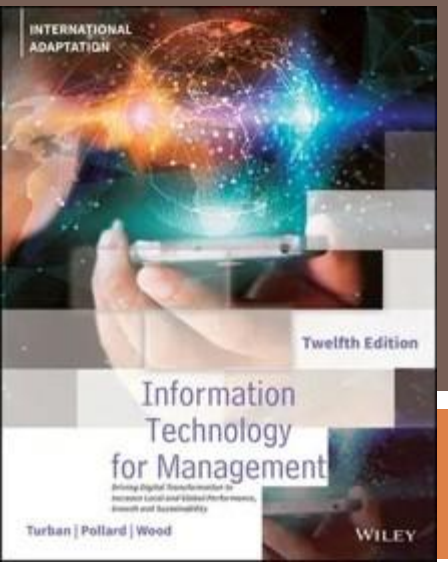
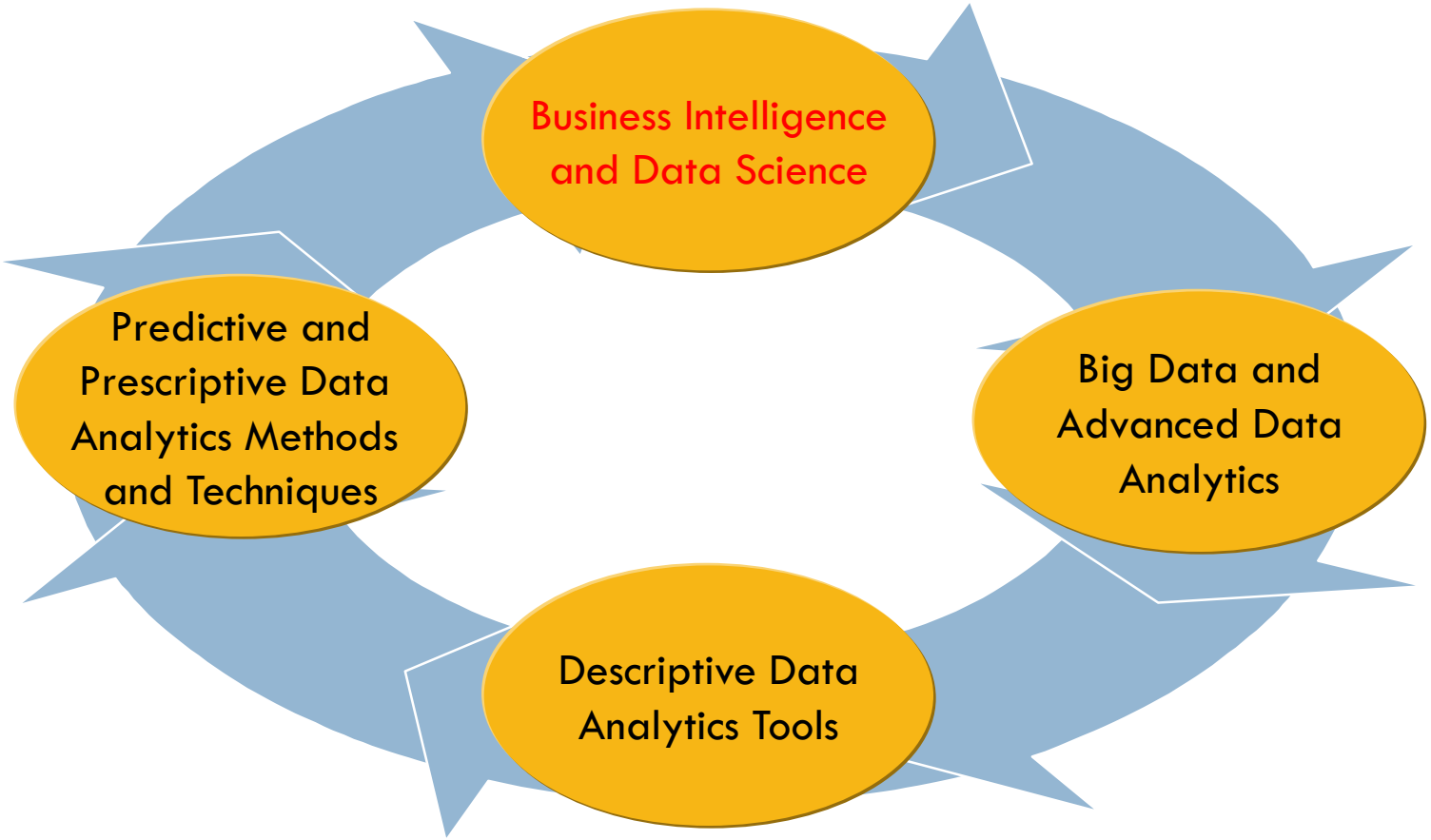
CHAPTER 6
BUSINESS INTELLIGENCE, DATA SCIENCE, AND DATA ANALYTICS



Chapter Outline (1 of 4)



Business Intelligence and Data Science

- Organizations are becoming increasingly dependent on analyzing data stored in IT systems to make important business-related decisions every day to make decisions about how to connect better with their customers, drive innovation and improve performance.
- Following a systematic step-by-step decision-making process helps managers make more deliberate and thoughtful decisions by collecting and organizing relevant data, gaining insights into the data, and developing acceptable alternative solutions to an identified problem or opportunity.
- Figure 6.1 Four phases of decision-making

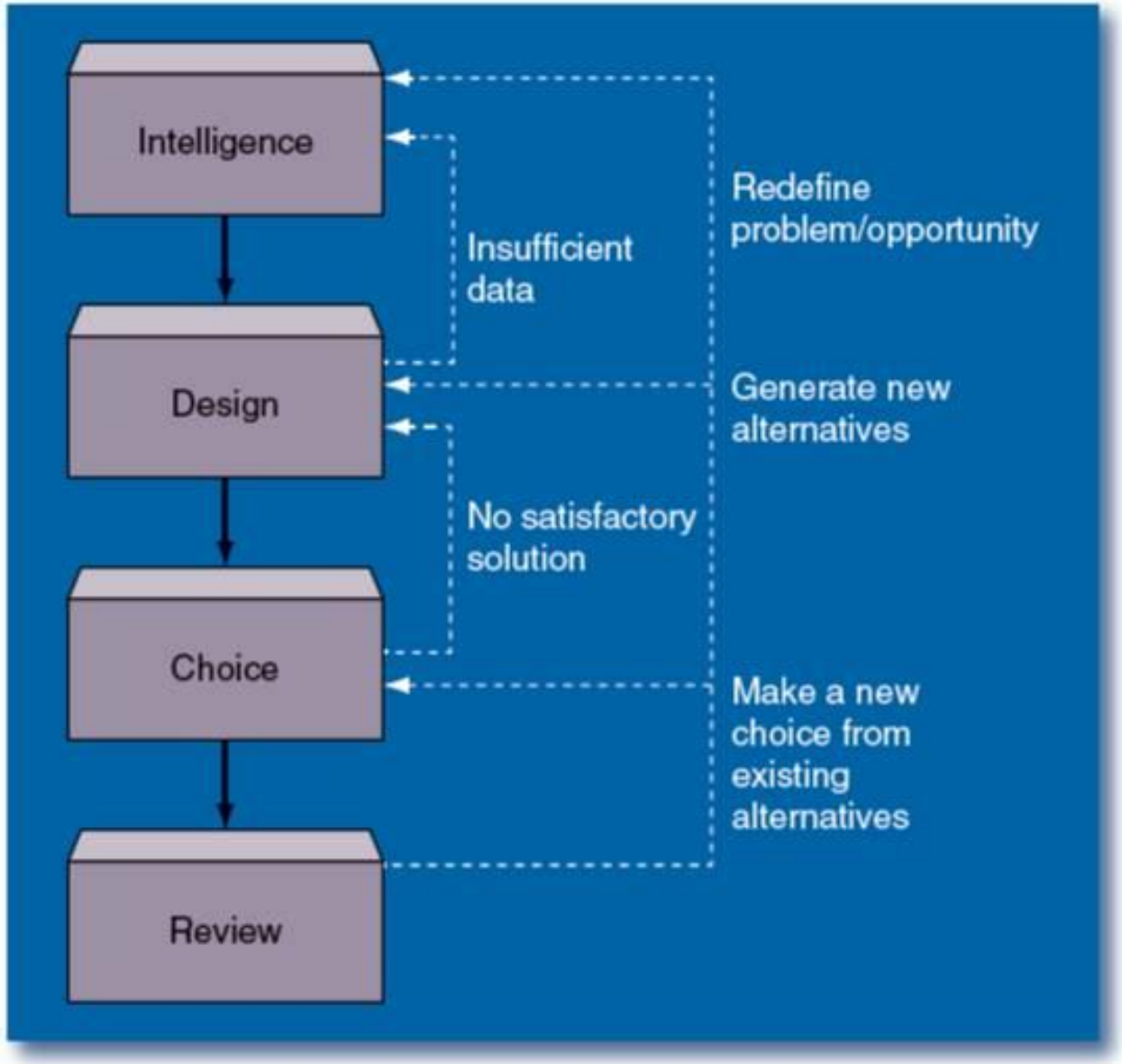


FIGURE 6.1 Four phases of decision-making

TABLE 6.0 Four Phases of Decision-Making

Intelligence Phase	• Identify the problem or opportunity. • Collect information. • Establish a goal and assessment criteria.
Design Phase	• Specify various courses of action to solve the problem or exploit the opportunity. • Analyze feasible alternatives. • Evaluate each alternative against the criteria established in the intelligence stage.
Choice Phase	• Select an alternative course of action.
Review Phase	• Sometimes called the monitor, control, or implementation stage. • Monitor and control the choice to ensure its proper execution. • Return to any previous phase, including redefining the problem or opportunity if the solution fails.

Four Phases of Decision-Making

Bounded Rationality (제한된 합리성)	• The idea that rationality is limited by the tractability of the decision, cognitive limitations of the mind, and time available to make the decision.
Satisficing (만족화)	• A decision-making strategy that involves searching through available alternatives until an acceptable solution is found. • It is a composite of the words "satisfy" and "suffice".
Optimizing (최적화)	• The process of finding an alternative that is most cost effective or produces best achievable performance under given constraints by maximizing desired effects and minimizing undesired effect.

Data-Driven Decision-Making with Data Analytics

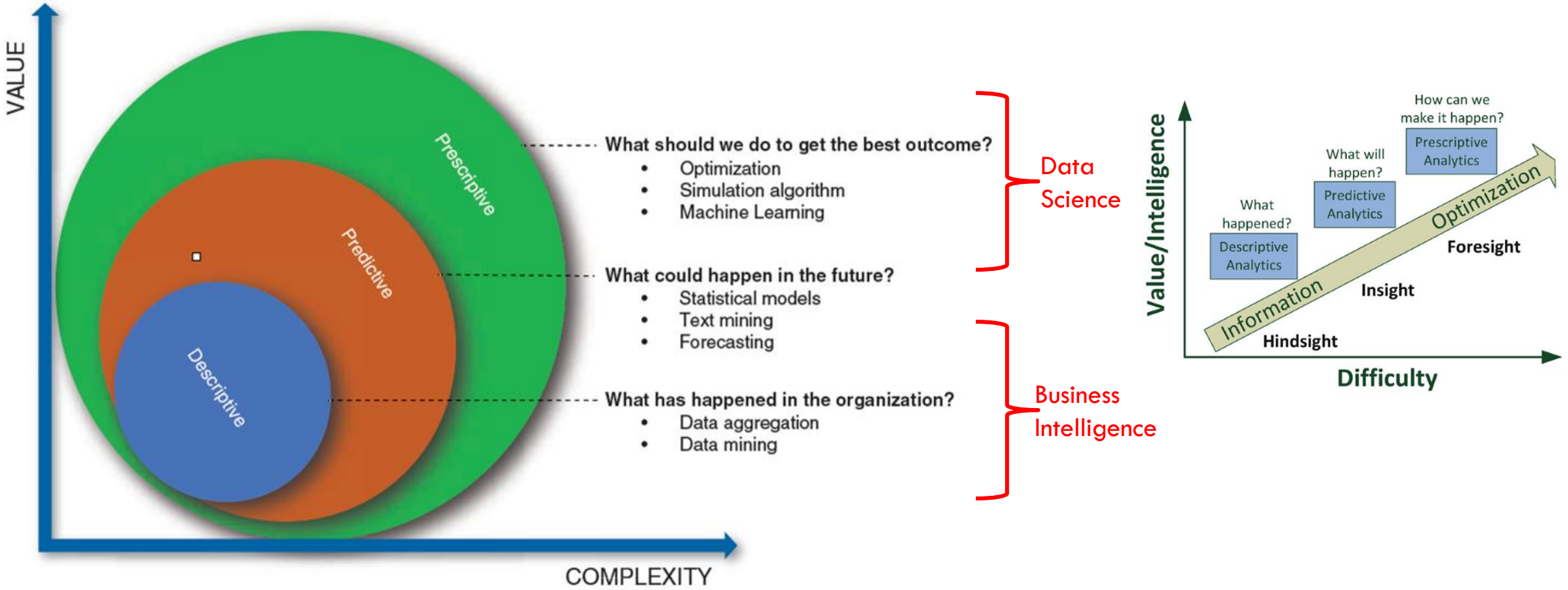


FIGURE 6.2 Three levels of data analytics

Data-Driven Decision-Making with Data Analytics

Descriptive Data Analytics	• Creates a summary of historical data to yield useful information and possibly prepare the data for future more sophisticated analysis.
Predictive Data Analytics	• Is the process of using data analytics methods and techniques to model and make predictions about unknown events from data.
Prescriptive Data Analytics	• Is dedicated to finding the best course of action among various choices given the known parameters.

