ima®

The Association of Accountants and Financial Professionals in Business

Statement on Management Accounting

DATA VISUALIZATION

TABLE OF CONTENTS

ABOUT THE AUTHOR

Ann C. Dzuranin, CPA, Ph.D., is the KPMG Endowed Professor of Accountancy at Northern Illinois University. She earned her B.S. from Fairleigh Dickinson University, her MBA from New York University, and her Ph.D. from the University of South Florida. Ann is a CPA with 15 years of experience in both public and corporate accounting. She conducts behavioral research in management accounting decision making and the ways in which accounting information systems interact with those decisions. Her work has appeared in publications including *Issues in Accounting Education*, *Journal of Information Systems*, *Journal of Business Ethics*, *Management Accounting Quarterly*, *Journal of Corporate Accounting and Finance*, and *Journal of Accounting Education*. Ann received the 2018 American Accounting Association's Innovation in Accounting Education Award for her work in data analytics curriculum development. Her work in data analytics has resulted in invitations to present on data analytics and accounting curriculum at both academic and professional conferences. Her presentations have reached more than 2,200 people, and her materials have been shared with more than 75 universities. She can be reached at **(815) 753-6212** or **adzuranin@niu.edu**.

EXECUTIVE SUMMARY

Data visualization skills are in high demand in the accounting and finance professions. The ability to use data visualization to explore data and to communicate findings is increasingly important in a data-driven business world. Many accounting and finance professionals lack formal training in data visualization yet are expected to be able to work with data and data visualization software. Furthermore, accounting and finance professionals are expected to be educated consumers of data visualizations. This IMA® (Institute of Management Accountants) Statement on Management Accounting (SMA) covers the data visualization skills needed to be successful. Specifically, it discusses explanatory data visualizations, how to create an effective visualization, and how to tell a story with your data. The SMA concludes with a discussion of ethics and data visualizations, highlighting the importance of being able to identify misleading visualizations. •

The Importance of **Data Visualization**

"A picture is worth a thousand words."

Although there is debate as to the origin of this familiar quotation, it speaks to the power of using a "picture" to convey meaning.[1] It is in this vein that data visualization is used in business. Data that has been visualized can help to communicate meaning more easily than would be possible with words or by the raw data. The proliferation of data combined with the development of easy-to-use data visualization software programs has led to data visualization quickly becoming a standard tool for management accountants. The concern is that most management accountants have not received data visualization training. When should you use data visualization? In what ways can visualization help inform decision making? How do you know if you have created a good visualization? How do you know if a visualization you are given is accurate and/or unbiased? These questions and others will be addressed in this IMA® (Institute of Management Accountants) Statement on Management Accounting (SMA), but let's start with a definition of data visualization.

What Is Data Visualization?

Data visualization can be defined as the process of displaying data in a meaningful fashion to provide insights that will support better decisions.[2] At its core, data visualization is a communication tool. The power of data visualization is that it helps make sense of large sets of data. For example, see sales by brand of a fictitious automobile manufacturer in Table 1.[3] By scanning the data in the table, it is easy to pick out which brand is performing best and which is performing worst.

TABLE 1: SALES BY BRAND

Brand	2019	2020
Apechete	$15,460,402	$16,938,230
Jackson	$8,451,502	$8,814,542
Tatra	$15,097,592	$15,885,281

What if instead you were given Table 2 and asked to identify which model is performing best?

[1] The quotation is sometimes attributed to a Chinese proverb or to Frederick R. Barnard (1921). See William Safire, "On Language; Worth a Thousand Words," *The New York Times Magazine*, April 7, 1996, p. 16, **nyti.ms/3BAJ2OT.**

[2] James R. Evans, *Business Analytics*, Third Edition, Pearson, 2020.

[3] The data used in this SMA are based on the case "Huskie Motor Corporation: Visualizing the Present and Predicting the Future" by Ann C. Dzuranin, Johan Perols, and Dana L. Hart, *IMA Educational Case Journal*, January 2018.

TABLE 2: SALES BY MODEL

Model	2019	2020
Advantage	$5,092,740	$5,321,829
Bloom	$3,063,047	$3,485,643
Brutus	$915,508	$1,074,344
Chare	$3,373,407	$4,255,434
Crux	$2,277,138	$2,205,425
Fiddle	$1,441,314	$1,511,390
Island	$3,162,502	$3,087,076
Jespie	$3,684,030	$3,531,304
Mortimer	$1,704,977	$1,728,847
Pebble	$4,368,798	$4,870,330
Rambler	$1,552,798	$1,817,659
Rebel	$1,785,481	$1,807,429
Robin	$1,877,014	$1,966,349
Summet	$2,678,681	$2,759,042
Wood	$2,032,062	$2,215,954

Clearly, this would take much longer to determine. But what if you were given Table 3 instead?

TABLE 3: SALES BY MODEL HIGHLIGHTS

Model	2019	2020	Gross Sales
Advantage	$5,092,740	$5,321,829	
Bloom	$3,063,047	$3,485,643	
Brutus	$915,508	$1,074,344	
Chare	$3,373,407	$4,255,434	
Crux	$2,277,138	$2,205,425	
Fiddle	$1,441,314	$1,511,390	
Island	$3,162,502	$3,087,076	
Jespie	$3,684,030	$3,531,304	
Mortimer	$1,704,977	$1,728,847	
Pebble	$4,368,798	$4,870,330	
Rambler	$1,552,798	$1,817,659	
Rebel	$1,785,481	$1,807,429	
Robin	$1,877,014	$1,966,349	
Summet	$2,678,681	$2,759,042	

In Table 3, the sales figures are highlighted using a graduated color scale. The color scale provides the viewer a quick way to identify the models that are performing well (Advantage and Pebble) and which are performing poorly (Brutus). The comparison of these tables illustrates the power of a simple visualization—the highlighting of the sales figures using a graduated color scale.

Visualizing data not only helps to make sense of large sets of data but can also help to uncover patterns or trends that might otherwise go undetected. Table 5 and Figure 1 were constructed in 1973 by the statistician Francis Anscombe.[4] He prepared the table and the graphs of the data from the table to demonstrate the importance of graphing before analyzing the data. The raw data are given in Table 4.

TABLE 4: ANSCOMBE'S QUARTET DATA

Set 1		Set 2		Set 3		Set 4	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Anscombe then provided the statistical analysis of the data sets (see Table 5).

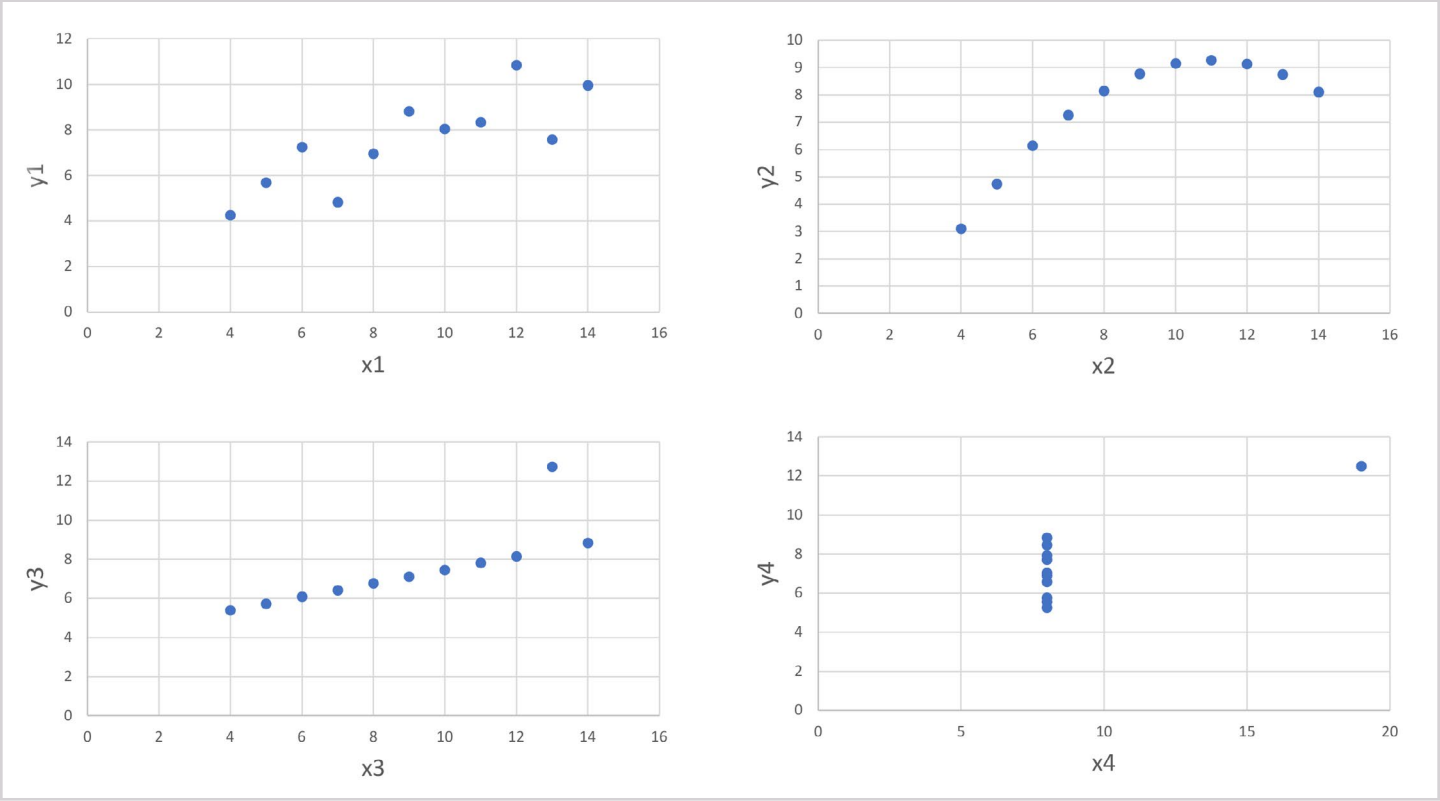
TABLE 5: ANSCOMBE'S QUARTET—STATISTICAL ANALYSIS

	Set 1	Set 2	Set 3	Set 4
Mean (x)	9.00	9.00	9.00	9.00
Mean (y)	7.50	7.50	7.50	7.50
Variance (x)	11.00	11.00	11.00	11.00
Variance (y)	4.13	4.13	4.12	4.12
Correlation	0.82	0.82	0.82	0.82
Intercept	3.00	3.00	3.00	3.00
Slope	0.50	0.50	0.50	0.50

[4] F.J. Anscombe, "Graphs in Statistical Analysis," *The American Statistician*, February 1973, pp. 17-21.

Note that in the statistical analysis, all the data sets appear to be equal. The mean, variance, correlation, intercept, and slope are identical for each data set. If you stopped your analysis here, you would conclude that there are no statistical differences among the data sets. But what about the distribution of the data? Understanding distribution of data is a critical aspect of analyzing data, as Anscombe shows by graphing the data points (see Figure 1).

FIGURE 1: ANSCOMBE'S QUARTET—SCATTERPLOTS



It is clear from the graphs in Figure 1 that there are very different interpretations for the data sets. Sets 1 and 3 appear to be linear. Set 2 appears to be curvilinear. Sets 3 and 4 appear to have outliers in the data set.

The takeaway from this illustration is that it was only by visualizing the data that we were able to determine that there are differences in the distributions and to detect potential outliers in the data. Herein lies the power of data visualization: the ability to identify what might have otherwise gone undetected. •

How Can Data Visualization Be Used in **Management Accounting?**

Management accounting involves partnering in management decision making, and decision making relies on data.[5] Given the speed and volume of data collection in business, managers need ways to make sense of large sets of data to help inform decision making and to help communicate results. Data visualization is a powerful tool that can be used to both facilitate and influence decision making.

There are two types of data visualizations used in management accounting:

1. **Exploratory** visualizations help provide insights into business performance.
2. **Explanatory** visualizations help managers communicate the results of their analyses to influence changes and improvements.

Exploratory Data Visualization

Exploratory data visualizations are used when you want or need to explore data to find insights. You use these types of visualizations to help better understand your underlying data. Data exploration is a discovery process, with the goal of providing insights about your data. Explore data to see if there are any interesting relationships, patterns, or trends. Exploratory data visualization can help to describe your data as well as help to identify anomalies and outliers. The activities in data exploration include:

1. Examine the data.
2. Investigate the distribution.
3. Investigate relationships.
4. Identify patterns.

Examine the Data

Examining the data should always be the first step in exploratory data visualization. Although visualizations can help you to identify errors, first be sure that you aren't missing an obvious problem in the data. Some questions to ask yourself include:

1. Do the columns make sense based on the purpose or question being investigated?
2. Do the values make sense?
3. Are the values the appropriate scale?
4. Are there any missing data points?

Answering these questions will help you to identify any initial problems in the data that can be corrected before beginning your analysis.

Investigate the Distribution

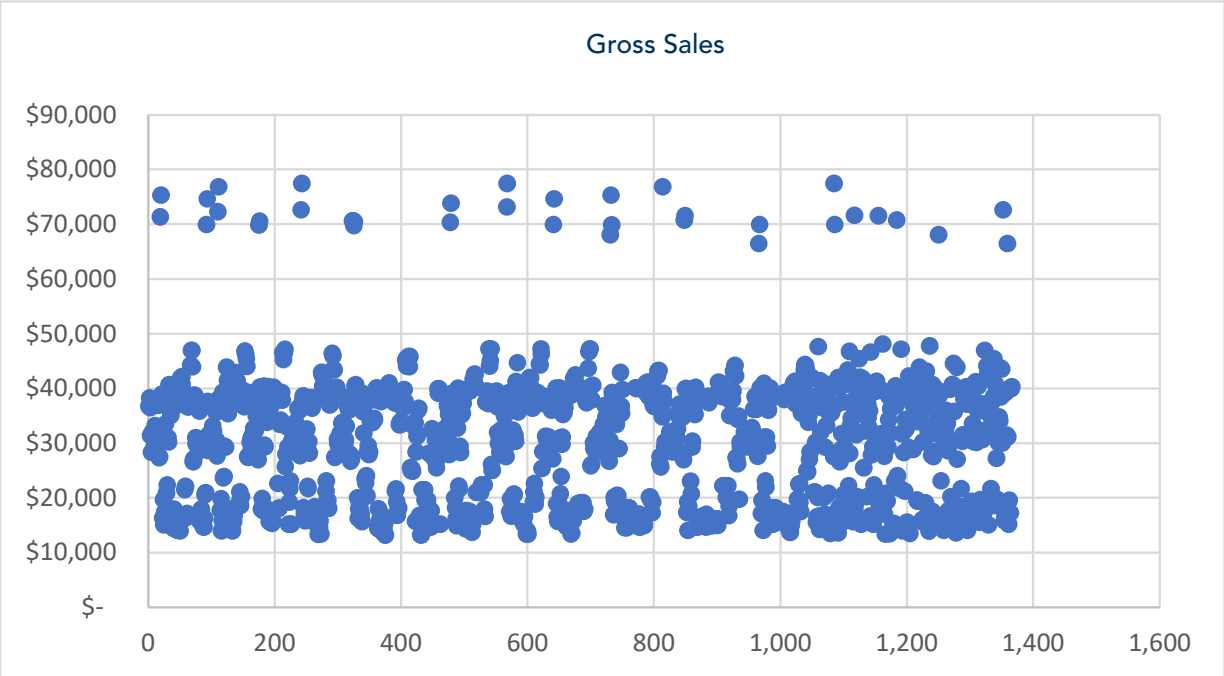
The second activity in exploratory data visualization is to investigate the distribution. What does this mean? By gaining an understanding of how your data is distributed, you will be able to quickly identify potential outliers, unexpected distributions, and potential measurement errors.

