate magnitude.

The following steps show how to create a heat map in Excel using the data in the file *SameStoreSales*.

Step 1. Select cells B2:M17

Step 2. Click the **Home** tab on the Ribbon

Step 3. Click **Conditional Formatting** in the **Styles** group, select **Color Scales**, and click on **Blue-White-Red Color Scale**

2.5 Build a Treemap

	A	B	C
1	Industry	Brand	Value (in \$ Billions)
2	Technology	Apple	205.5
3	Technology	Google	167.7
4	Technology	Microsoft	125.3
5	Technology	Amazon	97.0
6	Technology	Facebook	88.9
7	Beverages	Coca-Cola	59.2
8	Technology	Samsung	53.1
9	Leisure	Disney	52.2
10	Automotive	Toyota	44.6
11	Restaurants	McDonald's	43.8

The following steps show how to create a treemap in Excel using the data in the file *BrandValues*.

Step 1. Select cells A1:C11

Step 2. Sort the data by Industry by using the following steps:

Click Data on the Ribbon

Click the **Sort** button in the **Sort & Filter** group
In the **Sort** dialog box, select **Industry** from the drop-down menu

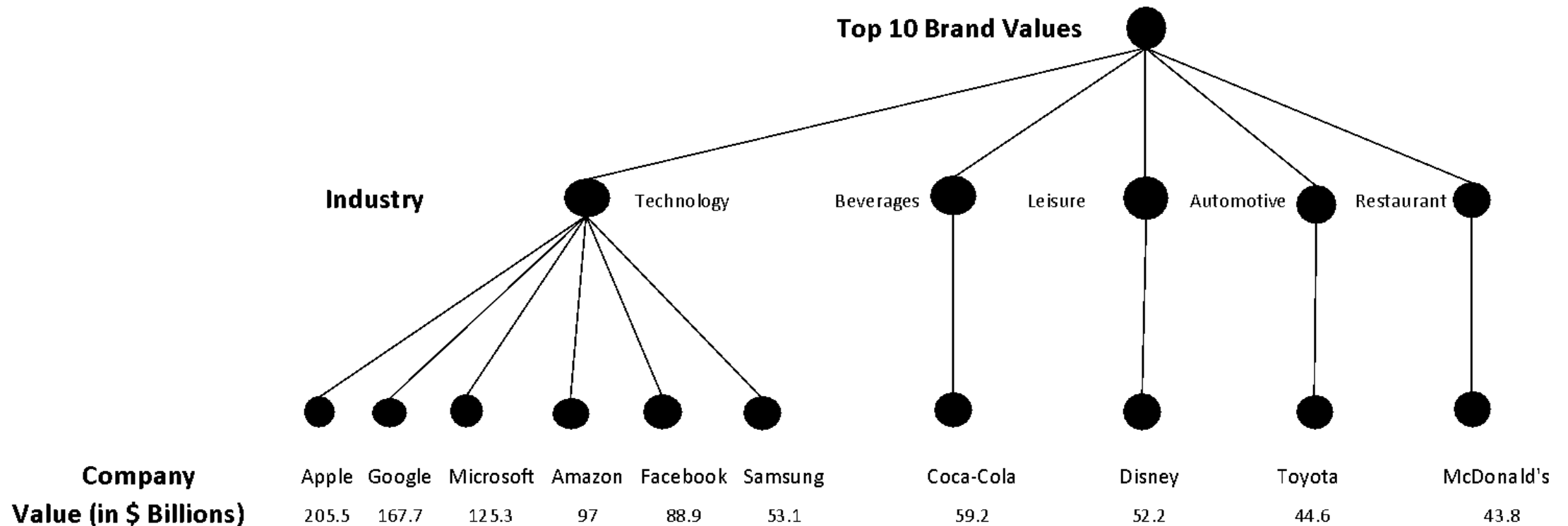
From the **Order** drop-down menu select **A to Z**

