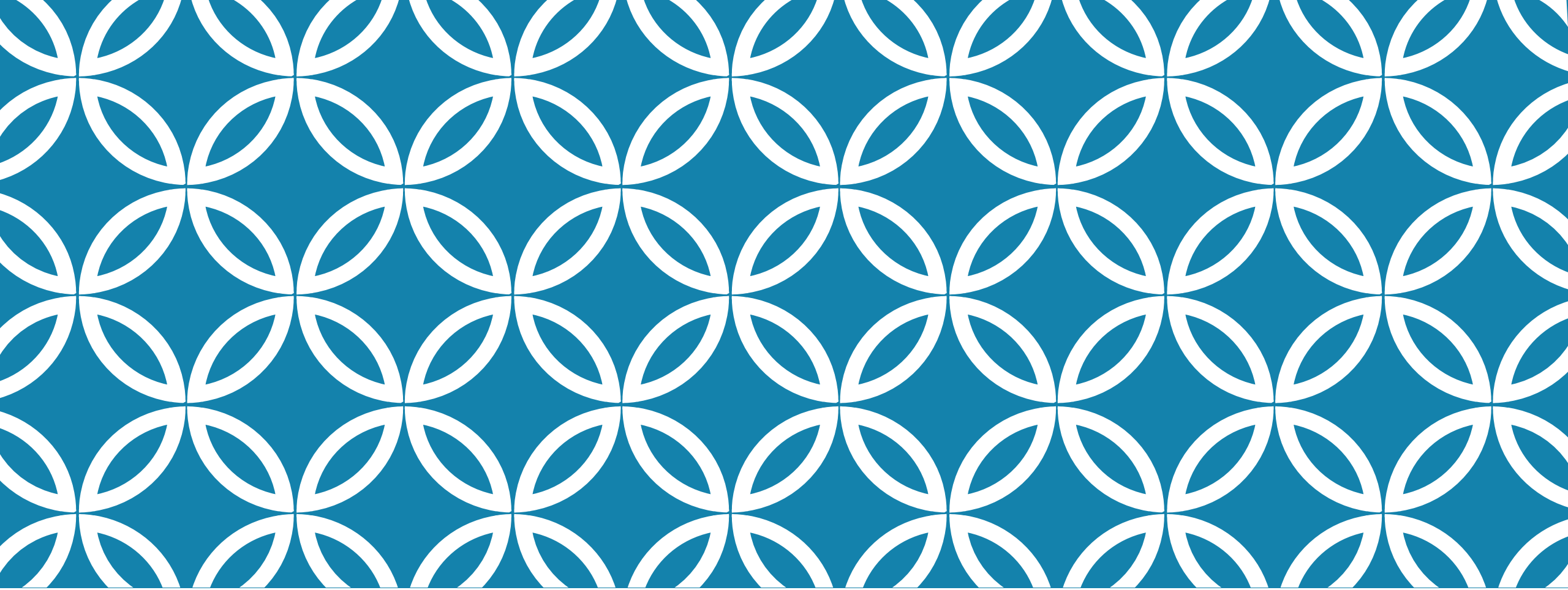
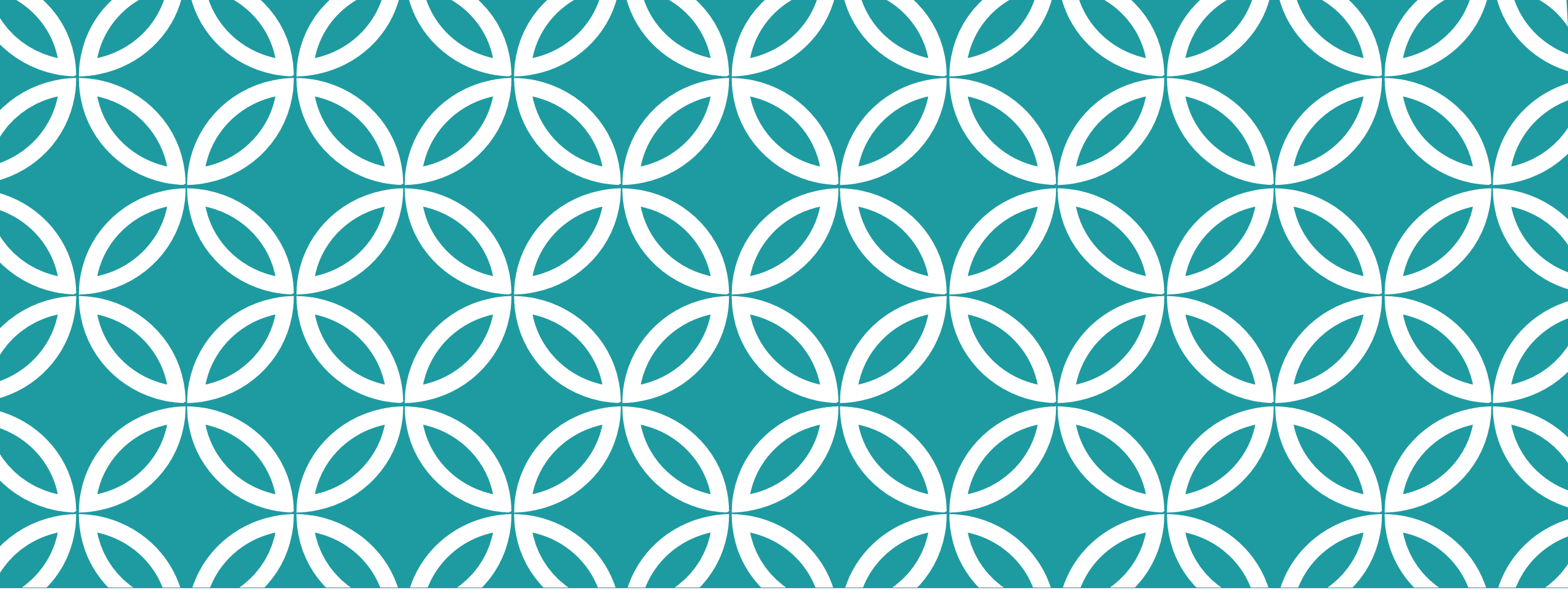


INTRODUCTION TO DATA ANALYTICS

Lesson 1.1



INTRODUCTION

Lesson 1.1.1

INTRODUCTION

- Three developments spurred recent explosive growth in the use of analytical methods in business applications:
- First development:
  - Technological advances, Internet social networks, and data generated from personal electronic devices, produce incredible amounts of data for businesses.
  - Businesses want to use these data to improve the efficiency and profitability of their operations, better understand their customers, price their products more effectively, and gain a competitive advantage.