The type of visualization you use to explore distribution depends on the type of data that you are exploring. Scatterplots (like the ones shown in Figure 1) are perfect for examining the distribution of

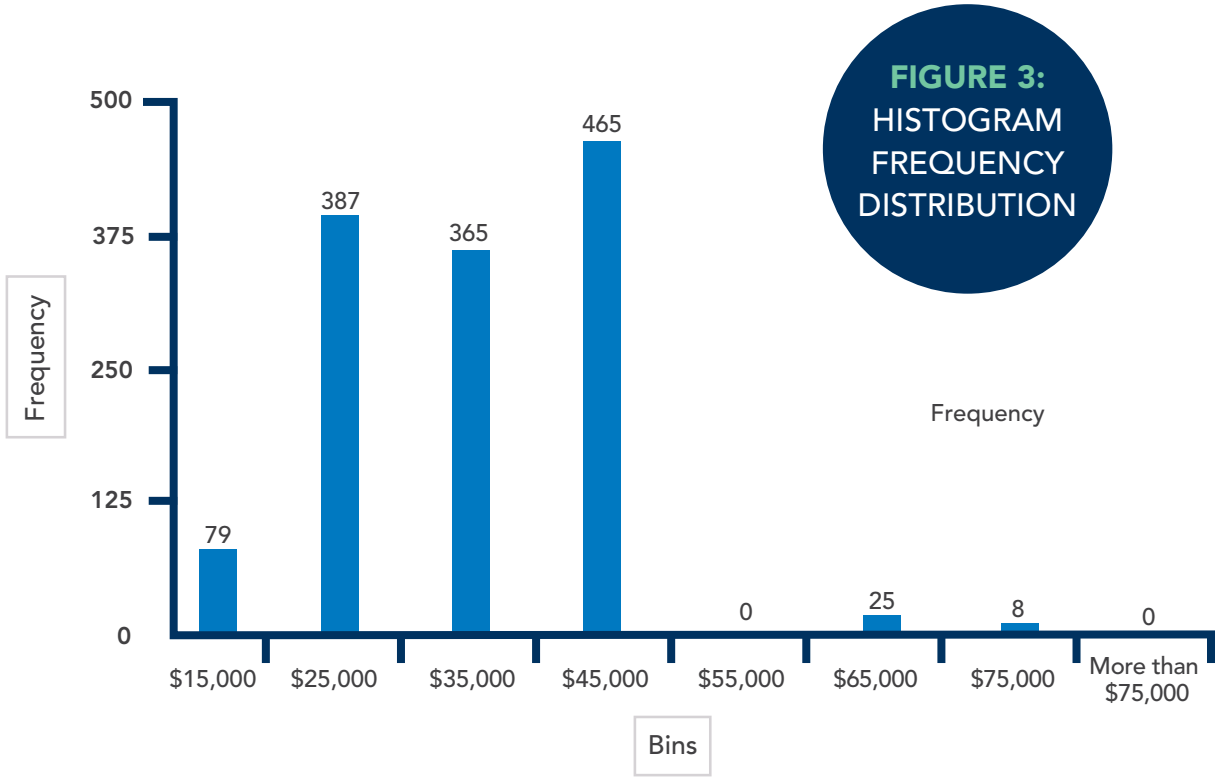
[5] IMA, *Definition of Management Accounting*, 2008, **bit.ly/3zFJoTH**.

numerical data. The scatterplot in Figure 2 displays the distribution of gross sales for an automobile manufacturing company. Each dot in the plot represents the sale of a vehicle. You can see from this visualization that you have two groupings of data. One set is grouped between $60,000 and $80,000, and the other set is grouped between $10,000 and $50,000.

FIGURE 2: DISTRIBUTION OF GROSS SALES



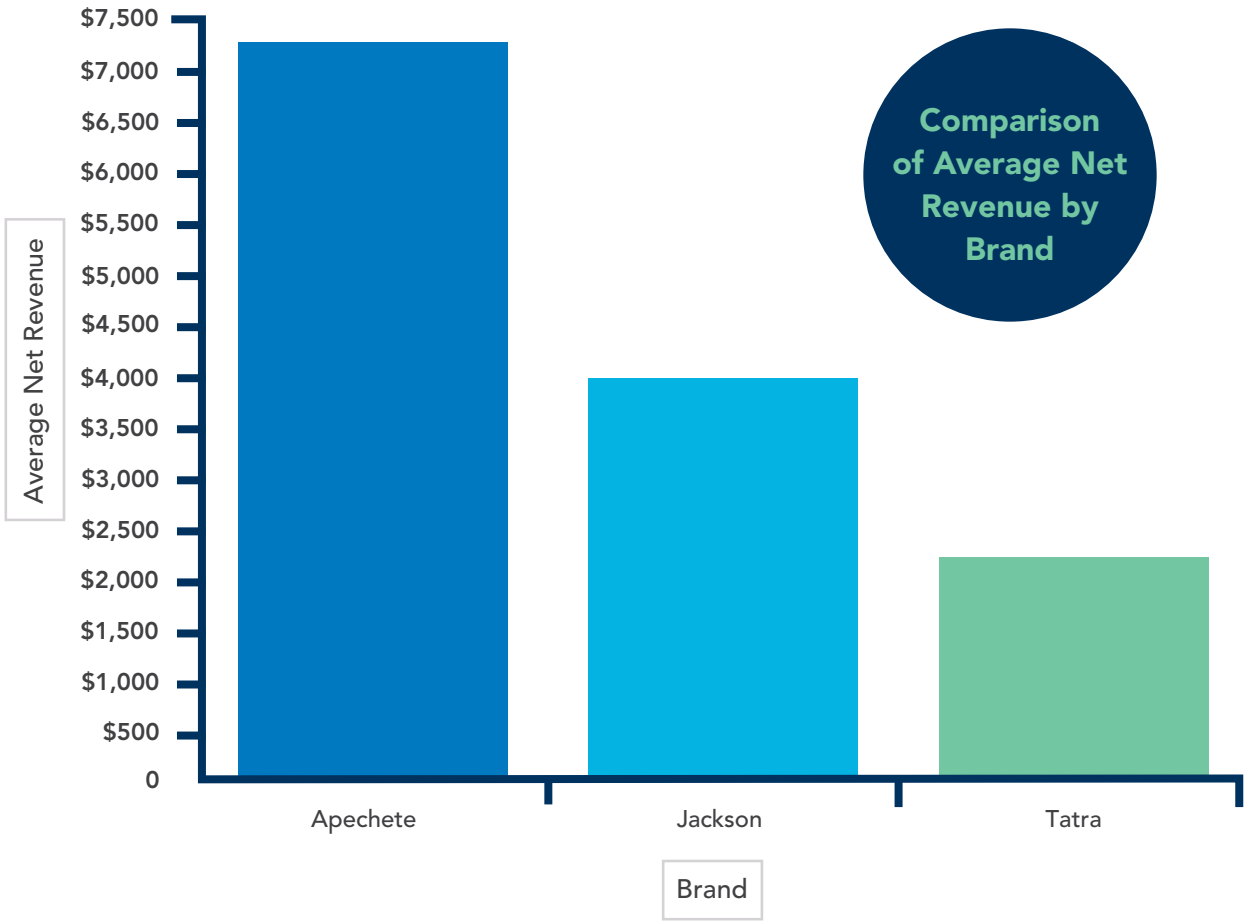
The data in Figure 2 are univariate data. In other words, you are visualizing one variable. Univariate data can also be visualized using a histogram. Suppose you are interested in the frequency of distribution of sales based on the following groupings of gross sales amount: $15,000, $25,000, $35,000, $45,000, $55,000, $65,000, and $75,000.



The visualization in Figure 3 provides further insight into gross sales. By visualizing the data in price groupings, you can see the total number of cars sold within each price range. The highest number of sales are in the $45,000 price range. Only eight vehicles are in the $75,000 range.

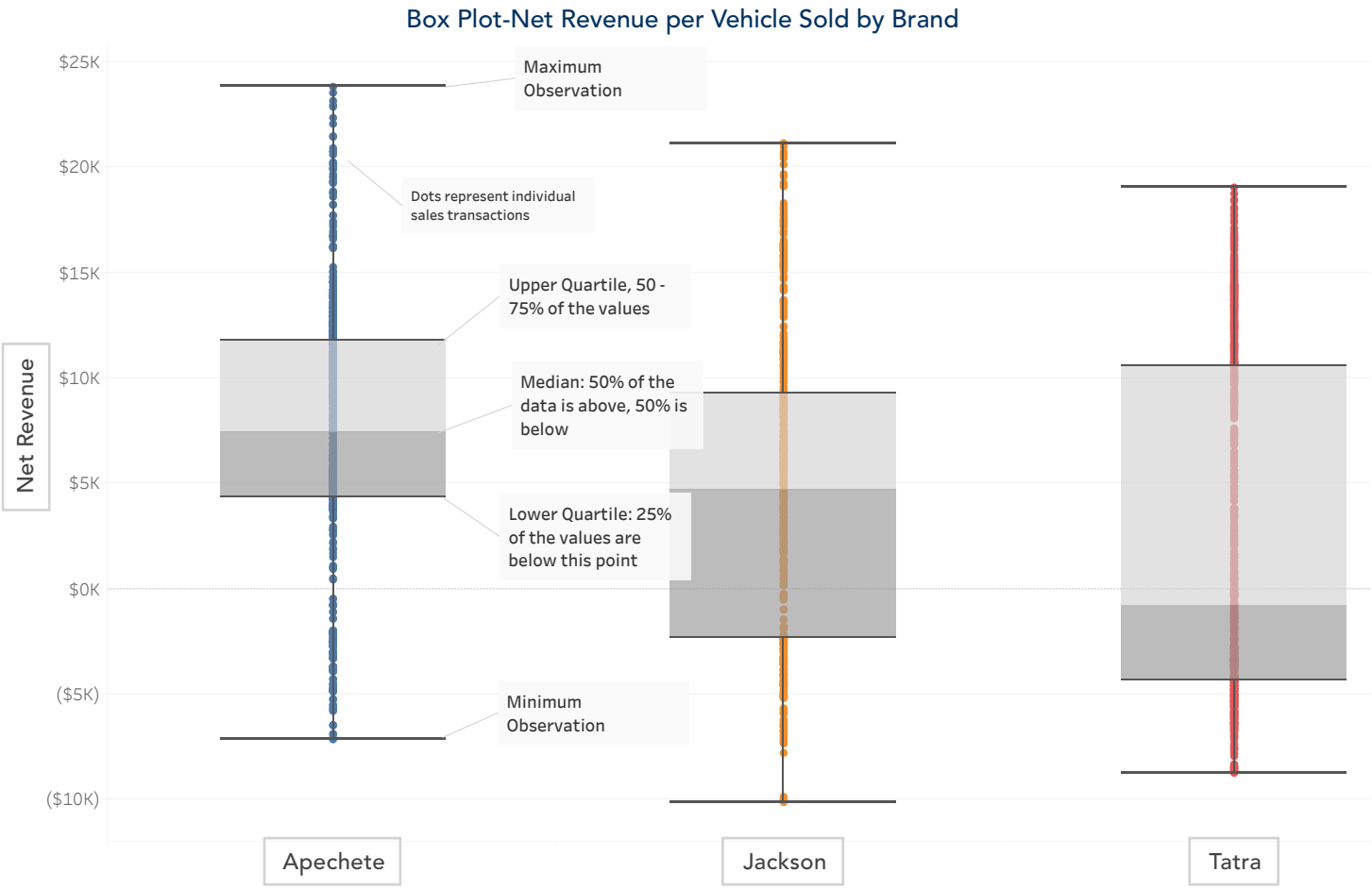
If the variable you are interested in is categorical, then you can use a bar chart to gain an understanding of the distribution. Continuing with the automobile manufacturer, if you want to understand the distribution of net revenue by brand, you can use a bar chart like the one in Figure 4. Using this visualization, you can see that the Apechete brand has the highest average net revenue.

FIGURE 4: DISTRIBUTION OF CATEGORICAL VARIABLES



Although Figure 4 does provide a quick visualization of the distribution of net revenue among the brands, it does not consider the distribution within the brands. A box plot is a great way to visualize distribution within categories. Also known as a box and whiskers plot, it provides a graphical illustration based on the minimum, lower quartile, median, upper quartile, and maximum values in the data. Figure 5 provides a box plot visualization of net revenue distribution within the brand categories.

