

Data Visualization

Exploring and Explaining with Data



Data Visualization, 1e

Chapter 3: Data Visualization and Design

Chapter Objectives (1 of 2)

After completing this chapter, you will be able to:

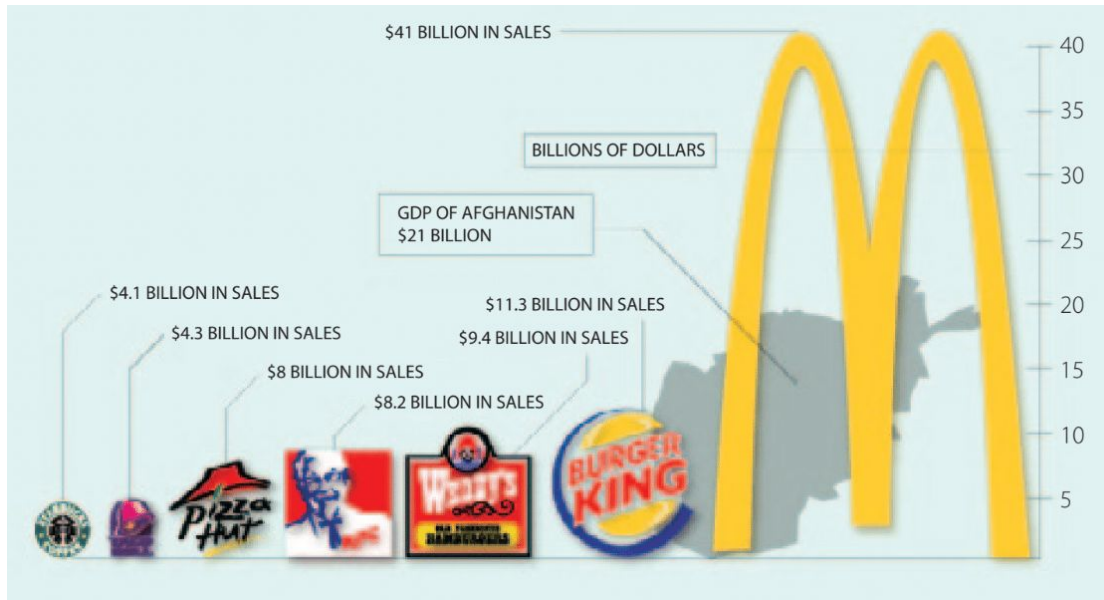
- LO 3.1a Define *preattentive attributes*
- LO 3.1b Explain how preattentive attributes associated with color, form, spatial positioning, and movement are used in data visualizations
- LO 3.2 Explain how the Gestalt principles of similarity, proximity, enclosure, and connection can be used to create effective data visualizations
- LO 3.3a Define *data-ink ratio*

Chapter Objectives (2 of 2)

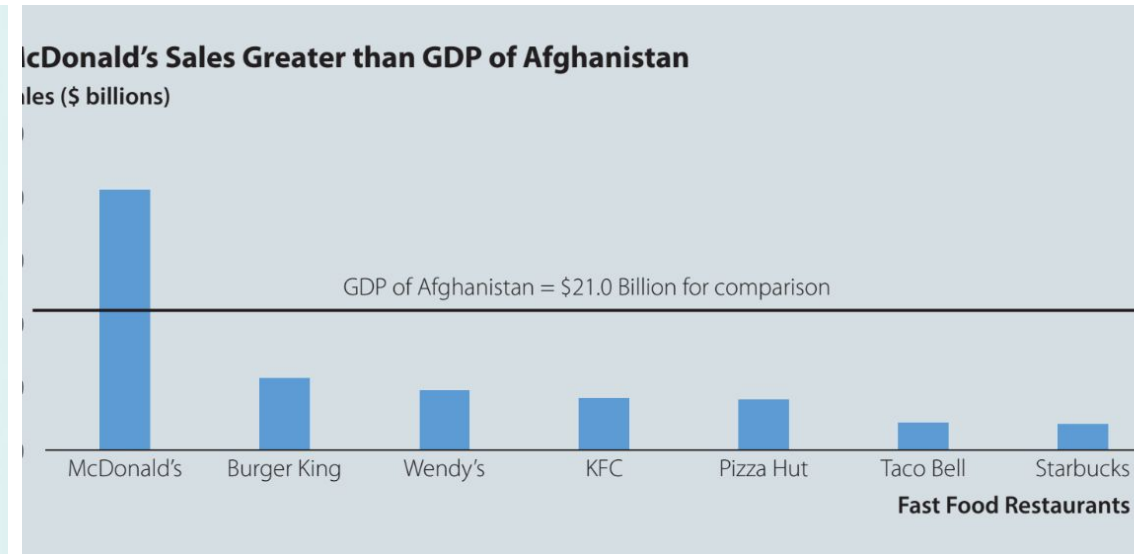
- LO 3.3b Explain how increasing the data-ink ratio through decluttering can create data visualizations that are easier to interpret
- LO 3.4 Create data visualizations that are easier to interpret by minimizing the required eye travel and applying the concepts of preattentive attributes, Gestalt principles, and decluttering
- LO 3.5 Explain why certain types of font are preferred over others for use of text in data visualizations
- LO 3.6 List several common mistakes in designing data visualizations and explain how these mistakes can be avoided

Data Visualization Makeover

Column chart using logos to compare fast-food restaurant sales



Improved column chart to compare fast-food restaurant sales



(Source: <http://www.princeton.edu/~ina/infographics/starbucks.html>)

*See additional comments in the notes.

3.1 Visual Perception and Preattentive Attributes

Visual perception is the brain process that interprets the reflections of the light entering our eyes.

Three forms of memory affect visual perception:

Iconic memory holds information for less than a second, and it is processed automatically.

Short-term memory is grouped in chunks of information and holds information for about a minute.

Long-term memory holds information for more extended periods through repetition, rehearsal, or story-telling.

Preattentive attributes are features processed by iconic memory.

3.1 Cognitive Load and Preattentive Attributes

Cognitive load is the effort needed to accurately and efficiently process information communicated by a data visualization.

- Proper use of preattentive attributes reduces the cognitive load in a data visualization.

The four preattentive attributes related to visual perception are:

- **Color**
- **Form**
- **Spatial positioning**
- **Movement**

3.1 Application of Preattentive Attributes

Count the number of 7s aided by the preattentive attributes of color and form (size)

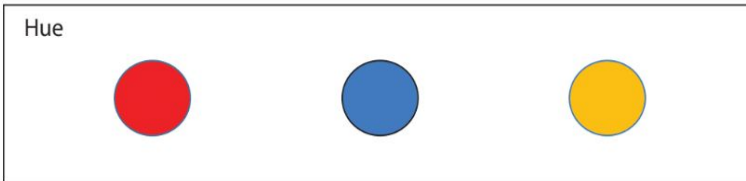
7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8

7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8

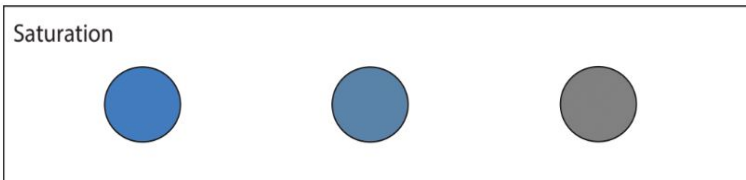
7	3	4	1	3	4	5	6	4	0
3	0	6	9	0	4	5	8	6	3
2	7	2	2	9	9	4	5	2	1
2	2	4	5	2	0	9	2	0	4
2	4	0	7	6	9	3	0	0	4
7	7	8	9	2	6	7	2	4	7
6	1	3	3	2	1	4	4	9	0
3	6	6	2	7	5	5	2	5	4
1	1	4	0	6	3	4	0	5	1
3	7	5	2	7	5	7	7	3	9
3	3	8	6	9	5	5	3	6	4
7	6	0	3	0	9	9	0	2	9
4	6	9	4	8	2	6	5	8	3
9	3	9	2	2	8	4	3	9	8
5	8	8	2	9	1	2	4	8	5
1	7	4	0	1	1	9	9	5	8

3.1 The Preattentive Attribute of Color

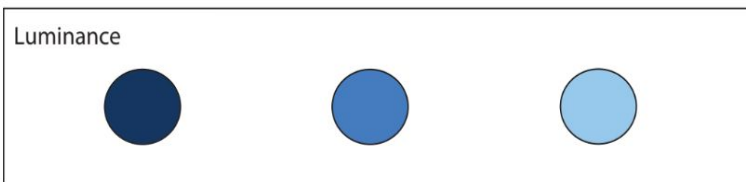
The preattentive attribute of color includes hue, saturation, and luminance



Hue is the position light occupies on the visible light spectrum. It is the basis of the different colors.