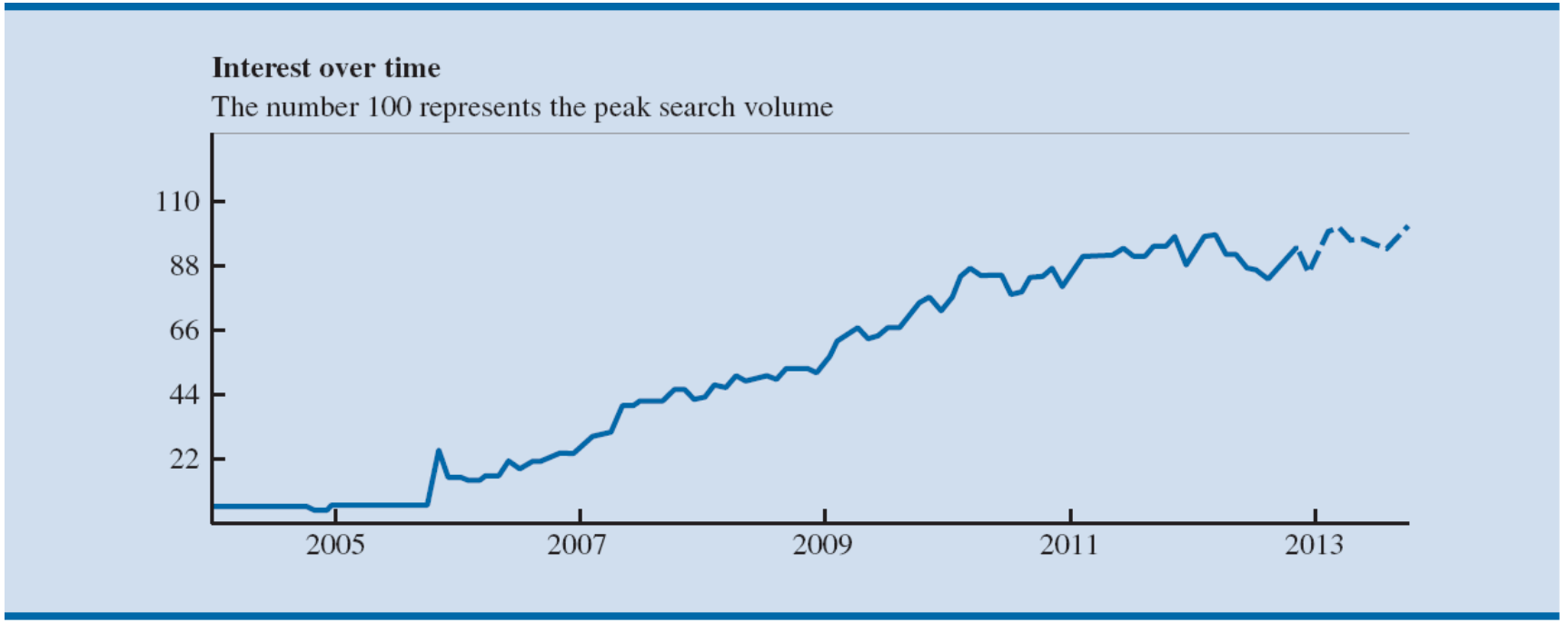
INTRODUCTION

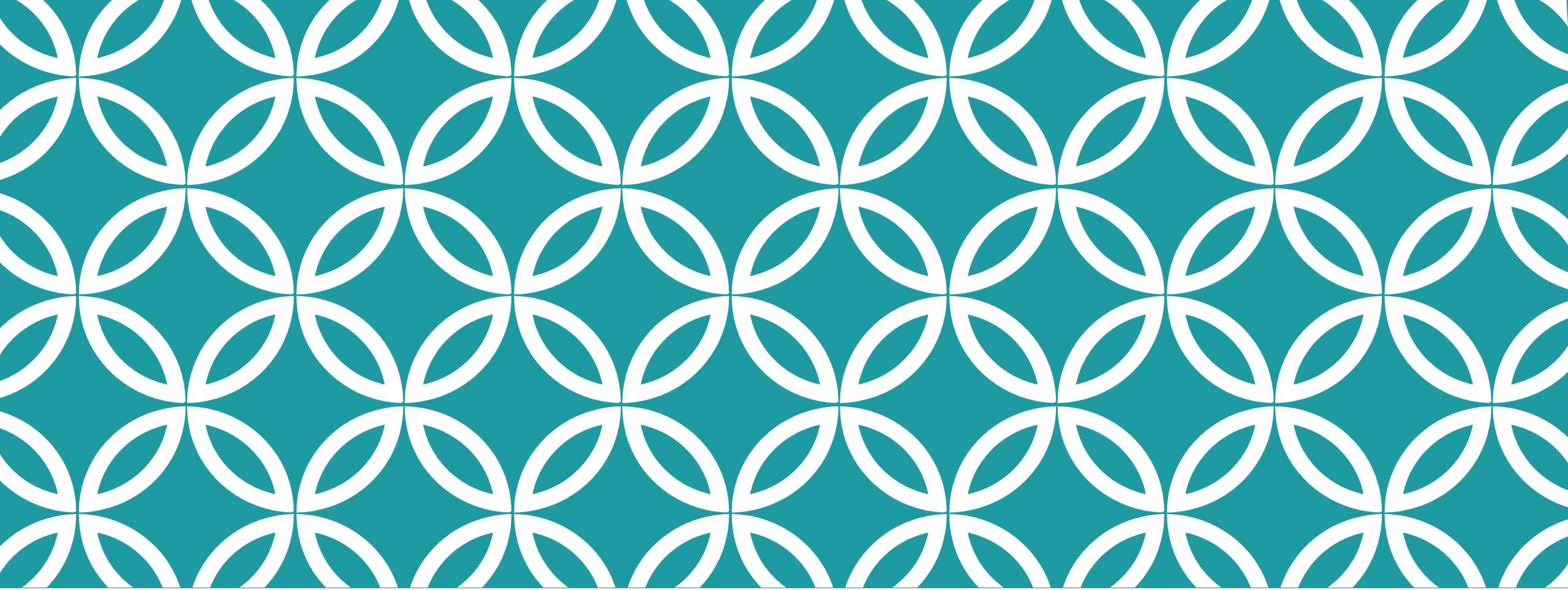
- Second development:
  - Ongoing research has resulted in numerous methodological developments, including:
    - Advances in computational approaches to effectively handle and explore massive amounts of data
    - Faster algorithms for optimization and simulation, and
    - More effective approaches for visualizing data.

INTRODUCTION

- Third development:
  - The methodological developments were paired with an explosion in computing power and storage capability.
  - Better computing hardware, parallel computing, and cloud computing have enabled businesses to solve big problems faster and more accurately than ever before.

FIGURE 1.1 - GOOGLE TRENDS GRAPH OF SEARCHES ON THE TERM *ANALYTICS*





DECISION MAKING

Lesson 1.1.2

DECISION MAKING

- Managers' responsibility:
  - To make strategic, tactical, or operational decisions.
- **Strategic decisions:**
  - Involve higher-level issues concerned with the overall direction of the organization.
  - These decisions define the organization's overall goals and aspirations for the future.

DECISION MAKING

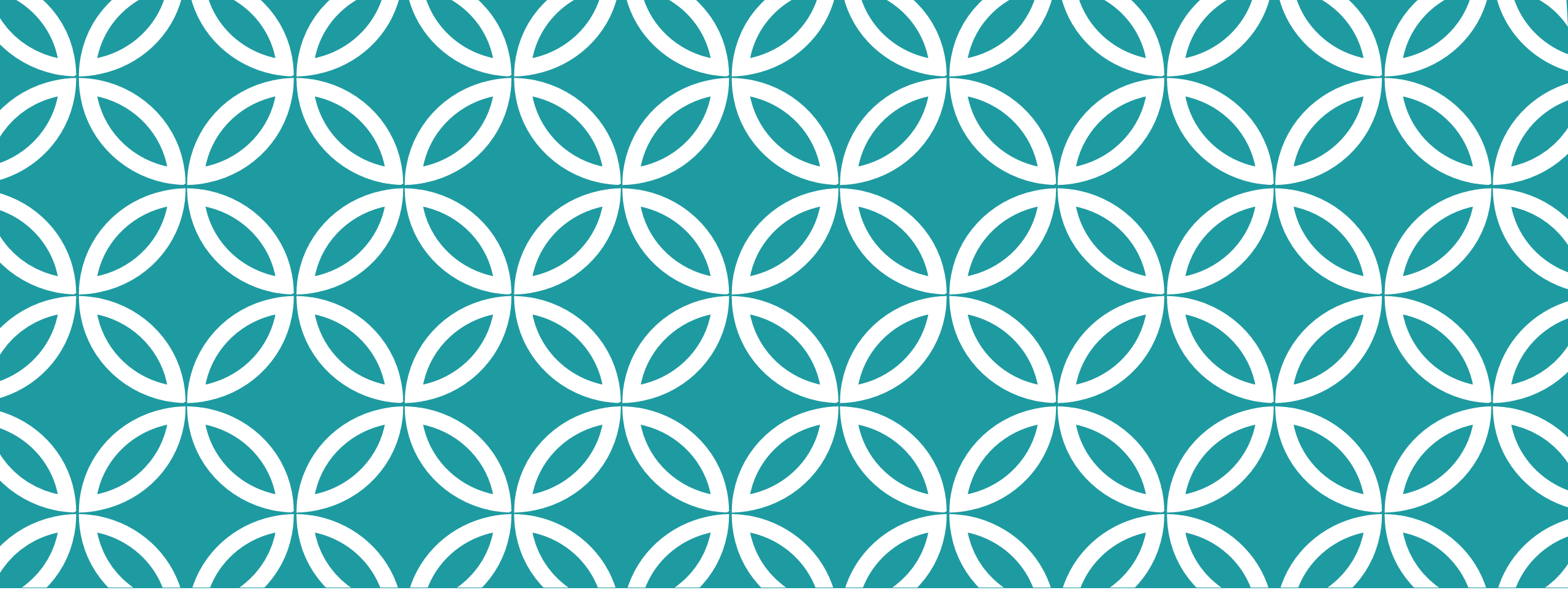
- **Tactical decisions:**
  - Concern how the organization should achieve the goals and objectives set by its strategy.
  - They are usually the responsibility of midlevel management.
- **Operational decisions:**
  - Affect how the firm is run from day to day.
  - They are the domain of operations managers, who are the closest to the customer.

DECISION MAKING

- Decision making can be defined as the following process
  1. Identify and define the problem
  2. Determine the criteria that will be used to evaluate alternative solutions
  3. Determine the set of alternative solutions
  4. Evaluate the alternatives
  5. Choose an alternative

DECISION MAKING

- Common approaches to making decisions
  - Tradition
  - Intuition
  - Rules of thumb
  - Using the relevant data available



BUSINESS ANALYTICS DEFINED

Lesson 1.1.3

BUSINESS ANALYTICS DEFINED

- **Business analytics:**
  - Scientific process of transforming data into insight for making better decisions.
  - Used for data-driven or fact-based decision making, which is often seen as more objective than other alternatives for decision making.

BUSINESS ANALYTICS DEFINED

- Tools of business analytics can aid decision making by:
  - Creating insights from data
  - Improving our ability to more accurately forecast for planning
  - Helping us quantify risk
  - Yielding better alternatives through analysis and optimization

A CATEGORIZATION OF ANALYTICAL METHODS AND MODELS

Lesson 1.1.4

DESCRIPTIVE ANALYTICS

- **Descriptive analytics:** It encompasses the set of techniques that describes what has happened in the past.

Examples - data queries, reports, descriptive statistics, data visualization (data dashboards), data-mining techniques, and basic what-if spreadsheet models.

Data query - It is a request for information with certain characteristics from a database.

DESCRIPTIVE ANALYTICS

- Data dashboards - Collections of tables, charts, maps, and summary statistics that are updated as new data become available.
  - Uses of dashboards
    - To help management monitor specific aspects of the company's performance related to their decision-making responsibilities.
    - For corporate-level managers, daily data dashboards might summarize sales by region, current inventory levels, and other company-wide metrics.
    - Front-line managers may view dashboards that contain metrics related to staffing levels, local inventory levels, and short-term sales forecasts.