Step 3. Click **Insert** on the Ribbon

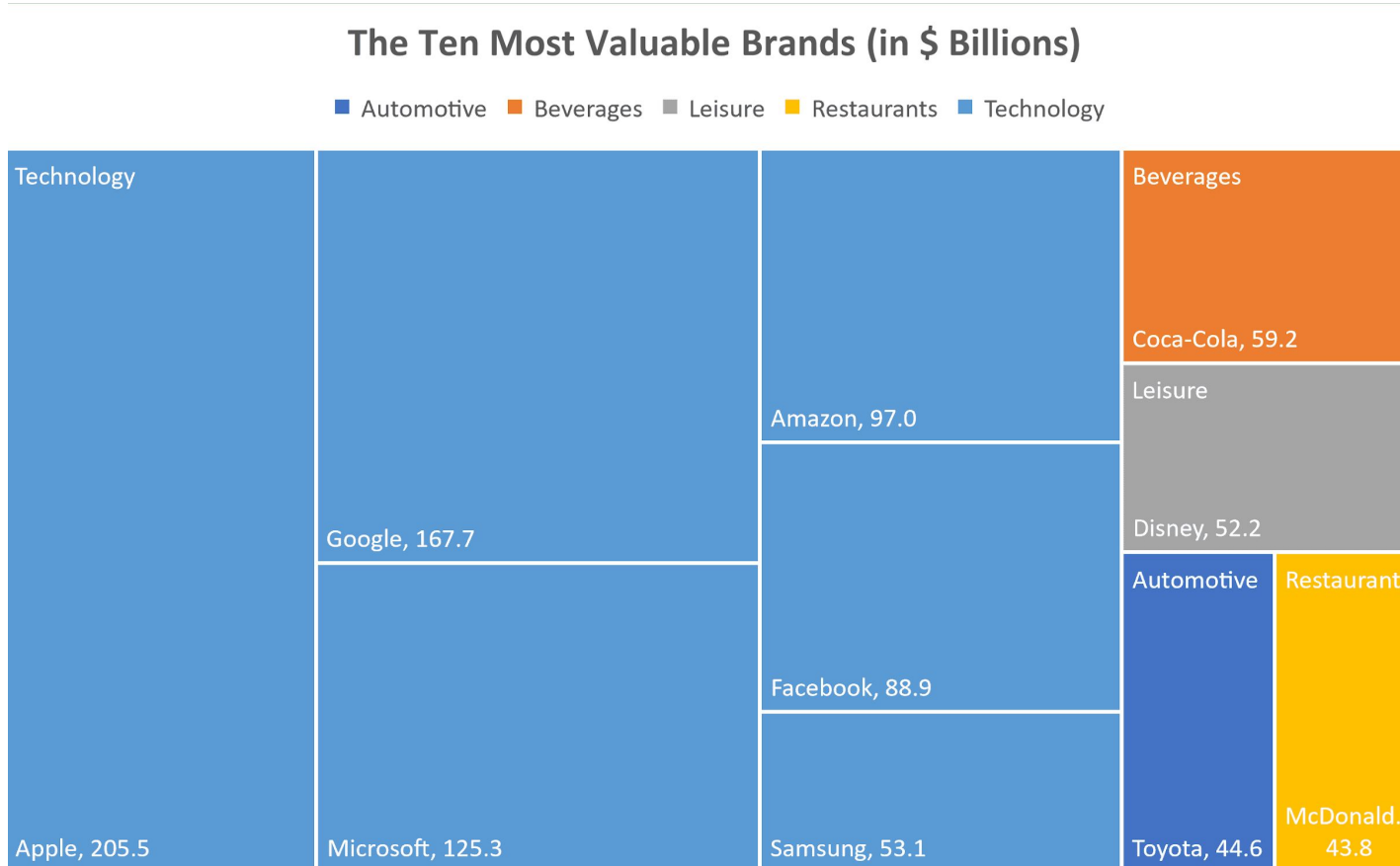
Step 4. Click on the **Insert Hierarchy Chart** button in the **Charts** group and select **Treemap** from the drop-down menu

2.5 Hierarchy of a Treemap

The Hierarchical Tree Structure of the Top Ten Brand Values Data



2.5 Edit a Treemap



A treemap is a chart that uses the size, color, and arrangement of rectangles to display the magnitudes of a quantitative variable for hierarchical data.

To display the brand values:

Step 5. Select any brand label and then right-click

Step 6. Select **Format Data Labels...** from the drop-down menu

Step 7. Select **Text Options**, **Label Options**, and select **Value**

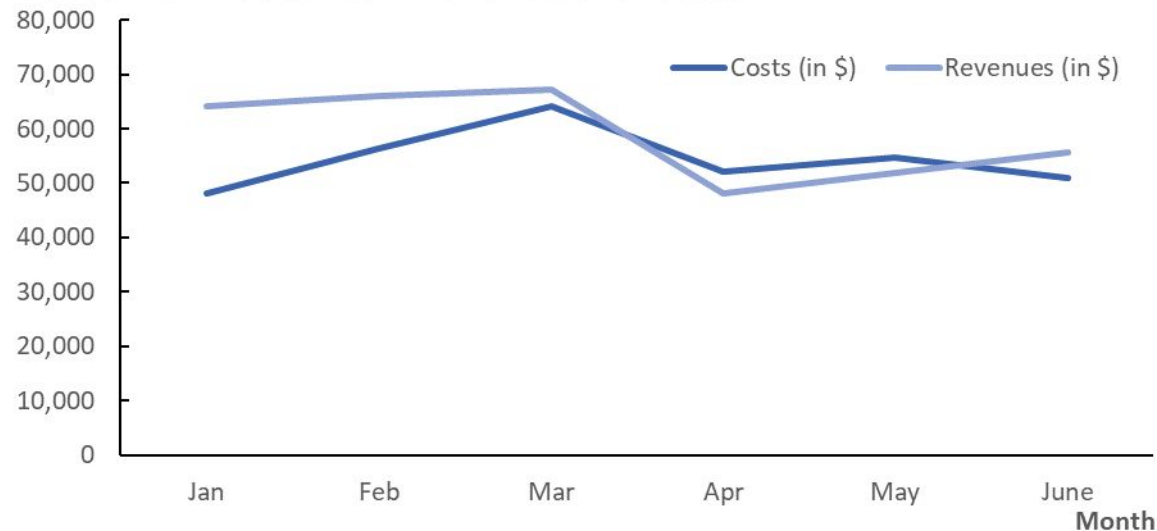
2.6 Tables versus Charts

Table Showing Exact Values for Costs and Revenues by Month for Gossamer Industries