FIGURE 5: BOX PLOT TO SHOW DISTRIBUTION AMONG AND WITHIN CATEGORIES

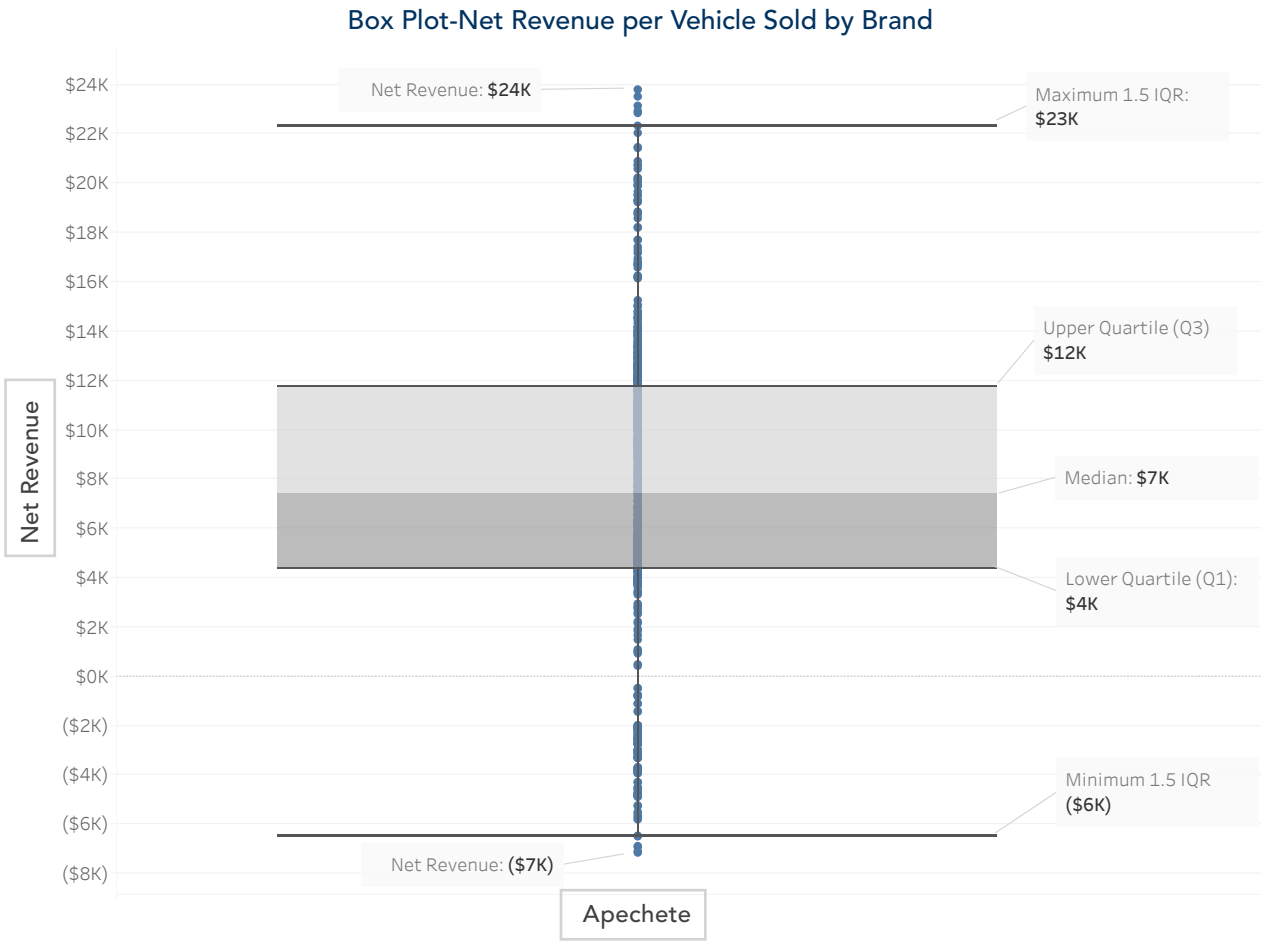


In Figure 5, the line at the top of the visual represents the highest observation, and the line at the bottom is the lowest observation. The box in the center of the visual represents the upper quartile, median, and lower quartile. The dots represent each observation in the data set. By examining Figure 5, you can see that all three brands have a large distribution of net revenue per sale.

It is also possible to create a box plot to help identify potential outliers in the distribution. By setting the maximum and minimum to 1.5 of the interquartile range (IQR) (instead of the maximum and minimum observation), you will be able to see if there are observations higher than the maximum IQR or lower than the minimum IQR.[6] Figure 6 displays the box plot for the Apechete brand using this method. You can now see that there are a few observations that are outside of the maximum and minimum IQR. It is likely that you will need to investigate these observations further.

[6] A commonly used rule is that a data point is an outlier if it is higher than 1.5 IQR or lower than 1.5 IQR.

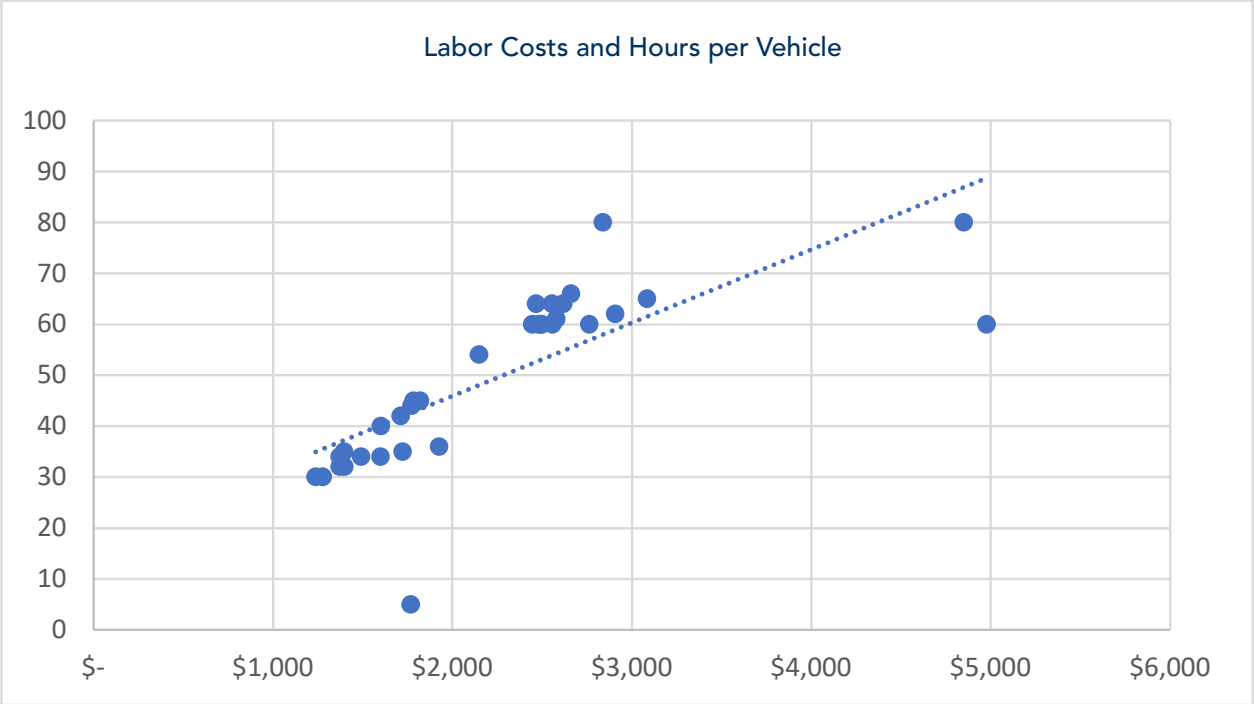
FIGURE 6: BOX AND WHISKERS PLOT WITH MAXIMUM AND MINIMUM INTERQUARTILE RANGES



Investigate Relationships

The third activity of exploratory data visualization is to investigate relationships. Data visualization can help you either to identify relationships you expect or to discover new relationships in the data. Whereas univariate data analysis involves examining the distribution of one variable, bivariate data analysis allows you to examine the relationship between two variables. Figure 7 is a visualization exploring the relationship between labor hours and labor costs for an automobile manufacturing company. The initial hypothesis is that there is a linear relationship between labor costs and labor hours.

FIGURE 7: BIVARIATE DATA ANALYSIS

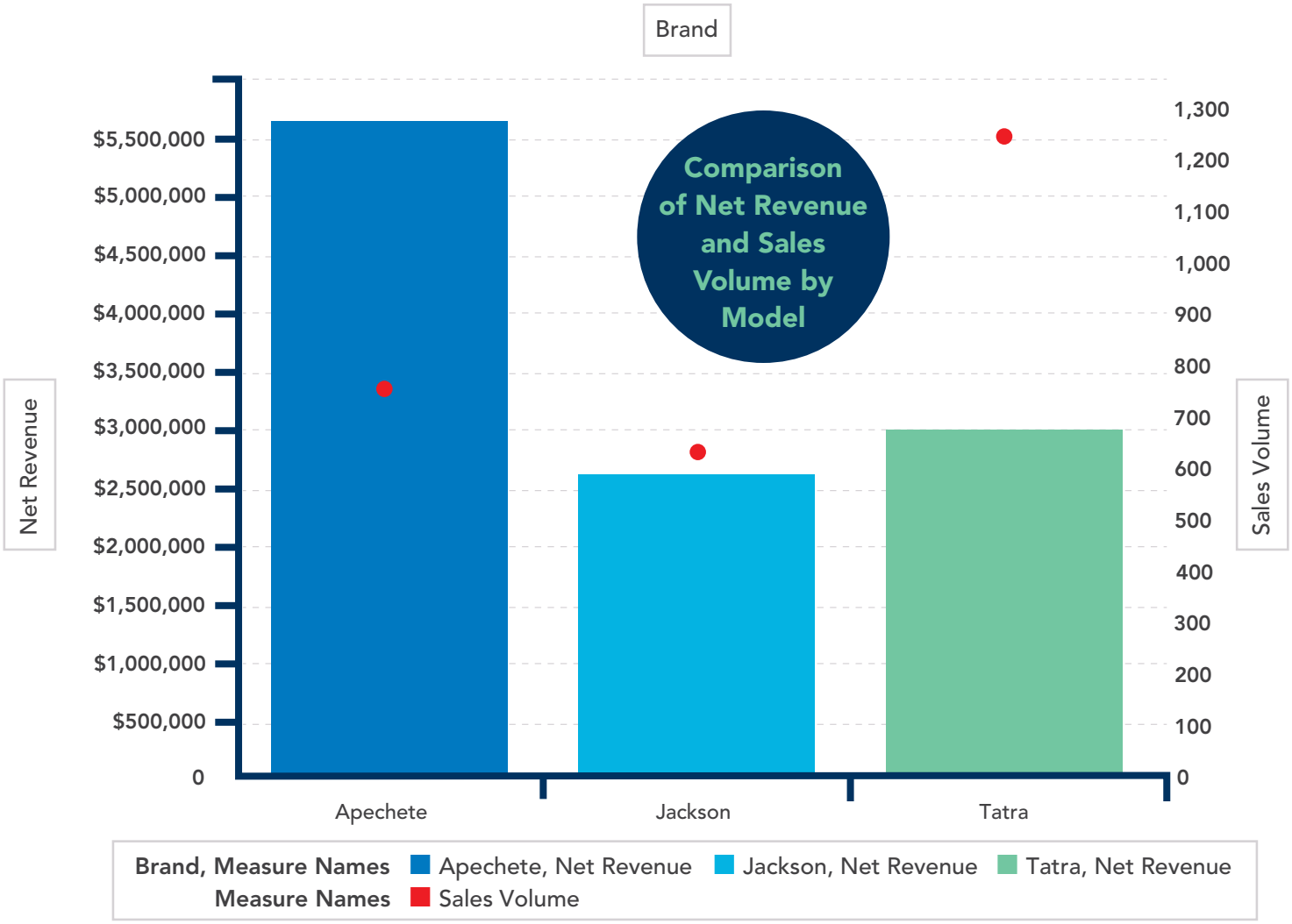


As can be seen in the visualization, there does appear to be a linear relationship, as most of the points are around the linear trendline. But there are two points that are well outside the linear trendline. By visualizing labor costs and labor hours per vehicle, you can confirm a linear relationship as well as identify two potential outliers for further investigation.

Oftentimes we would like to examine the relationship between two variables with different magnitudes and scales of measurement. A dual-axis chart is a useful tool to use in this situation.[7] The visualization in Figure 8 combines an analysis of net revenue and sales volume.

[7] Note that when using dual-axis visualizations for explanatory purposes, the reader might get confused. This is discussed in more detail later in the explanatory visualization section.

FIGURE 8: DUAL-AXIS CHART

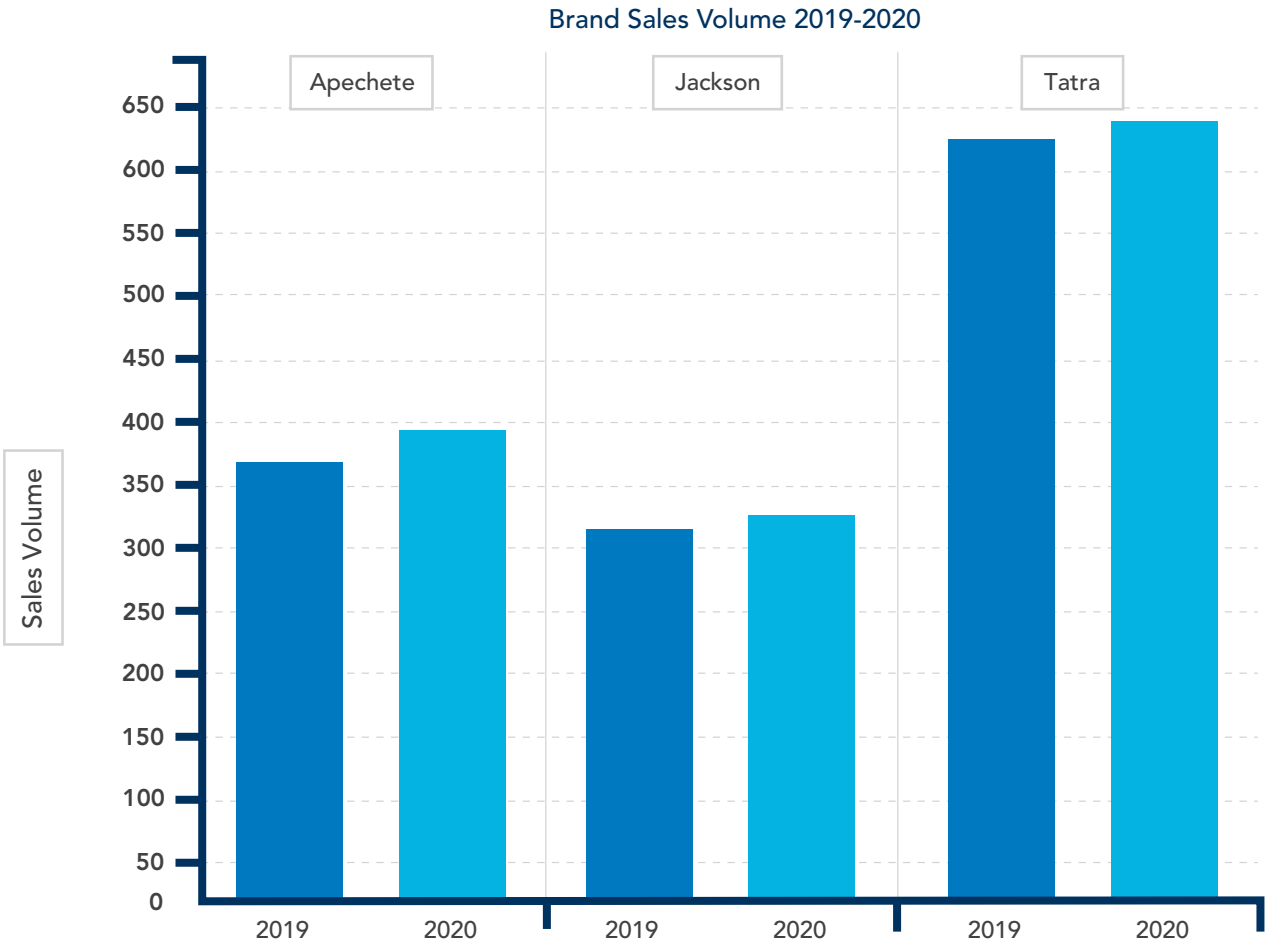


Identify Patterns

The fourth activity in data exploration is to explore the data to identify patterns or trends. Sometimes you may be exploring the data to see if an expected pattern exists (e.g., sales are highest in the fourth quarter) or you may be exploring the data to identify unexpected patterns. Two of the most common types of visualizations for investigating patterns are bar charts and line charts. Both types of visualizations can show changes over time and help to see if patterns exist.