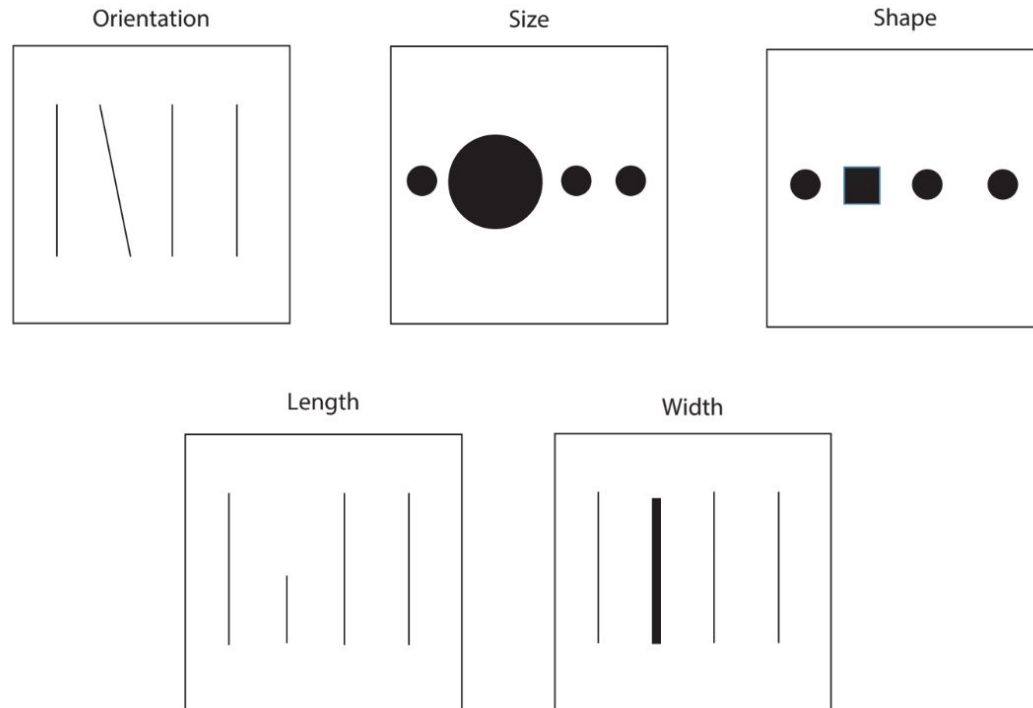
Saturation is the amount of gray present in color. It refers to the intensity or purity of the color.



Luminance refers to the amount of black versus white within the color.

3.1 The Preattentive Attributes of Form

The preattentive attributes of form include orientation, size, shape, length, and width



Orientation refers to the relative positioning of an object within a data visualization. It is common in line graphs.

Size refers to the relative amount of space an object occupies in a visualization. Most people struggle to estimate relative size differences.

Shape refers to the form of objects used in data visualization to distinguish between groups.

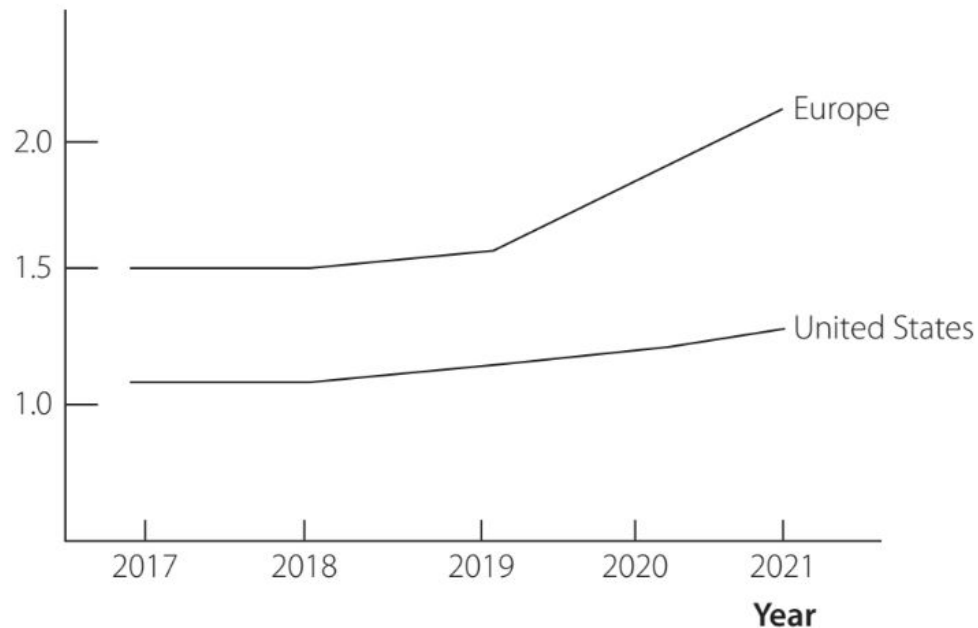
Length and **Width** refer to the distance and thickness of a line or bar/column, respectively.

3.1 The Preattentive Attribute of Orientation

Demonstration of the preattentive attribute of orientation

Sales of Insulin Syringes Increasing Faster in Europe than in United States

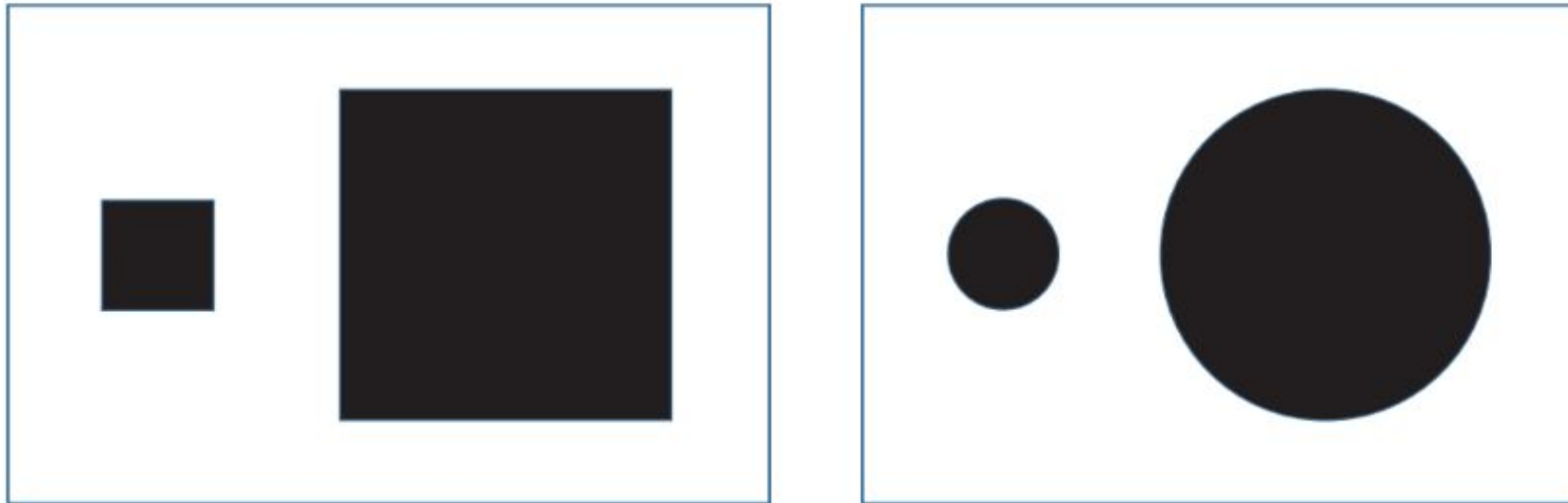
Sales (\$ millions)



The difference in the orientation of these lines makes it easy for the audience to perceive that sales in Europe are increasing at a much faster rate than the United States for the years 2019 and 2020.

3.1 The Preattentive Attribute of Size

Demonstration of the preattentive attribute of size

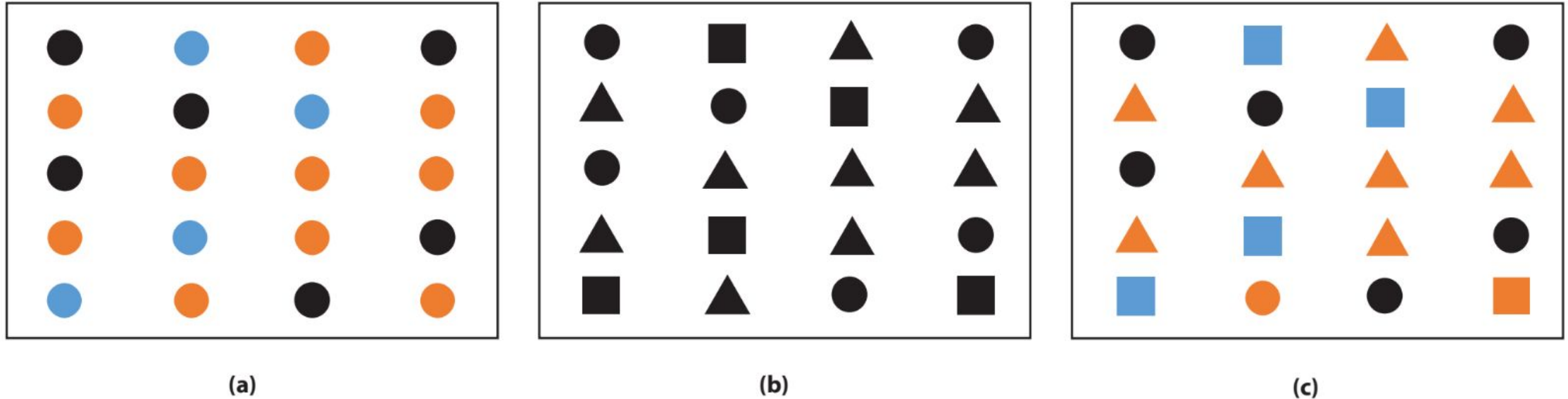


The heights and widths of both the bigger square and bigger circle are three times larger than their smaller counterparts, but their respective areas are nine times bigger.

Most people are not good at estimating relative size differences using the attribute of size.