	Month						Total
	Jan	Feb	Mar	Apr	May	June	
Costs (in \$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (in \$)	64,124	66,125	67,125	48,178	51,785	55,678	353,015

Line Chart of Monthly Costs and Revenues

Gossamer Industries: Revenue and Costs

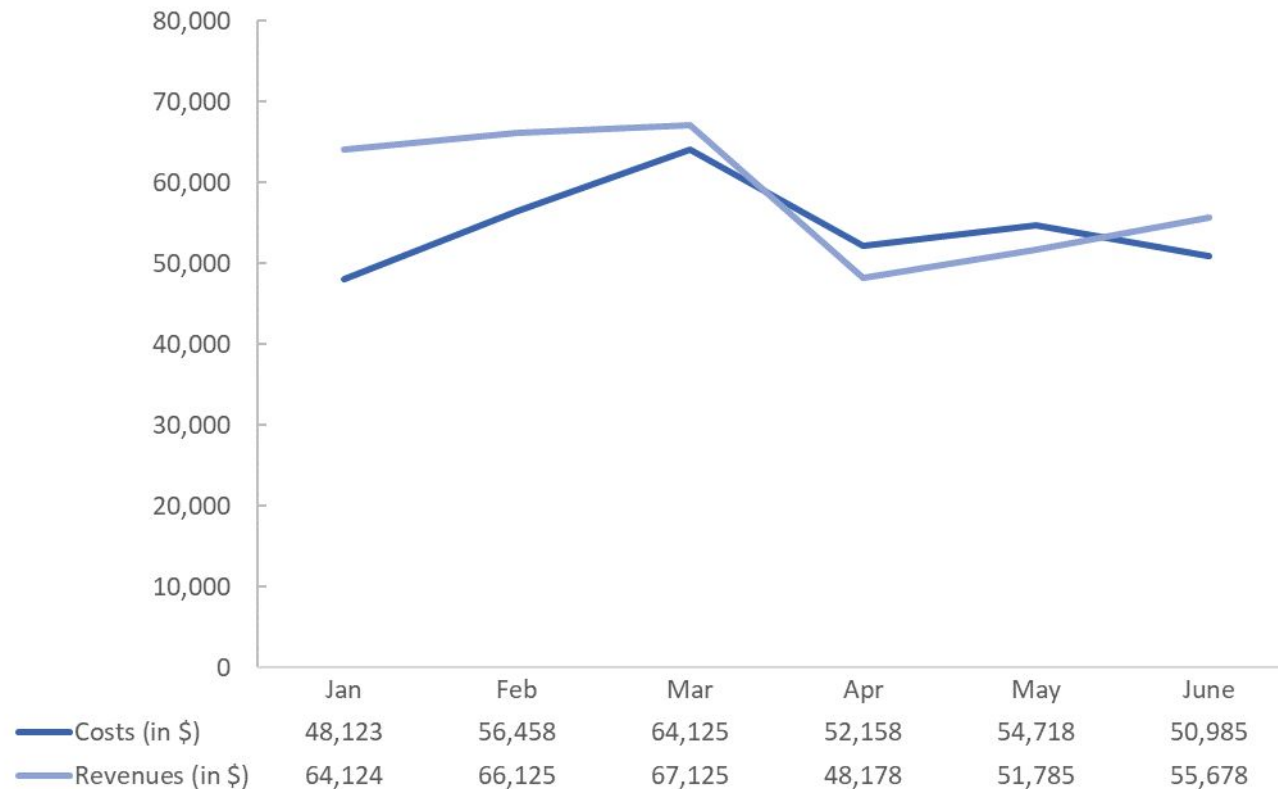


Tables should be used when:

1. The reader needs to refer to specific numerical values
2. The reader needs to make precise comparisons between different values and not just relative comparisons
3. The values being displayed have different units or very different magnitudes

2.6 Combined Line Charts and Tables

Gossamer Industries: Revenue and Costs



The following steps show how to create a combined line chart and table in Excel using the data in the file *Gossamer*.