Figure 9 is an example of a bar chart used to identify sales trends for the three car models we have been investigating. The purpose of analysis is to determine if sales are increasing or decreasing.

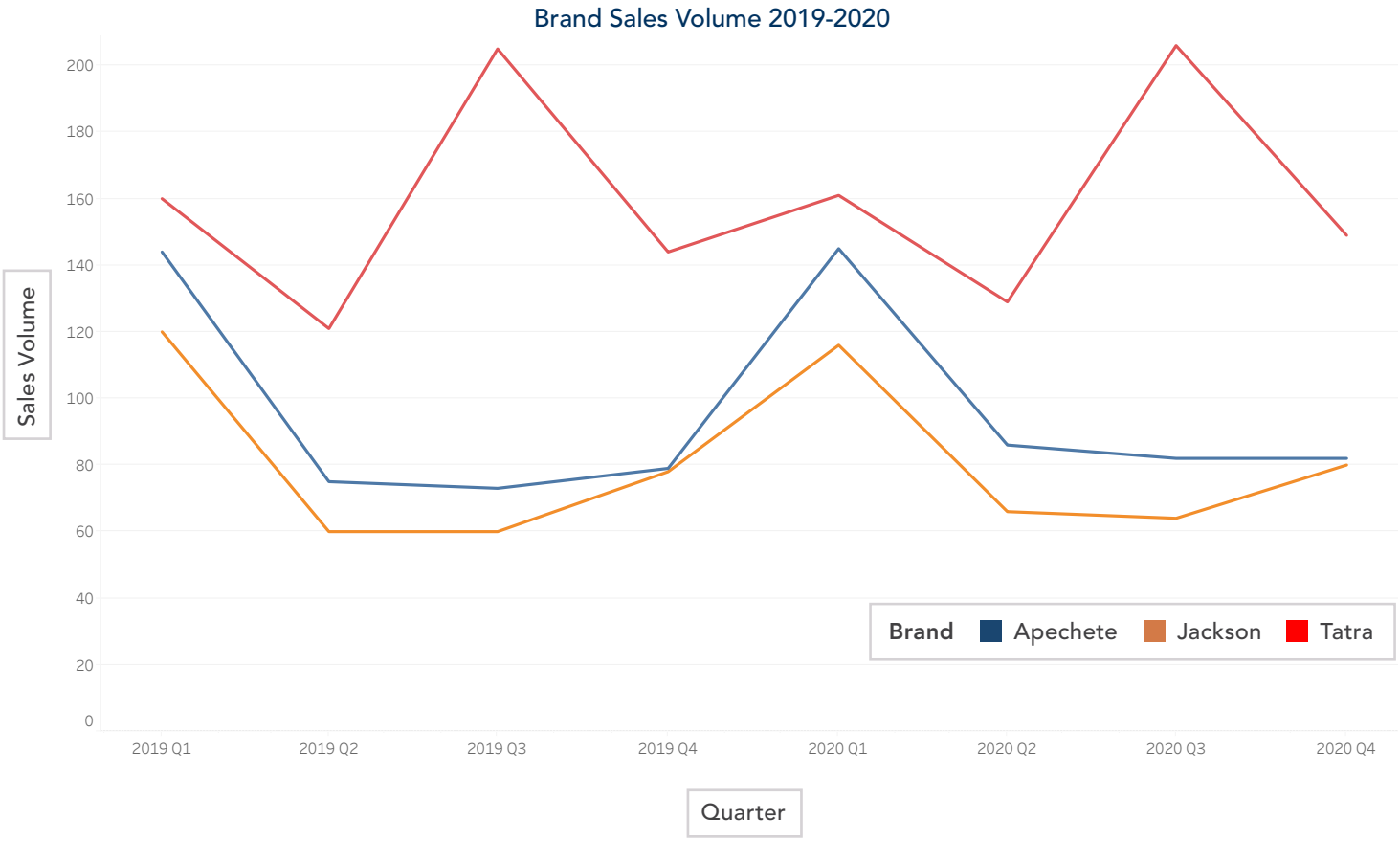
FIGURE 9: BAR CHART TO VISUALIZE SALES VOLUME TRENDS



As you can see in Figure 9, sales volume is somewhat flat year to year, with a slight increase in each of the brands in 2020.

Although bar charts can be useful to identify patterns, if you want to see patterns in greater granularity, a line chart is usually a better choice. For example, if you want to know what sales volume is over time by quarter, a line chart would be a good choice. Figure 10 provides an example of a line chart using sales volume by brand and quarter.

FIGURE 10: LINE CHART TO VISUALIZE SALES VOLUME PATTERNS



The line chart allows you to see which quarters have high sales volume and which ones have low sales volume. Figure 10 shows that Tatra has the highest sales volume compared to the Apechete and Jackson models. You can also see that Tatra sales are highest in the third quarter of each year, whereas the Apechete and Jackson models have a peak in the first quarter of each year.

In summary, exploratory data visualization comprises four activities. The first is to examine the data to be sure that it makes sense and is accurate. After that, you can use data exploration to investigate the distribution of your data, investigate relationships among the data, and identify patterns and trends in the data.

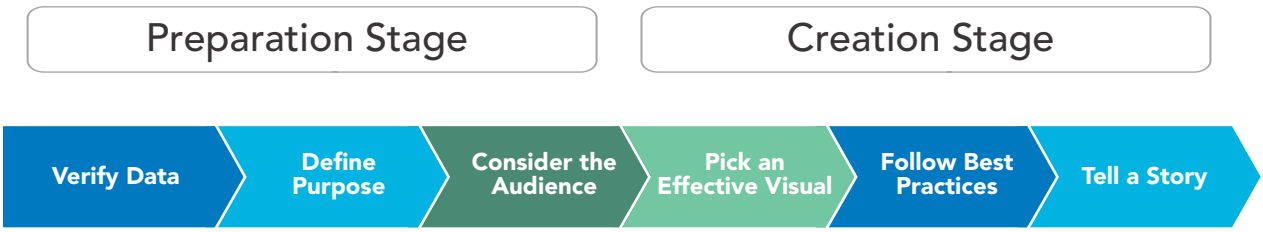
As depicted in Figure 11, exploratory data visualization helps you to discover insights about your data. To share those insights, you need to use explanatory data visualization.

FIGURE 11: MOVING FROM EXPLORATORY TO EXPLANATORY DATA VISUALIZATION

Explanatory Data Visualization

Explanatory data visualizations are used to communicate results of your analyses. Well-designed explanatory visualizations convey the results of your analyses in a clear and concise manner. The goal of explanatory data visualization is to communicate findings and inspire action. The process of explanatory data visualization is depicted in Figure 12.

FIGURE 12: EXPLANATORY DATA VISUALIZATION PROCESS



The first three steps in the process can be thought of as the preparation stage. Here you are making sure you have the right data and that the data meet the criteria for the analysis. You are also considering the purpose of the analysis and for whom the analysis is being prepared. Once you have addressed those steps, you can move to the creation stage. In this stage, you are choosing an effective visual, using best practices, and telling a story to effectively convey your results.

Preparation Stage

As just discussed, this stage is where you verify the data, define the purpose, and consider the audience for your explanatory data visualization.

Step 1: Verify the Data. This step in the process cannot be overemphasized. The adage "garbage in, garbage out" is relevant in all data analyses including data visualization. If the data you are using in your visualizations is incorrect or incomplete, your visualizations will also be incorrect and potentially misleading. Data should meet the following criteria:

- Accuracy: Are the data error-free?
- Completeness: Do you have all the data?
- Consistency: Are the data in a consistent format?
- Freshness: Are the data the most recent data available?
- Timeliness: Will the data you need be accessible and available?

Accuracy of the data means that the data are error-free, reliable, and representative of the phenomenon you are visualizing. Completeness of data means that you have all the data that you are supposed to have. In other words, you are not missing any data. Consistency of data means there are no format inconsistencies. For example, dates are formatted the same way from period to period, and values are in the same denomination across time periods. Freshness of data means that you are using the most recent data in your visualizations. Finally, timeliness means that the data you need for your visualization will be available and accessible when you need it.

Step 2: Define the Purpose. Before beginning your explanatory data visualizations, you need to be sure you have specifically identified the purpose of your analysis. During the exploratory phase of analysis, you likely will have identified your overall purpose and prepared many visualizations as part of your discovery process. It is very tempting to use every visualization that you created during the exploratory phase. Avoid falling into this trap. The purpose of exploration is to find insights. You should share the insights with your audience, rather than sharing the entire journey you took to get there.

Stephen Few, in his book *Show Me the Numbers*, identifies four purposes for tables and graphs of quantitative business information:[8]

- Analyzing
- Communicating
- Monitoring
- Planning

Think about what your purpose is and choose only the visualizations that align with that purpose. In combination with purpose, you should consider the audience for your data visualizations. Is the purpose to inform the audience or to persuade the audience?

Step 3: Consider the Audience. When preparing visualizations, always keep in mind the audience that will be viewing the visualization. Who will be consuming the information? How much background do they have regarding the topic? Will they be the decision maker? Answering these questions will help you focus your visualizations so that they achieve maximum impact. For example, if you have prepared an analysis that will be shared internally to colleagues who know the background information related to the analysis, you will not need to include a lot of background information. But if the analysis is prepared for an audience unfamiliar with the topic, you will likely need to provide more detailed background information for them to understand the analysis. An important aspect of providing an analysis that your audience will understand is to be sure you pick an effective visualization.

Creation Stage

Once you are confident about your data, have clearly defined the purpose of the analysis, and know your audience, you can begin to create the visualizations you will use to convey your results.

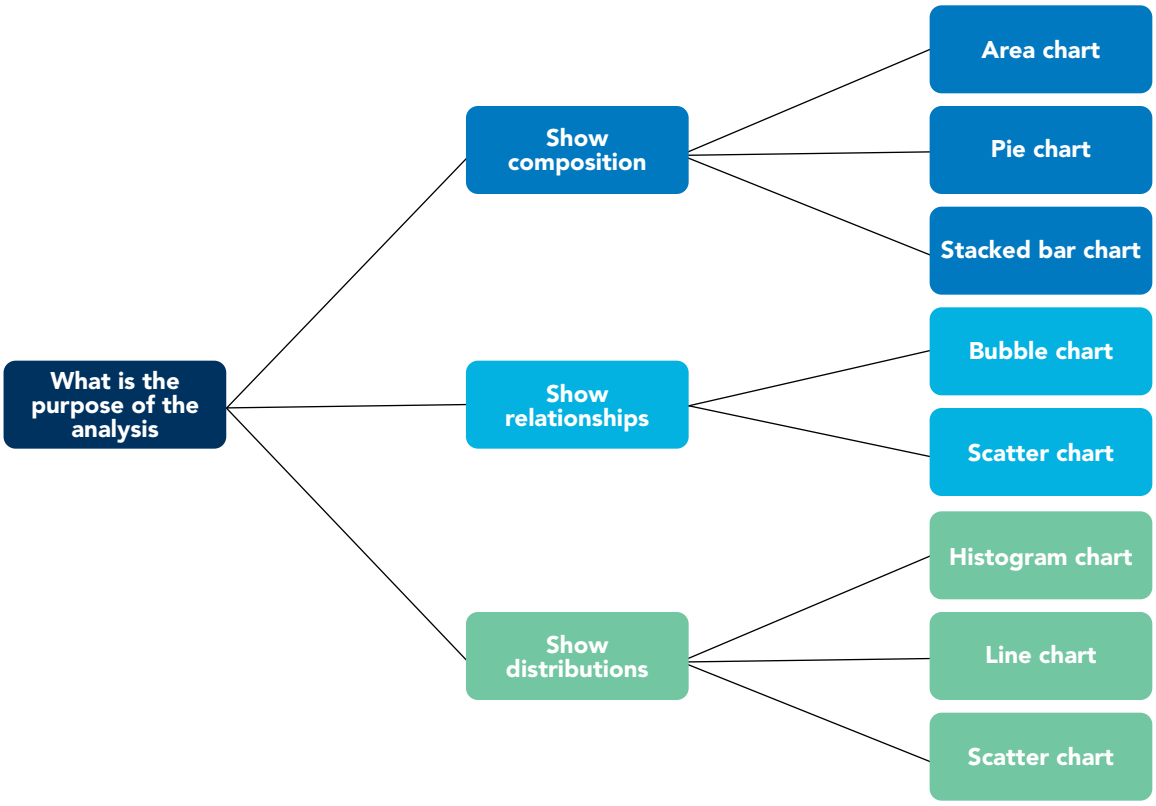
[8] Stephen C. Few, *Show Me the Numbers*, Analytics Press, El Dorado Hills, Calif., 2004.

Step 4: Pick an Effective Visual. This part of data visualization is easier said than done. Most of us have never been trained in data visualization and have little knowledge as to what type of visual is most effective. It helps to break this down into the following steps:

a. Consider the purpose of your analysis.
b. Identify visuals that will meet that purpose.
c. Understand the data that is needed to create the visual you would like.
d. Match the purpose and the data to the appropriate visual.
e. Be sure to follow best practices for your chosen visual.

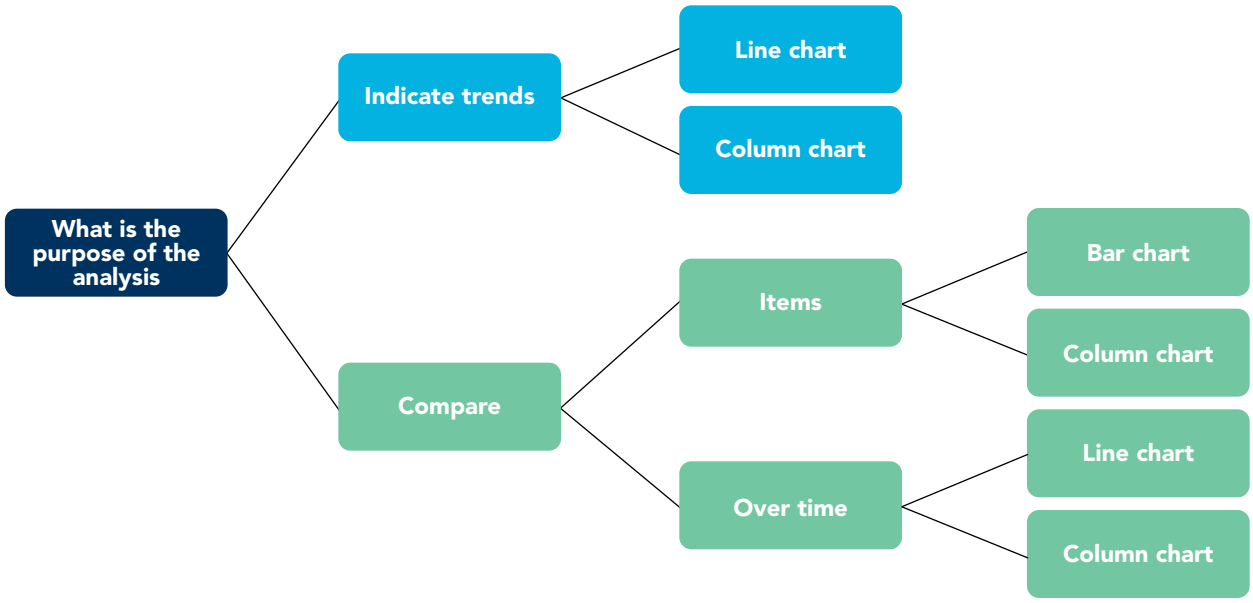
Determining the purpose of the analysis is critical so that you can pick an effective visual. There are many potential purposes for data analysis. Figure 13 is a decision tree that provides a way to identify appropriate visualizations if the purpose of the analysis is to show composition, relationships, or distributions. Figure 14 provides guidance when the purpose is to show trends or comparisons.