3.1 The Preattentive Attribute of Shape

Grouping items by color and shape

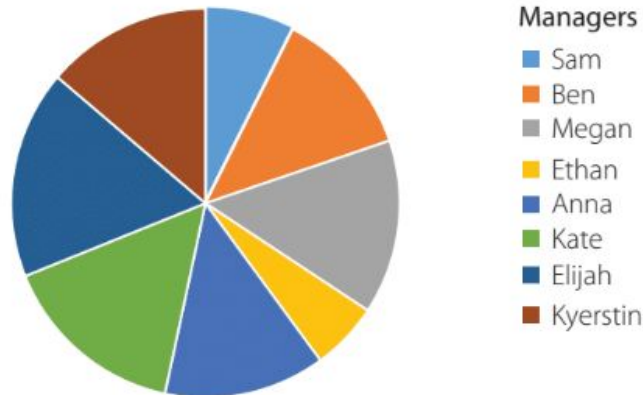


- The preattentive attribute of color divides the items into three different groups: orange, blue, and black.
- The preattentive attribute of shape divides the items into three groups: circle, square, and triangle.
- The combination of preattentive attributes of color and shape divides the items into nine groups, each with one of the three colors combined with one of the three shapes

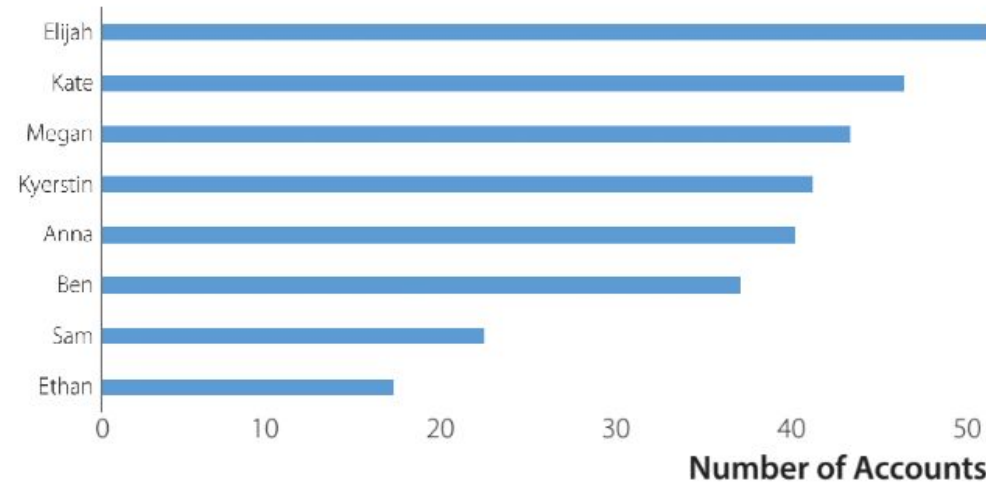
3.1 The Preattentive Attribute of Length

The bar chart (preattribute of length) makes it easier to display the accounts managed data than the pie chart (preattributes of size and color)

Comparison of Account Mangers
(by Number of Accounts)



Comparison of Account Managers
Managers

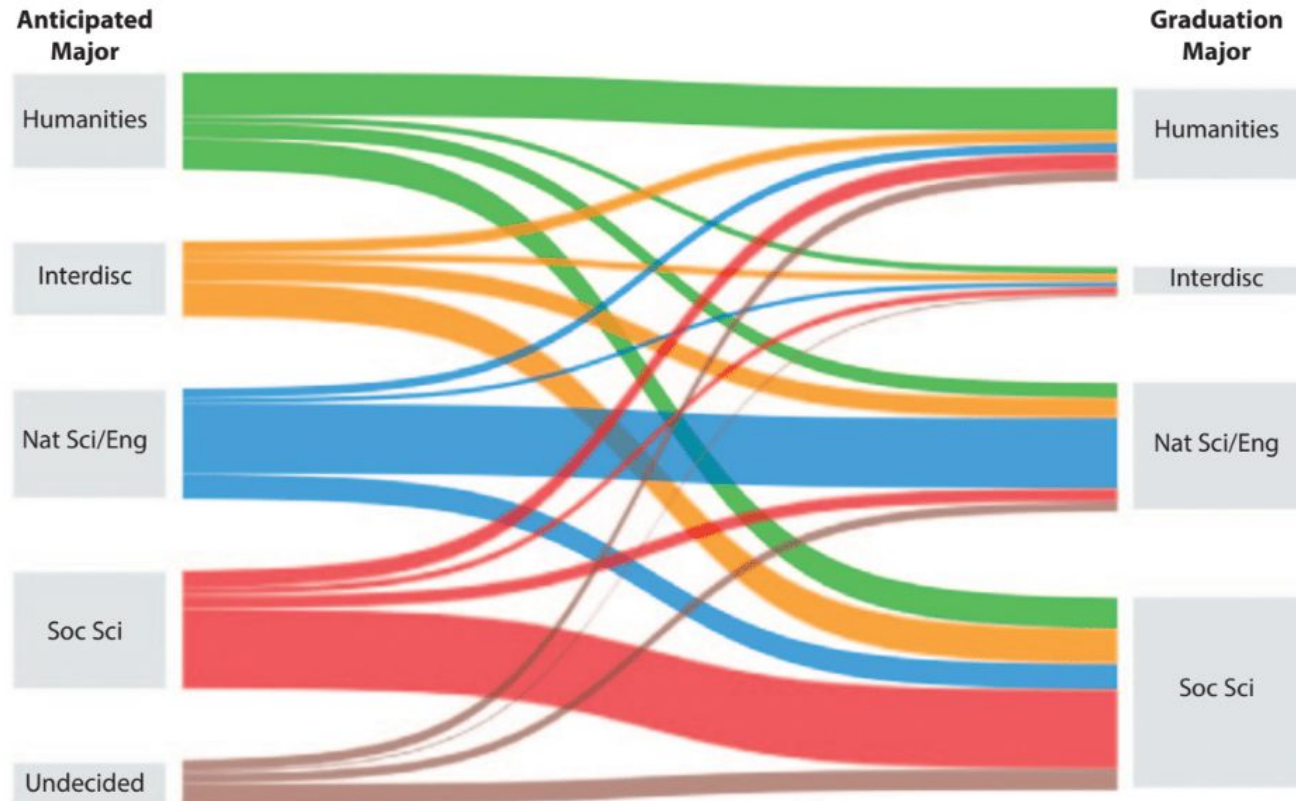


Line, bar, and column charts often employ the preattentive attributes of length and width:

- The **Length** refers to the horizontal, vertical, or diagonal distance of a line or bar/column.
- The **Width** refers to the thickness of the line or bar/column.

3.1 The Preattentive Attribute of Width

A Sankey chart demonstrating the use of the line width attribute



A **Sankey chart** typically depicts the proportional flow of entities where the width of the line represents the relative flow rate compared to the widths of the other lines.

In the example shown, the graduation major for students in a liberal arts college flows from the anticipated major.

The line width attribute is used much less frequently in data visualizations.

3.1 The Attributes of Spatial Positioning and Movement

Spatial positioning refers to the location of an object within some defined space.

- The most used spatial positioning in tables and charts is 2-D, such as the data points in a scatter chart.

Movement can be categorized into two main attributes:

- **Flicker** refers to effects such as flashing to draw attention.
- **Motion** involves directed movement and can be used to show changes within a visualization.

Movement is not possible in most data visualizations because tables and charts are static.

3.2 Gestalt Principles

Gestalt principles refer to the guiding principles of how people interpret and perceive what they see.

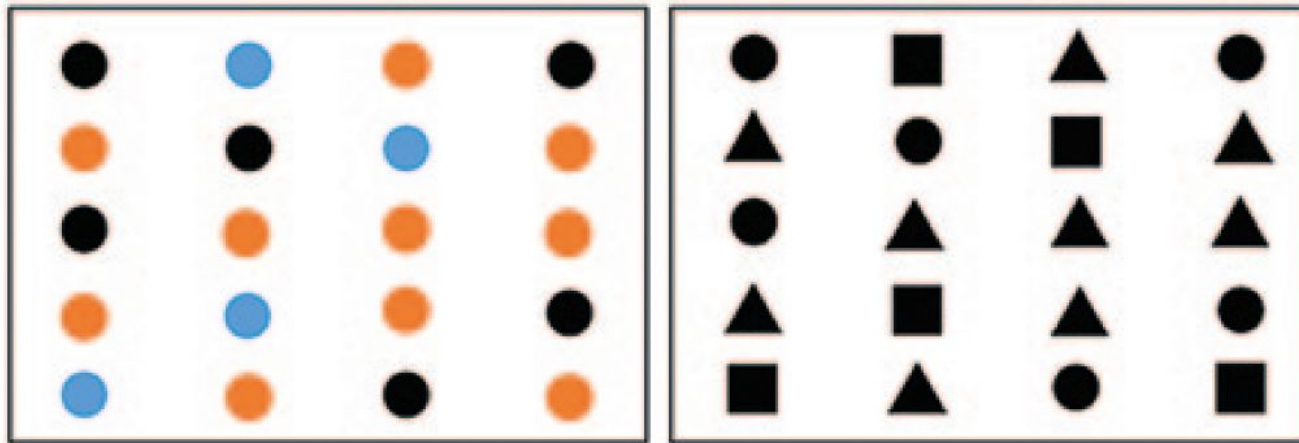
- These principles generally describe how people define order and meaning in the things they see.

The four main Gestalt principles related to the design of data visualizations are:

- **Similarity**
- **Proximity**
- **Enclosure**
- **Connection**

3.2 Gestalt Principle of Similarity

People consider objects with similar characteristics as belonging to the same group

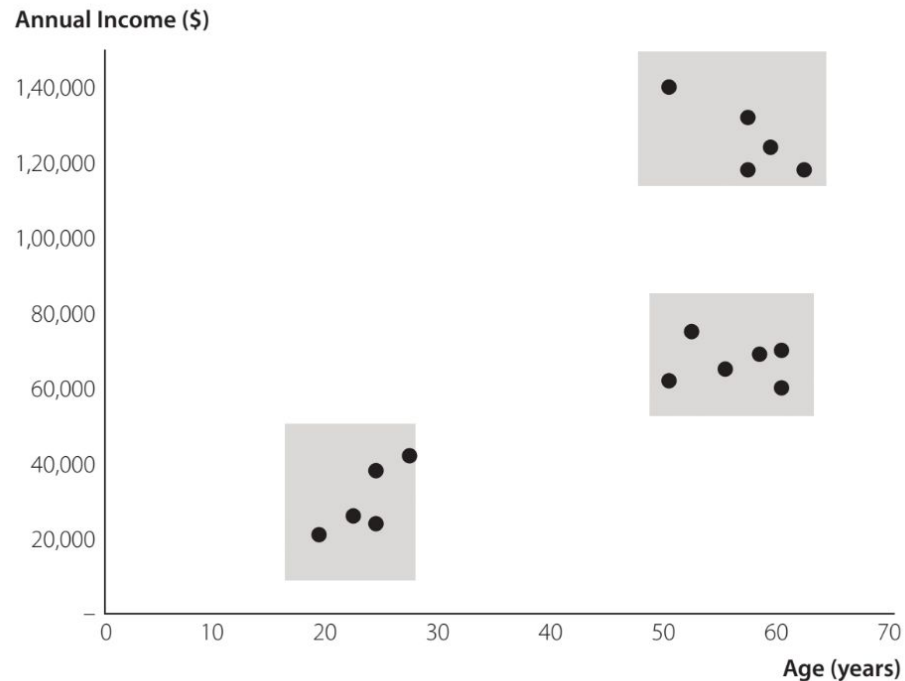


The audience will perceive objects of the same color, or the same shape, as belonging to the same group.