PREDICTIVE ANALYTICS

- **Predictive analytics**: It consists of techniques that use models constructed from past data to predict the future or ascertain the impact of one variable on another.
  - Survey data and past purchase behavior may be used to help predict the market share of a new product.

PREDICTIVE ANALYTICS

- Techniques used in Predictive Analytics: *contd.*

Data mining

- Used to find patterns or relationships among elements of the data in a large database; often used in predictive analytics.

Simulation

- It involves the use of probability and statistics to construct a computer model to study the impact of uncertainty on a decision.

PRESCRIPTIVE ANALYTICS

- **Prescriptive Analytics:** It indicates a best course of action to take
  - Models used in prescriptive analytics:

Optimization models

- Models that give the best decision subject to constraints of the situation.

Simulation optimization

- Combines the use of probability and statistics to model uncertainty with optimization techniques to find good decisions in highly complex and highly uncertain settings.

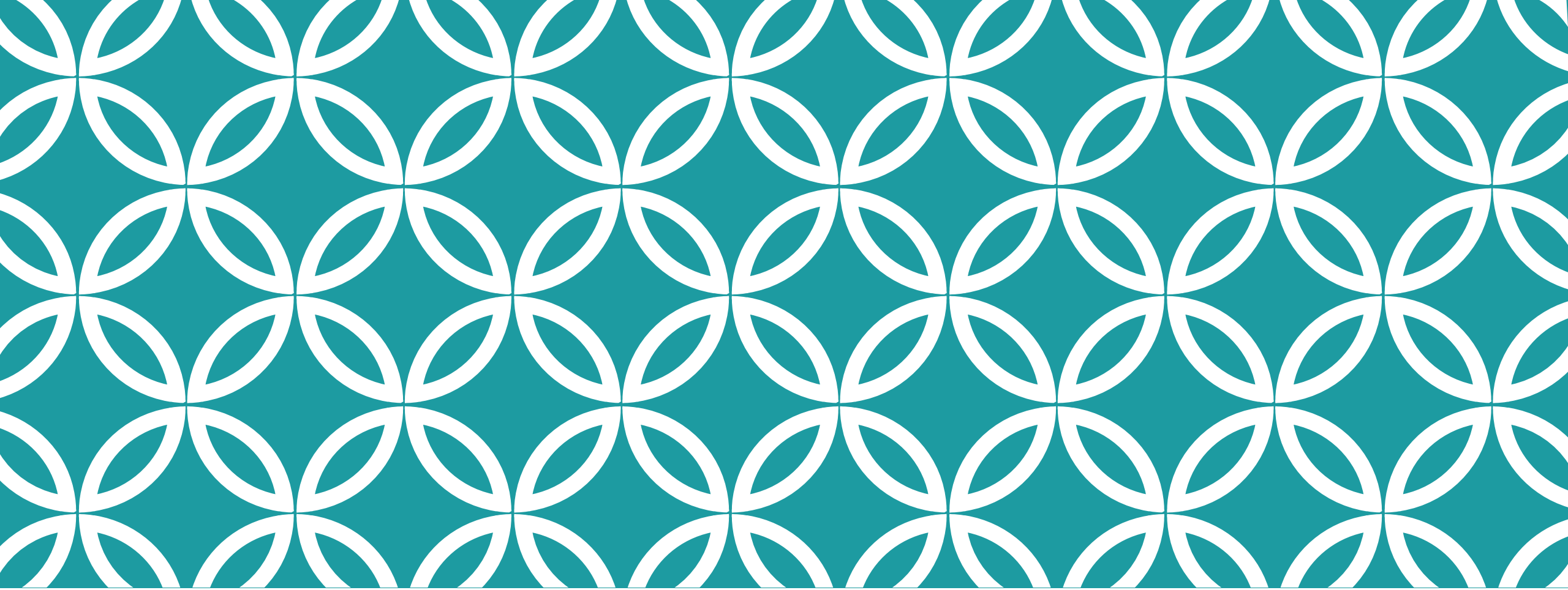
Decision analysis

- Used to develop an optimal strategy when a decision maker is faced with several decision alternatives and an uncertain set of future events.
- It also employs utility theory, which assigns values to outcomes based on the decision maker's attitude toward risk, loss, and other factors.

PRESCRIPTIVE ANALYTICS

- Optimization models

Model	Field	Purpose
Portfolio models	Finance	Use historical investment return data to determine the mix of investments that yield the highest expected return while controlling or limiting exposure to risk.
Supply network design models	Operations	Provide the cost-minimizing plant and distribution center locations subject to meeting the customer service requirements.
Price markdown models	Retailing	Uses historical data to yield revenue-maximizing discount levels and the timing of discount offers when goods have not sold as planned.