Data-Driven Decision-Making with Data Analytics

Business Intelligence (BI)	• A set of best practices, software, infrastructure, and tools to acquire and transform raw highly structured data into actionable insights to help managers at all levels of the organization make informed business decisions.
Data Science	• A multidisciplinary field that uses domain expertise, scientific methods, programming skills, algorithms and statistics to extract knowledge and insights from structured, semi-structured and unstructured big data sets to predict future behavior and prescribe actions.

Data-Driven Decision-Making with Data Analytics

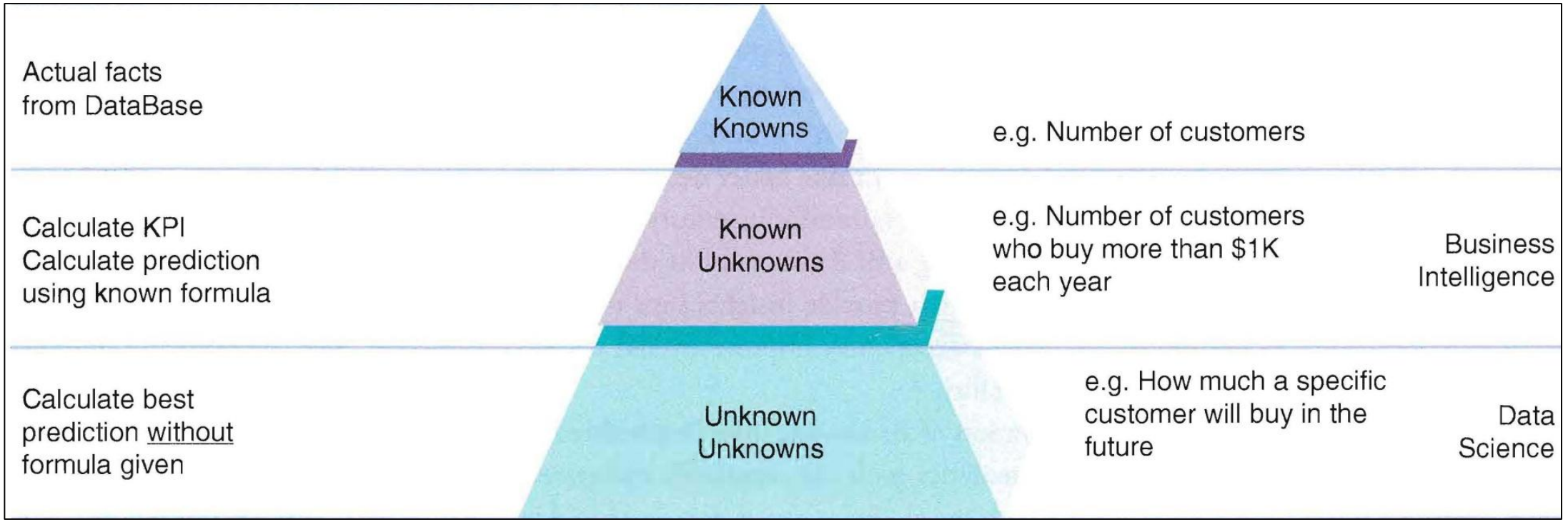


FIGURE 6.3 Different focus of BI and data science

Traditional and Modern Business Intelligence (BI)

Dashboard	• A graphical user interface that provides at-a-glance views of relevant KPIs to an organization or department.
Data Mashup	• The integration of two or more data sets from various business systems and external sources without relying on the middle step of ETL (extract, transform, and load) into a data warehouse or help from IT.
Data Visualization	• The process of representing abstract business or scientific data as images, diagrams, graphs, or animations that can aid in understanding the meaning of the data.

Traditional and Modern Business Intelligence (BI)

Modern BI	• Allows users to produce reports and analysis on-the-fly and share data with other users to make decisions and optimize business results.
Self-Service Analytics	• A form of BI that enables and encourages managers and other users to perform queries and generate reports with nominal (명목상의) IT support.
Embedded BI	• The integration of self-service analytics tools and capabilities within commonly used business software apps.

➢ on-the-fly: quickly and often without preparation

Adding Value with Traditional and Modern BI

- Quick Detection and Decisions in Stock Markets
  - Wall Street firms, traders, wealth managers, risk analysts, and regulators rely on their ability to process and capitalize on market anomalies in real time.
- Prompt Disaster Response by the Insurance Industry
  - The effectiveness of an insurer's response to a devastating hurricane or other catastrophic event depends on its ability to combine large amounts of data to fully understand the impact.

Finding and Hiring BI Professionals

- One of the challenges organizations are struggling with is finding and hiring qualified data analytics professionals.
- BI professionals include BI analysts, BI developers, BI managers, BI consultants, and business analysts.
- For example, a BI analyst produces financial and market intelligence by querying data repositories and generating period reports.

Software to Support BI Professionals

Speed	• You can ask and answer questions in real-time even when working with massive and diverse data sets.
Visualization	• Modern BI enables self-service analytics so users can ask more questions, drill down into the data and generate shareable dashboards.
Single Source of Truth	• Massive volumes of data from different sources can easily be combined with different data sources in real-time with no upfront (선불) integration costs.
Real-Time Collaboration	• Data is always live. Users can filter, sort, discuss, transform, and share data instantaneously.

Software to Support BI Professionals

Comprehensive Governance	• Views data governance as an important step in creating a safe and trusted environment for self-service analytics resulting in accurate, available, and audited dashboards and reports.
Scalability	• Start small and scale as needed. • Modern BI supports all stages of an organization's analytical evolution.
Mobility	• Smartphone and tablet compatible to allow users to view, interact, and share analytics regardless of device and location.

Software to Support BI Professionals

- Figure 6.5 shows how BI has evolved from performance monitoring and KPI analysis to sophisticated self-service analytics platforms driven by artificial intelligence (AI).
- The major feature of modern BI that is proving to be a key competitive differentiator for vendors is augmented analytics.
- Augmented analytics uses machine learning and artificial intelligence to automate data preparation, insight generation, and insight explanation to augment how business managers and analysts explore, interpret, and share data.

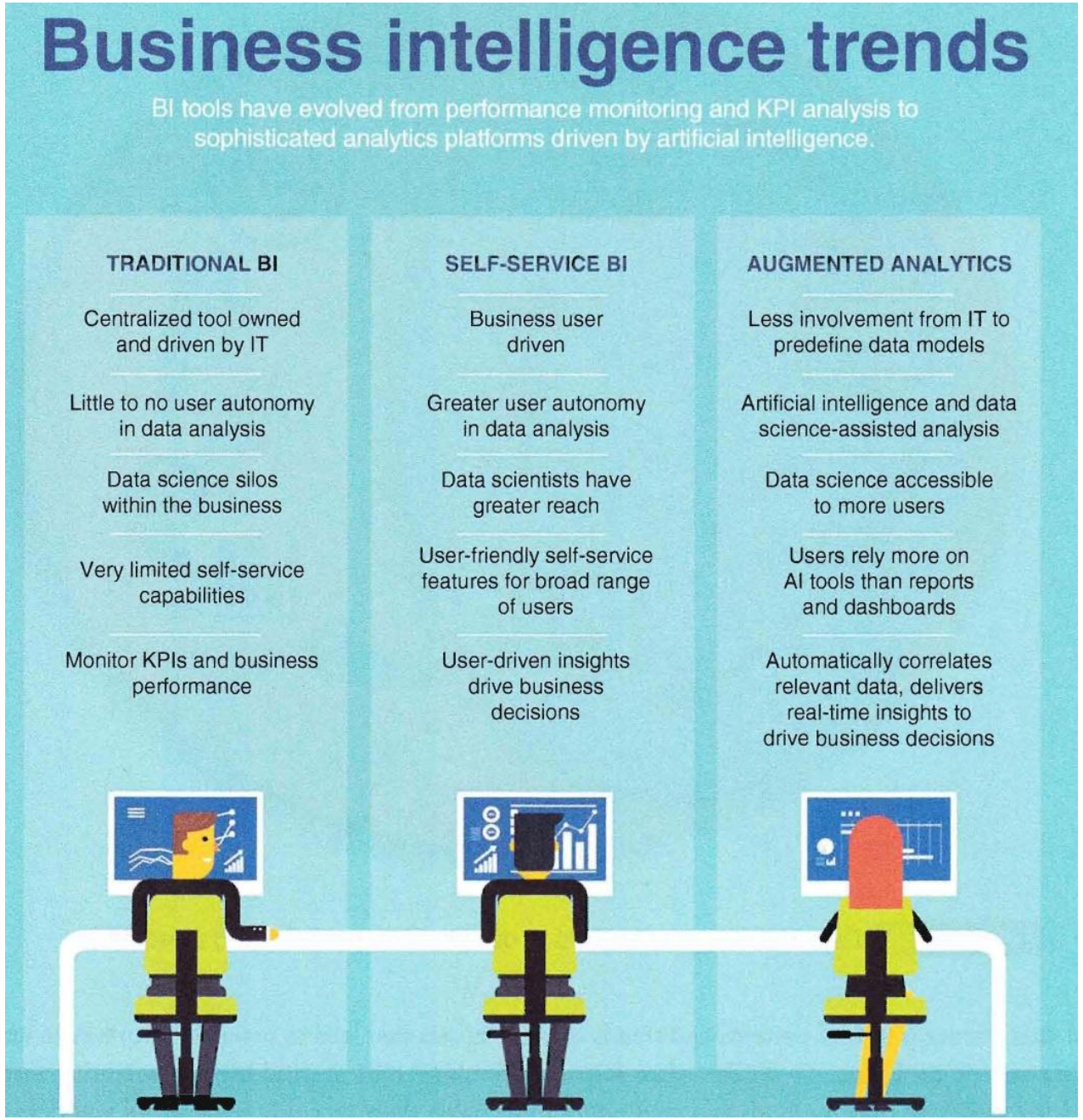


FIGURE 6.5 The evolution of BI

- Augmented analytics is the use of machine learning and AI in BI tools to automate data preparation and help users discover and share insights.
- Machine learning is scientific algorithms that identify patterns in big data to learn from the data and create Insights based on the data.

Data Science

- Oftentimes, BI is not enough.
- Instead of just describing the current status of primarily structured data, data science goes beyond the reporting functions of BI by using scientific methods and data modeling to predict future behavior and prescribe action to optimize processes based on unstructured, semi-structured, and structured data.
  - Figure 6.6
- Typically, historical data is used to build a mathematical model based on important trends.
- In data science, the model is then used on current data to predict what will happen next and prescribe actions to take for optimal outcomes.

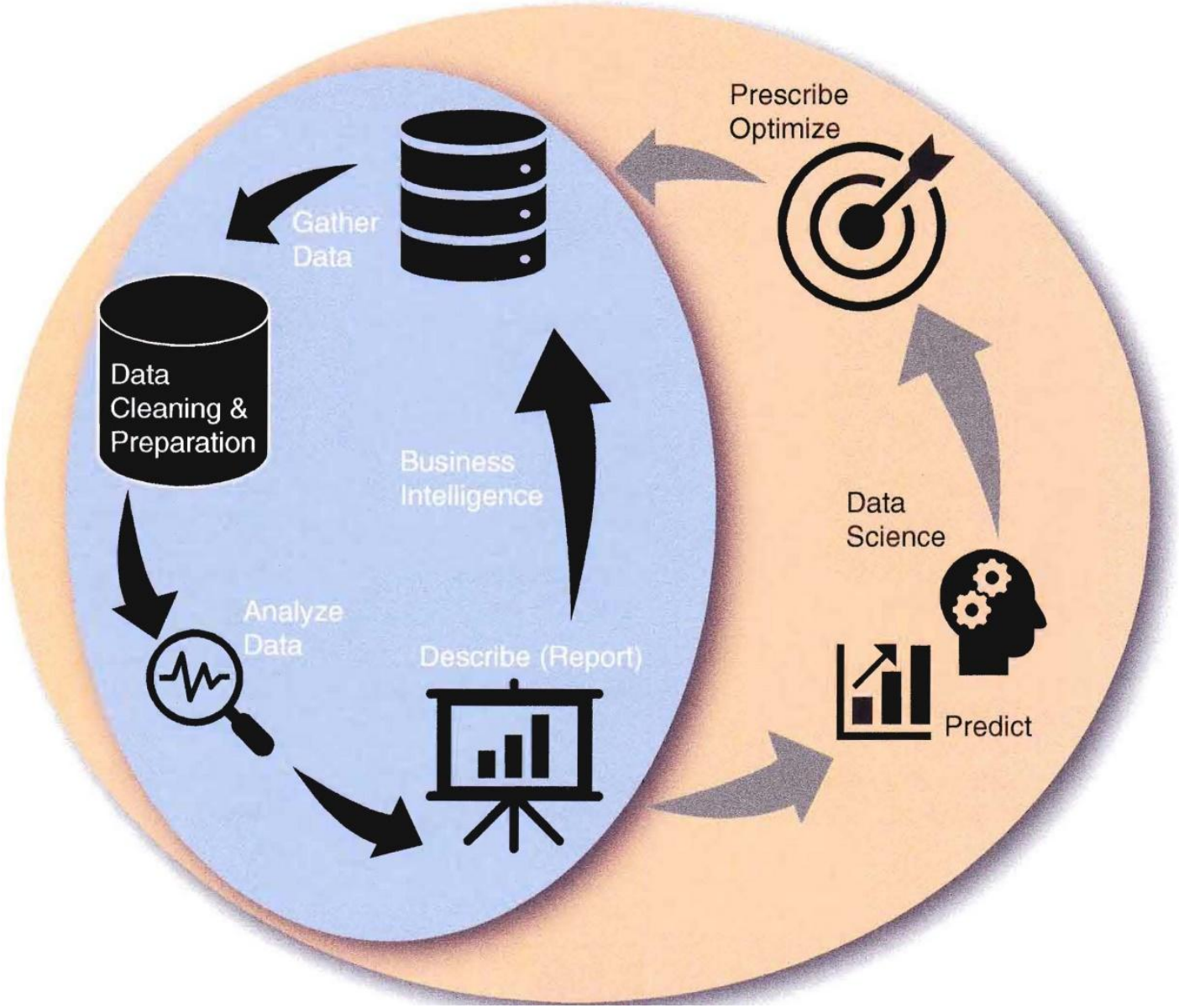


FIGURE 6.6 Data Science goes beyond BI descriptions to predict and prescribe.