Step 1. Select cells A2:G4

Step 2. Click the **Insert** tab on the Ribbon

Step 3. Click the **Insert Line or Area Chart** button in the **Charts** group

Step 4. When the list of column and bar charts subtypes appears, click the **Line** button

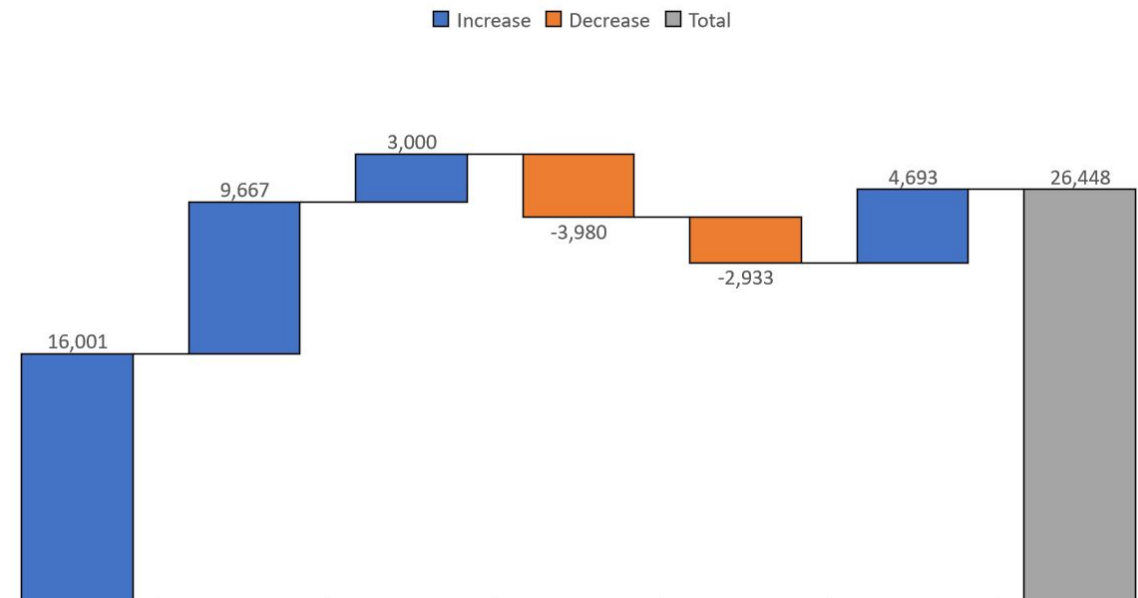
Step 5. Click anywhere on the chart
Click the **Chart Elements** button
Select the check box for **Data Table**, and edit the chart as outlined in Section 2.2.

2.7 Waterfall Charts

GossamerGP data file for Gossamer costs, revenues, and gross profit

1				Month				
2		Jan	Feb	Mar	Apr	May	June	Total
3	Costs (in \$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
4	Revenues (in \$)	64,124	66,125	67,125	48,178	51,785	55,678	353,015
5	Gross Profit (in \$)	16,001	9,667	3,000	-3,980	-2,933	4,693	26,448

A waterfall chart for the gossamer gross profit data



*See the notes for step-by-step instructions on how to build a waterfall chart.