FIGURE 13: VISUALIZATION CHOICES GUIDE TO SHOW COMPOSITION, RELATIONSHIPS, OR DISTRIBUTIONS[9]



[9] Ann C. Dzuranin, Guido Geerts, and Margarita Lenk, *Data and Analytics in Accounting: An Integrated Approach*, John Wiley & Sons, Hoboken, N.J., forthcoming.

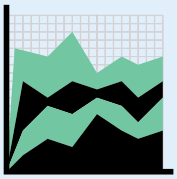
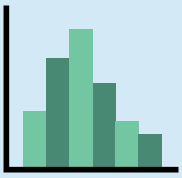
FIGURE 14: VISUALIZATION CHOICES GUIDE TO SHOW TRENDS OR COMPARISONS[10]



Once you have identified your purpose (step a), you then can narrow down to the types of appropriate visualizations (step b). At this point, you will need to be sure you understand the type of data required for the possible visualizations (step c). The best visualization will be the one that matches the purpose and works with the data that you need to visualize (step d). Finally, you should be sure to follow best practices for the type of visualization that you decide to use (step e). Table 6 provides an explanation of each type of visualization, how it is used, the type of data needed, and best practices.

[10] *Ibid.*

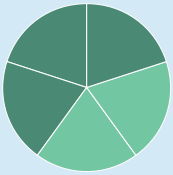
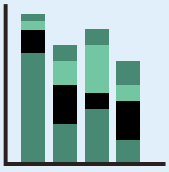
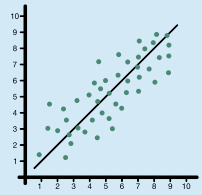
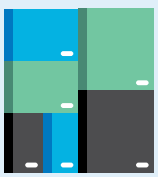
TABLE 6: COMMON DATA VISUALIZATIONS, DATA REQUIRED, AND BEST PRACTICES[11]

Visualization	Use	Data Required	Best Practices
Area chart 	To represent changes in volume over time	• Date field • At least one quantitative measure	• To avoid confusion and clutter, do not use with data that have more than four categories. • Start the y-axis at zero or above. • Put highly variable data on the top and low variability on the bottom.
Bar chart  	To compare parts to a whole, highlight categories, or show changes over time	• Horizontal bars: zero or more categories, one or more measures • Vertical bars: one or more categories, one or more quantitative measures	• Compare two to seven categories with vertical bars. • Use horizontal bars if more than seven categories or long category labels. • Use horizontal labels for better readability. • Space bars appropriately and consistently. • Use color sparingly or as an accent color. • Always have a zero baseline.
Bubble chart 	To compare independent values that have distinct gaps or outliers	• One or more categories • One or more quantitative measures	• Label bubbles and make sure they are visible. • Scale bubble size by area and not diameter. • Do not use bubbles if they are all similar in size.
Histogram chart 	To show frequency distributions	• Must be numerical data	• Use a zero baseline. • Choose an appropriate number of bins. - Bins are numbers that represent the intervals into which you want to group the data. - Bins define the groups used for the frequency distribution. - There are generally between five and 15 bins.

Continued to page 25

[11] *Ibid.*

Continued from page 24

Line graph 	To display one or more series of data. Allows for the use of multiple data series and data points	• For continuous lines, you need one date, zero or more categories, and one or more quantitative measures.	• Time runs from left to right. • Be consistent in plotting time points. • Use solid lines, not dotted. • Use a zero baseline. • Do not plot more than four lines; instead use multiple charts.
Pie chart 	Use for illustrating simple part-to-whole relationships. Not a good visual to use to make precise comparisons	• One or more categories • One or two quantitative measures	• Most impactful with small data sets • Best to use when you want to show differences within groups based on one variable • Make sure the data adds to 100%. • Limit the chart to a maximum of five segments. • Start the first segment at the 12 o'clock position.
Stacked bar chart 	Use when comparing multiple part-to-whole relationships.	• One or more categories • One or more quantitative measures	• Can be vertical or horizontal • Follow same best practices as bar charts.
Scatter chart 	Use to highlight correlation and distribution of large amounts of data.	• Zero or more categories • One or more quantitative measures	• Data set should be in pairs with an independent variable (x-axis) and a dependent variable (y-axis). • Use if order is not relevant–otherwise use a line graph. • Do not use if there are only a few pieces of data or if there is no correlation.
Tree map 	Use to visualize a part-to-whole relationship among a large number of categories.	• One or more categories • One or two quantitative measures	• Appropriate when precise comparisons are not important • Use bright, contrasting colors so that each box is easily defined. • Label boxes with text or numbers.

Once you have chosen the type of visualization you will use and have followed the best practices, you still have some work to do to create an effective visualization. What is an effective visualization? Edward Tufte, a pioneer in the field of data visualization, states that "excellence in statistical graphics consists of complex ideas communicated with clarity, precision and efficiency."[12] Tufte goes on to say that a well-designed data visualization should give the viewer "the greatest number of ideas in the shortest time with the least ink in the smallest space."

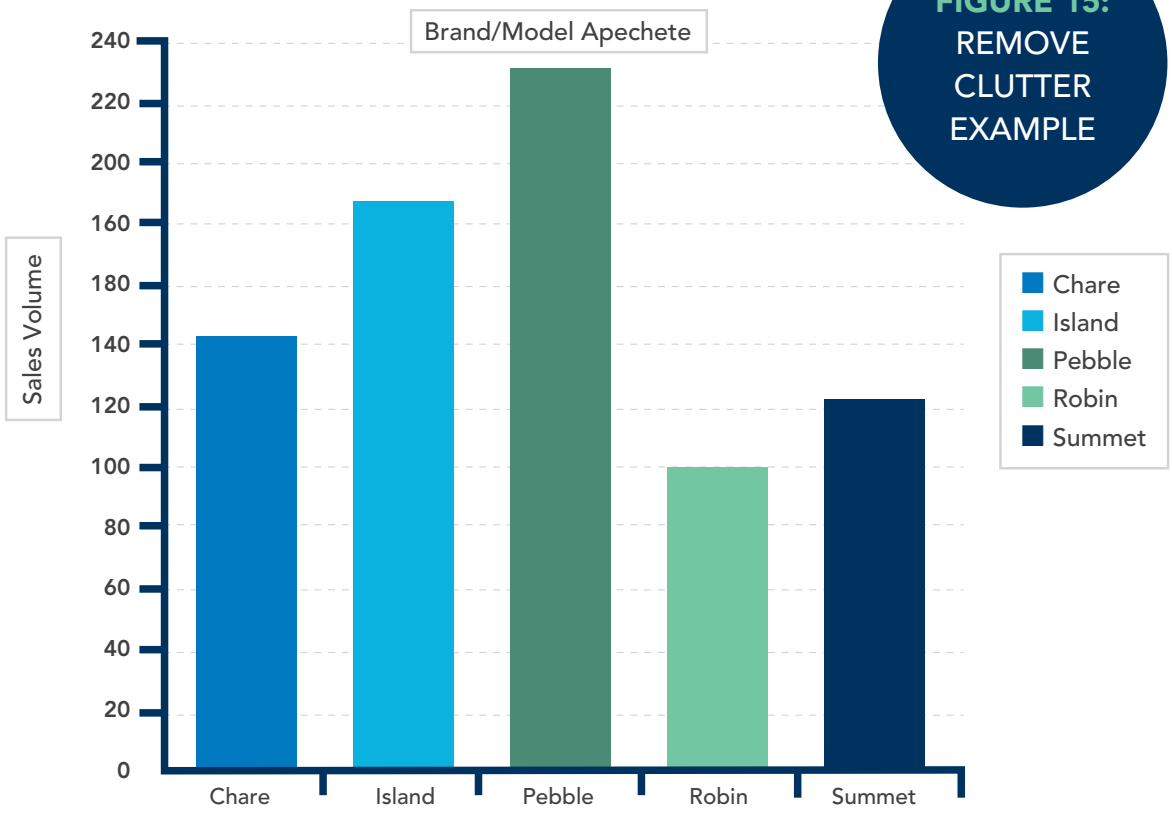
Step 5: Follow Best Practices. Step 5 in the exploratory data visualization process is to consider visualization best practices. Along with following the best practices given in Table 6 based for the type of visualization, there are other best practices to consider. The first is to avoid cluttering your visual with nonessential information. The second is to use best practices to help the audience focus its attention on what is most important in your visualization.

• **Avoid Clutter**. Clutter is the enemy of a good visualization. Another piece of advice from Tufte is to maximize the data-to-ink ratio, meaning you should remove any nondata-related ink and any redundant ink. Figure 15 provides a before-and-after view following the elimination of nondata and redundant ink. The purpose of the visual is to understand sales volume of the models offered for the Apechete model.

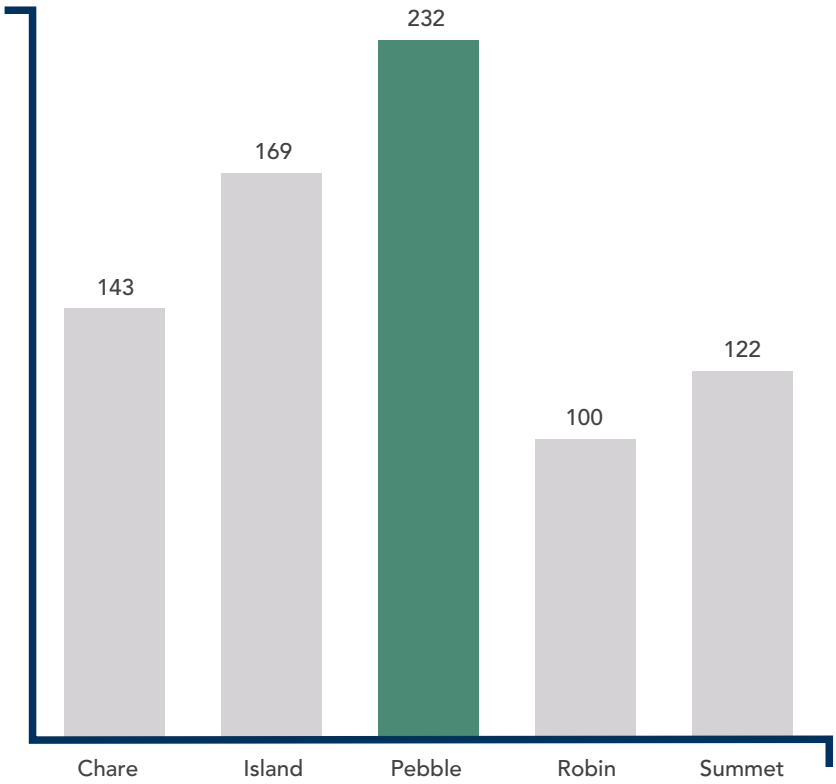
[12] Edward R. Tufte, *The Visual Display of Quantitative Information*, Second Edition, Graphics Press, Cheshire, Conn., 2001.

FIGURE 15: REMOVE CLUTTER EXAMPLE

BEFORE: Apechete Models Sales Volume



AFTER: Apechete Models Sales Volume



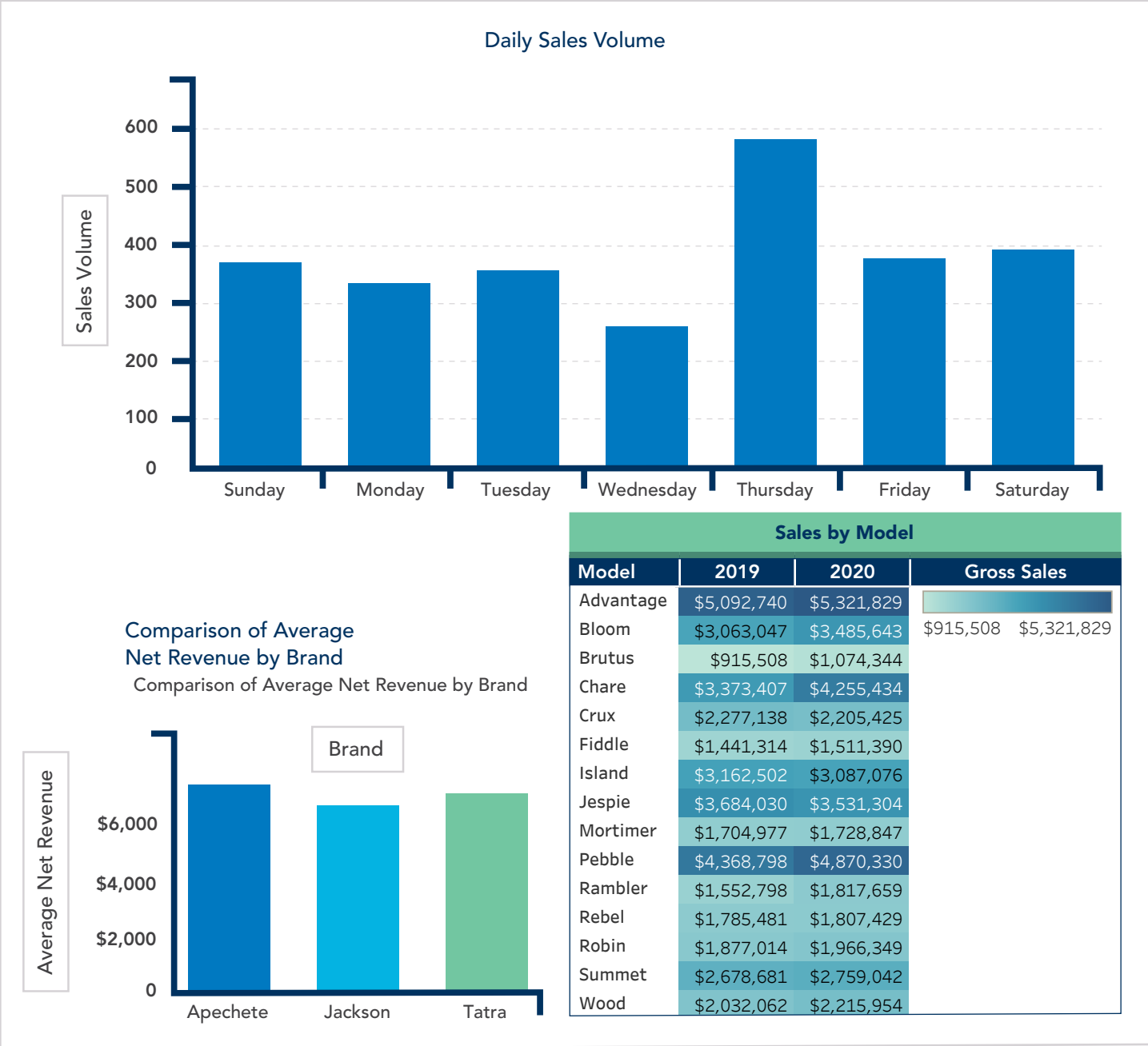
Recall that the purpose of the visualization is to show how the various models are performing for the Apechete model. The "before" graph has too much color, and the legend is redundant because the columns are labeled. The "after" graph relays the same information in a much more concise and clear way. The legend has been removed because the title of the graph provides that context. The sales volume amounts are shown at the top of the columns, eliminating the need for the scale on the left of the graph. Highlighting the best-selling model by graying out the remaining models quickly brings the reader's attention to the highest-selling model. The background gridlines and shading have been removed. In general, when it comes to the amount of ink in a visualization, less is more.