Data Science

- The goal of data science is to find patterns and trends in big data sets that lead to practical solutions to real-life data centric problems.
- To do this, data scientists combine critical thinking and scientific methods with a variety of data methods and techniques that enable them to easily understand and derive real meaning from the data.
- There are seven stages in data science life cycle.
  - Figure 6.7

FIGURE 6.7 Seven-stage data science life cycle

Data Science

TABLE 6.1.1 Seven-stage data science life cycle

Stage	Description
1. CAPTURE DATA	• Understand business requirements. • Gain domain knowledge and understanding. • Gather, enter, and extract data.
2. STORE DATA	• Clean data. • Ensure data security and integrity.
3. MODEL DATA	• Create a model. • Evaluate model performance.
4. ANALYZE DATA	• Perform exploratory data analysis. • Conduct confirmatory statistical data analytics (descriptive, predictive, and prescriptive) and perform visualization tests to discover patterns in the underlying data.

TABLE 6.1.2 Seven-stage data science life cycle

Stage	Description
5. COMMUNICATE INFORMATION	• Use business intelligence methods and techniques to report results to stakeholders. • Distribute data visualization reports and dashboards. • Get validation of how the model performs in a wide variety of business situations from all project stakeholders. • Manage stakeholder expectations.
6. DEPLOY DATA MODEL	• Collaborate with data scientist, data engineers, software developers depending on the nature of the project. • Make sure there are no conflicts with other apps that are inter-dependent. • Test the model in the real-world production environment. • Gain business buy-in.
7. REITERATE (REPEAT) PROCESS	• Operate and optimize the data model as changes in the business degrade the effectiveness of the current model. • Monitor KPIs continuously-possibly through a visualization dashboard. • If model performance degrades, retrain the model using updated data.

Data Science

- Throughout its life cycle, data science combines knowledge from many STEM (Science, Technology, Engineering, and Mathematics, 과학·기술·공학·수학 융합교육) disciplines including mathematics, statistics, and computer programming and supplements it with domain knowledge about the topic of interest to use algorithmically supported scientific methods.
- Figure 6.8 shows the components of data science knowledge base.

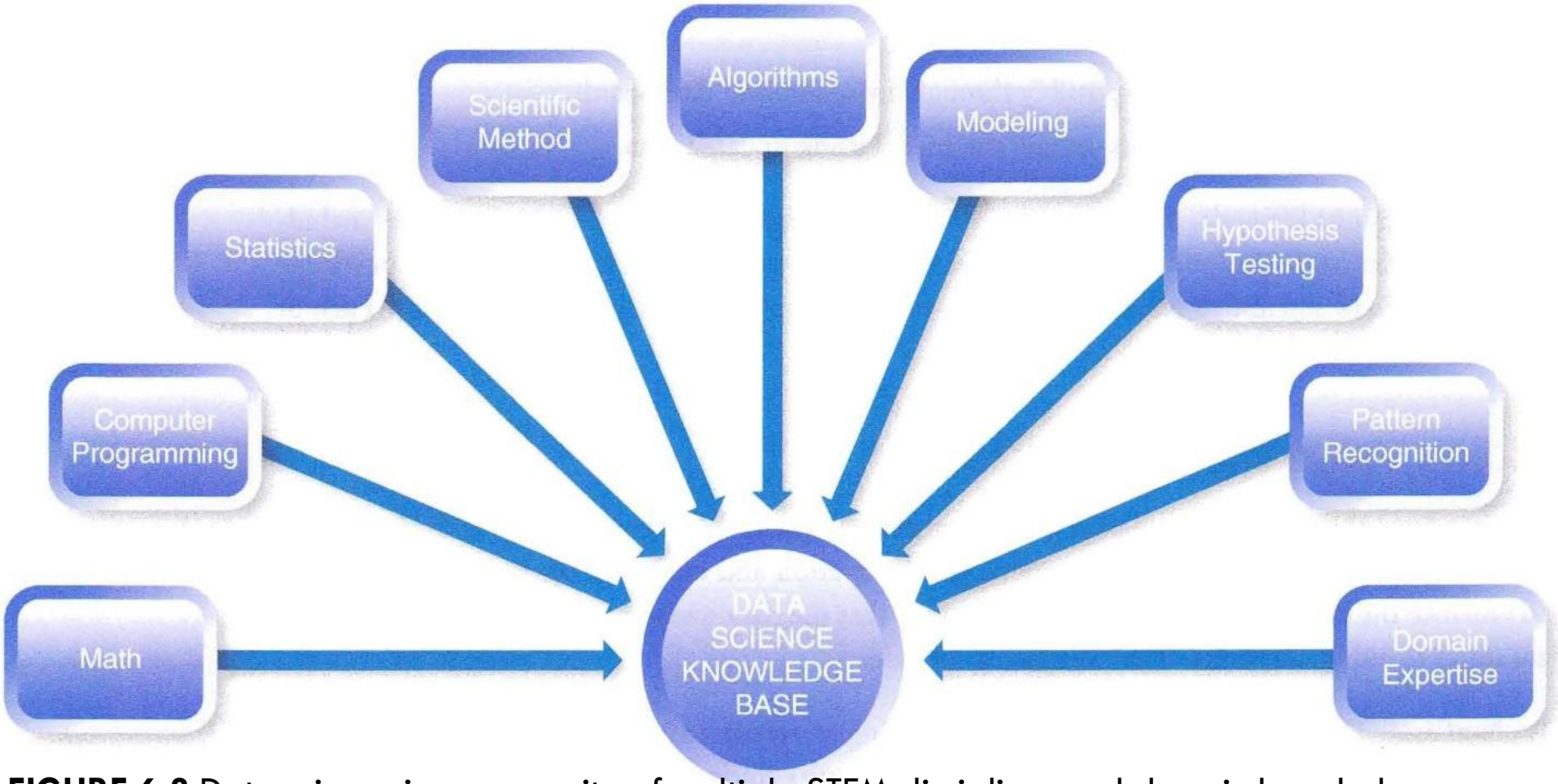


FIGURE 6.8 Data science is a composite of multiple STEM disciplines and domain knowledge.

Adding Value with Data Science

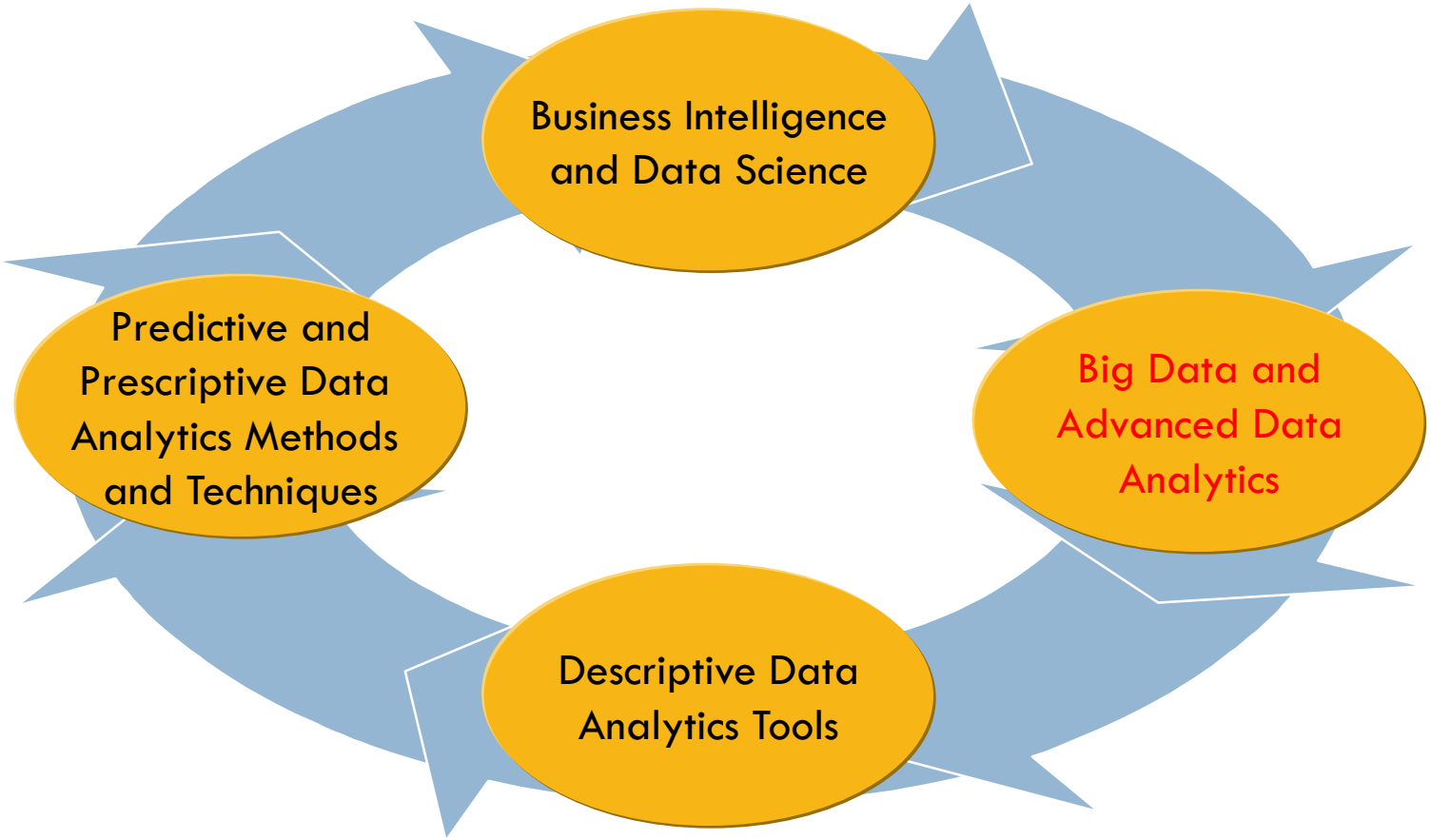
- Organizations that use data science to discover insights in their data stores to aid in corporate decision-making include:

Netflix	• Data mining uncovers movie viewing patterns to understand what drives user interest and uses the algorithmically generated results to decide on topics for future Netflix original series.
Target	• It identifies major customer segments within its customer base and analyzes the unique shopping behaviors of each segment to help them guide message content for different market audiences.
Procter and Gamble	• It uses time series models to understand future consumer demand and help plan for optimal production levels across their extensive range of products.

Adding Value with Data Science

- Data science methods and techniques also develop "data products" like those created by recommendation engines that utilize user data to make personalized suggestions such as items to buy presented by Amazon, movie suggestions from Roku, and music tips from Pandora.
- Data product
  - A technical function that encapsulates an algorithm and is designed to integrate directly into core applications.

Chapter Outline (2 of 4)



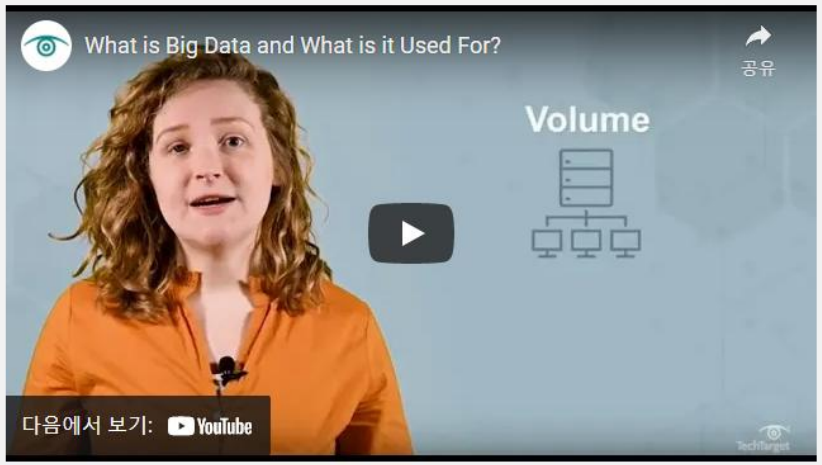
Big Data

- An extremely large data set that is too large or complex to be analyzed using traditional data processing techniques.
- Big data is the major factor driving the increased importance of advanced data analytics today.
- Advanced Data Analytics is the process of examining data sets to draw conclusions about the information they contain, usually with the help of computer software.
- Table 6.2 shows the characteristics that differentiate "big data" from traditional or "small data."