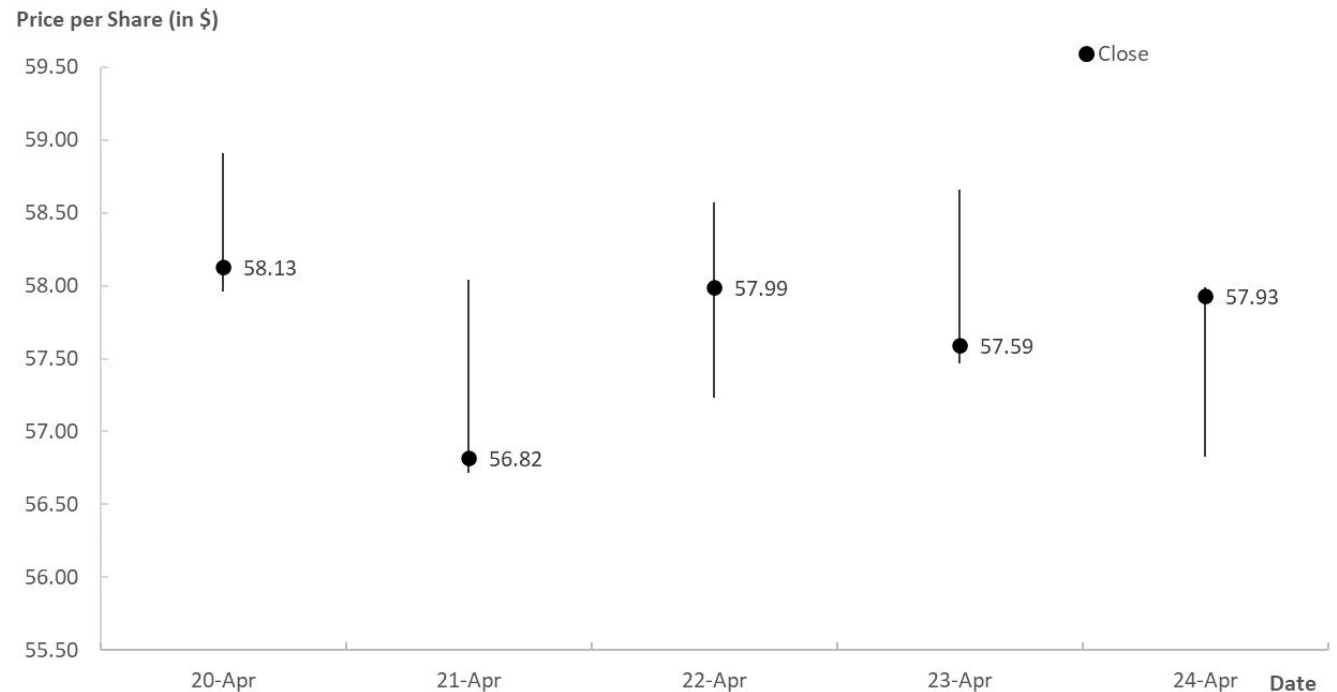
2.7 Stock Charts

Verizon stock price data file

	A	B	C	D	E
1	Date	Open	High	Low	Close
2	20-Apr	58.10	58.91	57.96	58.13
3	21-Apr	57.39	58.04	56.72	56.82
4	22-Apr	57.41	58.57	57.23	57.99
5	23-Apr	58.12	58.66	57.47	57.59
6	24-Apr	57.64	57.99	56.83	57.93

*See the notes for step-by-step instructions on how to build a high-low-close stock chart.

A high-low-close stock chart for Verizon Wireless

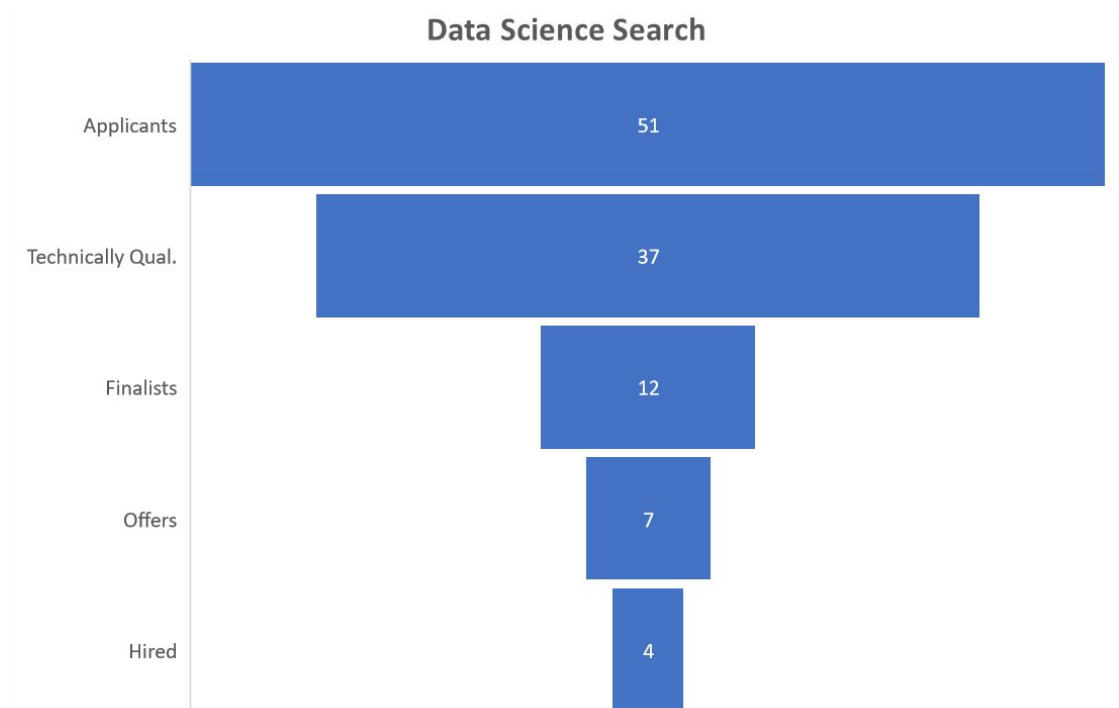


2.7 Funnel Charts

DataScienceSearch data for hiring

	A	B
1	Stage	Number of Applicants
2	Applicants	51
3	Technically Qualified	37
4	Finalists	12
5	Offers	7
6	Hired	4

A funnel chart for Data Scientist hiring



The following steps may be used to create the funnel chart shown for Data Scientist hiring:

Step 1. Select cells A1:B6.

Step 2. Click the **Insert** tab on the Ribbon

Step 3. Click the **Insert Waterfall, Funnel, Stock, Surface or Radar Chart** button in the **Charts** group

Step 4. When the list of subtypes appears, click the **Funnel Chart** button

Then edit the chart using steps outlined in Section 2.2.