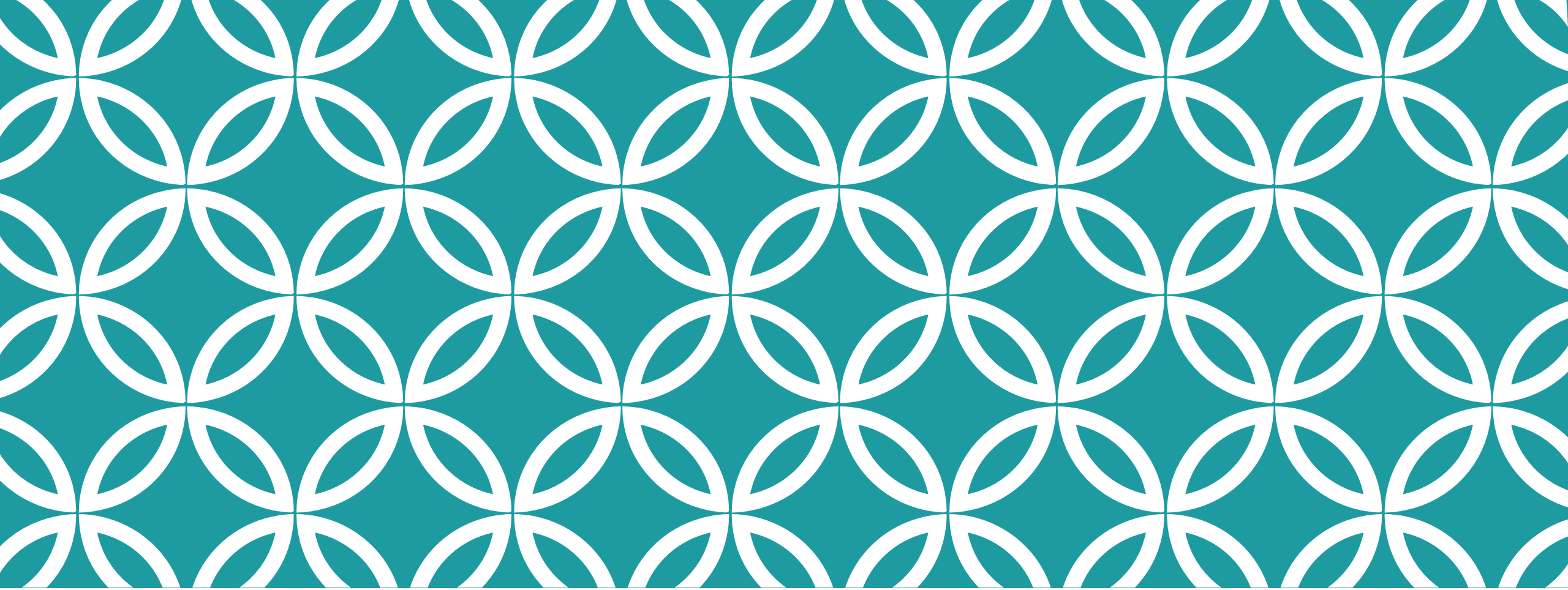


BIG DATA

Lesson 1.1.5

BIG DATA

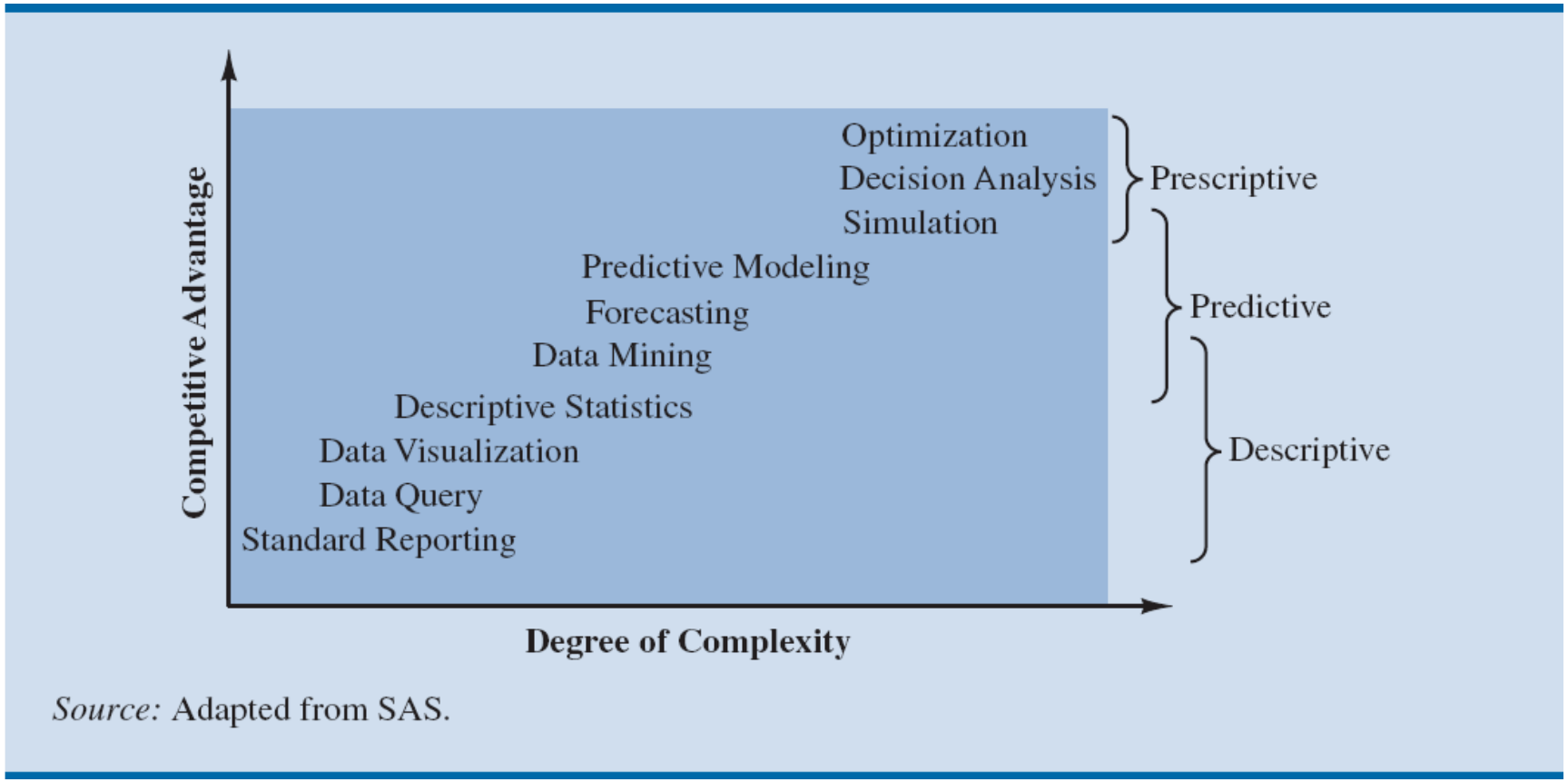
- **Big data:** A set of data that cannot be managed, processed, or analyzed with commonly available software in a reasonable amount of time.
  - Big data represents opportunities.
  - It also presents analytical challenges from a processing point of view and consequently has itself led to an increase in the use of analytics.
  - More companies are hiring data scientists who know how to process and analyze massive amounts of data.



BUSINESS ANALYTICS IN PRACTICE

Lesson 1.1.6

FIGURE 1.2 - THE SPECTRUM OF BUSINESS ANALYTICS



Source: Adapted from SAS.

BUSINESS ANALYTICS IN PRACTICE

Retail Markdown Decisions

- Most department stores clear seasonal inventory by reducing prices.
- The question is:
- When to reduce the price and by how much?
- Descriptive analytics: examine historical data for similar products (prices, units sold, advertising, …)
- Predictive analytics: predict sales based on price
- Prescriptive analytics: find the best sets of pricing and advertising to maximize sales revenue

BUSINESS ANALYTICS IN PRACTICE

Harrah's Entertainment

- Harrah's owns numerous hotels and casinos
- Uses analytics to:
  - forecast demand for rooms
  - segment customers by gaming activities
- Uses prescriptive models to:
  - set room rates
  - allocate rooms
  - offer perks and rewards to customers

BUSINESS ANALYTICS IN PRACTICE

- Types of applications of analytics by application area
- Financial analytics
  - Use of predictive models
    - To forecast future financial performance
    - To assess the risk of investment portfolios and projects
    - To construct financial instruments such as derivatives

BUSINESS ANALYTICS IN PRACTICE

- Financial analytics (contd.)
  - Use of prescriptive models
    - To construct optimal portfolios of investments
    - To allocate assets, and
    - To create optimal capital budgeting plans.
  - Simulation is also often used to assess risk in the financial sector

BUSINESS ANALYTICS IN PRACTICE

- Human resource (HR) analytics
  - New area of application for analytics
  - The HR function is charged with ensuring that the organization
    - Has the mix of skill sets necessary to meet its needs
    - Is hiring the highest-quality talent and providing an environment that retains it, and
    - Achieves its organizational diversity goals.

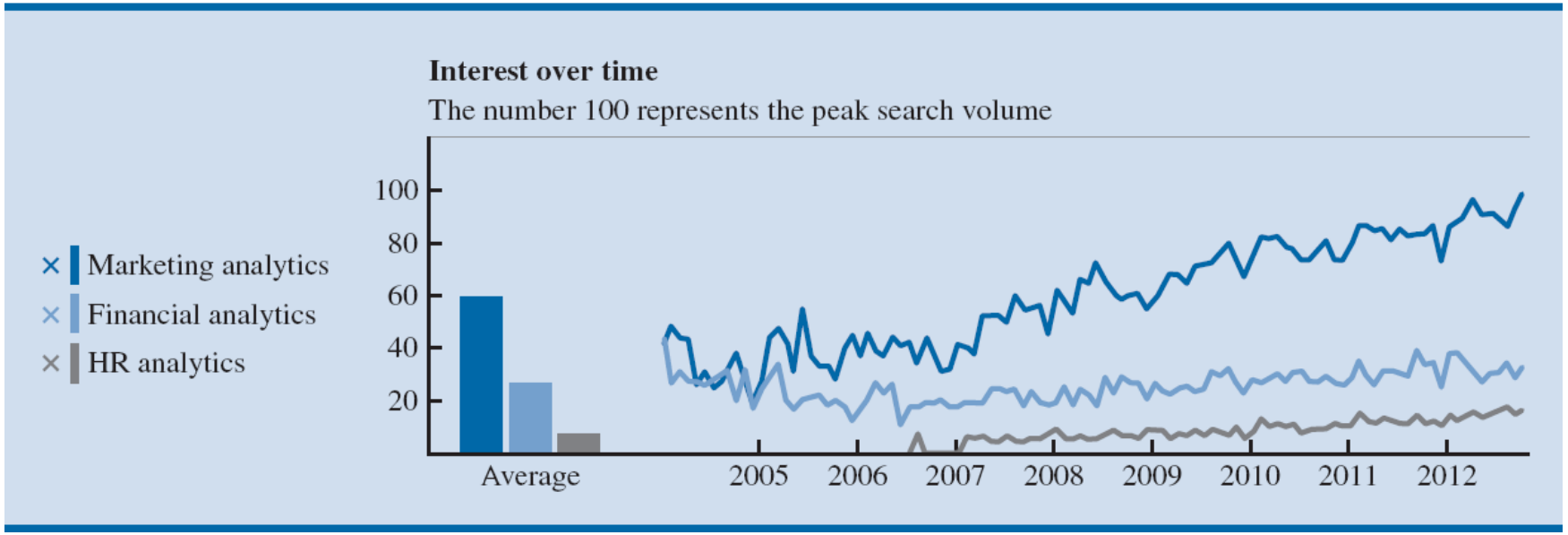
BUSINESS ANALYTICS IN PRACTICE

- Marketing analytics
  - Marketing is one of the fastest growing areas for the application of analytics.
  - A better understanding of consumer behavior through the use of scanner data and data generated from social media has led to an increased interest in marketing analytics.

BUSINESS ANALYTICS IN PRACTICE

- Marketing analytics (contd.)
  - A better understanding of consumer behavior through marketing analytics leads to:
    - The better use of advertising budgets
    - More effective pricing strategies
    - Improved forecasting of demand
    - Improved product line management, and
    - Increased customer satisfaction and loyalty

FIGURE 1.3 - GOOGLE TRENDS FOR MARKETING, FINANCIAL, AND HUMAN RESOURCE ANALYTICS, 2004–2012



BUSINESS ANALYTICS IN PRACTICE

- Health care analytics
  - Descriptive, predictive, and prescriptive analytics are used:
    - To improve patient, staff, and facility scheduling
    - Patient flow
    - Purchasing
    - Inventory control
  - Use of prescriptive analytics for diagnosis and treatment

BUSINESS ANALYTICS IN PRACTICE

- Supply chain analytics
  - The core service of companies such as UPS and FedEx is the efficient delivery of goods, and analytics has long been used to achieve efficiency.
  - The optimal sorting of goods, vehicle and staff scheduling, and vehicle routing are all key to profitability for logistics companies such as UPS, FedEx, and others like them.
  - Companies can benefit from better inventory and processing control and more efficient supply chains.