TABLE 6.2 The Differences Between "Big" Data and "Small" Data

Category	"Big" Data	Traditional or "Small" Data
Data Sources	• Data generated outside the enterprise from non-traditional data sources: ✓ Social media ✓ Sensor data ✓ Log data ✓ Device data ✓ Video ✓ Images, etc.	• Traditional enterprise data such as: ✓ Enterprise resource planning transactional data ✓ Customer relationship management (CRM) systems ✓ Web transactions ✓ Financial data, e.g., general ledger data
Volume	• Terabytes (10^{14}) • Petabytes (10^{15}) • Exabytes (10^{18}) • Zettabytes (10^{21})	• Gigabytes (10^{9}) • Terabytes (10^{12})
Velocity	• Often real-time • Requires immediate response	• Batch or near real-time • Does not always require immediate response
Variety	• Structured • Unstructured • Multi-structured	• Structured • Unstructured
Value	• Complex, advanced, predictive business analysis, and insights	• Business intelligence, analysis, and reporting

Big Data


What is Big Data and What is it Used For?
공유
Volume
다음에서 보기: YouTube


시사기획 창
고품격 · 탐사 · 다큐멘터리


빅데이터 Big Data
세상을 바꾸다

The Four Vs of Big Data

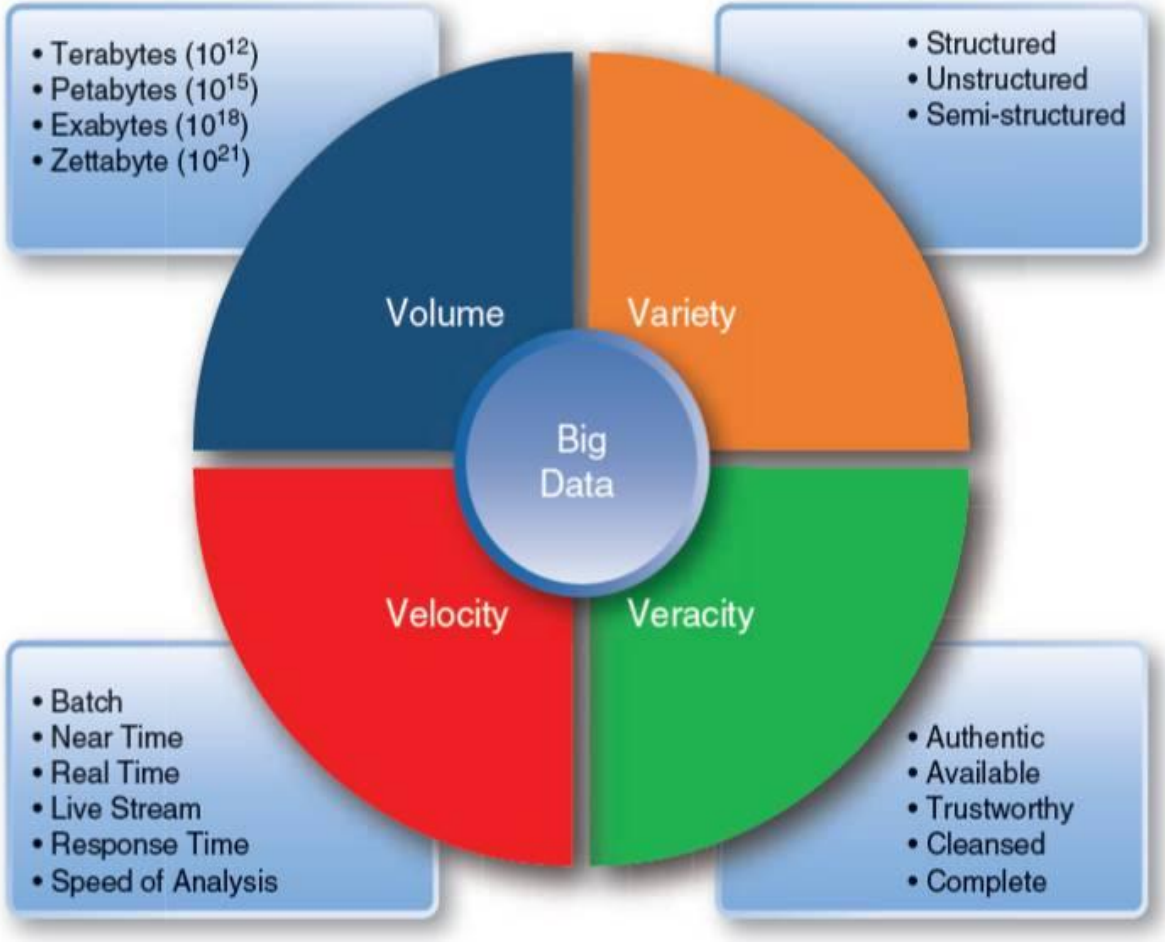


FIGURE 6.9.1 The four Vs of big data

The Four Vs of Big Data

Volume (양)	• Large volumes of structured and unstructured data are analyzed.
Variety (다양성)	• The analytic environment has expanded from pulling data from enterprise systems to include big data and unstructured sources.
Velocity (속도)	• Speed of access to reports that are drawn from data defines the difference between effective and ineffective analytics.
Veracity (정확성)	• Validating data and extracting insight that manager and workers can trust are key factors successful analytics. • Trust in analytics has grown more difficult with the explosion of data sources.

The Four Vs of Big Data

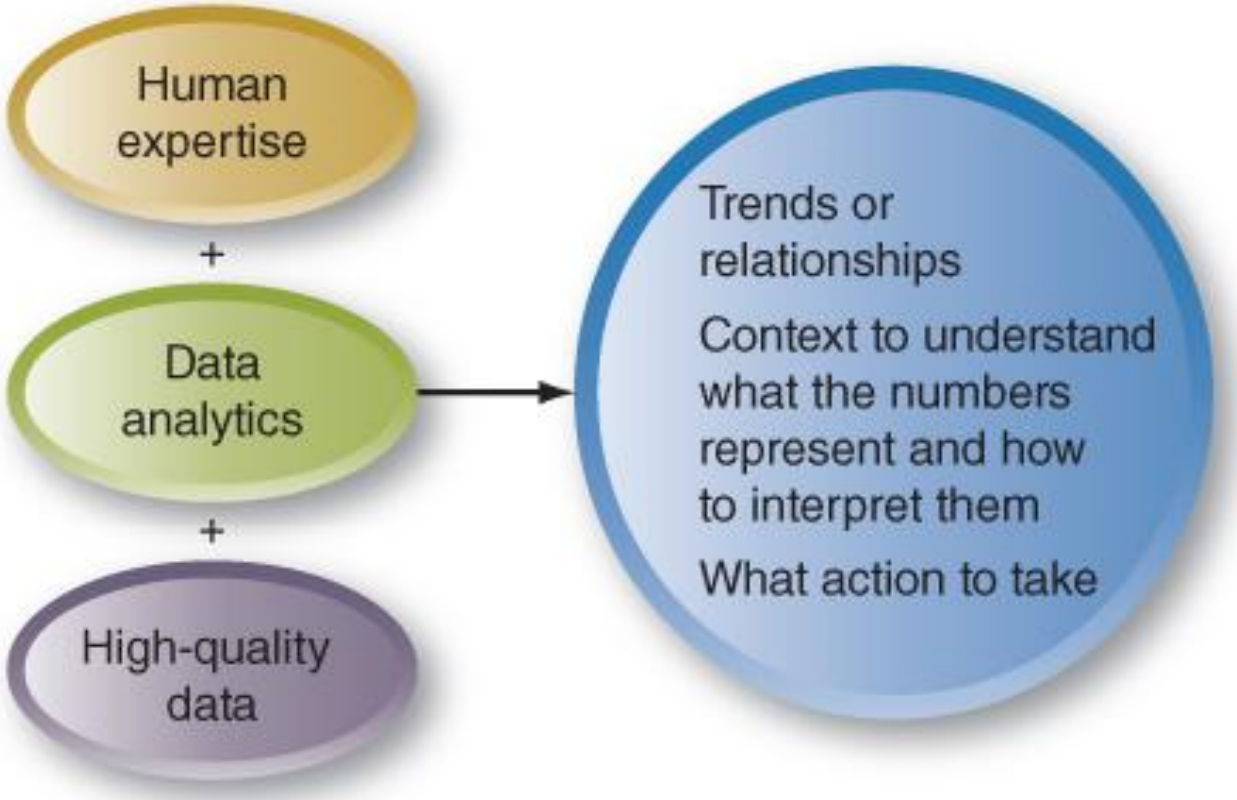


FIGURE 6.9.2 Data analytics, human expertise, and high quality data are needed to obtain actionable information.

Big Data Goals and Challenges

1. Cut costs.
2. Gain market share by accurately predicting the future.
3. Establish a data-driven culture.
4. Create new ways to innovate and disrupt with technology.
5. Accelerate speed of offering new capabilities and services.
6. Launch new products and services.
7. Improve processes.

Big Data Goals and Challenges

1. Cultural
   - Encourage business units to share information across organizational silos.
   - Determine what internal and external, structured and unstructured data to use for different business decisions.
   - Find and hire experienced data science professionals.
   - Build high levels of trust between the data science team and the functional managers.

Big Data Goals and Challenges

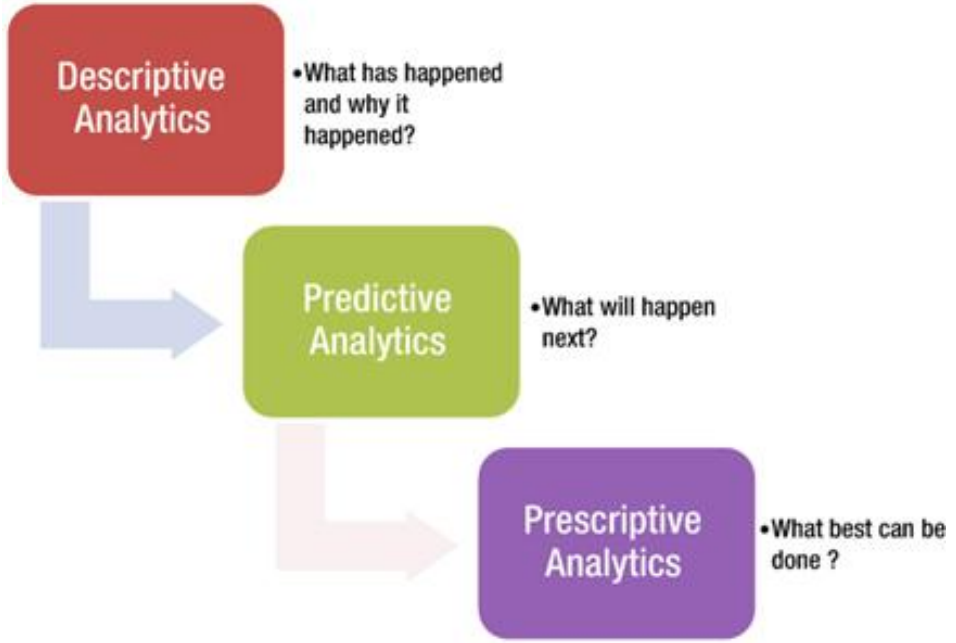
1. Cultural
   - Gain top management support for investments in big data and training.
   - Create optimal way to organize big data programs.
   - Understand where big data investments should be focused in the organization.
   - Determine how to apply insights created from big data.
2. Technology-related
   - Effectively handle the four Vs of big data.
   - Determine best way of presenting data analysis results (e.g., visualization, dashboards, augmented reality) to facilitate actionable decision-making.

Big Data Goals and Challenges

- To analyze big data sets, organizations turn to predictive and prescriptive data analytics--that go beyond the capabilities of traditional BI--to solve complex problems.
- Taken together, these two high-level data analytics methods and techniques are referred to as advanced data analytics.
- Advanced Data Analytics is the examination of data using sophisticated methods and techniques to discover deeper insights, make predictions, and/or generate recommendations.

Predictive Data Analytics

- Data scientists use predictive analytics to create a predictive model for forecasting future events based on probabilities of what might happen based on probabilities.
- Predictive model is based on several factors likely to influence future behavior and predicts at some confidence level the outcome of an event.