2.8 Guidelines for Selecting a Chart

Chart	Relationship	Distribution	Composition	Ranking
Scatter	✓	✓		
Bubble	✓	✓		
Line	✓			
Stock	✓			
Column	✓	✓		✓
Bar	✓	✓	✓	✓
Heat Map	✓			
Choropleth Map		✓		
Stacked Bar			✓	
Stacked Column			✓	
Treemap			✓	
Waterfall			✓	
Funnel				✓

2.8 Charts to Avoid

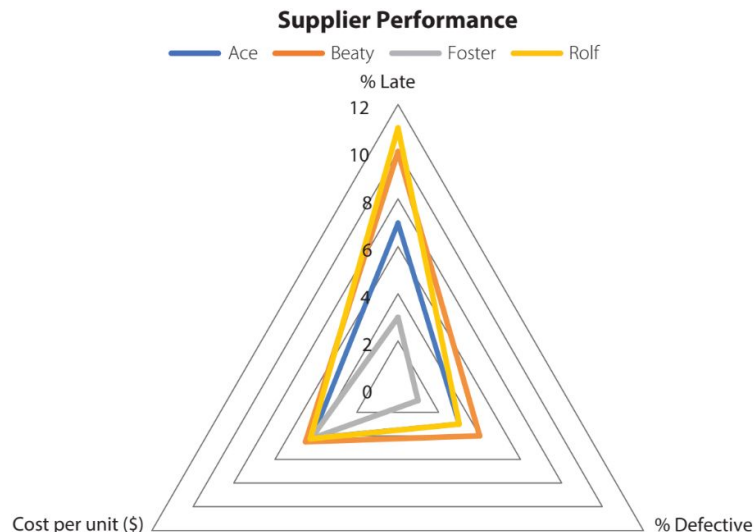
- Here is a summary of some types of charts data visualization experts agree should be avoided:
 - To visualize the contribution of each category to a whole entity, use a pie chart in which the slices represent each category contribution.
 - When there are more than a few categories, pie charts tend to be cluttered and hard to interpret. Consider us a bar chart instead.
 - Science has shown that we are better at determining differences in length than angle and area. The bar chart makes it much easier to tell which category has the larger allocation.
 - Color should only be used to distinguish between categories when necessary. For the bar chart, color is not needed, as the length of the bars is used to quantify allocation by category.
 - In a horizontal bar chart, the category labels, a few of which are rather lengthy, are easier to read.
 - Avoid combo charts and 3D representations of a 2D chart.

2.8 Charts to Avoid – Radar Chart

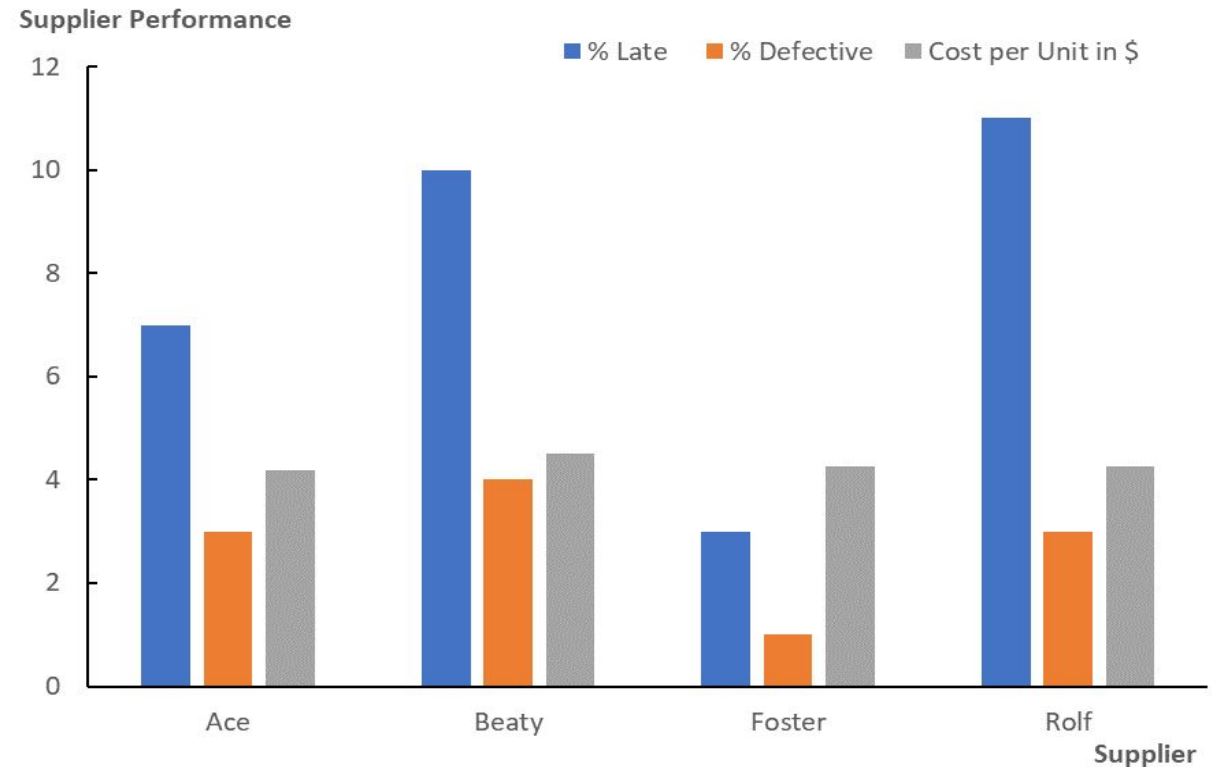
NewtonSupplier data file for a component performance