3.2 Gestalt Principle of Proximity

People consider objects physically close to one another as belonging to the same group

Marketing Segmentation by Customer Age and Income

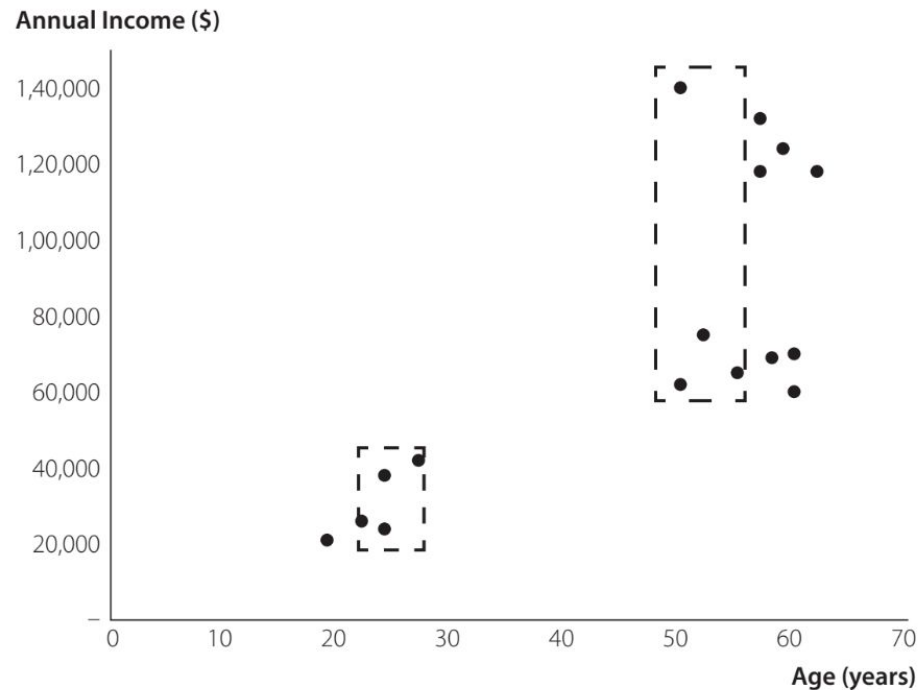


People will seek to group objects near each into the same group and objects that are far from one another into different groups.

3.2 Gestalt Principle of Enclosure

People consider objects physically enclosed together as belonging to the same group

Marketing Segmentation by Customer Age and Income



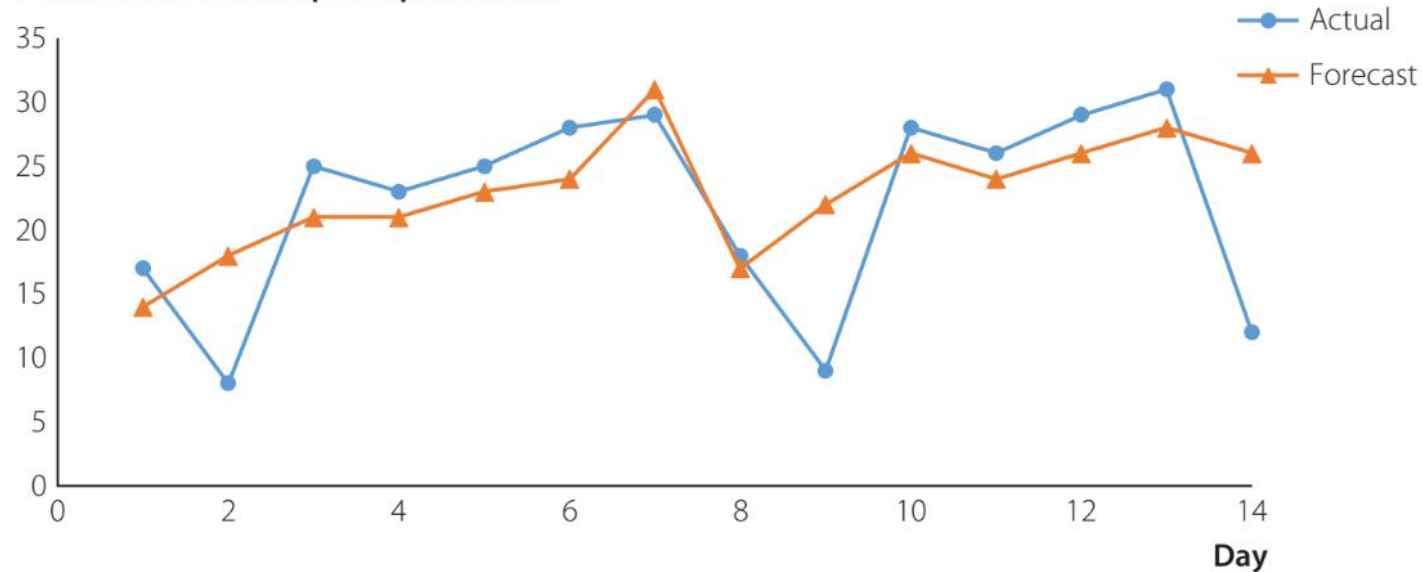
The presence of a third attribute besides age and income, such as educational background, can be visually indicated with enclosures.

3.2 Gestalt Principle of Connection

People interpret connected objects as belonging to the same group

Server Load Forecast vs. Actual For Past 14 Days

Peak Server Load (requests per second)



The principle of connection in data visualization is often used for time-series data.

Connecting the markers emphasizes trends and makes it easier to separate forecast values from actual values.

3.3 Data-Ink Ratio

Data-ink ratio is the proportion of ink used for data to the total amount of ink in a table or chart.

- Helpful for creating effective tables and charts for data visualization
- **Data-ink**: ink used in a table or chart that is necessary to convey the meaning of the data to the audience
- **Non-data-ink**: ink used in a table or chart that serves no useful purpose in conveying the data to the audience
- **Decluttering** is the process of increasing the data-ink ratio in a chart

3.3 Data-Ink Ratio in Tables (1 of 2)

Example of low data-ink table

Scarf Sales			
Day	Sales (units)	Day	Sales (units)
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140
9	140	19	150
10	200	20	230

Example of high data-ink table

Scarf Sales			
Day	Sales (units)	Day	Sales (units)
1	150	11	170
2	170	12	160
3	140	13	290
4	150	14	200
5	180	15	210
6	180	16	110
7	210	17	90
8	230	18	140
9	140	19	150
10	200	20	230

White space is the portion of a data visualization that is devoid of markings. Its use is equivalent to increasing the data-ink in a visualization.

*See notes for step-by-step instructions on how to increase data-ink for this example.

3.3 Different Levels of Data-Ink Ratio in Tables

Comparing different table designs with different data-ink ratios

Design A:

	Month						
	1	2	3	4	5	6	Total
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design B:

	Month						
	1	2	3	4	5	6	Total
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design C:

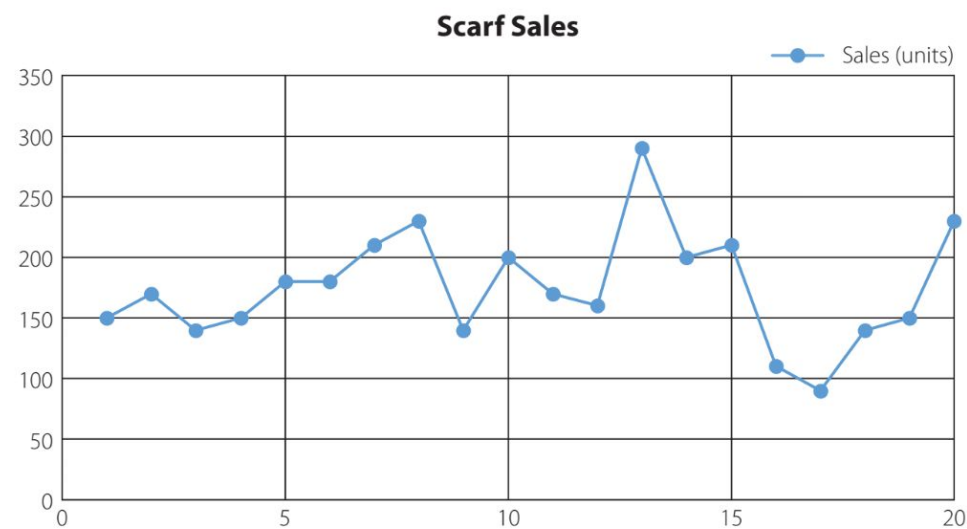
	Month						
	1	2	3	4	5	6	Total
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

Design D:

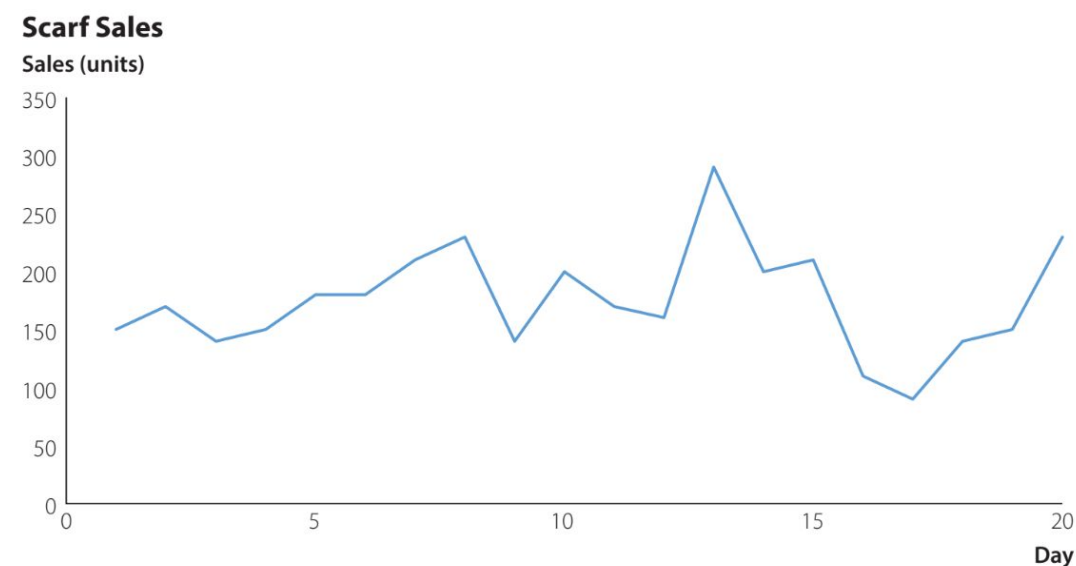
	Month						
	1	2	3	4	5	6	Total
Costs (\$)	48,123	56,458	64,125	52,158	54,718	50,985	326,567
Revenues (\$)	64,124	66,128	67,125	48,178	51,785	55,687	353,027
Profits (\$)	16,001	9,670	3,000	(3,980)	(2,933)	4,702	26,460

3.3 Increase Data-Ink Ratio in Charts to Increase Clarity

Example of low data-ink chart



Example of high data-ink chart

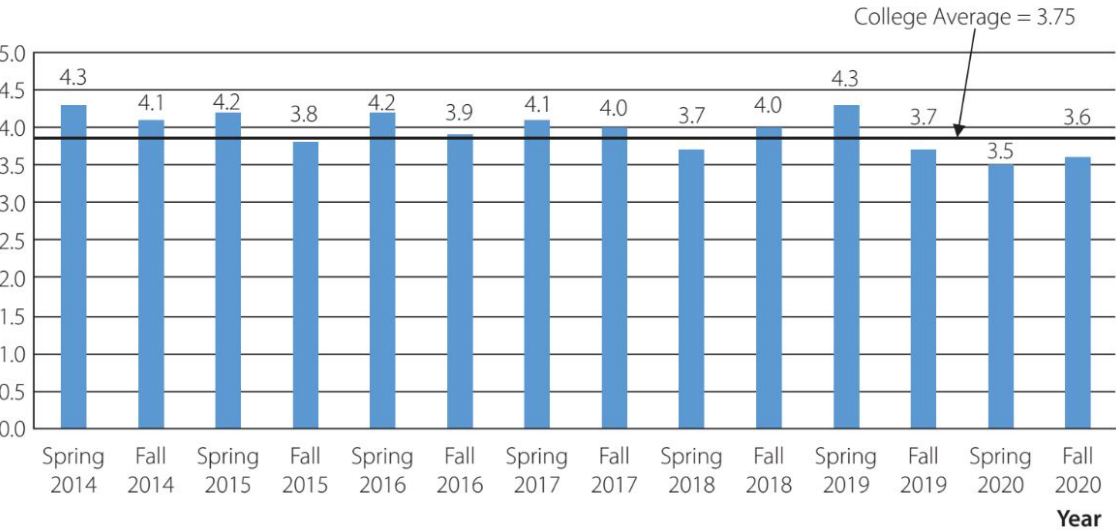


3.3 Increase Data-Ink Ratio in Charts to Decrease Clutter

A cluttered column chart for the STAT 7011 course evaluation data

Professor Smith STAT 7011 Course Evaluations

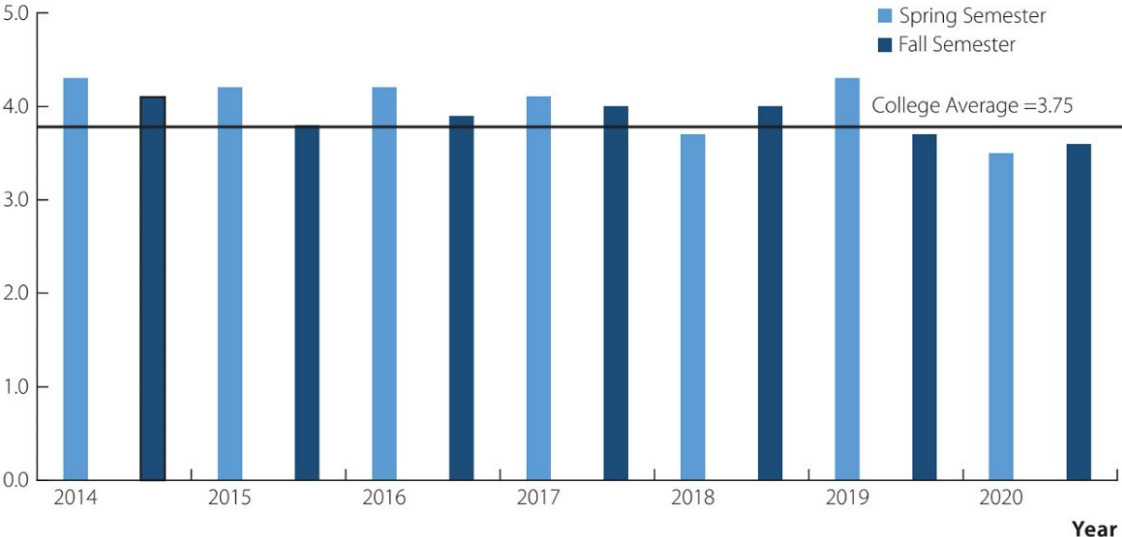
How effective is this instructor? (1 = Not effective; 5 = Extremely effective)



An improved column chart for the STAT 7011 course evaluation data

Professor Smith STAT 7011 Course Evaluations

How effective is this instructor? (1 = Not effective; 5 = Extremely effective)



*See step-by-step instructions on how to improve cluttering for this example in the notes.

3.3 Suggestions for Text Alignment in Tables

- Columns of numerical values in a table should usually be right-aligned.
- All columns showing values should include the same number of decimal digits.
- Only include the necessary number of decimal digits. Additional digits that are not meaningful to the audience increase clutter.
- For extremely large numbers, display data rounded to the nearest thousand, ten thousand, or million.
- It is generally most effective to left-align text values within a column in a table.

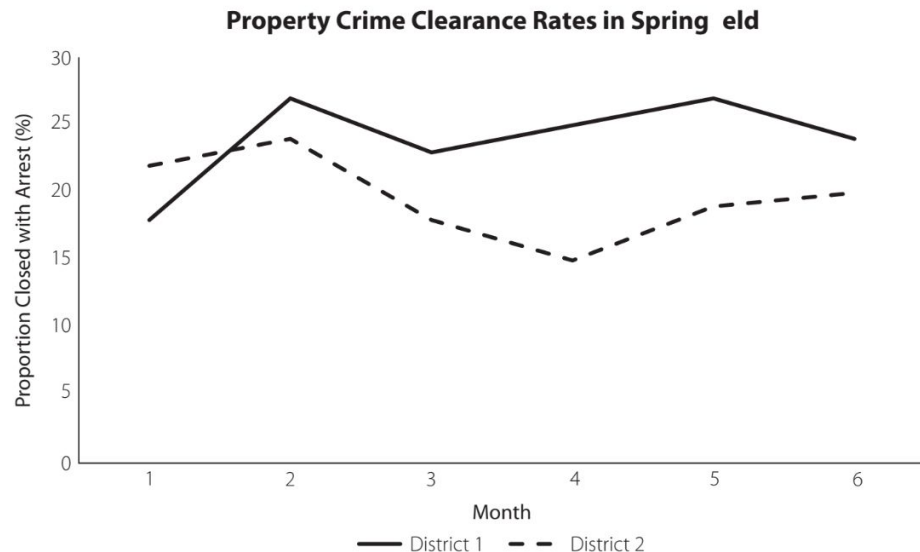
3.4 Minimizing Eye Travel

- We can minimize the cognitive load needed to view and interpret data visualization by:
 - using preattentive attributes
 - using Gestalt principles
 - increasing the data-ink ratio
 - **minimizing eye travel**

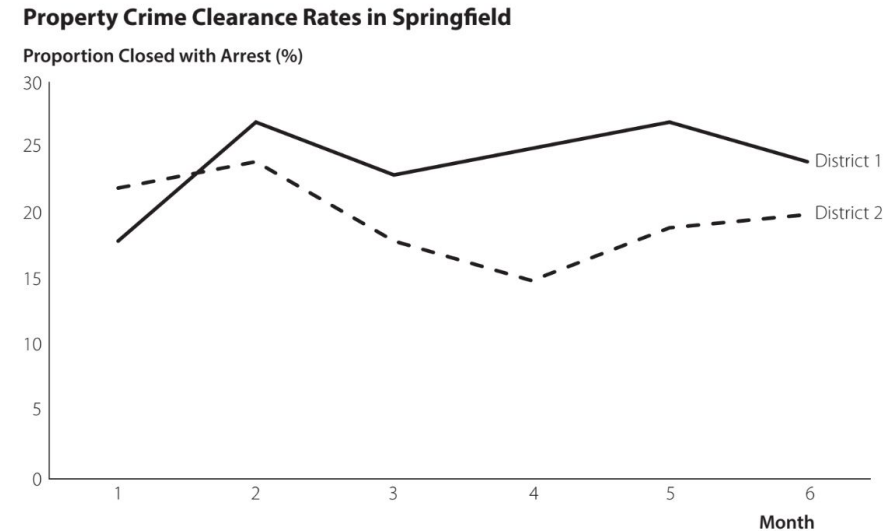
*See additional comments in the notes.