Predictive Data Analytics

- Predictive modeling develops a predictive model by statistically analyzing aggregated massive amounts of data from functional systems, enterprise wide systems, and connected systems and applying data and text mining, linear or time-series regression, and machine learning.
- Predictive modeling is a process that uses data mining and probabilities to forecast outcomes to create a statistical model to predict outcomes.
  - Figure 6.10

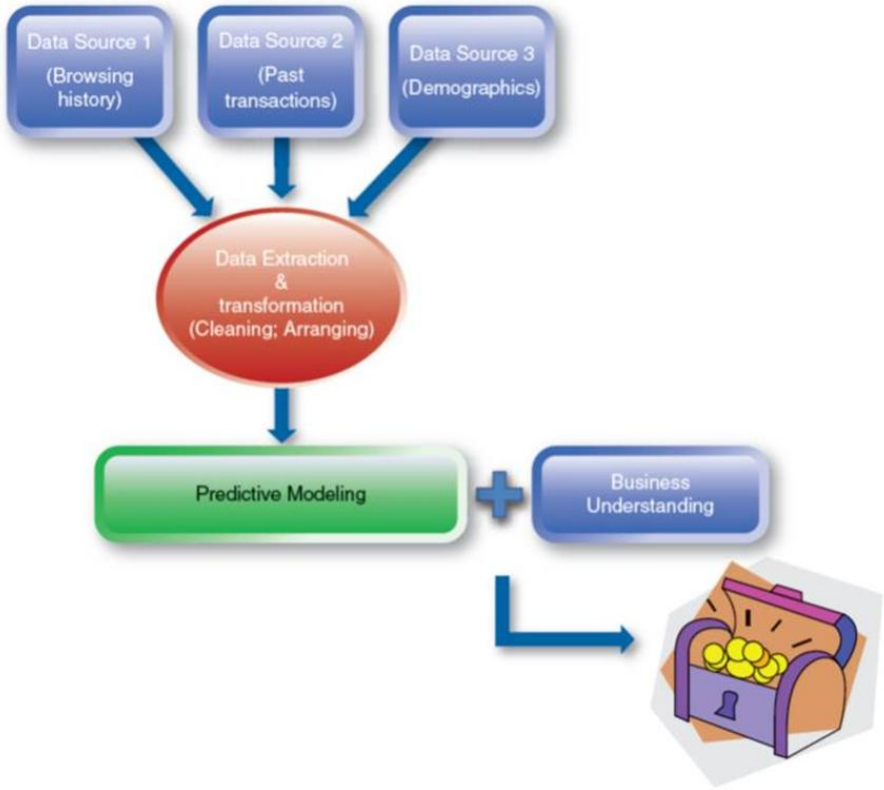


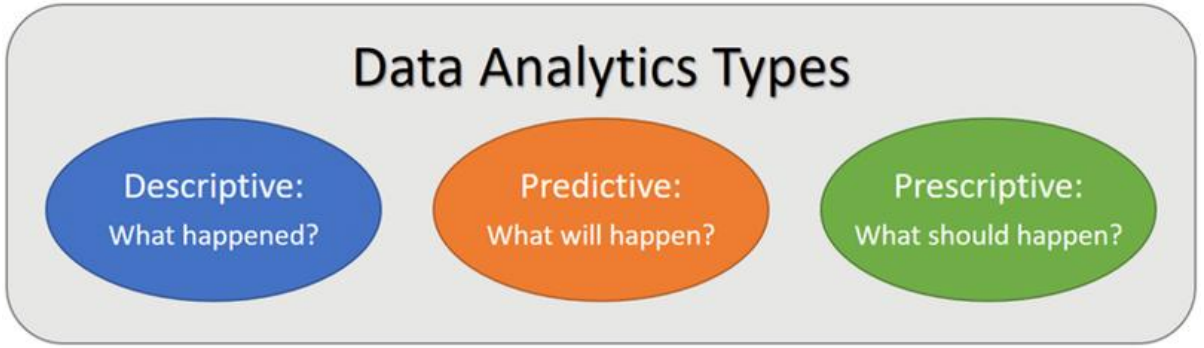
FIGURE 6.10 Predictive modeling

Predictive Data Analytics

Amazon	• Recommends products and services to users based on their past behavior. • It is estimated that predicting modeling accounts for approximately 30% of all Amazon sales.
Macy's	• Combines browsing behavior within product categories and sends targeted emails for each customer segment. • Macy's saw an 8-12% increase in online sales from their predictive modeling efforts.
Harley Davidsons	• Targets potential customers, generates leads, and closes sales by identifying potential high-value customers who are ready to make a purchase and following up with sales rep contact to walk them through the sales process.
Sprint	• Has significantly reduced its churn rate and increased its customer satisfaction ratings by identifying customers at risk of leaving and proactively providing personalized retention offers using an AI-power algorithm.

Prescriptive Data Analytics

- Prescriptive analytics is the third level of data analytics and the most powerful.
- Just as predictive analytics anticipate *what* will happen next, prescriptive analytics goes one step further to advise organizations *how to* react in the best way possible based on the prediction.
- Just a few examples of different industry sectors and the ways they can benefit from prescriptive analytics include:



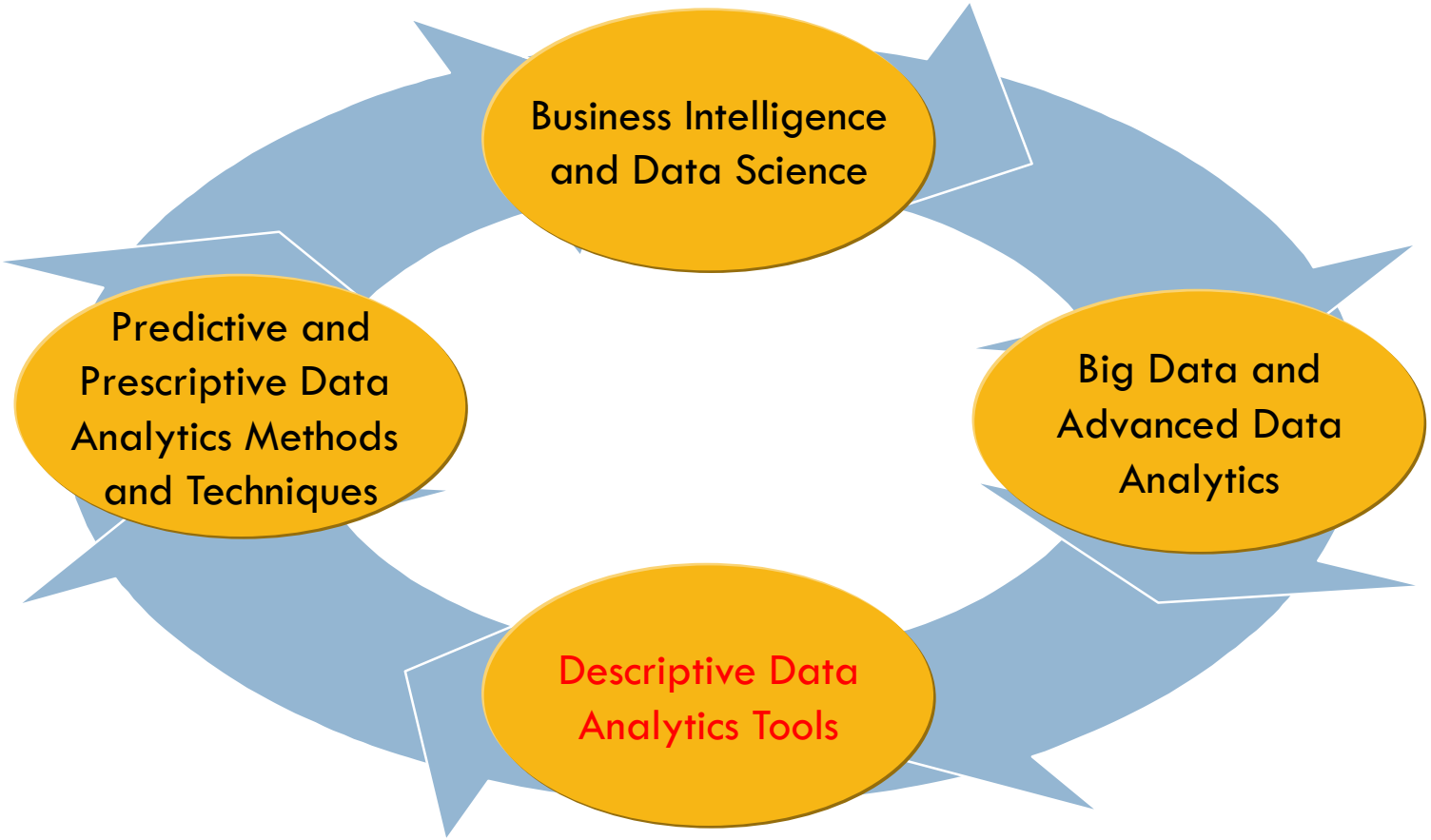
Prescriptive Data Analytics

Retail	• Minimize customer churn. • Gauge customer reaction to company actions. • Optimize product offerings; inventory levels; customer satisfaction to increase customer retention and satisfaction; increase profits and reduce costs.
Travel and transportation	• Optimize driver and route planning every 10 minutes to eliminate thousands of miles of unnecessary driving and improve driver retention. • Optimize fuel consumption to reduce costs and be more environmentally responsible. • Optimize crew schedules to increase employee morale and retain more pilots.

Prescriptive Data Analytics

Manufacturing	• Identify process improvements and optimize production planning, scheduling, inventory, and supply chain logistics to meet business requirements to offer major time and cost savings, increase agility, and provide a greater return on investment (ROI).
Healthcare	• Significantly improve patient transportation by optimizing hospital (location, specialization, and available beds) and transport data to help dispatchers plan, manage, and execute hundreds of daily patient transport requests.

Chapter Outline (3 of 4)



Descriptive Data Analytics Tools

Data Mining	• The process of using software to analyze unstructured, semi-structured, and structured data from various perspectives, categorize them, and derive correlations or patterns among fields in the data.
Data Visualization	• The presentation of data in a graphical format to make it easier for decision-makers to grasp difficult concepts or identify new patterns in the data.
Digital Dashboard	• A static or interactive electronic interface used to acquire and consolidate data across an organization.
Mashups	• Combine business data and applications from multiple sources—typically a mix of internal data and applications with externally sourced data to create an integrated experience.

Data Mining

- Adding Value with Data Mining
  1. Making more informed decision at the time they need to be made.
  2. Discovering unknown insights, patterns, or relationships.
  3. Automating and streamlining or digitizing business processes.
- Affinity analysis (친밀도 분석) is a data mining technique that discovers co-occurrence relationships among activities performed by specific individuals or groups.

Data Visualization

- Drill down
  - Searching for something on a computer moving from general information to more detailed information by focusing in on something of interest, e.g., quarterly sales, monthly sales, daily sales.
- Geospatial data
  - An explicitly geographic component, ranging from vector (크기와 방향)and raster (화상) data to tabular data with site locations.
  - Figure 6.11

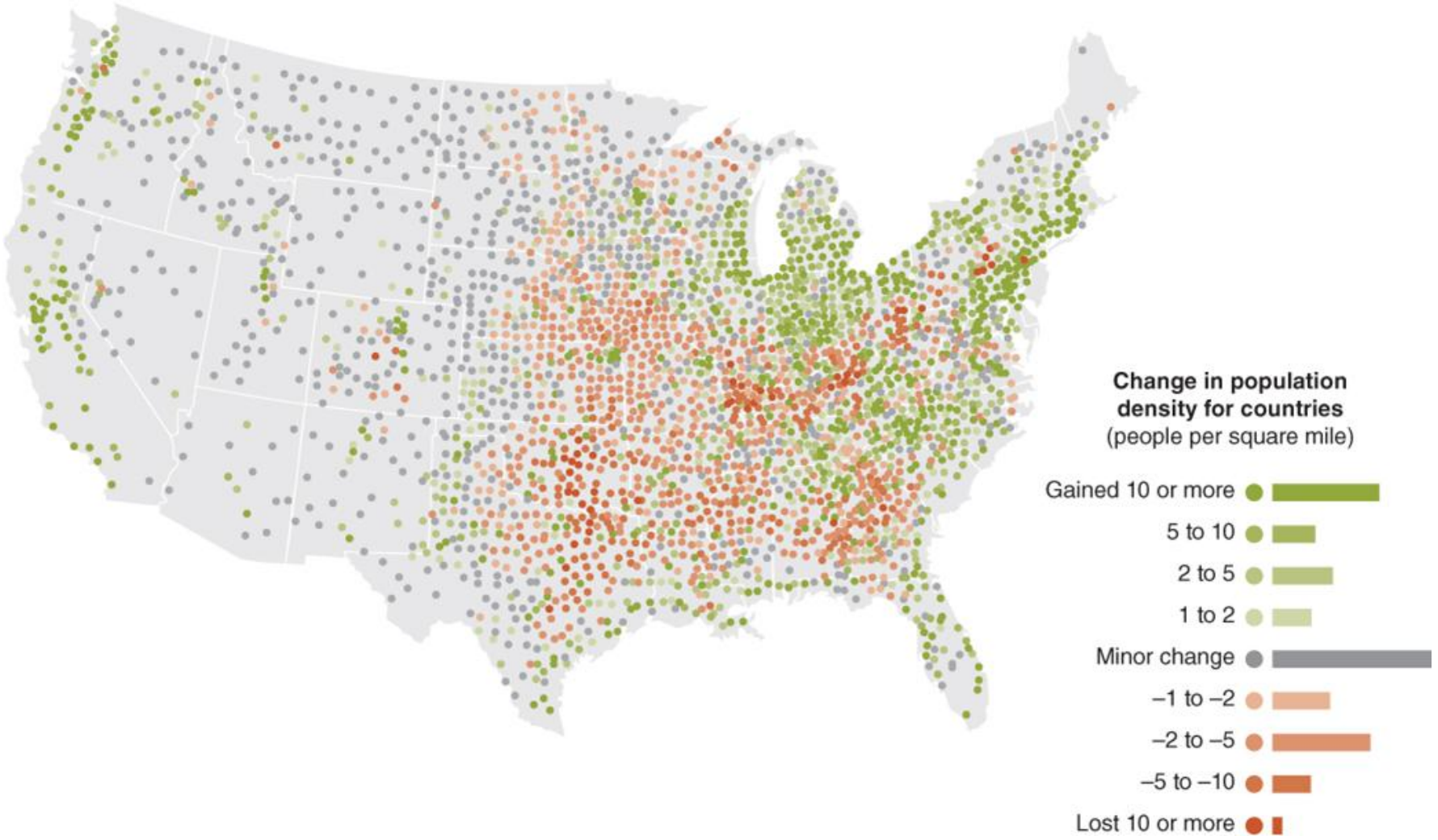


FIGURE 6.11 U.S. Census Bureau map shows easily identifiable changes in county population density. Different colors are used to indicate areas that gained and lost population. Intensity of color indicates extent of gain/loss.

Data Visualization

- Some useful business applications for data visualization include the following:
  1. Identifying areas that need attention or improvement.
  2. Clarifying which factors influence customer behavior.
  3. Helping understand which products to place where.
  4. Predicting sales volumes.