	A	B	C	D
1	Supplier	% Late	% Defective	Cost per Unit in \$
2	Ace	7	3	\$4.19
3	Beaty	10	4	\$4.50
4	Foster	3	1	\$4.25
5	Rolf	11	3	\$4.26

A radar chart of supplier performance for a component for Newton Industries

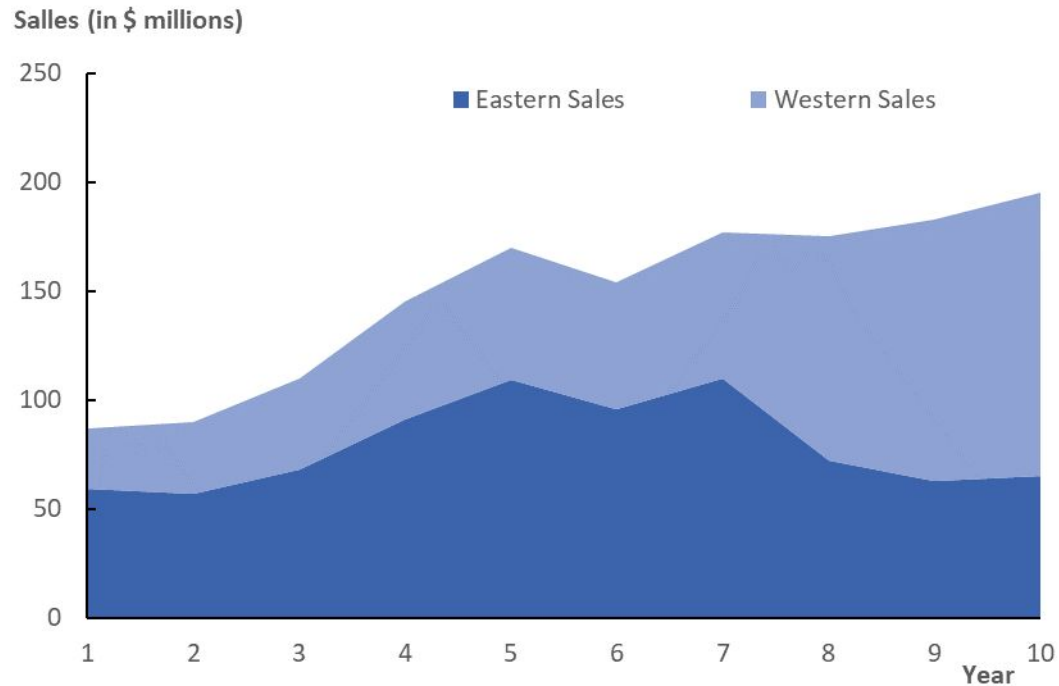


Better choice: a clustered column chart of supplier performance for a component for Newton Industries