3.4 Minimizing Eye Travel in a Chart

A default line chart created in Excel for property crime clearance rates



Improved line chart for property crime clearance rates – Minimized eye travel

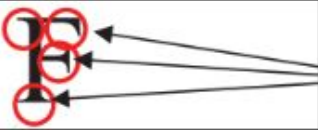





Required eye travel is minimized by:

- Aligning the chart title and vertical-axis title with the vertical axis
- Moving the line labels adjacent and to the right of the chart lines
- Aligning the horizontal-axis label with the line labels at the end of the horizontal axis.

*See step-by-step instructions on how to improve the line chart in the notes.

3.4 Choosing a Font for Text

Serif Font		Sans-Serif Font	
			

Serif Font

- Preferred for text in printed work
- Includes end-of-stroke features
- Less legible at small sizes
- Common fonts include:
 - Times, Times New Roman, and Courier

Sans-Serif Font

- Preferred for text in data visualization
- Does not Includes end-of-stroke features
- More legible at small sizes
- Common fonts include:
 - Arial, Calibri, Myriad Pro, and Verdana

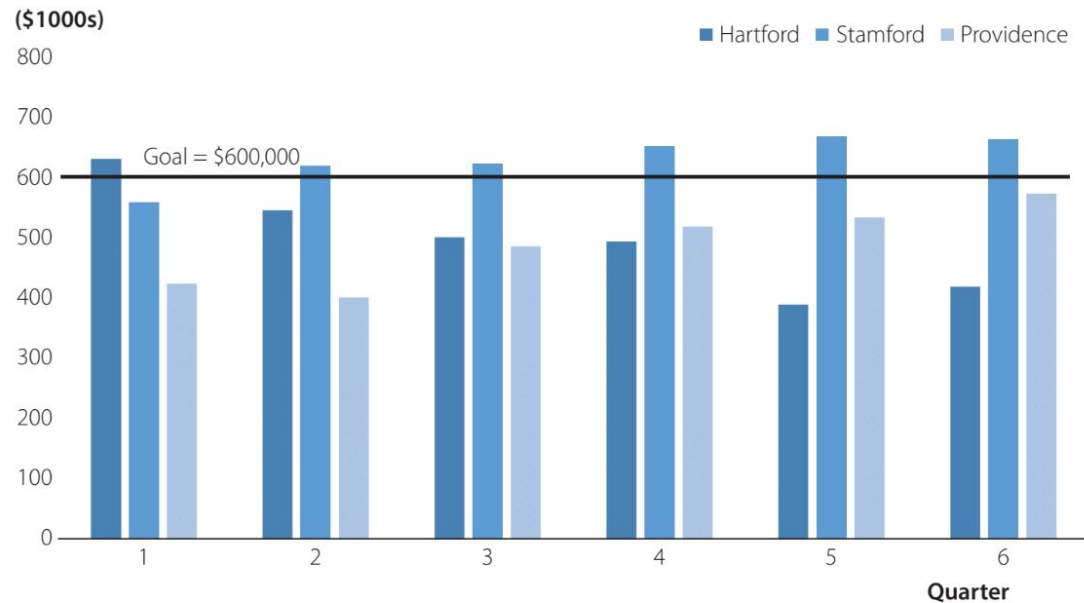
3.4 Suggestions for Choice of Font for Text

- The choice of the font in data visualization should be consistent and restrained.
- Most data visualizations should use a single type of font for text to minimize cognitive load.
- Horizontal- and vertical-axis titles should have the same font type and size.
- The use of different font sizes, **bold**, *italic*, and **color** enhancements, help differentiate between features and direct audience attention.

3.5 The Wrong Type of Visualization

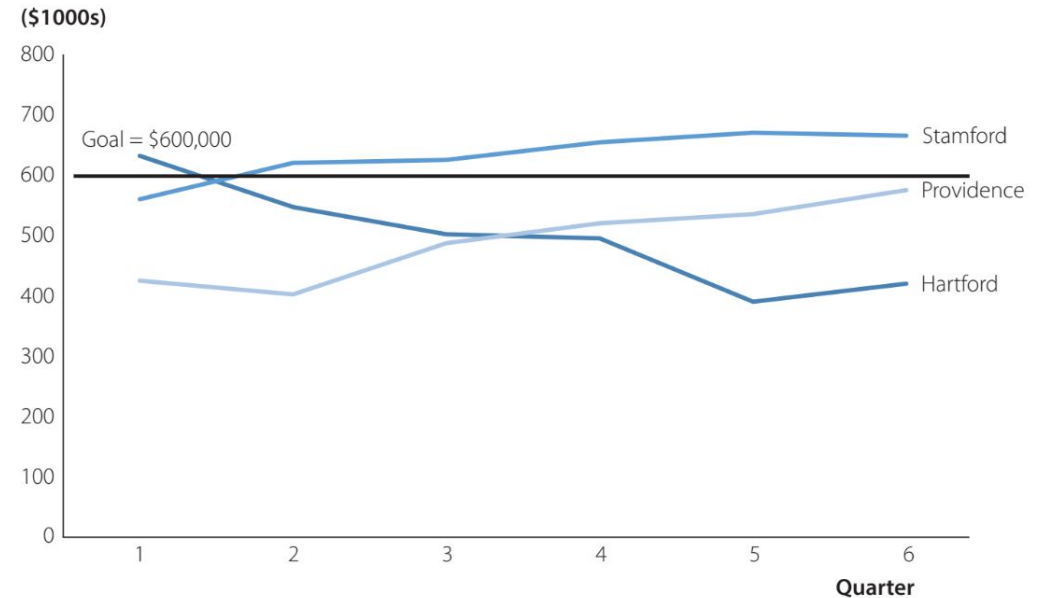
A cluttered clustered column chart for the booked revenue data comparison

Booked Revenue Comparison



An improved line chart for the booked revenue data comparison

Booked Revenue Comparison



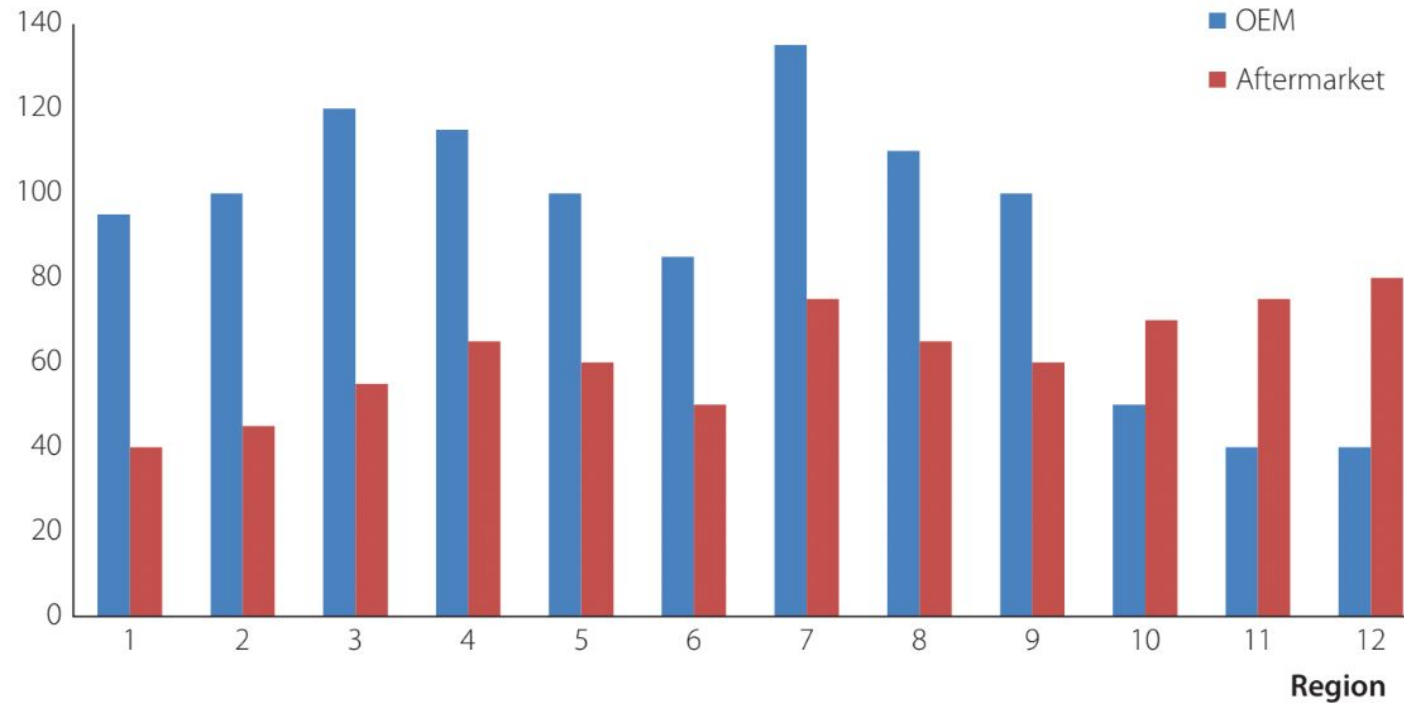
*See step-by-step instructions on how to improve the chart in the notes.

3.5 Displaying Excessive Information

A cluttered clustered column chart showing OEM and aftermarket sales for several regions.