Data Visualization

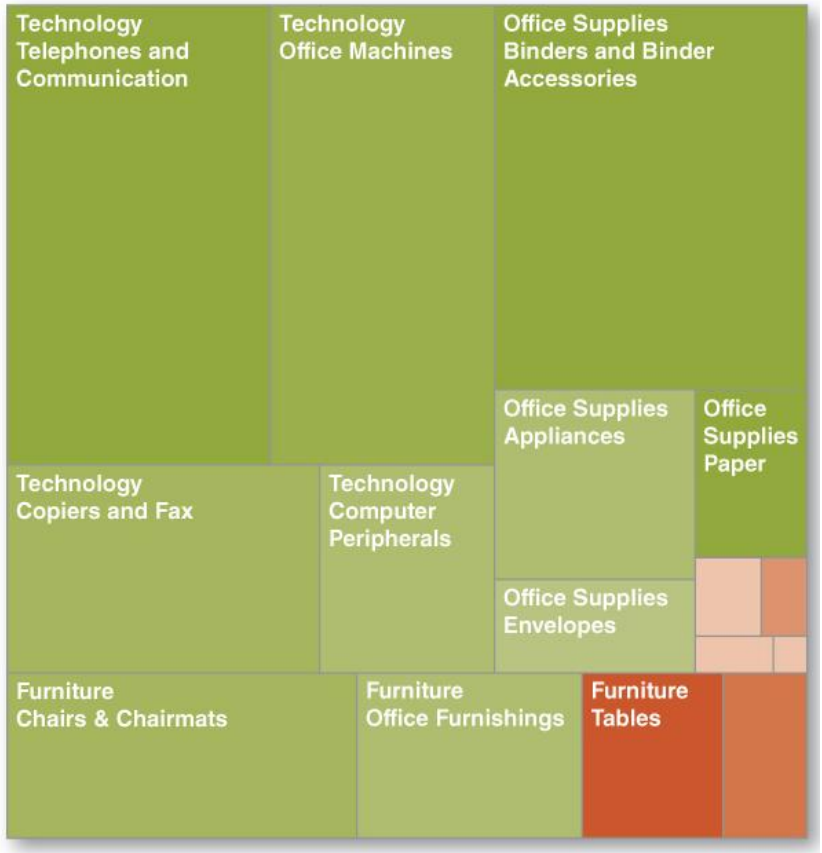
- Adding Value through Learning, Exploration, and Discovery with Data Visualization
  - Data visualization enables learning that is the basis for continuous improvement.
  - Data visualization is also used as a data explorer and data discovery tool.
  - Data Discovery is the process of using BI to collect data from various databases and consolidate it into a single source that can be easily and instantly evaluated.

Data Visualization

- Heat Maps
  - This heat map uses three colors to convey information at a glance.
  - The heat map is like a spreadsheet whose cells are formatted with colors instead of numbers.
  - Figure 6.12

		Region			
Product Cate...	Product Sub-Category	Central	East	South	West
Furniture	Bookcases	73	–10,151	–22,417	–676
	Chairs & Chairmats	37,920	33,583	34,026	44,409
	Office Furnishings	26,293	14,523	25,121	30,941
	Tables	–19,777	–50,677	26,172	–16,990
Office Supplies	Appliances	22,950	16,812	26,986	31,276
	Binders and Binder Accessories	73,951	71,420	69,530	92,273
	Envelopes	10,825	7,482	19,182	11,222
	Labels	2,429	4,041	3,479	3,740
	Paper	11,047	13,510	10,997	10,433
	Pens & Art Supplies	2,781	2,856	1,397	518
	Rubber Bands	–174	–238	156	178
	Scissors, Rulers and Trimmers	–1,765	–1,179	–2,903	–1,953
	Storage & Organization	–68	–7,233	11,836	–2,018
Technology	Computer Peripherals	11,971	14,808	30,475	37,280
	Copiers and Fax	513	67,254	63,598	35,997
	Office Machines	38,876	47,277	129,060	61,377
	Telephones and Communication	79,393	73,715	78,985	84,860

(a)



(b)

FIGURE 6.12 These heat maps represent the same data set using different colors (usually red and green) and color intensity to show the profitability of three product categories and their subcategories. In (a), data labels show detailed profit, while in (b), the area of each segment is used to make comparisons between profitability of product categories.

Data Visualization

- Augmented Reality (AR)
  - The highest level of data visualization currently available is augmented reality (AR) that compacts data into an easily digestible graphic or chart that can be visually projected with AR technology to fully engage the user.
  - Augmented Reality (AR)
    - ✓ The use of more contemporary 3-D visualization methods and techniques to illustrate the relationships within data including smart mapping, smart routines, machines learning, and natural language processing.
  - Some uses of augmented reality include:

Data Visualization

Coca-Cola	• Has developed an AR application that assists retailers in visualizing how a beverage cooler would fit into their stores.
Toys R Us	• in Australia created a Virtual Easter Egg Hunt for its young customers. • Using a computer table a child could follow Easter Bunnies around the store looking for digital eggs that were hidden around the store.
IKEA	• Uses augmented reality to assist customers with AR visualizations of how furniture will look in different living spaces.
Amazon View	• Lets online shoppers see how a product will look in their homes by clicking on the camera icon in the Amazon app and selecting products across categories like furniture, appliances, kitchenware, and home decor.

Data Visualization

Google	• Is trialing a new augmented reality feature for Google Maps. • In this new feature, the app pickups up a person's location via GPS and uses Street View data to narrow it down to the user's exact location. • It then displays big arrows and directions on the screen for the user to follow (Figure 6.14).
Bareburger	• Projects their new meatless Impossible Burgers onto their guests' virtual plate through the Snapchat app on their smartphones allowing Bareburger customers to view their lunch before ordering.

FIGURE 6.14 Google Maps is trialing an augmented reality feature.

Dashboards

- Dashboards improve information integration by collecting multiple, disparate data feeds and sources, extracting features of interest, and manipulating the data, so the information is in a more accessible format.
- Experts and novices can collect data quickly from disparate sources and then explore the data set with easy-to-use interactive dashboards (Figure 6.15).

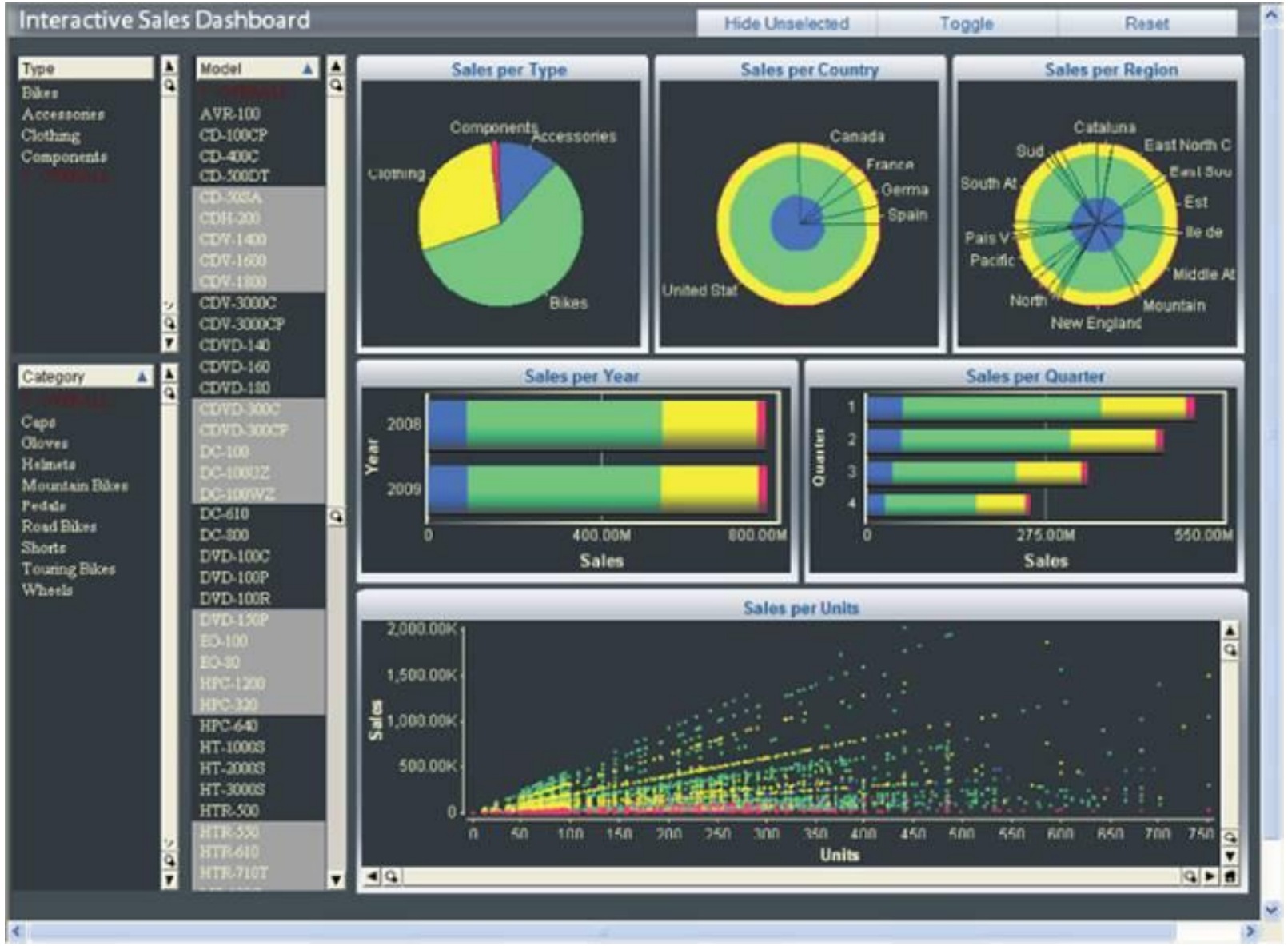


FIGURE 6.15 Dashboards allow users to interact with multiple corporate data sources on a single screen.

Dashboards: Major Components

Design	• The visualization method and descriptive captions to convey information so that they are correctly understood. • Infographics are widely used because they convey information in interesting and informative ways.
Performance Metrics	• KPIs and other real-time content displayed on the dashboard. • All dashboard data should reflect the current value of each metric.
API (Application Program Interface)	• APIs connect disparate data sources and feeds to display on the dashboard. • The alternative is for users or IT to manually enter data to the dashboard. • Dashboards created in this manner tend to fail because of the risk of incomplete, outdated, or wrong data, which users learn not to trust.
Access	• Preferred access is via a secure Web browser from a mobile device.

Dashboards

- Dashboards are Real Time
  - The purpose of dashboards is to give users a clear view of the current state of KPIs, real-time alerts, and other metrics about operations.
  - Dashboards work by connecting to business systems, such as accounting software, ERP, CRM, SCM, e-mail systems, website analytics programs, and project management software via APIs.
  - Table 6.3 lists some metrics that are commonly displayed on dashboards by function.

Dashboards

TABLE 6.3.1 Example of Metrics Displayed on Dashboards by Function

Dashboard Type	Metrics	
Accounting	• Net income • Net profits • Cash balance, actual vs. expected	• Profit, current month projection • Changes in A/R and A/P
E-commerce	• Daily website visitors by traffic source • Trend of mobile vs. tablet traffic • Location where visitors are located	• Top referring websites • Top keywords referring traffic • Revenue per website visitor
Finance	• Sales per day per channel • How revenue is trending	• Days with the strongest sales, weakest sales • Products selling the best, worst

Dashboards

TABLE 6.3.2 Example of Metrics Displayed on Dashboards by Function

Dashboard Type	Metrics	
Sales team	• Sales by lead source; which leads are most and least effective • Number of leads and proposals per salesperson • Proposal close percentage	• Salesperson closing percentages • Point at which prospective customers are being lost
Advertising	• Number of leads generated by advertising; which advertising is most and least effective • Cost per lead, by advertising source	• Advertising expense, as a percent of sales • Which advertising sources directly lead to sales
Order fulfilment	• Number of products manufactured, reworked • On-time completion percent	• Changes in inventory levels • Percent of on-time delivery per week, month

Dashboards

- Adding Value with Digital Dashboards
  1. Visibility
     - Blind spots are minimized or eliminated.
     - Threats and opportunities are detected as soon as possible.
  2. Continuous improvement
     - Executive dashboards are custom designed to display the user's critical metrics and measures.
  3. Single sign-on (통합 인증)
     - Single sign-on dashboards save time and effort.

Dashboards

- Adding Value with Digital Dashboards
  4. Deviations from what was budgeted or planned
     - Any metrics can be programmed to display deviations from targets, such as comparisons of actual and planned or budgeted.
  5. Accountability
     - When employees know that their performance is tracked in near real-time and can see their results, they tend to be motivated to improve their performance.

Data Mashups for Actionable Dashboards

- Using mashups, digital dashboards pull data from disparate data sources and feeds to report KPIs and operational or strategic information on intuitive dashboards and interactive displays.
  - Figure 6.16
- Enterprise mashups combine data from internal business sources (e.g., sales records, customer information, etc.) and/or information from external sources for enhanced usefulness and productivity.

Figure 6.16 Dashboards pull data from disparate data sources and feeds, manipulate the data, and display the metrics.

Data Mashups for Actionable Dashboards: Examples

Example	Description
Customer	• A customer data mashup provides a quick view of customer data for a salesperson in preparation for a customer site visit. • Data can be pulled from internal data stores and Web sources, such as contact information, links to related websites, recent customer orders, lists of critical situations, and more.
Logistics	• A logistics mashup displays inventory for a group of department stores based on specific criteria. • For example, you can mash current storm information onto a map of store locations and then wire the map to inventory data to show which stores located in the path of storms are low on generators.
Human Resource	• An HR mashup provides a quick glance at employee data such as profiles, salary, ratings, benefits status, and activities. • Data can be filtered to show custom views, for example, products whose average quarterly sales are lower than last quarter.