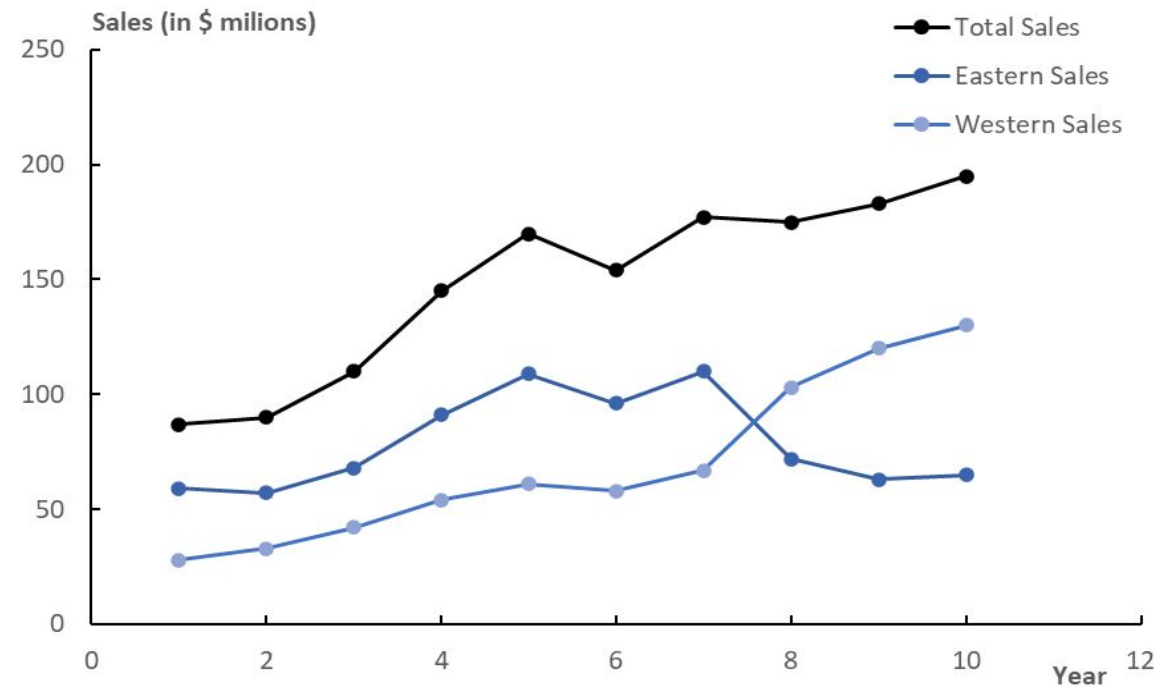


2.8 Charts to Avoid – Area Chart

An area chart for the Cheetah Sports regional sales data

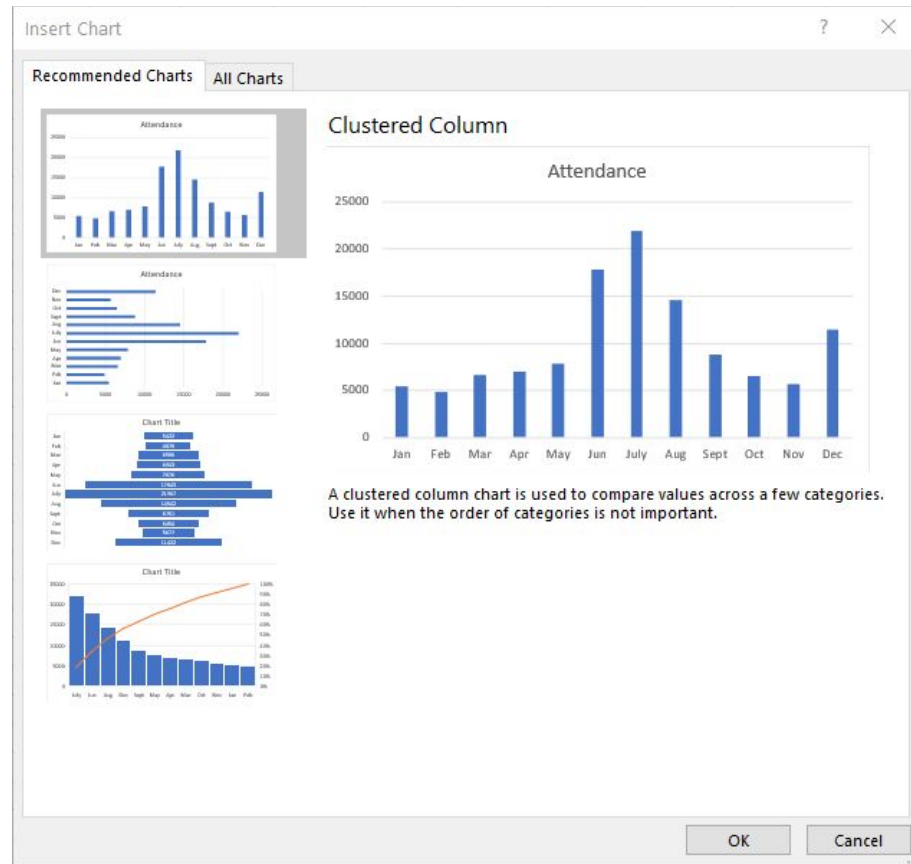


Better choice: a multiple line chart for the Cheetah Sports regional sales data



2.8 Excel's Recommended Charts Tool

The Insert Chart task pane for the Zoo Attendance data



The following steps show how to use the Recommended Charts tool using data in the file Zoo.

Step 1. Select cells A1:B13

Step 2. Click the Insert tab on the Ribbon

Step 3. Click the **Recommended Charts** button in the **Charts** group

Step 4. When the **Insert Chart** task pane appears, select any of the four **Recommended Charts** to the left.

Step 5. Select the Clustered Column chart and click OK. Then edit the chart as outlined in slide 6.

*See additional instructions in the notes.

Check Your Knowledge

1. Which of these is an appropriate alternative to a radar chart?
 - a. Choropleth
 - b. 3D bubble chart
 - c. Treemap
 - d. Clustered column chart

2. To achieve the colored and shading effect for a heat map in Excel, use the options in the _____ menu.
 - a. Format Data Labels
 - b. Chart Options
 - c. Conditional Formatting
 - d. Insert Chart

Summary

In this chapter, you should have learned how to:

- Select a suitable chart, given the goals of the analysis and the type of data.
- Create and edit through step-by-step instructions the following charts in Excel:
 - Relationship charts, such as the scatter chart, bubble chart, and line chart.
 - Composition charts, such as clustered/stacked column and bar charts.
 - Maps, such as choropleth geographical maps, heat maps, and treemaps.
 - Specialized charts, such as waterfall, stock, and funnel charts.
 - Charts to avoid, such as radar and area charts.