BUSINESS ANALYTICS IN PRACTICE

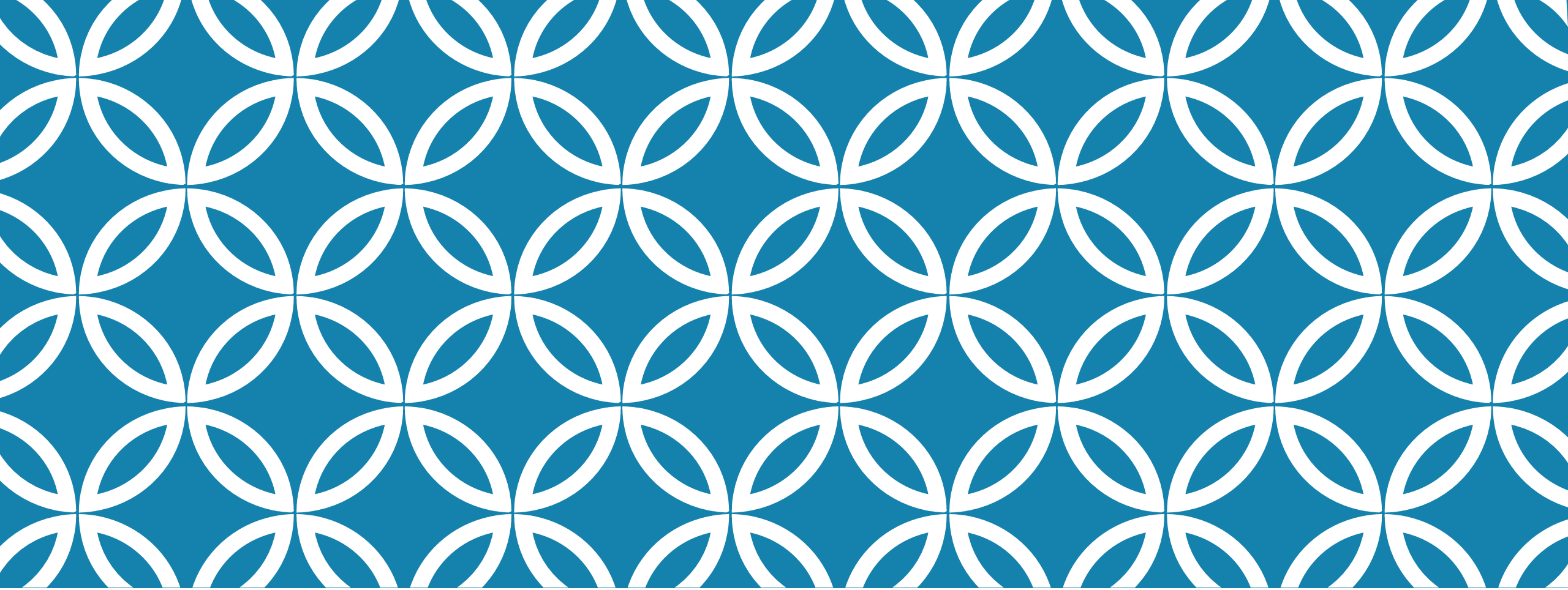
- Analytics for government and nonprofits
  - To drive out inefficiencies
  - To increase the effectiveness and accountability of programs
  - Analytics for nonprofit agencies
    - To ensure their effectiveness and accountability to their donors and clients.

BUSINESS ANALYTICS IN PRACTICE

- Sports analytics
  - Used for player evaluation and on-field strategy in professional sports.
    - To assess players for the amateur drafts and to decide how much to offer players in contract negotiations.
    - Professional motorcycle racing teams that use sophisticated optimization for gearbox design to gain competitive advantage.

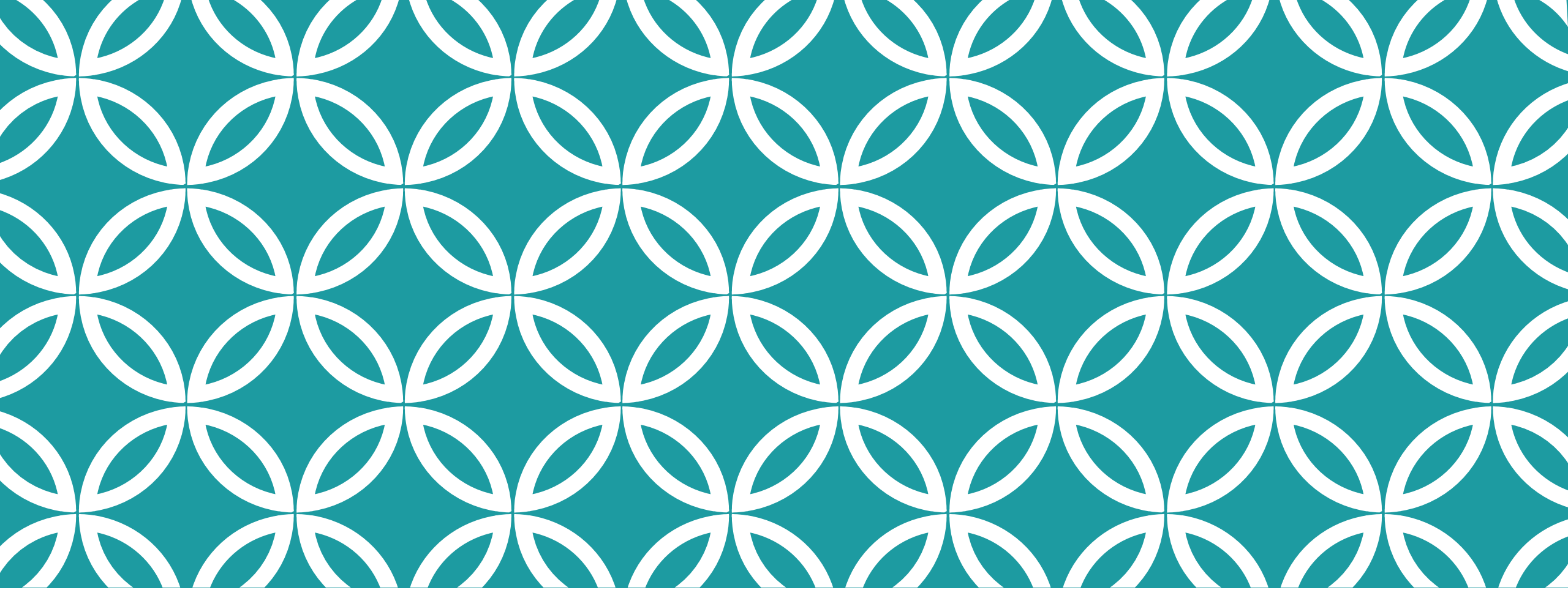
BUSINESS ANALYTICS IN PRACTICE

- Sports analytics (contd.)
  - The use of analytics for off-the-field business decisions is also increasing rapidly.
    - Using prescriptive analytics, franchises across several major sports dynamically adjust ticket prices throughout the season to reflect the relative attractiveness and potential demand for each game.



OVERVIEW OF DATA

Lesson 2.1



OVERVIEW OF USING DATA: DEFINITIONS AND GOALS

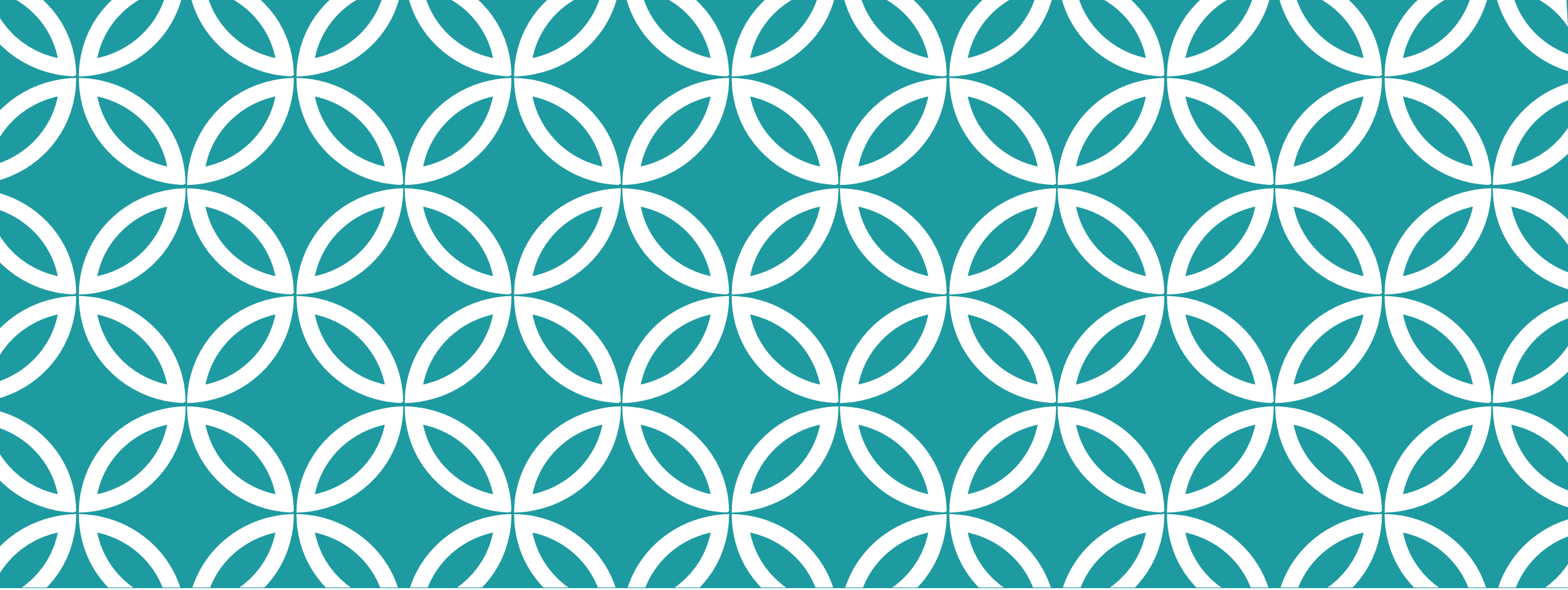
Lesson 2.1.1

OVERVIEW OF USING DATA: DEFINITIONS AND GOALS

- **Data**: The facts and figures collected, analyzed, and summarized for presentation and interpretation.
- **Variable**: A characteristic or a quantity of interest that can take on different values.
- **Observation**: Set of values corresponding to a set of variables.
- **Variation**: The difference in a variable measured over observations.
- **Random variable/uncertain variable**: A quantity whose values are not known with certainty.