Keeland Sales Comparison

Sales (\$1000s)

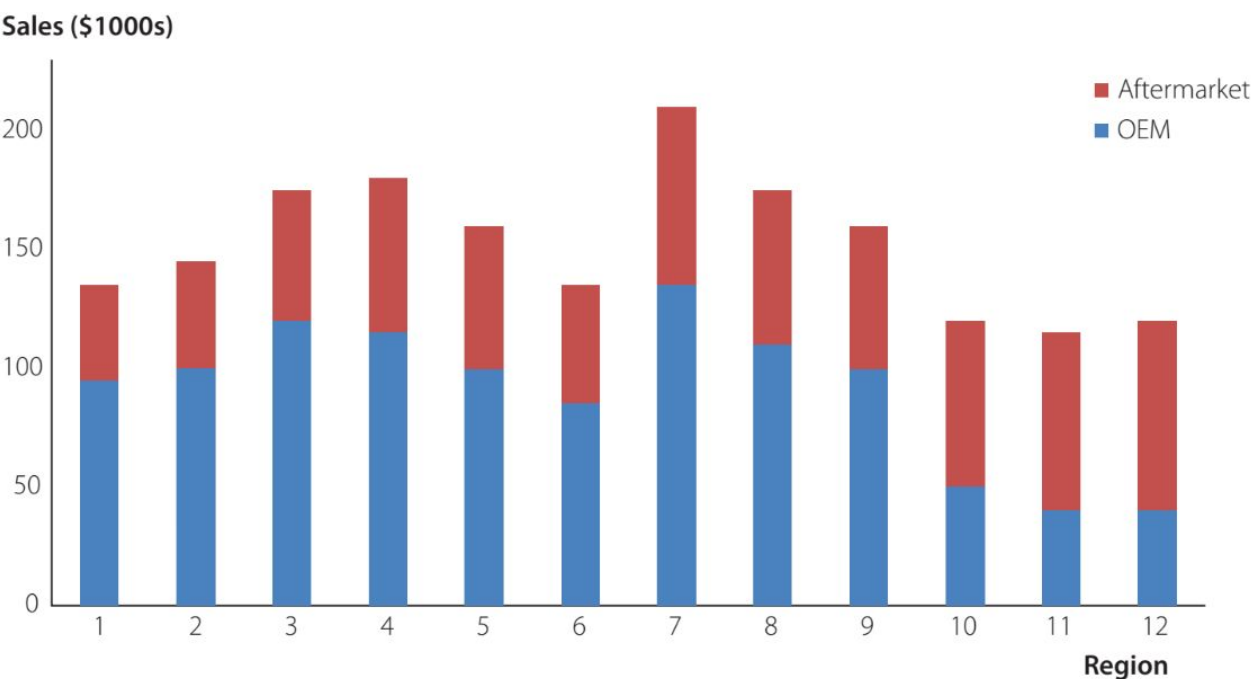


3.5 Displaying Excessive Information – Two Solutions

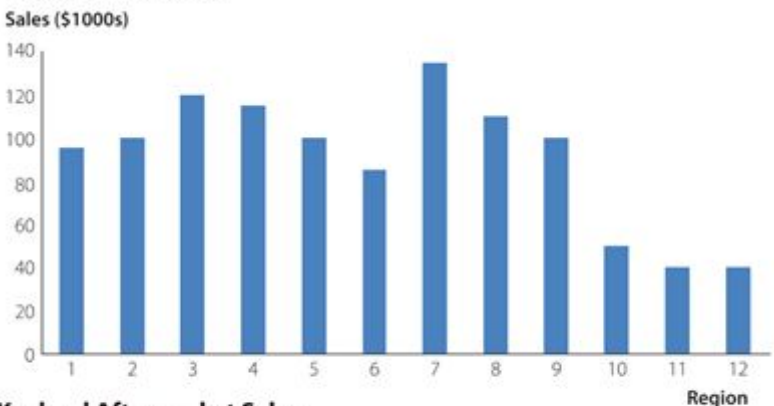
A stacked clustered column chart showing OEM and aftermarket sales for several regions.

Improved column charts showing OEM and aftermarket sales separately

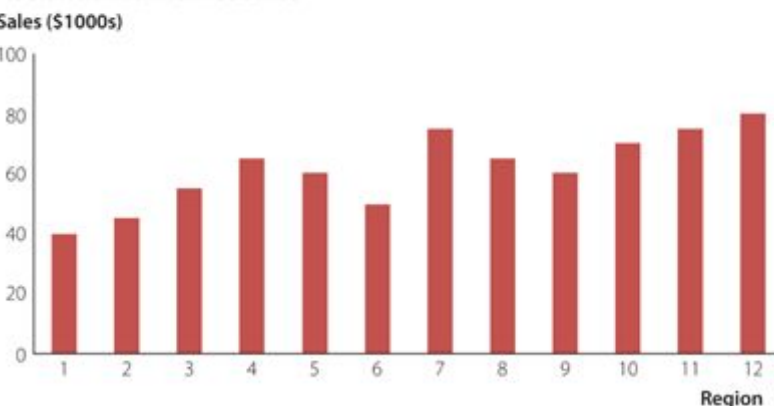
Keeland Sales Comparison



Keeland OEM Sales

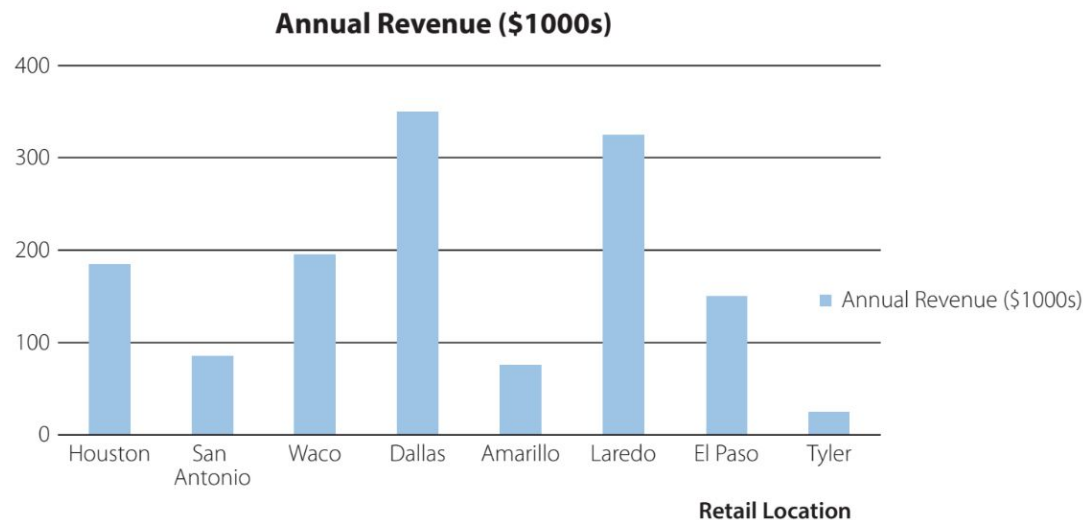


Keeland Aftermarket Sales

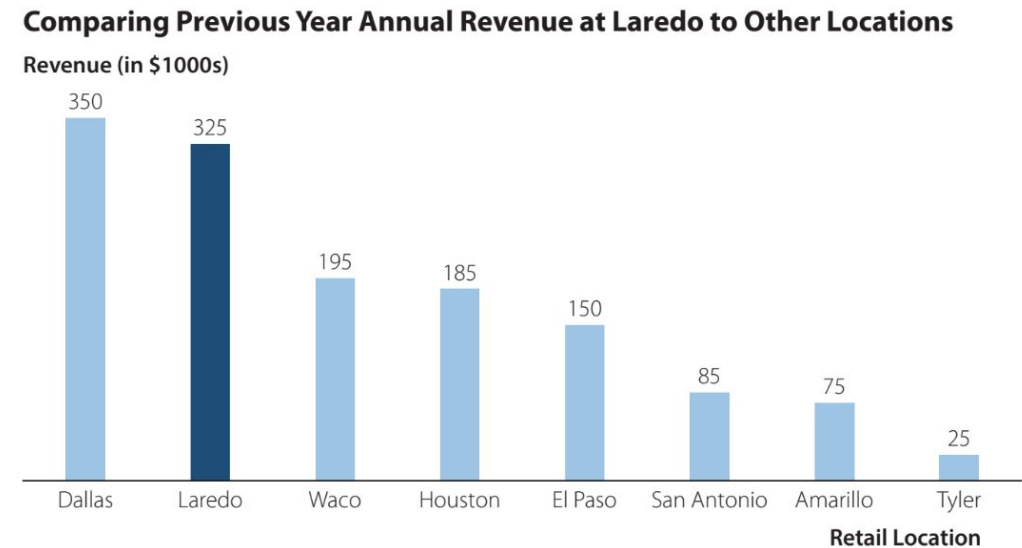


3.5 Using Excel Default Settings for Charts

Default Column Chart Comparing Annual Revenue at Eight Retail Locations



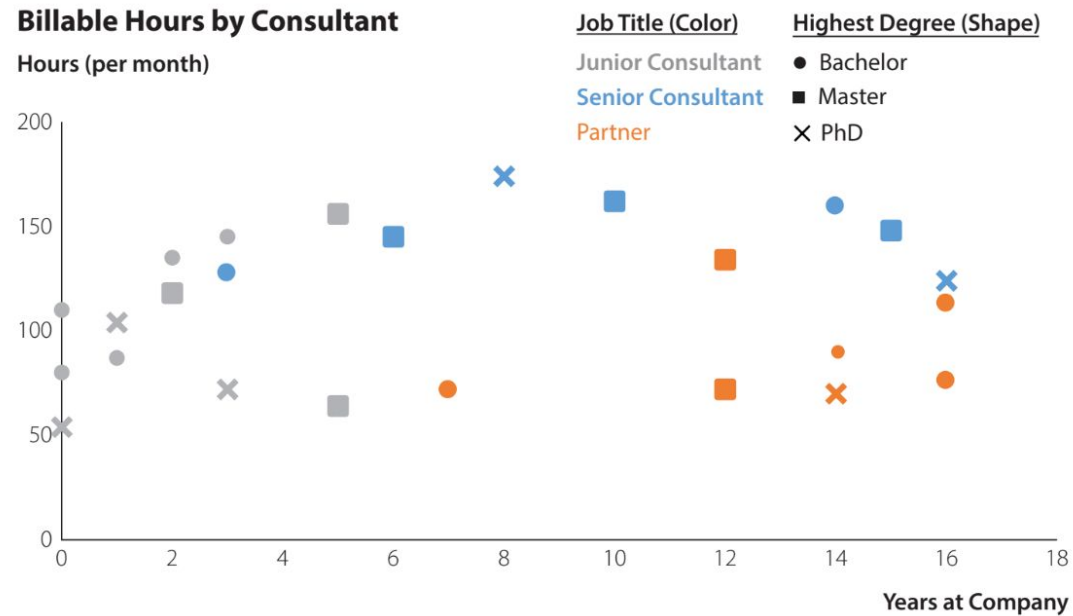
Improved Column Chart Comparing Annual Revenue at Eight Retail Locations