• **Focus Attention**. The "after" graph in Figure 15 is an example of how to create a visualization that will focus your audience's attention on the best-selling brand. It is important to keep in mind preattentive attributes when creating explanatory data visualizations. Preattentive attributes are visual properties that we notice without realizing we are doing so. Attributes such as size, color, and position can be leveraged to help direct your audience's attention to where you want them to focus. Studies have shown that an audience will spend only three to eight seconds to decide whether they want to continue to look at a visualization or turn their attention to something else.[13]

Size is an indication of importance. When creating visualizations, keep in mind that relative size will be interpreted as relative importance. This is true in individual visualizations as well as in dashboards. If one aspect of the dashboard is larger than the other aspects, your focus will be on the larger of the images (see Figure 16). In the dashboard, the visualization showing daily sales volume is the largest visual in the dashboard. Users will be drawn to that aspect of the dashboard first, and, given its size relative to the other two visuals, it will appear to be the most important.

[13] Cole Nussbaumer Knaflic, *Storytelling with Data*, John Wiley & Sons, Hoboken, N.J., 2015.

FIGURE 16: DASHBOARD



Not only will the size of the visualization indicate importance, the size of the text in the illustration relative to other text will indicate importance as well.

Color should be used sparingly and consistently. A good rule of thumb to consider is that if adding color does not add to the interpretation of the visualization, you likely do not need it. Also keep in mind that color evokes emotion, so be thoughtful of what the tone of the color conveys. For example, red evokes a sense of urgency. In general, avoid using too many colors together in a visualization because it will make it more difficult for the audience to discern what you are trying to convey. For example,

Table 7 is the same as the "sales by model" visual in the dashboard in Figure 16. Notice how much more difficult it is to interpret the table when more colors are added.

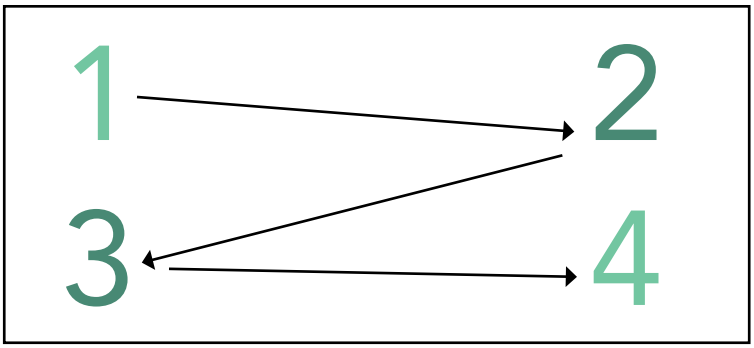
TABLE 7: OVERUSE OF COLOR EXAMPLE

Sales by Model			
Model	2019	2020	Gross Sales
Advantage	$5,092,740	$5,321,829	$915,508 $5,321,829
Bloom	$3,063,047	$3,485,643	
Brutus	$915,508	$1,074,344	
Chare	$3,373,407	$4,255,434	
Crux	$2,277,138	$2,205,425	
Fiddle	$1,441,314	$1,511,390	
Island	$3,162,502	$3,087,076	
Jespie	$3,684,030	$3,531,304	
Mortimer	$1,704,977	$1,728,847	
Pebble	$4,368,798	$4,870,330	
Rambler	$1,552,798	$1,817,659	
Rebel	$1,785,481	$1,807,429	
Robin	$1,877,014	$1,966,349	
Summet	$2,678,681	$2,759,042	
Wood	$2,032,062	$2,215,954	

Also keep in mind the view of colorblind users when creating visualizations (avoid red and green shades).

Position of where you place items in your visualizations and dashboards matters. Most of your audience is going to start at the top left of the visual and then scan with their eyes in a zigzag motion (see Figure 17). Therefore, be sure to put the most important information at the top left of the visualization.

FIGURE 17: THE WAY INFORMATION IS VIEWED ON A SCREEN OR PAGE

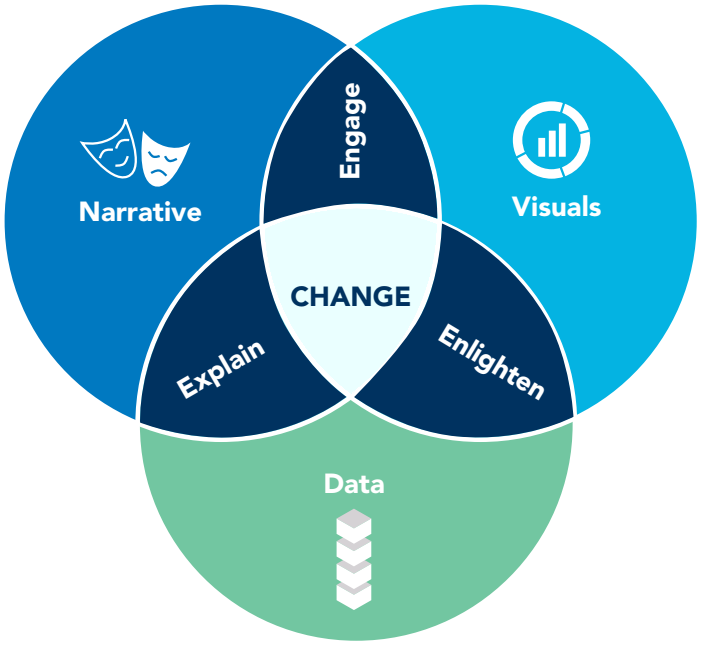


Following these best practices will help you to create effective visualizations. The true power of data visualization does not end here, however. The real power comes when great data visualizations are used to tell a story that the reader not only will remember but will act upon.

Step 6: Tell a Story. The last step in explanatory data visualization is to tell a story. Storytelling with data is an essential skill for finance professionals.[14] Research has shown that storytelling helps us to remember. Research by cognitive psychologist Jerome Bruner indicated that we are 22 times more likely to remember facts that have been told in a story.[15] The reason is that stories are memorable and help us to understand an idea more quickly. Research has shown that when we hear or read a story, our brains are activated at many levels. The emotional part of the brain releases chemicals to stimulate feelings of connection, reward, and recognition.[16] The powerful combination of data into an engaging story will help to ignite your audience into action.[17]

• **Elements of a Data Story**. There are three elements to a data story: data, narrative, and visuals. The author of *Effective Data Storytelling*, Brent Dykes, describes how these elements explain, enlighten, and engage the audience. The combination of all three elements is what leads to change. Figure 18 portrays the connection among the three elements.

FIGURE 18: ELEMENTS OF EFFECTIVE STORYTELLING[18]



[14] Fatema El-Wakeel, Loreal Jiles, and Raef Lawson, "Storytelling with Data Visualization," *Strategic Finance*, December 2020, **bit.ly/3kR1Tyt**.
[15] Vanessa Boris, "What Makes Storytelling So Effective for Learning?" Harvard Business Publishing, Corporate Learning, December 20, 2017, **s.hbr.org/3h2oMhs**.
[16] Pamela Rutledge, "The Psychological Power of Storytelling," *Psychology Today*, January 16, 2011, **bit.ly/2WMAtlw**.
[17] Kate Harrison, "A Good Presentation Is about Data and Story," *Forbes*, January 20, 2015, **bit.ly/3zKJBVJ**.
[18] Brent Dykes, *Effective Data Storytelling*, John Wiley & Sons, Hoboken, N.J., 2020.

The intersection of data and narrative explains your data story. It provides the context and commentary needed to understand the results of your analysis. The intersection of data and visuals enlightens the reader to the insights they may not have had were it not for the visualizations. In fact, humans process visual images 60,000 times faster than text.[19]

The combination of narrative with your visuals engages your audience in the story. Throughout our lives much of our learning is based on a combination of narrative and visuals. A good story can hold the attention of the reader and increases the likelihood of action.

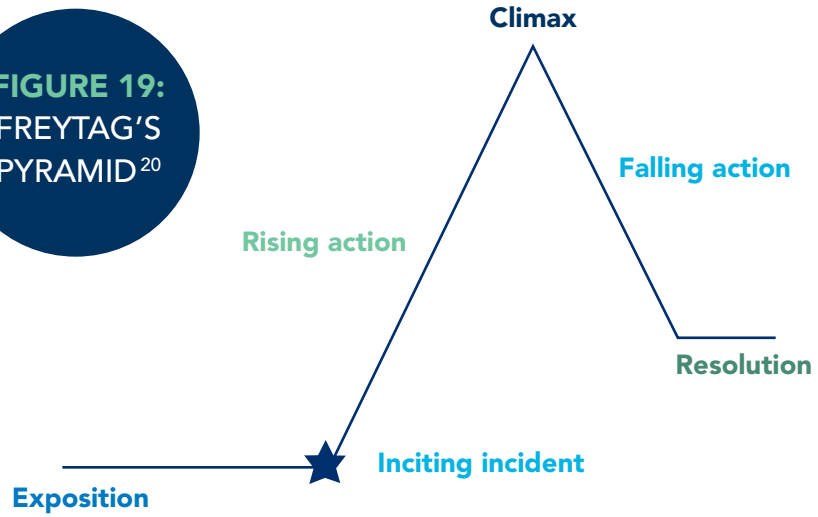
The key to a good data story is to include all three elements and then to structure your story effectively.

• **Structure of a Data Story**. How you structure your story is extremely important. You want to be sure you create a memorable and relatable story that leads to action. A best practice for developing your data story is to first create a visual outline of the content you plan to create/use. Create the outline on paper or a whiteboard rather than starting with creating slides. That will help you avoid getting sucked into creating slide after slide or visual after visual without direction. Although you may believe every visual you created and every step you took in a data analysis is critical, only the key analyses and visuals should be included in the data story.

A common structure used in storytelling is called Freytag's Pyramid, sometimes referred to as a storytelling arc. The pyramid was developed by the German playwright and novelist Gustav Freytag in 1863. The basic structure is displayed in Figure 19.

The exposition is the introduction to the characters and sets the scene for the story. Next, the rising

FIGURE 19: FREYTAG'S PYRAMID[20]



action is introduced via a series of events to build up to the most climactic or important point of the story. After the climax, you see the rest of the events that unfold after the main conflict has occurred. Finally, you see the conclusion of the story, where all conflicts are resolved.

[19] Harris Eisenberg, "Humans Process Visual Data Better," Thermopylae Sciences + Technology, September 15, 2014, **bit.ly/3h1fKRB**.
[20] Sean Glatch, "The 5 Elements of Dramatic Structure: Understanding Freytag's Pyramid," writers.com, May 12, 2020, **bit.ly/3kU4RlM**.

How do we apply this structure to a data story? Let's use the fictitious automobile manufacturing company, Huskie Motor Corporation (HMC), as an example (see Table 8). Imagine you have been asked to do an analysis of the model performance for the Tatra brand. The Tatra brand is the lowest average net revenue HMC brand. You need to determine why and make recommendations for improvement.

TABLE 8: APPLYING FREYTAG'S PYRAMID

Freytag's Pyramid	Data Story Equivalent	HMC Example
Exposition	Introduction: Introduce the problem or issue that the story addresses.	Briefly discuss the background information relevant to the analysis. Keep it interesting so that you get the reader's attention. For example, "Four of our Tatra models had a total net loss of more than $2.3 million in 2020."
Rising action	Demonstration of the issue: Explore the subject of the analysis deeper.	Methodically build up to the reveal of the four models that are operating at a loss.
Climax	Reveal significant finding(s): This is the "aha" moment. Provide ideas for overcoming the issue or problem.	Identify the models that constitute the highest net loss. There are two models that have very high sales volume but have net losses.
Falling action	Describe the idea: Build toward your recommendation.	Identify what is causing the loss: rising variable costs, labor and materials, etc.
Resolution	Final recommendations/call to action: Conclude your story with your recommendations.	Provide suggestions for reducing costs. Explain the next steps needed.