TABLE 2.1 - DATA FOR DOW JONES INDUSTRIAL INDEX COMPANIES

Company	Symbol	Industry	Share Price ($)	Volume
Alcoa	AA	Manufacturing	8.03	8,360,181
American Express	AXP	Financial	66.83	5,020,965
Boeing	BA	Manufacturing	87.82	3,377,781
Bank of America	BAC	Financial	11.67	85,564,239
Caterpillar	CAT	Manufacturing	80.60	4,418,069
Cisco Systems	CSCO	Technology	20.47	37,824,927
Chevron Corporation	CVX	Chemical, Oil, and Gas	116.21	4,331,463
DuPont	DD	Chemical, Oil, and Gas	48.97	5,610,522
Walt Disney	DIS	Entertainment	61.28	5,893,711
General Electric	GE	Conglomerate	21.81	74,030,249
The Home Depot	HD	Retail	74.04	5,627,195
Hewlett-Packard	HPQ	Technology	19.68	20,229,367
IBM	IBM	Technology	190.29	13,890,330
Intel	INTC	Technology	22.38	33,303,641
Johnson & Johnson	JNJ	Pharmaceuticals	84.04	6,094,620
JPMorgan Chase	JPM	Banking	47.28	12,334,210
Coca-Cola	KO	Food and Drink	42.60	8,916,978
McDonald's	MCD	Food and Drink	99.94	5,571,538
3M	MMM	Conglomerate	105.79	1,850,264
Merck	MRK	Pharmaceuticals	47.18	6,601,636
Microsoft	MSFT	Technology	29.77	76,918,154
Pfizer	PFE	Pharmaceuticals	30.91	16,917,714
Procter & Gamble	PG	Consumer Goods	81.42	7,894,506
AT&T	T	Telecommunications	38.28	14,762,872
Travelers	TRV	Insurance	84.79	1,284,813
UnitedHealth Group Inc.	UNH	Healthcare	59.94	3,730,520
United Technologies Corporation	UTX	Conglomerate	92.92	2,466,956
Verizon Communications	VZ	Telecommunications	52.04	9,643,848
Wal-Mart	WMT	Retail	78.07	4,766,959
ExxonMobil	XOM	Chemical, Oil, and Gas	87.02	9,446,864



TYPES OF DATA

Lesson 2.1.2

TYPES OF DATA

- **Population**: All elements of interest
- **Sample**: Subset of the population
  - **Random sampling** - A sampling method to gather a representative sample of the population data.
- **Quantitative data**: Data on which numeric and arithmetic operations, such as addition, subtraction, multiplication, and division, can be performed.
- **Categorical data**: Data on which arithmetic operations cannot be performed.

TYPES OF DATA

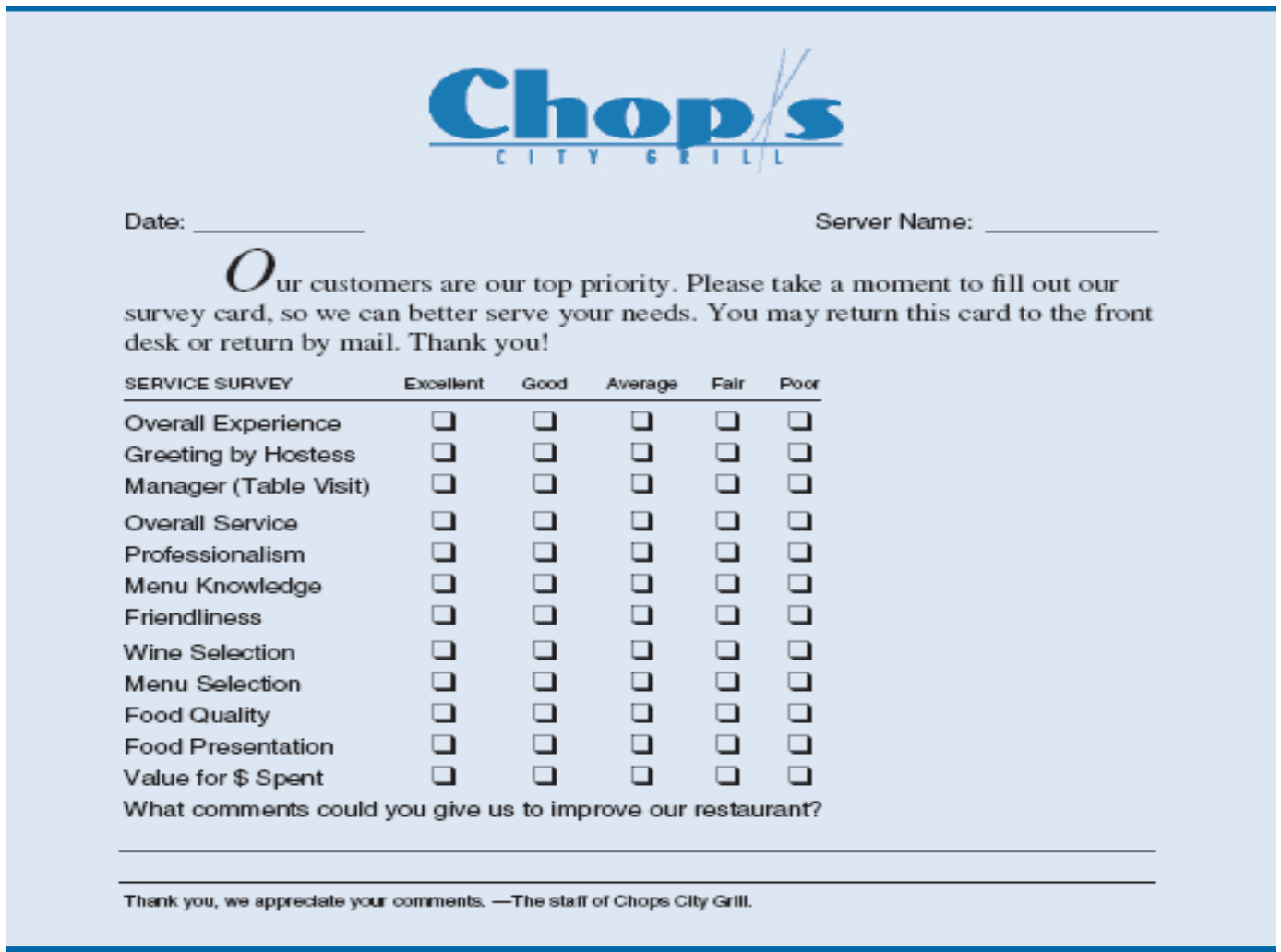
- **Cross-sectional data**: Data collected from several entities at the same, or approximately the same, point in time.
- **Time series data**: Data collected over several time periods.
  - Graphs of time series data are frequently found in business and economic publications.
    - Help analysts understand what happened in the past, identify trends over time, and project future levels for the time series.

FIGURE 2.1 - DOW JONES INDEX VALUES SINCE 2002

SOURCES OF DATA

- Sources of data
  - Experimental study - A variable of interest is first identified.
    - Then one or more other variables are identified and controlled or manipulated so that data can be obtained about how they influence the variable of interest.
  - Nonexperimental study or observational study - Make no attempt to control the variables of interest.
    - A survey is perhaps the most common type of observational study.

FIGURE 2.2 - CUSTOMER OPINION QUESTIONNAIRE USED BY CHOPS CITY GRILL RESTAURANT



Chop's
CITY GRILL

Date: _____ Server Name: _____

Our customers are our top priority. Please take a moment to fill out our survey card, so we can better serve your needs. You may return this card to the front desk or return by mail. Thank you!