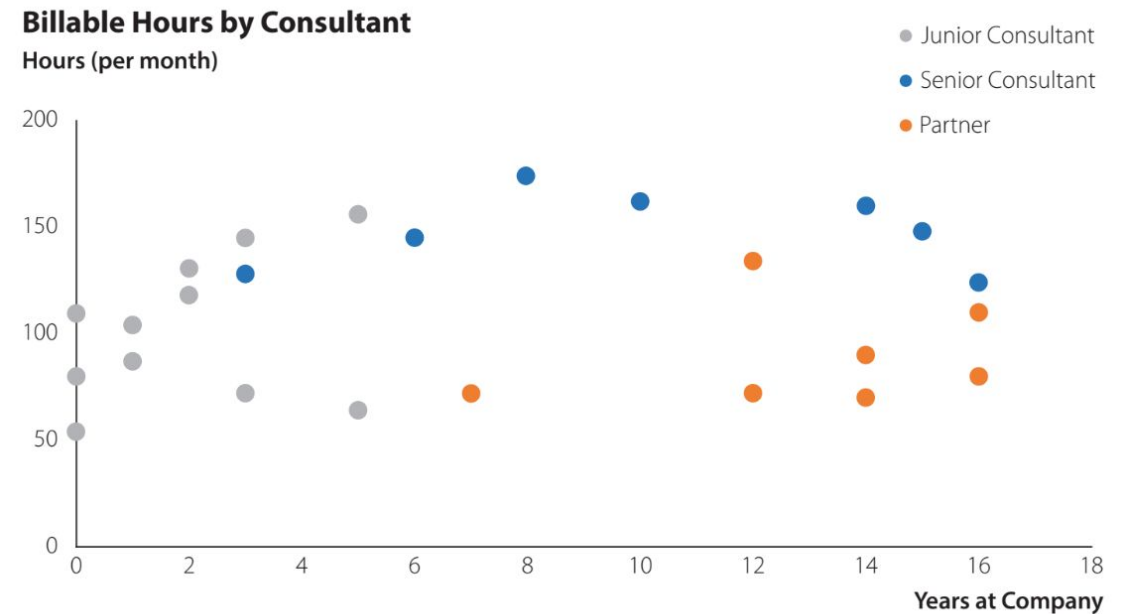
*See step-by-step instructions on how to improve the chart in the notes.

3.5 Too Many Attributes

Chart displaying billable hours association with years at the company, job title, and highest degree attained



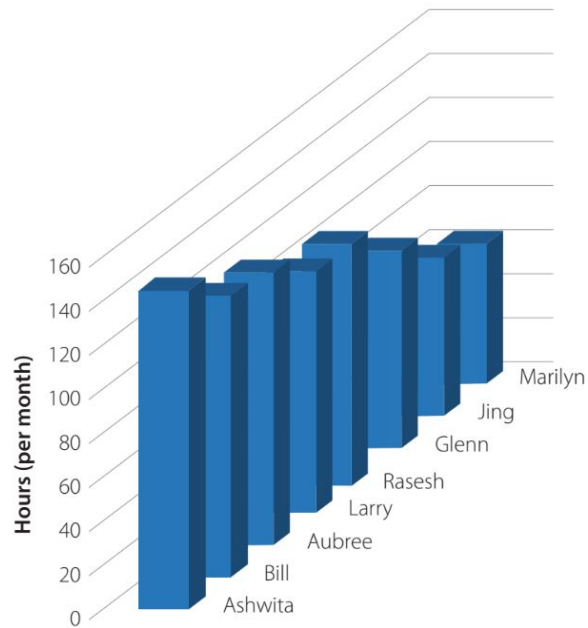
Simplified chart displaying billable hours association with years at the company and job title



3.5 Unnecessary Use of 3D

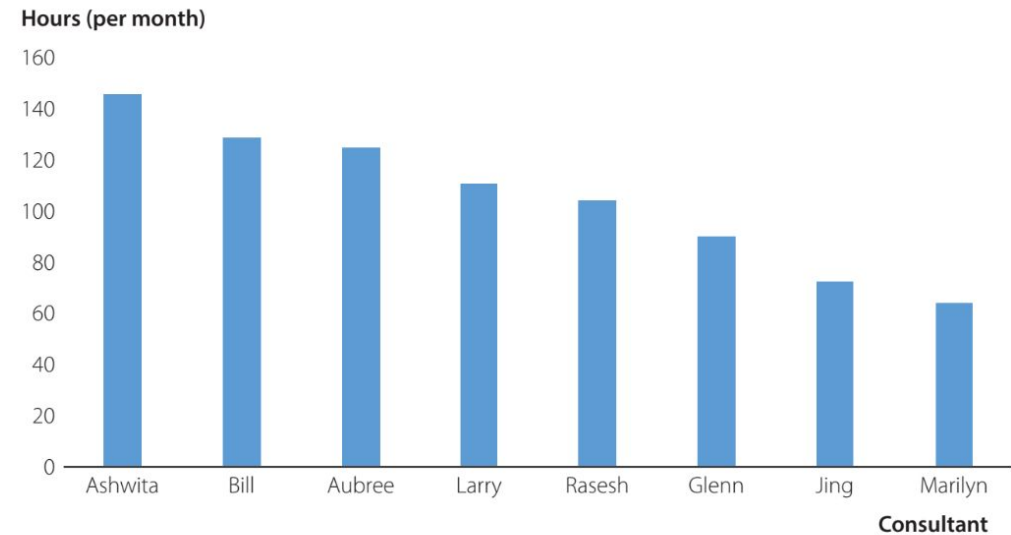
A 3D column chart displaying billable hours – the third dimension is not useful

Billable Hours By Consultant



An improved 2D column chart displaying billable hours

Billable Hours by Consultant



Discussion Activity 1

- Consider the matrix of 160 single digits (16 rows x 10 columns) shown in the left image of slide 6. The textbook describes how we can aid in counting the number of 7s in the matrix using the preattentive attributes of red hue (middle image of slide 6) and size (right image of slide 6.)
 - Regardless of the specific hue and size used in the two mentioned examples, list at least three more ways you could use individual preattentive attributes to aid in counting the number of 7s in the matrix of first digits.
 - Reassess the previous question, but this time answer it using the individual Gestalt principles of similarity, proximity, enclosure, and connection, rather than preattentive attributes. Be creative.

Discussion Activity 2

- Consider the Scarf Sales line chart included in the *ScarfSalesChart* data file (Slide 24, left chart). Default charts created in Excel often have a low and suboptimal data-ink ratio. You can build the default line chart from the *ScarfSalesChart* data file in Excel by selecting the A1:B21 cell range and selecting the ribbon's **Insert** tab. Then, in the **Charts** group, select the **Insert Line or Area Chart** icon and click on the top-left option under **2-D Line**.
 - When building the default line chart, try increasing its data-ink ratio by removing any redundant and unnecessary feature to the data visualization. Do not add any features. List all the steps you implemented to increase the data-ink ratio.
 - Once you have removed all unnecessary and redundant features, it is now time to make the chart more readable by adding features so that it looks like the right chart in Slide 24. List all the steps you implemented.

Summary

In this chapter, you should have learned how:

- The preattentive attributes of color, form, spatial positioning, and movement direct attention in data visualization.
- The Gestalt principles of similarity, proximity, enclosure, and connection help the audience interpret objects in data visualization.
- Data visualization strategies can help to:
 - increase the data-ink ratio
 - minimize eye travel for the audience
 - choose an effective font for text in data visualization.
- To avoid common mistakes in data visualization design such as:
 - choosing the wrong type of visualization
 - trying to display too much information
 - using the default format for Excel charts
 - using 3D charts unnecessarily

Data Fluency - Common Pitfalls

- Report Proliferation - Too Many Reports can distract, define what's important and prioritize
- Balkanized Data - Silos emerge on department or functional lines
- Data Elitism - If control of data is too concentrated there can be gaps
- The Supermodel - Avoid perfectionism and unnecessary animation
- Searching for Understanding - Align metrics with strategy
- Data Care - Curiosity includes care
- Metric Fixation - Consider top level metrics that span strategies and divisions/departments

Data Fluency - Finding Balance

Needs of Data Users	Interests of Data Producers
Letting Data Speak	Allowing data author to define message
Business users who use data to make decisions	Technologists who manage and deliver data
Focusing on a few key metrics	Flexibility to change metrics as needs evolve
Deep analysis by Data Scientists	Direct reporting in simple but most important formats
Putting understanding first	Capturing audience's imagination with aesthetics