Following a structure like Freytag's Pyramid is a clear way to structure your story. Once you have a structure, you can then apply the elements of a story (data, narrative, and visuals) to bring your story to life. •

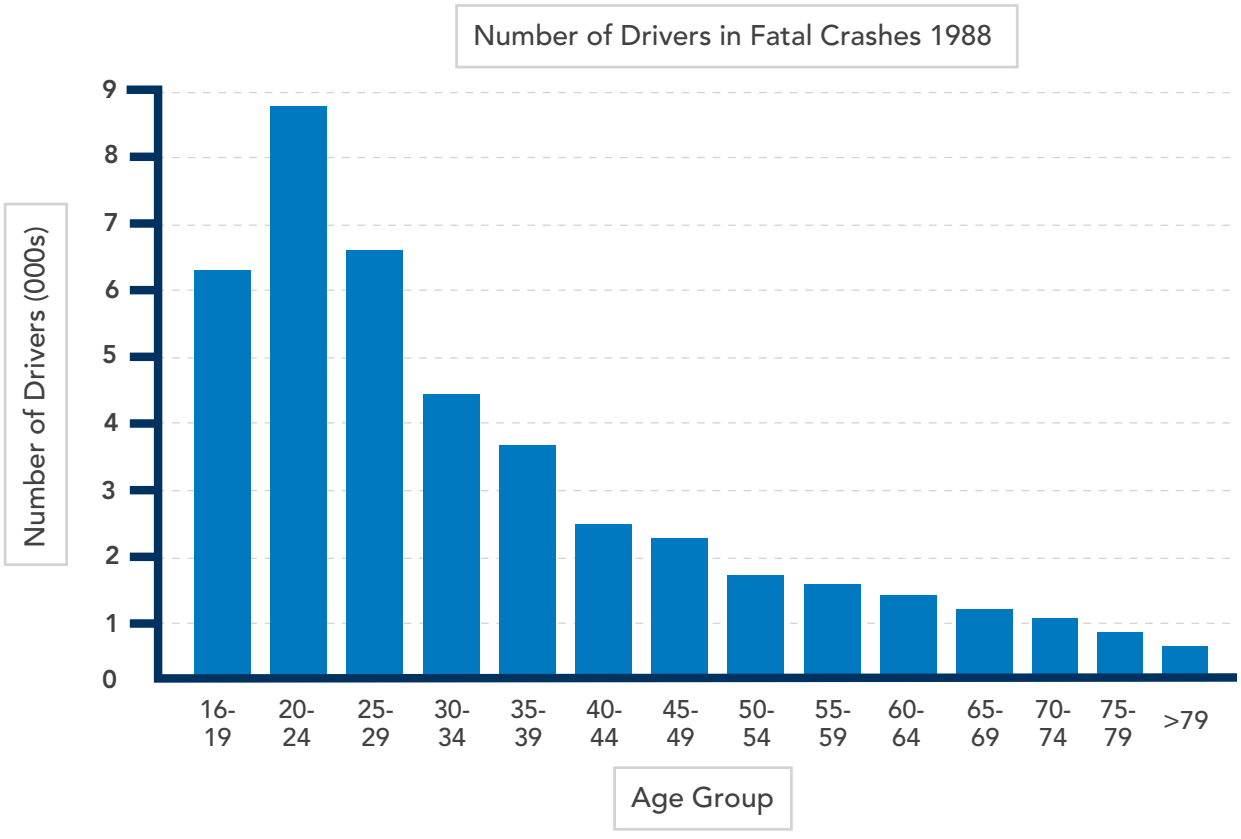
Other **Important Considerations**

In addition to applying best practices for data visualization and storytelling, be sure your data visualizations are not misleading. This last section discusses ethical issues to consider and how to avoid creating misleading visualizations.

Ethics and Data Visualization

Always avoid creating a visualization that will mislead the viewer. This could be done unintentionally or intentionally. Unintentional misleading visualizations are often the result of not following best practices or not providing relative context to the visual. For example, the visualization in Figure 20 was published in the *American Journal of Public Health* in 1989. This was a peer-reviewed research study, and the data are valid and accurate.

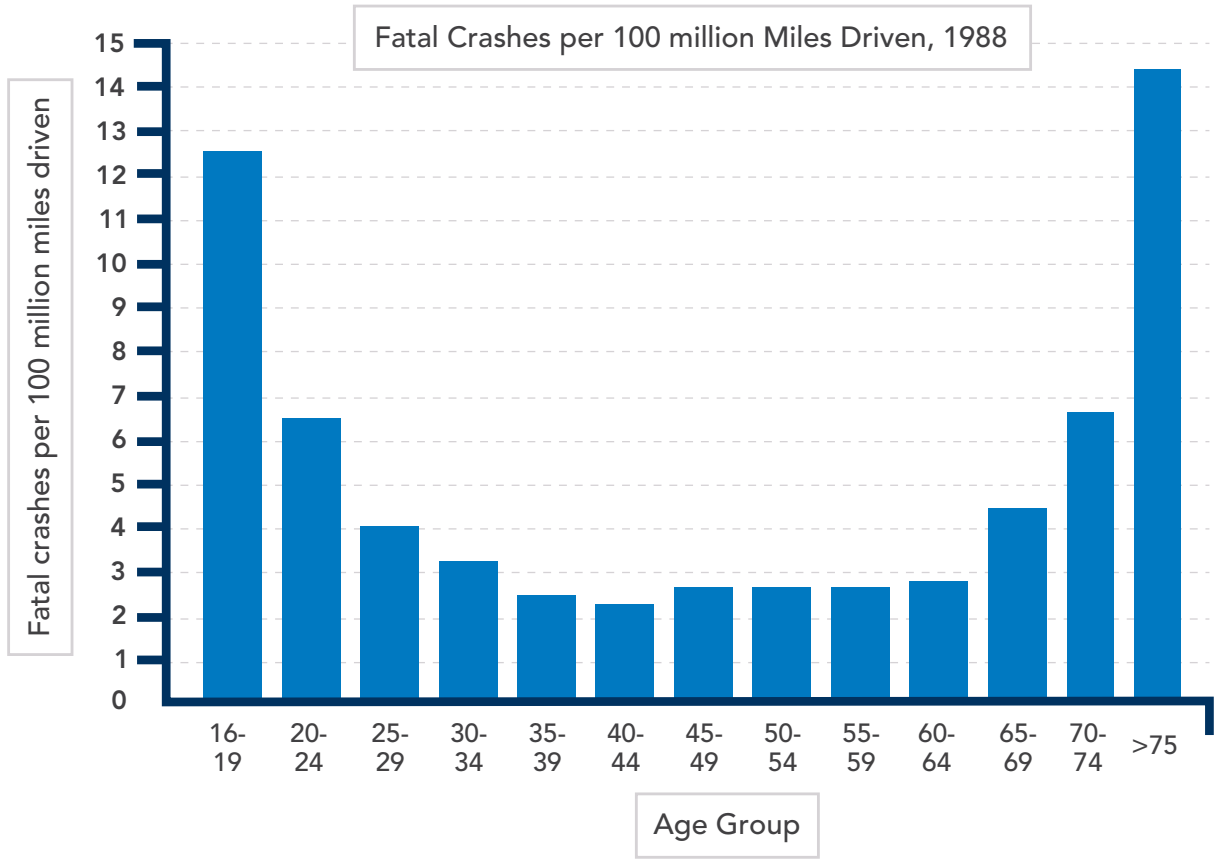
FIGURE 20: UNINTENTIONAL MISLEADING VISUALIZATION



Graph is based on data from Allan F. Williams and Oliver Carston, "Driver Age and Crash Involvement," *American Journal of Public Health*, March 1989, pp. 326-327.

If you were asked to determine which age group has the worst drivers, you would say that younger drivers are the worst and that as we age, we improve—so much so that the 79 and older age group appears to be the best. Logically, however, this does not make sense. The reason Figure 20 is misleading is that it doesn't show fatal crashes relative to the number of miles driven. The visual in Figure 21 puts the data into better context.

FIGURE 21: IMPROVED VISUALIZATION



Graph is based on data from Allan F. Williams and Oliver Carston, "Driver Age and Crash Involvement," *American Journal of Public Health*, March 1989, pp. 326-327.

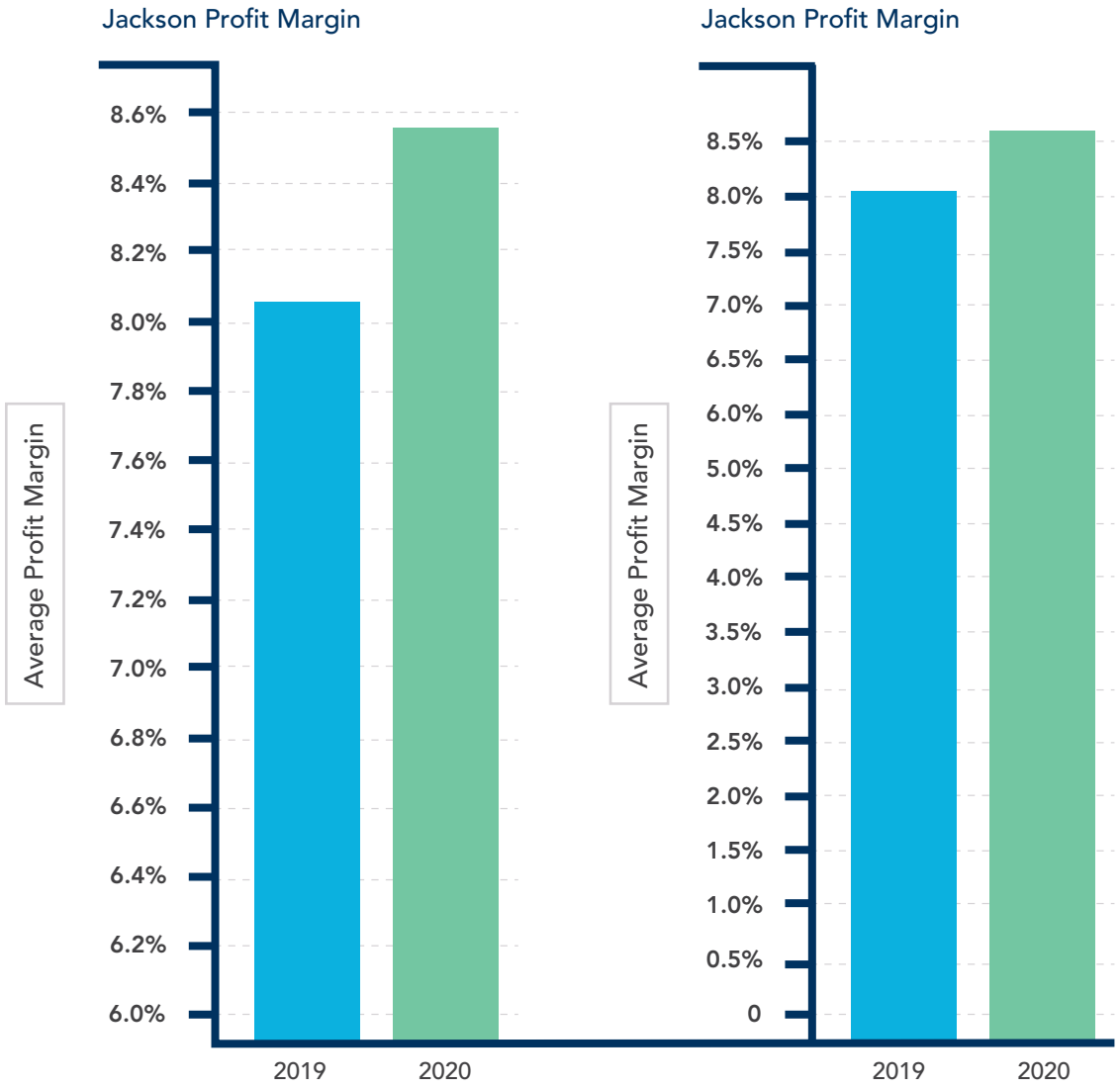
These visuals are an excellent example of how correct data can be used to create a misleading visualization. You can see evidence of misleading visualizations every day in the news and on social media. Visualizations have become a powerful way to capture the public's attention and are often used to promote a point of view rather than an unbiased view. Knowing how to interpret visualizations and how to spot potentially misleading visualizations are increasingly important skills. The next section covers best practices to avoid creating misleading visualizations. Knowing these best practices will not only help you to prepare fair and unbiased visualizations, it will also help you to spot those that are misleading.

Best Practices to Avoid Misleading the Reader

Avoid the following that can mislead the reader:

- Omitting the baseline
- Manipulating the y-axis
- Selectively picking the data
- Using the wrong type of graph
- Going against conventions

FIGURE 22: OMITTING THE BASELINE

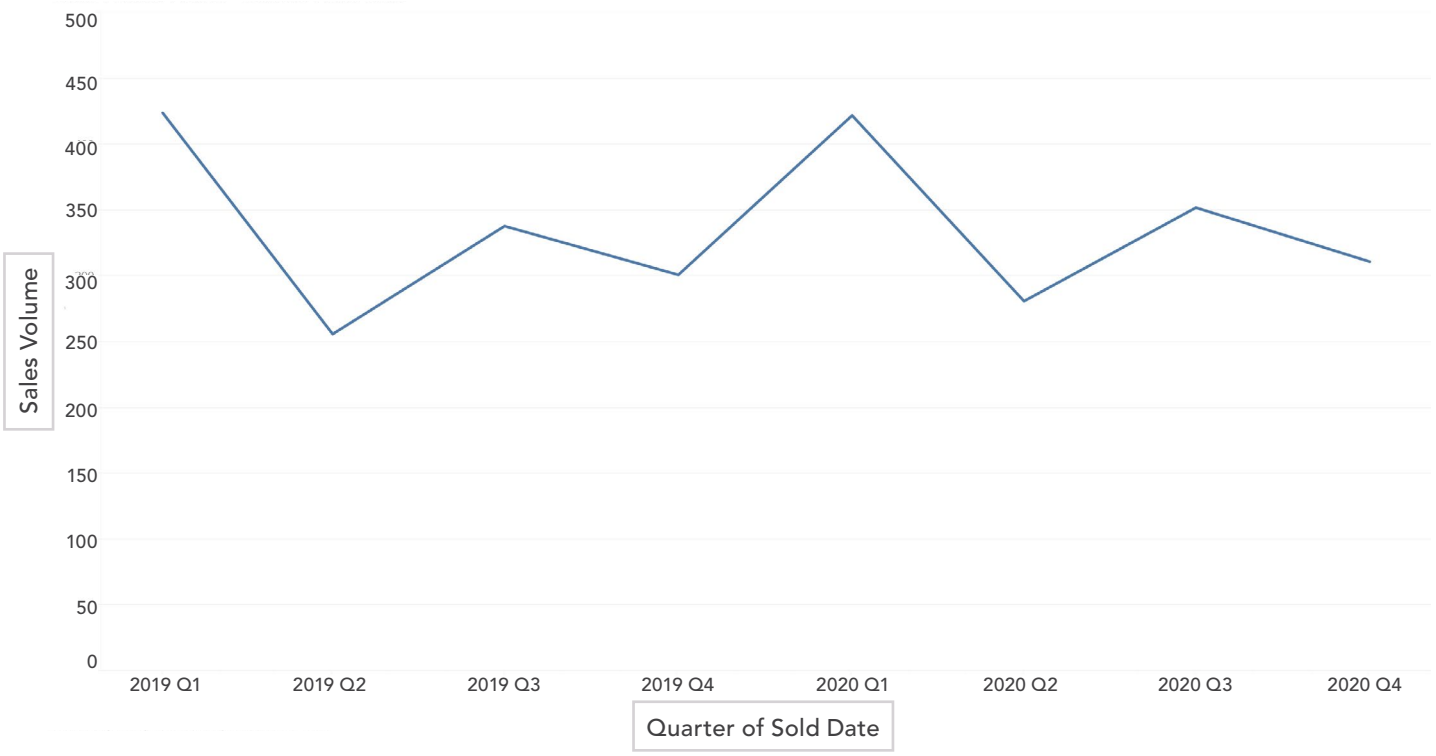


The image on the left of Figure 22 makes it look as if there was a large increase in the profit margin from 2019 to 2020. But the baseline starts at 6.0%. The image on the right uses a zero baseline, and you can see that the change from 2019 to 2020 was not as dramatic.