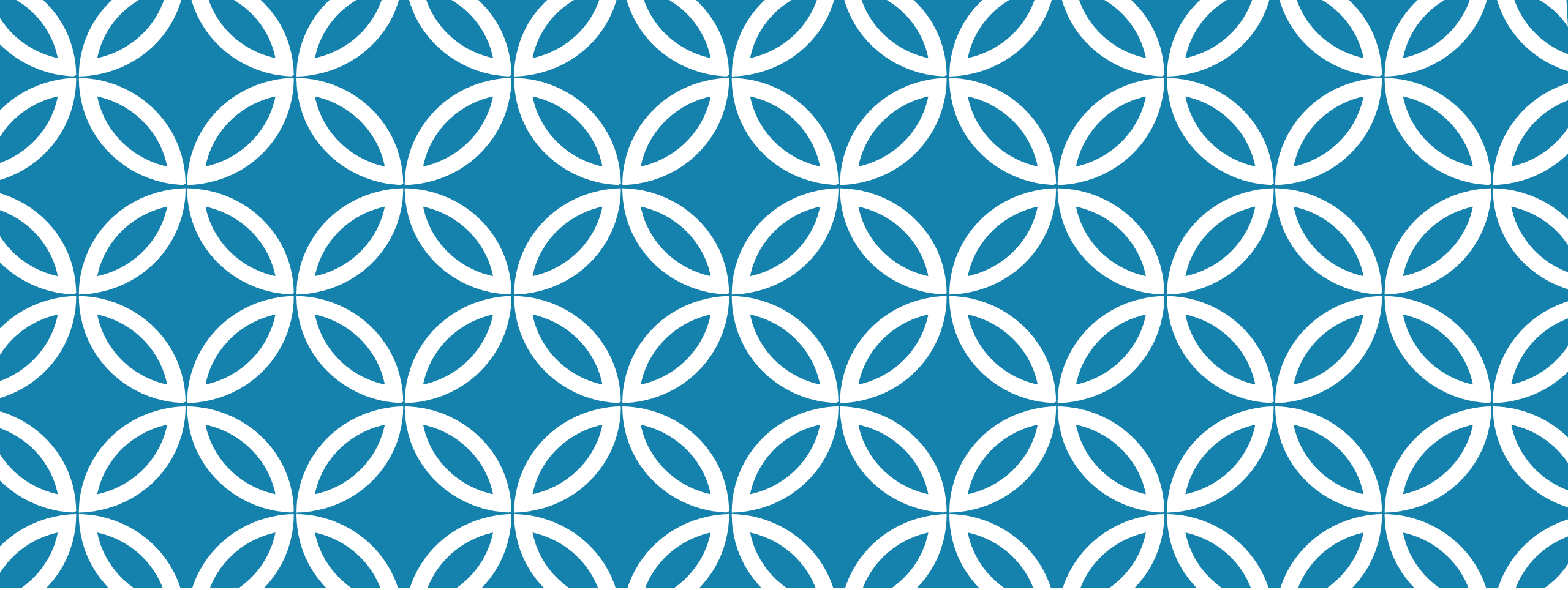
SERVICE SURVEY	Excellent	Good	Average	Fair	Poor
Overall Experience	❑	❑	❑	❑	❑
Greeting by Hostess	❑	❑	❑	❑	❑
Manager (Table Visit)	❑	❑	❑	❑	❑
Overall Service	❑	❑	❑	❑	❑
Professionalism	❑	❑	❑	❑	❑
Menu Knowledge	❑	❑	❑	❑	❑
Friendliness	❑	❑	❑	❑	❑
Wine Selection	❑	❑	❑	❑	❑
Menu Selection	❑	❑	❑	❑	❑
Food Quality	❑	❑	❑	❑	❑
Food Presentation	❑	❑	❑	❑	❑
Value for $ Spent	❑	❑	❑	❑	❑

What comments could you give us to improve our restaurant?

_____

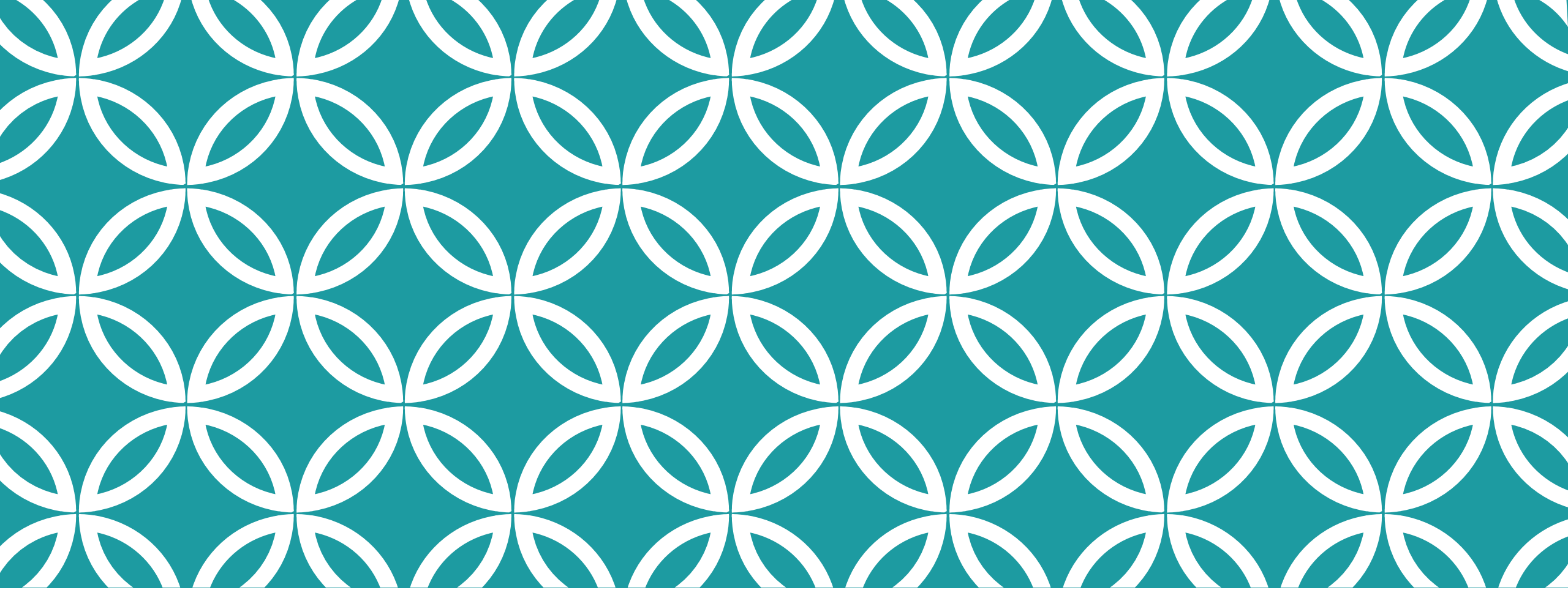
_____

Thank you, we appreciate your comments. —The staff of Chops City Grill.



CREATING DISTRIBUTIONS FROM DATA

Lesson 2.2



MODIFYING DATA IN EXCEL

Lesson 2.2.1

TABLE 2.2 - TOP 20 SELLING AUTOMOBILES IN UNITED STATES IN MARCH 2011

Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)
1	Honda	Accord	33616	29120
2	Nissan	Altima	32289	24649
3	Toyota	Camry	31464	36251
4	Honda	Civic	31213	22463
5	Toyota	Corolla/Matrix	30234	29623
6	Ford	Fusion	27566	22773
7	Hyundai	Sonata	22894	18935
8	Hyundai	Elantra	19255	8225
9	Toyota	Prius	18605	11786
10	Chevrolet	Cruze/Cobalt	18101	10316
11	Chevrolet	Impala	18063	15594
12	Nissan	Sentra	17851	8721
13	Ford	Focus	17178	19500
14	Volkswagon	Jetta	16969	9196
15	Chevrolet	Malibu	15551	17750
16	Mazda	3	12467	11353
17	Nissan	Versa	11075	13811
18	Subaru	Outback	10498	7619
19	Kia	Soul	10028	5106
20	Ford	Fiesta	9787	0

MODIFYING DATA IN EXCEL

- Sorting and filtering data in excel

Illustration - To sort the automobiles by March 2010 sales

Step 1: Select cells A1:F21

Step 2: Click the **DATA** tab in the Ribbon

Step 3: Click **Sort** in the **Sort & Filter** group

Step 4: Select the check box for **My data has headers**

Step 5: In the first **Sort by** dropdown menu, select **Sales (March 2010)**

Step 6: In the **Order** dropdown menu, select **Largest to Smallest**

Step 7: Click **OK**

FIGURE 2.4 - USING EXCEL'S SORT FUNCTION TO SORT THE TOP SELLING AUTOMOBILES DATA

	A	B	C	D	E	F	G
1	Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)	Percent Change in Sales from 2010	
2	1	Honda	Accord	33616	29120	15.4%	
3	2	Nissan	Altima	32289	24649	31.0%	
4	3	Toyota	Camry	31464	36251	–13.2%	
5	4	Honda	Civic	31213	22463	39.0%	
6	5	Toyota	Corolla/Matrix	30234	29623	2.1%	
7	6	Ford	Fusion	27566	22773	21.0%	
8	7	Hyundai					
9	8	Hyundai					
10	9	Toyota					
11	10	Chevrolet					
12	11	Chevrolet					
13	12	Nissan					
14	13	Ford					
15	14	Volkswagon					
16	15	Chevrolet					
17	16	Mazda					
18	17	Nissan	Versa	11075	13811	–19.8%	
19	18	Subaru	Outback	10498	7619	37.8%	
20	19	Kia	Soul	10028	5106	96.4%	
21	20	Ford	Fiesta	9787	0	-----	

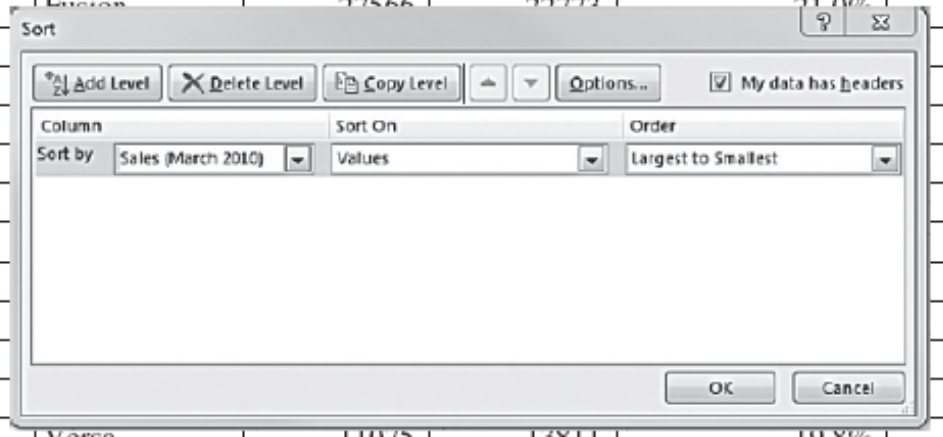


FIGURE 2.5 - TOP SELLING AUTOMOBILES DATA SORTED BY SALES IN MARCH 2010 SALES

	A	B	C	D	E	F
1	Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)	Percent Change in Sales from 2010
2	3	Toyota	Camry	31464	36251	–13.2%
3	5	Toyota	Corolla/Matrix	30234	29623	2.1%
4	1	Honda	Accord	33616	29120	15.4%
5	2	Nissan	Altima	32289	24649	31.0%
6	6	Ford	Fusion	27566	22773	21.0%
7	4	Honda	Civic	31213	22463	39.0%
8	13	Ford	Focus	17178	19500	–11.9%
9	7	Hyundai	Sonata	22894	18935	20.9%
10	15	Chevrolet	Malibu	15551	17750	–12.4%
11	11	Chevrolet	Impala	18063	15594	15.8%
12	17	Nissan	Versa	11075	13811	–19.8%
13	9	Toyota	Prius	18605	11786	57.9%
14	16	Mazda	3	12467	11353	9.8%
15	10	Chevrolet	Cruze/Cobalt	18101	10316	75.5%
16	14	Volkswagon	Jetta	16969	9196	84.5%
17	12	Nissan	Sentra	17851	8721	104.7%
18	8	Hyundai	Elantra	19255	8225	134.1%
19	18	Subaru	Outback	10498	7619	37.8%
20	19	Kia	Soul	10028	5106	96.4%
21	20	Ford	Fiesta	9787	0	-----

MODIFYING DATA IN EXCEL

- Sorting and filtering data in excel

Illustration - Using Excel's Filter function to see the sales of models made by Toyota.