Data Mashups for Actionable Dashboards

- Enterprise Mashup Architecture
  - Figure 6.17 shows the general architecture of an enterprise mashup app.
  - Data from operational data stores, business systems, external data (economic data, suppliers, information, and competitors' activities), and real-time news feeds are integrated to generate an enterprise mashup.

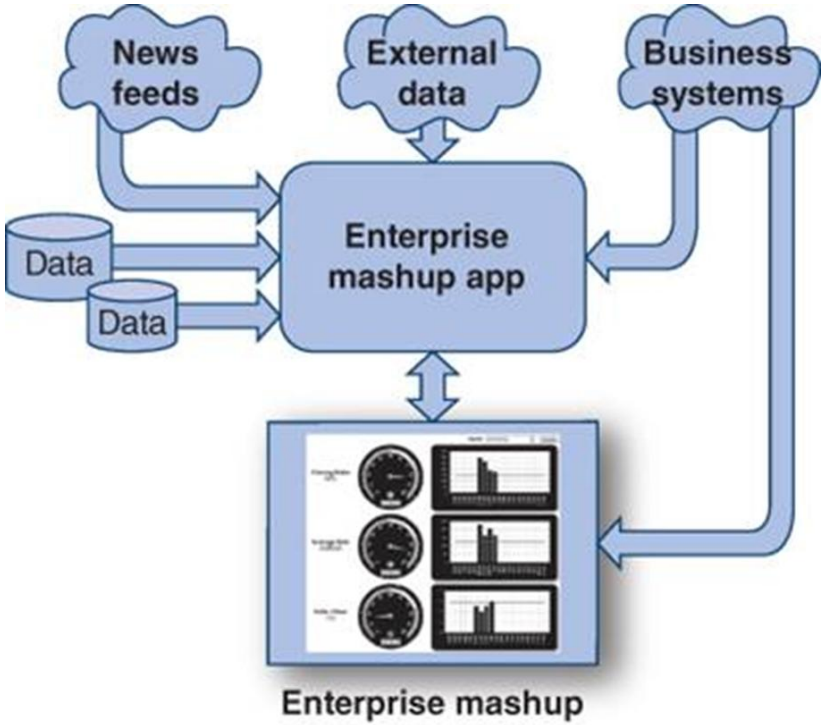
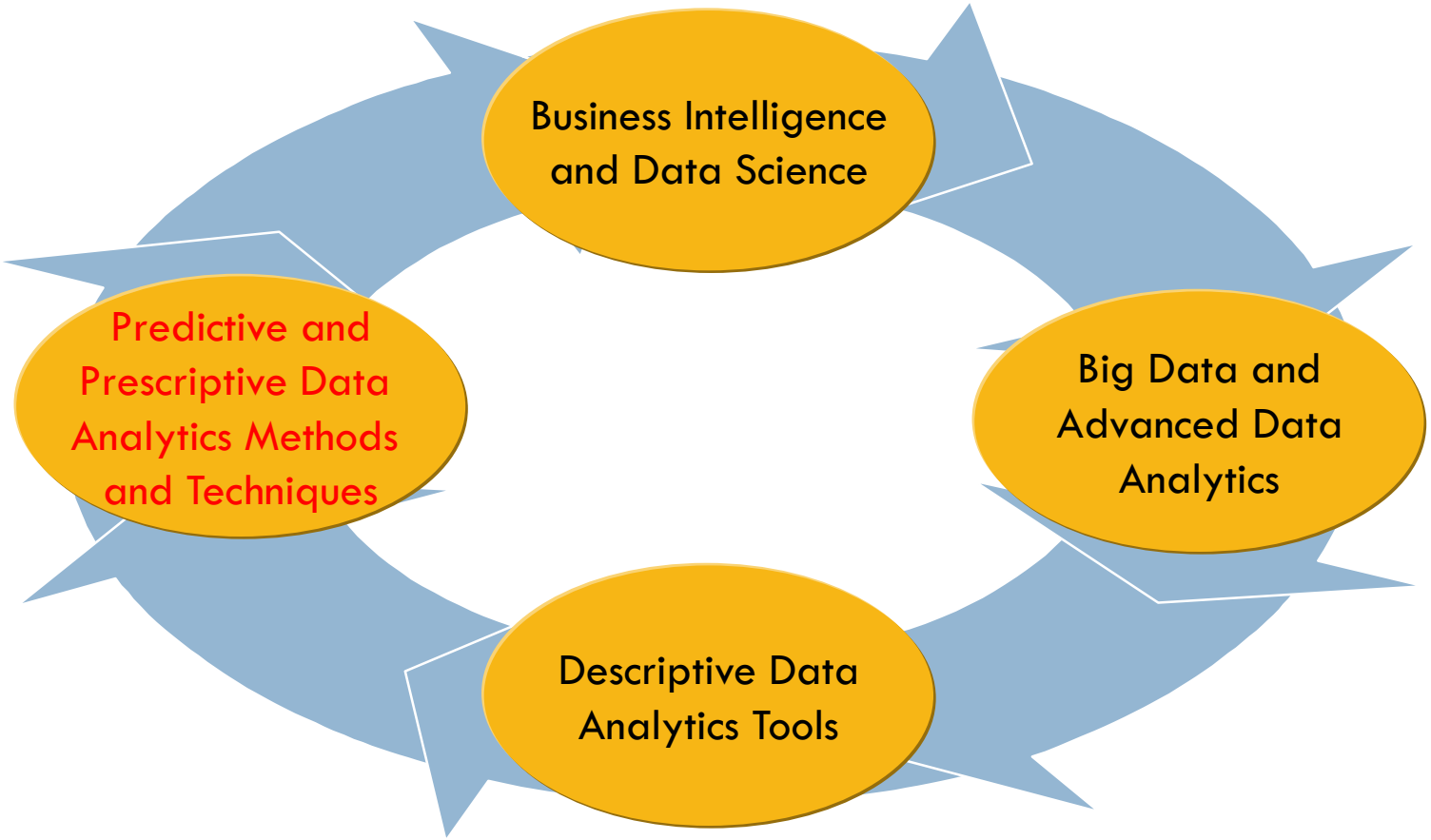


FIGURE 6.17 Architecture of enterprise mashup application

Data Mashups for Actionable Dashboards

- Adding Value with Mashups
  1. Dramatically reduces time and effort needed to combine disparate data sources.
  2. Users can define their own data mashups by combining fields from different data sources that were not previously modeled.
  3. Users can import external data sources, for example, spreadsheets and competitor data, to create new dashboards.
  4. Enables the building of complex queries by non-experts with a drag-and-drop query building tool.

Chapter Outline (4 of 4)



Predictive and Prescriptive Data Analytics Methods and Techniques

Text Mining

- Text Mining
  - The process of deriving high quality information from text aided by software that can identify concepts, patterns, topics, keywords, and other attributes in the unstructured data.
- Sentiment Analysis
  - Uses natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, and quantify affective stages and subjective information.

Text Mining

- Adding Value with Text Mining
  - Walmart analyzed clickstream data from its 45 million monthly online shoppers; then combined that data with product-and category-related popularity scores.
  - McDonald's bakery operation replaced manual equipment with high-speed photo analyses to inspect thousands of buns per minute for color, size, and sesame seed distribution.
  - Infinity Insurance discovered new insights that it applied to improve the performance of its fraud operation.

Spatial Data Mining

- A Geographic information system (GIS) connects data with geography to understand *what* belongs *where*.
- Geographic information system (GIS)
  - A computer-based tool that captures, manages, analyzes, and displays multidimensional geographic data, also called geospatial data.
- For example, it's really difficult to visualize the locations of towns by their latitude (위도) and longitude (경도) coordinates (좌표) listed in a spreadsheet, but it's easy to know where they are when you show these positions on a map.
  - Figure 6.18

	name	latitude	longitude
	Seattle	47.5700	–122.3400
	New York	40.7500	–73.9800
	Miami	25.7876	–80.2241
	Los Angeles	33.9900	–118.1800
	Dallas	32.8200	–96.8400
	Washington DC	38.9072	–77.0365

FIGURE 6.18 Latitude and Longitude Coordinates on a spreadsheet are much more difficult to visualize than when they are displayed on a map.

Spatial Data Mining

- Geocoding
  - In many cases, locations are already in existing data stores, but not in a format suitable for analytics.
  - A simple process called geocoding can convert postal addresses to geospatial data that can then be measured and analyzed.
  - By tapping into this resource, decision-makers can use the geographic or spatial context to detect and respond to opportunities.

Spatial Data Mining

- Case in Point: GM
  - General Motors (GM) spends a staggering (very large) $2 billion a year on marketing.
  - In the past, it shot-gunned its ads at the general public.
  - Now, it maps out which types of households will buy new cars, more accurately determines locations where people buy certain models, and channels its ads specifically to those areas.
  - As a result, GM spends less money to generate higher sales.

Spatial Data Mining

- Case in Point: GM
  - GM managers use ESRI's ArcGIS software to view local demographics, location characteristics, regional differences, and the competitive brand environment to determine how a given dealership should be performing compared to actual results.
  - The GIS makes it possible for GM to isolate demand, target its marketing efforts to local preferences, and position its dealerships to improve sales.
  - With the intelligence provided by the GIS, GM has increased sales despite cutting the advertising budget.

Spatial Data Mining

- GIS Is Not Your Grandfather's Map
  - Unlike a traditional flat map, a GIS-generated map is made up of many layers of information that provides users different ways to view a geographic space.
    - ✓ Figure 6.19

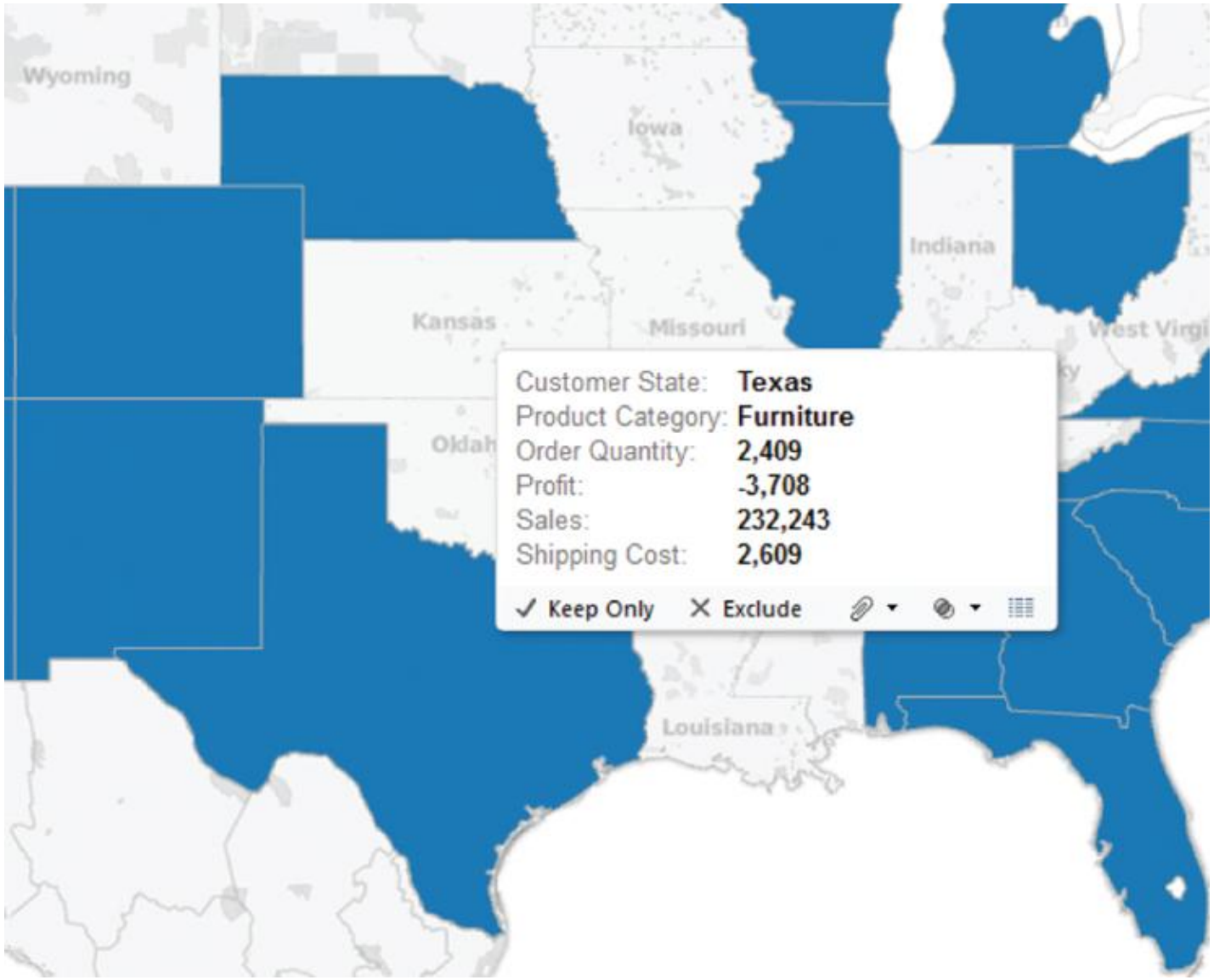


FIGURE 6.19 An example of a GIS-generated map. By hovering over a state, such as Texas, another layer of sales and financial data appears.

Spatial Data Mining

- Infrastructure and Location-Aware Collection of Geospatial Data
  - The infrastructure needed to collect geospatial data continues to expand.
  - Cellular and Internet service providers, sensors, Google Earth, GPS, and RFID systems know the location of each connected user or object.
  - Foursquare, Google Maps, and other mobile apps rely on GPS locations.