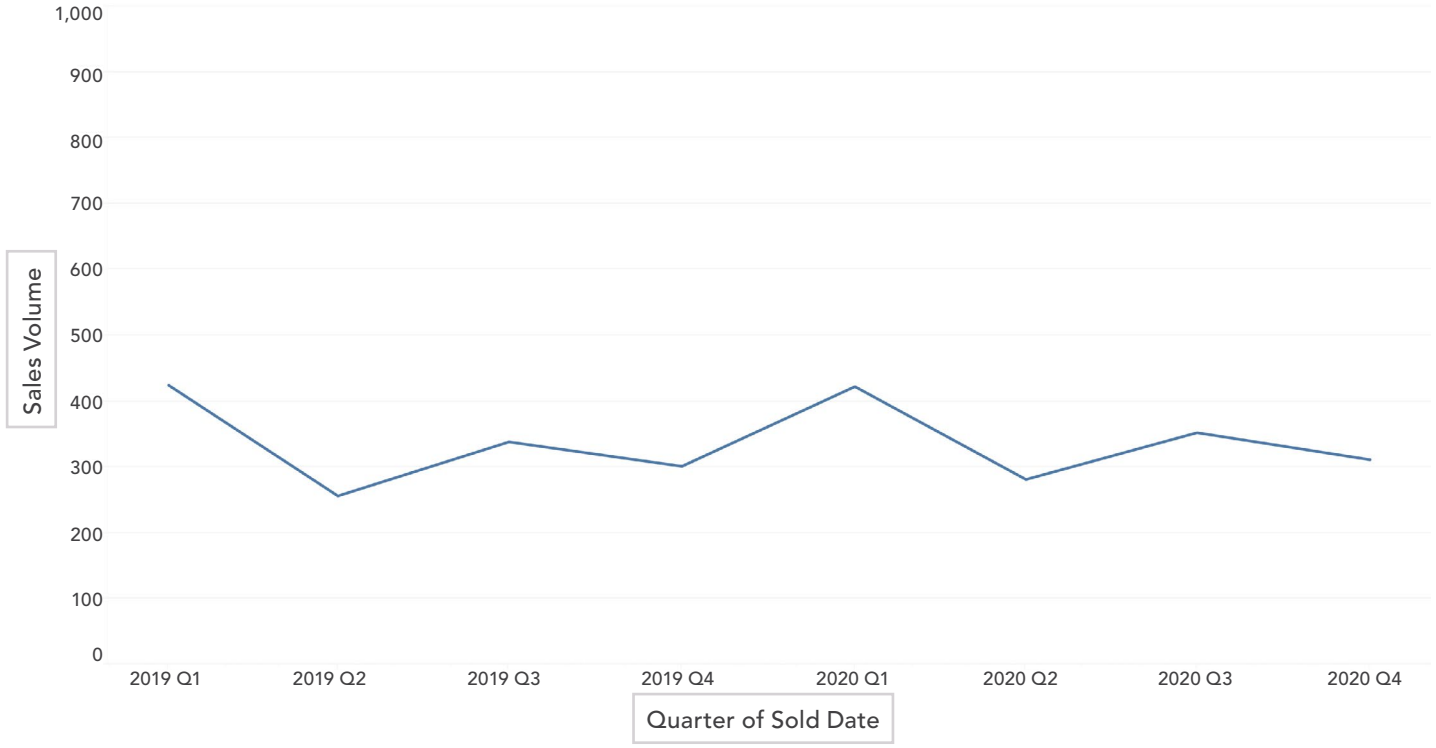
The y-axis can also be manipulated in other ways to make changes in the data seem to be more or less significant. The y-axis scale should be proportionate to the data. The graphs in Figure 23 show how manipulating the y-axis can make the change in sales volume look less significant.

FIGURE 23: MANIPULATING THE Y-AXIS

Sales Volume Trends—Correct Y-Axis Scale



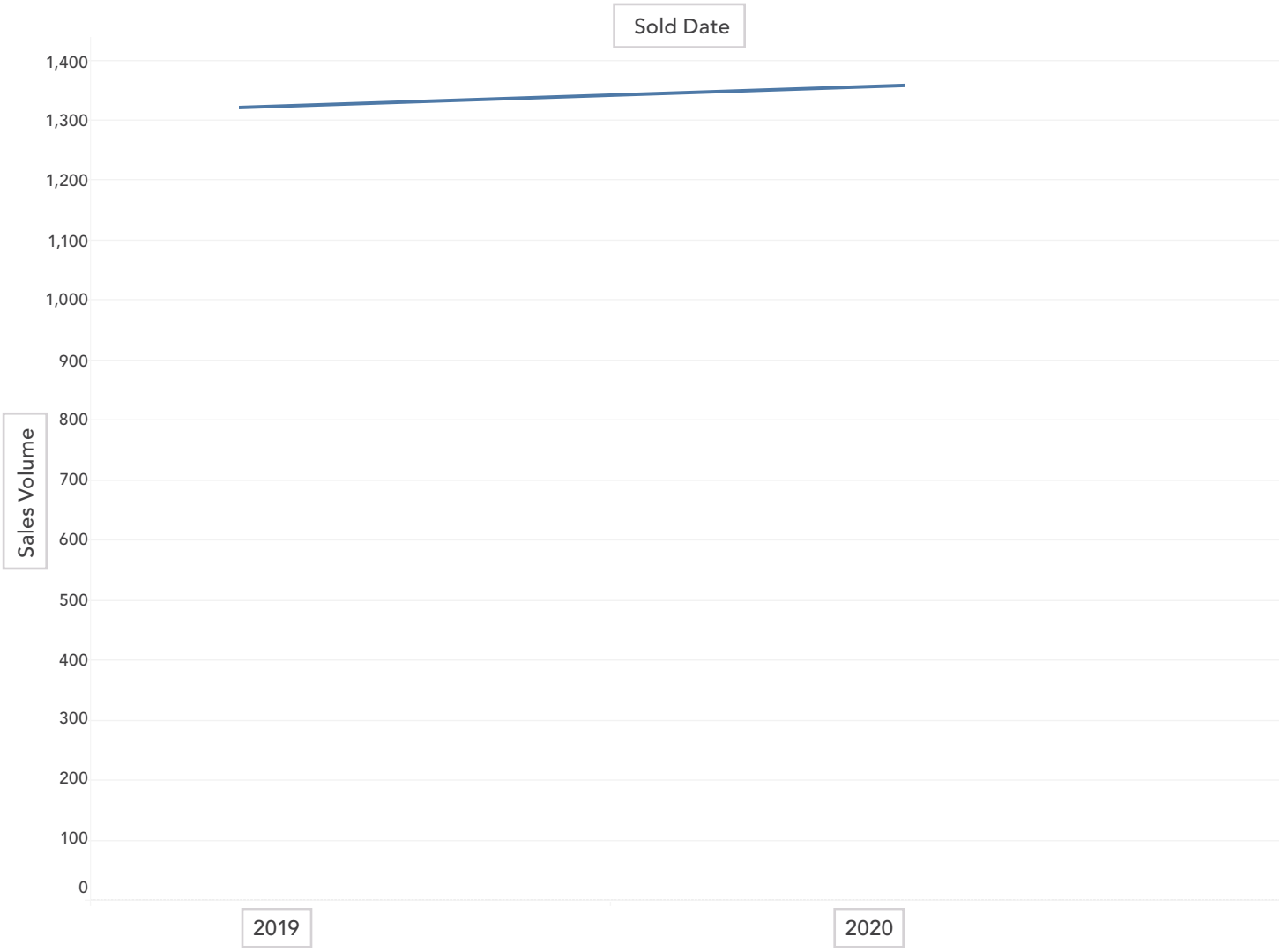
Sales Volume Trends—Manipulated Y-Axis to Show Less Volatility



The y-axis in the first line graph begins at zero and is proportionate to the data. The y-axis in the second graph has a larger scale, making the trend seem less volatile than in the first graph.

Another way to mask volatility in a visualization is to pick only the data points that will provide the story you want to tell. The same graph in Figure 23 is provided in Figure 24, but this time only the years are shown.

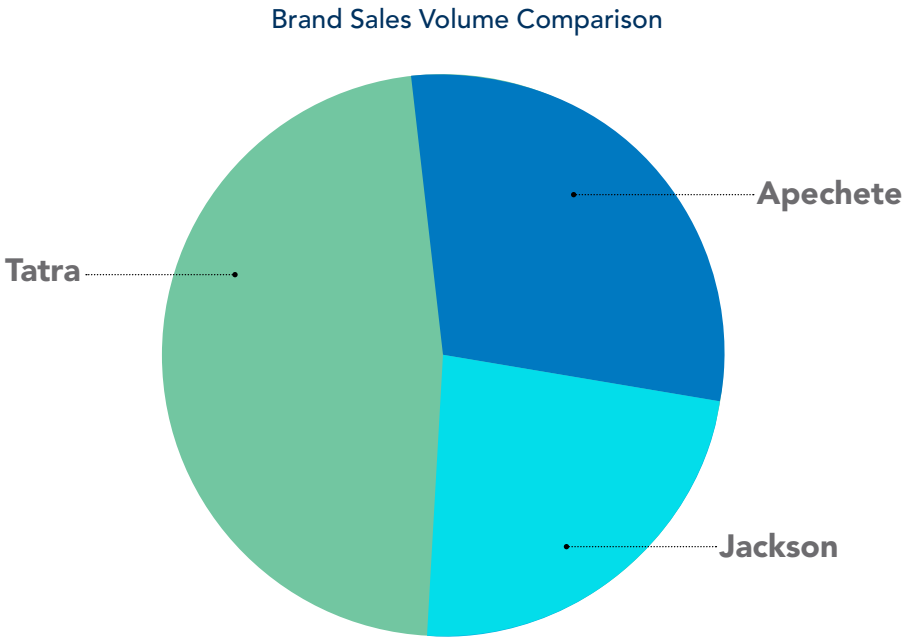
FIGURE 24: CHERRY-PICKING THE DATA



If the creator wanted to mask the volatility of sales, the graph in Figure 24 makes the sales volume trend look as if it is consistent and slightly rising. This masks a potential issue with volatile sales.

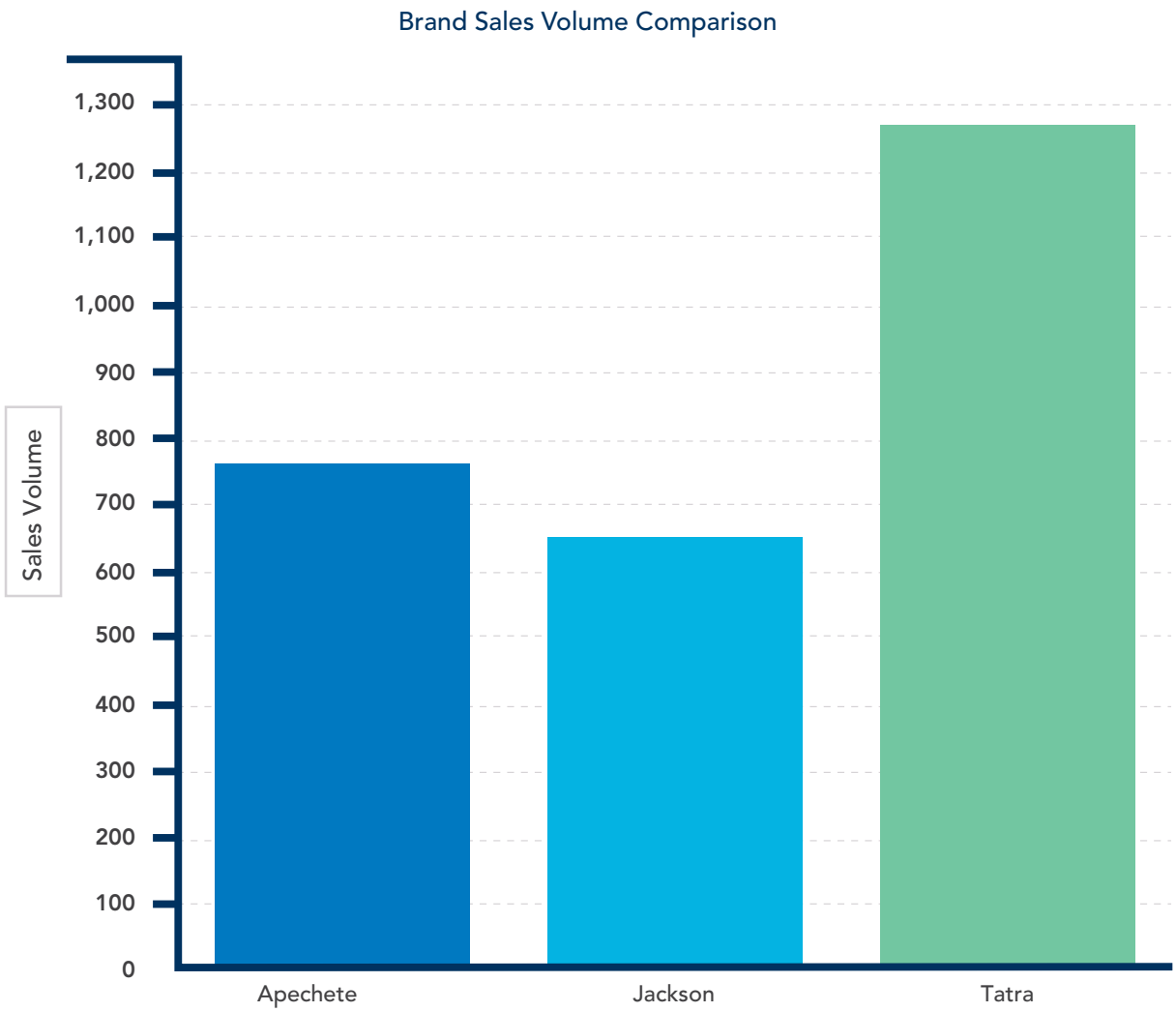
Using the wrong graph is another way to mislead the reader. As discussed in best practices for data visualizations, comparisons are best shown with bar or column charts. But if we use the wrong visual for comparison, the reader will have difficulty comparing the data. Figure 25 provides an example of this type of misleading visual.

FIGURE 25: USING THE WRONG GRAPH



In this visual, you are comparing the sales volume for all three brands of HMC vehicles. Although it is clear that the Tatra has the largest sales volume, it is much more difficult to estimate the difference between the Apechete and the Jackson. The bar chart in Figure 26 provides a much better comparison.

FIGURE 26: IMPROVED COMPARISON VISUAL



Finally, always be cognizant of standard conventions when preparing a visualiization so that you do not mislead the reader. For example, in the accounting and finance profession, the color red is associated with losses. You would not want to use red to illustrate gains in a visualization, because your readers will assume red means losses.

CONCLUSION

The ability to visualize data is a critical skill in today's data-driven business world. Data visualization helps you not only to explore data for insights but also to share those insights with stakeholders. Following best practices for creating data visualizations and then using those visualizations to tell a story with data can help to create change in your organization and to provide value to your skateholders.

For more information, please visit **www.imanet.org/thought_leadership.**