GPS

문서 토론

위키백과, 우리 모두의 백과사전.

다른 뜻에 대해서는 GPS (동음이의) 문서를 참고하십시오.

GPS(Global Positioning System 글로벌 포지셔닝 시스템[*]) 또는 **범지구위치결정시스템**은 현재 GLONASS와 함께 완전하게 운용되고 있는 범지구위성항법시스템 중 하나이다. 미국 국방부에서 개발되었으며 공식 명칭은 NAVSTAR GPS(NAVSTAR는 약자가 아니지만 종종 NAVigation System with Timing And Ranging 이라고 하기도 한다.)[1]이다. 무기 유도, 항법, 측량, 지도 제작, 측지, 시각 동기 등의 군용 및 민간용 목적으로 사용되고 있다.

Spatial Data Mining

- Adding Value with Spatial Data Mining
  - Analysts can pinpoint the geographic areas where the highest performing stores are established.
  - Retailers can learn how store sales are impacted by population or the proximity to competitors' stores.
  - A retail chain with plans to open a hundred new stores can use GIS to identify relevant demographics, proximity to highways, public transportation, and competitors' stores to select the best location options.
  - Food and consumer products companies can chart locations of complaint calls, enabling product traceability in the event of a crisis or recall.
  - Sales reps might better target their customer visits by analyzing the geography of sales targets.

Regression Modeling

- Linear Regression
  - A statistical method that analyzes and finds relationships between a dependent variable and one or more independent (or explanatory) variables.
  - Figure 6.20 presents an example of a simple linear regression model that predicts expected test scores (dependent variable) dependent on hours studied (independent/explanatory variable) for a class of 20 students.

Regression Modeling

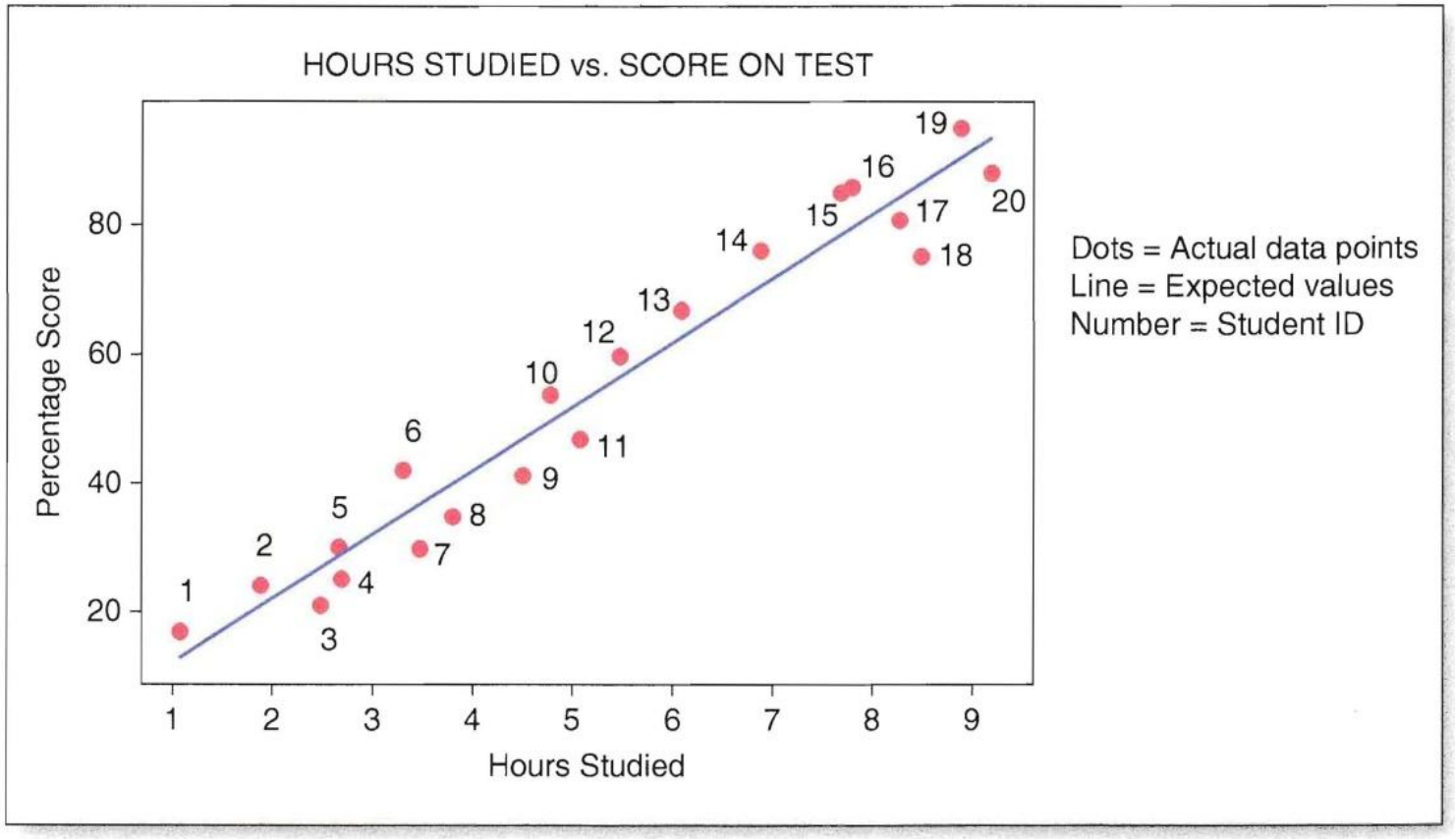


FIGURE 6.20 Simple linear regression model

Regression Modeling

- A Time-Series Regression
  - A time series is a collection of data values over time.
  - Time-series regression is performed by plotting a series of well-defined data points and attempting to predict what will happen to it in the future based on measuring the data at consistent time intervals over a specific period of time, such as monthly, quarterly, or annually.
  - The trend line shows the direction in which a variable is moving as time passes.

Regression Modeling

- A Time-Series Regression
  - Three ways data can be analyzed using a time-series regression are:
    1. Trend--series of data points go up, down, or stay flat over time.
    2. Rate of change--the extent of relative change between data points over time.
    3. Cycles--regularly repeating patterns in the data, such as at the end of a quarter when sales reps typically close sales out and see if they have made their target.

Regression Modeling

- Figure 6.21 shows two types of time series regressions-constant times series and trended time series.
- Constant time series is a time series in which the mean value of the time series is constant over time.
- Trended time series is a time series in which the mean value of the time series can fluctuate by season.

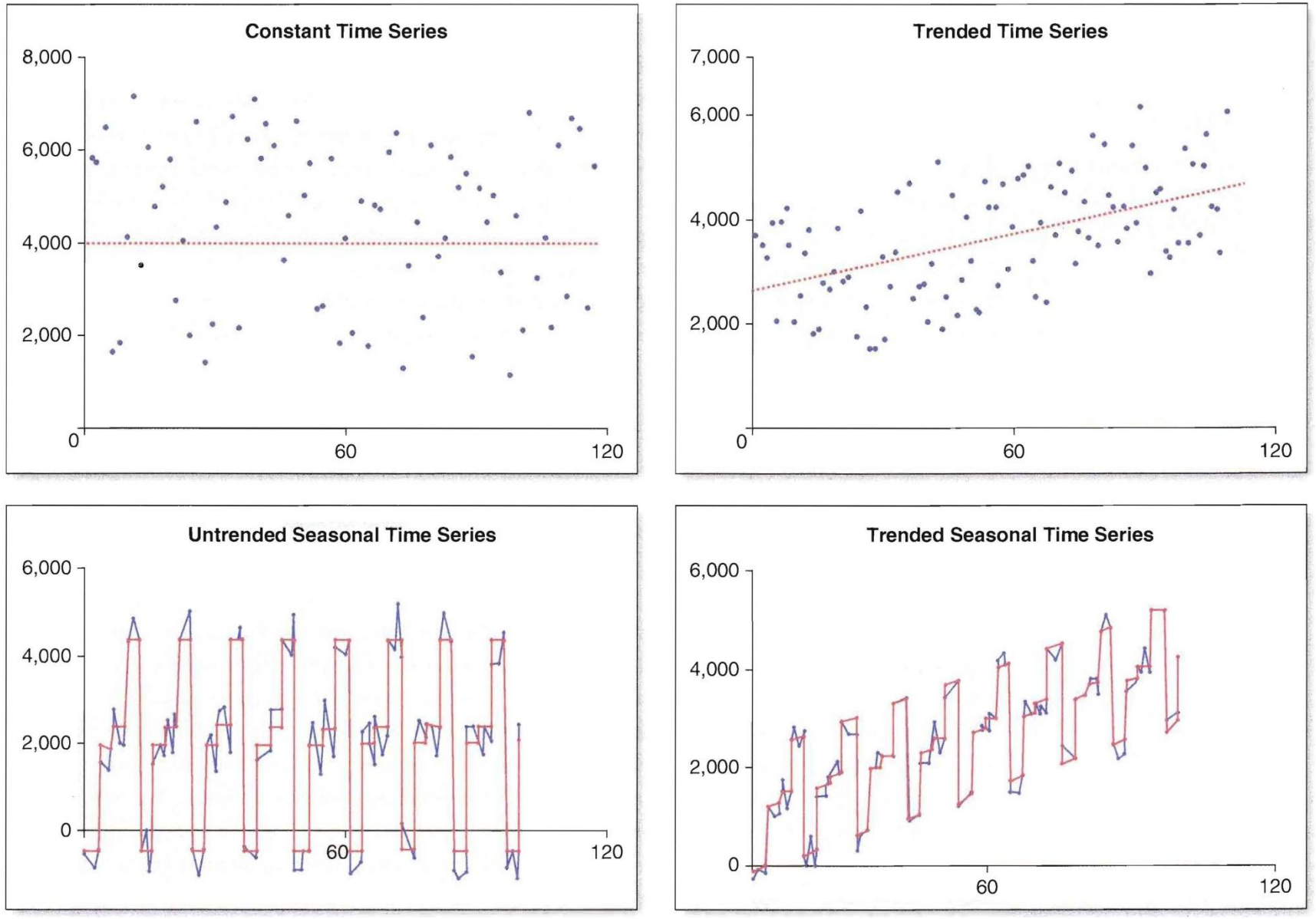


FIGURE 6.21 Example of time-series regression models

Regression Modeling

- Adding Value with Regression
  - Regression modeling is a key component in making data-driven decisions at just about every level of an organization.
  - It helps organizations understand the relationships between data points in their big data sets to make better decisions such as predicting sales or managing inventory levels based on supply on demand.

Regression Modeling

- Decision Optimization and Rules-Based Decision-Making
  - Decision optimization
    - ✓ The process of calculating values of variables that lead to an optimal value of the event under investigation.
  - Rules-based decision-making
    - ✓ Decision-making that helps novices make decisions like an expert.
  - With optimization and rules-based decision-making tools, organizations can run complex models, analyze scenarios, and factor in business rules and constraints to find the best course of action.

➢ Factor in: ~을 고려하다[감안하다]

Regression Modeling

- Adding Value with Optimization and Rules-Based Decision-Making
  - Decision-makers in all industry sectors are faced with decisions like this every day.
  - Other examples include:
    - ✓ Factory managers who must determine the best sequence for the manufacture of a product.
    - ✓ Airlines that must optimize schedules for aircraft, crews, and maintenance personnel.
    - ✓ Retailers who need to optimize their distribution networks.
  - These decisions are far too complex for spreadsheets or descriptive analytics apps.

Machine Learning

- Another data analytics tool that is becoming popular for use with big data sets is machine learning.
- The technology that is used to implement machine learning is known as cognitive computing.
- Cognitive computing
  - The technology that uses machine learning algorithms.

Machine Learning

- Adding Value with Machine Learning
  - Machine learning helps organizations identify images, personalize marketing campaigns, analyze genomics, and "steer" autonomous vehicles.
  - Machine learning creates models that provide accurate forecasts such as demand forecasts and equipment failure predictions by manipulating real-time inputs and historical data.

Machine Learning

- Adding Value with Machine Learning
  - The four main tasks that machine learning applies known rules to include:
    1. Categorizing people or things.
    2. Predicting likely outcomes or actions based on identified patterns.
    3. Identifying previously unknown patterns and relationships.
    4. Detecting unexpected behaviors.

Machine Learning: Under Armour

- An American sportswear company that manufactures footwear and apparel, uses IBM Watson Cognitive Computing to provide the "Cognitive Coaching System"--a personal health assistant that provides users with real-time, data-based coaching using sensor and manually input data for sleep, fitness, activity, and nutrition.
- The app has a rating of 4.5 stars from users and Under Armour has reported more than 51% growth in its customer base since implementing the Cognitive Coaching System.

Machine Learning: Coca-Cola Amatil

- The largest bottler and distributor of non-alcoholic bottled beverages in the Asia Pacific region, uses Trax Retail Execution--an image-based technology--let's sales reps take photos of store shelves with their mobile devices.
- The images are then sent to the Trax Cloud and analyzed, returning actionable reports within minutes to sales reps who can quickly identify performance gaps and apply corrective actions in stores and providing more detailed online assessment to management.
- Within five months after implementing Trax Retail Execution, Coca-Cola Amatil reported a gain of 1.3% market share in the Asia Pacific region.

Machine Learning: Peter Glenn

- An outdoor apparel and gear retailer and wholesaler, uses AgilOne Advanced Analytics to gain insight into its customer profile and act on their findings.
- When they looked at trends between buyer groups, they were able to make better segmentation decisions and when the company learned that more than 80% of its customer base had left, they were able to use that information to retarget and re-engage lost customers.