Step 1: Select cells A1:F21

Step 2: Click the **DATA** tab in the Ribbon

Step 3: Click **Filter** in the **Sort & Filter** group

Step 4: Click on the **Filter Arrow** in column B, next to **Manufacturer**

Step 5: Select only the check box for **Toyota**. You can easily deselect all choices by unchecking (**Select All**)

FIGURE 2.6 - TOP SELLING AUTOMOBILES DATA FILTERED TO SHOW ONLY AUTOMOBILES MANUFACTURED BY TOYOTA

	A	B	C	D	E	F
1	Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)	Percent Change in Sales from 2010
2	3	Toyota	Camry	31464	36251	–13.2%
3	5	Toyota	Corolla/Matrix	30234	29623	2.1%
13	9	Toyota	Prius	18605	11786	57.9%

MODIFYING DATA IN EXCEL

- Conditional Formatting of Data in Excel: Makes it easy to identify data that satisfy certain conditions in a data set.

Illustration - To identify the automobile models in Table 2.2 for which sales had decreased from March 2010 to March 2011.

Step 1: Starting with the original data shown in Figure 2.3, select cells F1:F21

Step 2: Click on the **HOME** tab in the Ribbon

MODIFYING DATA IN EXCEL

Illustration (contd.)

Step 3: Click **Conditional Formatting** in the **Styles** group

Step 4: Select **Highlight Cells Rules**, and click **Less Than** from the dropdown menu

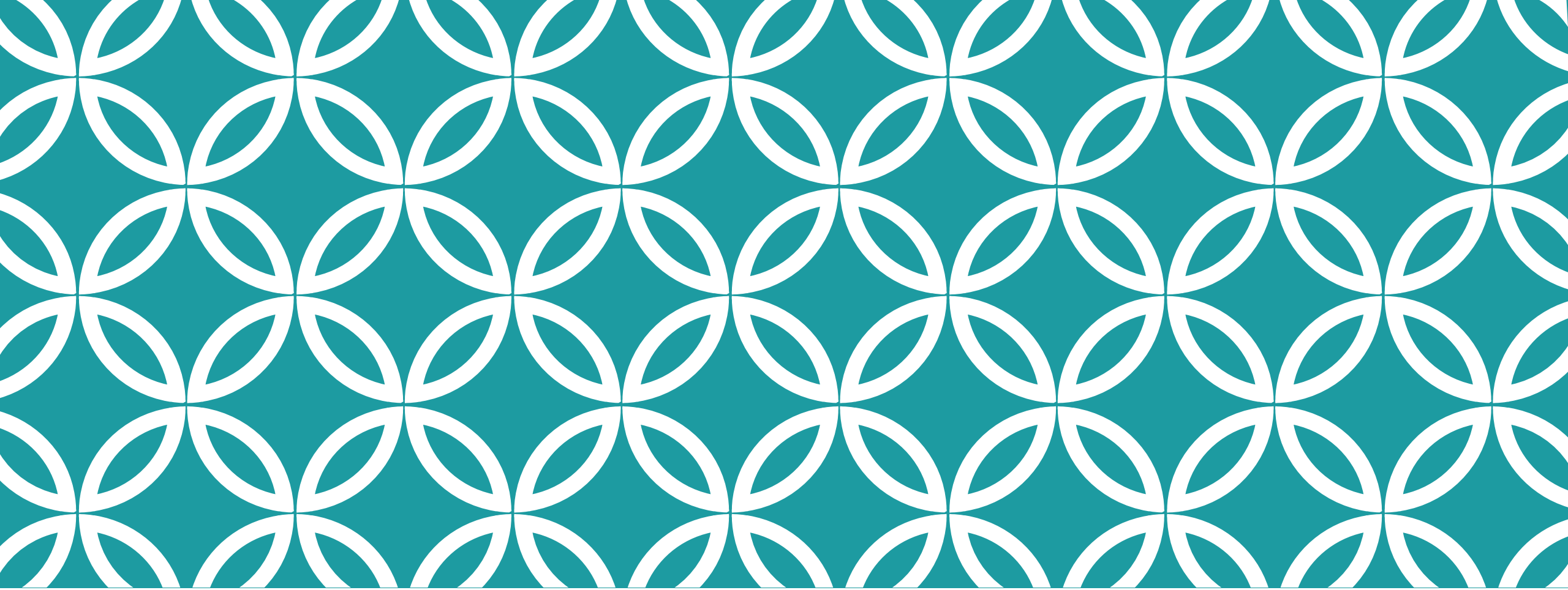
Step 5: Enter *0%* in the **Format cells that are LESS THAN:** box Step 6: Click **OK**

FIGURE 2.7 - USING CONDITIONAL FORMATTING IN EXCEL TO HIGHLIGHT AUTOMOBILES WITH DECLINING SALES FROM MARCH 2010

	A	B	C	D	E	F
1	Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)	Percent Change in Sales from 2010
2	1	Honda	Accord	33616	29120	15.4%
3	2	Nissan	Altima	32289	24649	31.0%
4	3	Toyota	Camry	31464	36251	–13.2%
5	4	Honda	Civic	31213	22463	39.0%
6	5	Toyota	Corolla/Matrix	30234	29623	2.1%
7	6	Ford	Fusion	27566	22773	21.0%
8	7	Hyundai	Sonata	22894	18935	20.9%
9	8	Hyundai	Elantra	19255	8225	134.1%
10	9	Toyota	Prius	18605	11786	57.9%
11	10	Chevrolet	Cruze/Cobalt	18101	10316	75.5%
12	11	Chevrolet	Impala	18063	15594	15.8%
13	12	Nissan	Sentra	17851	8721	104.7%
14	13	Ford	Focus	17178	19500	–11.9%
15	14	Volkswagon	Jetta	16969	9196	84.5%
16	15	Chevrolet	Malibu	15551	17750	–12.4%
17	16	Mazda	3	12467	11353	9.8%
18	17	Nissan	Versa	11075	13811	–19.8%
19	18	Subaru	Outback	10498	7619	37.8%
20	19	Kia	Soul	10028	5106	96.4%
21	20	Ford	Fiesta	9787	0	-----

FIGURE 2.8 - USING CONDITIONAL FORMATTING IN EXCEL TO GENERATE DATA BARS FOR THE TOP SELLING AUTOMOBILES DATA

	A	B	C	D	E	F
1	Rank (by March 2011 Sales)	Manufacturer	Model	Sales (March 2011)	Sales (March 2010)	Percent Change in Sales from 2010
2	1	Honda	Accord	33616	29120	15.4%
3	2	Nissan	Altima	32289	24649	31.0%
4	3	Toyota	Camry	31464	36251	–13.2%
5	4	Honda	Civic	31213	22463	39.0%
6	5	Toyota	Corolla/Matrix	30234	29623	2.1%
7	6	Ford	Fusion	27566	22773	21.0%
8	7	Hyundai	Sonata	22894	18935	20.9%
9	8	Hyundai	Elantra	19255	8225	134.1%
10	9	Toyota	Prius	18605	11786	57.9%
11	10	Chevrolet	Cruze/Cobalt	18101	10316	75.5%
12	11	Chevrolet	Impala	18063	15594	15.8%
13	12	Nissan	Sentra	17851	8721	104.7%
14	13	Ford	Focus	17178	19500	–11.9%
15	14	Volkswagon	Jetta	16969	9196	84.5%
16	15	Chevrolet	Malibu	15551	17750	–12.4%
17	16	Mazda	3	12467	11353	9.8%
18	17	Nissan	Versa	11075	13811	–19.8%
19	18	Subaru	Outback	10498	7619	37.8%
20	19	Kia	Soul	10028	5106	96.4%
21	20	Ford	Fiesta	9787	0	-----



CREATING DISTRIBUTIONS FROM DATA

Lesson 2.2.2

FREQUENCY DISTRIBUTION FOR CATEGORICAL DATA

- Frequency distributions for categorical data
  - **Frequency distribution:** A summary of data that shows the number (frequency) of observations in each of several nonoverlapping classes, typically referred to as **bins,** when dealing with distributions.

TABLE 2.3 - DATA FROM A SAMPLE OF 50 SOFT DRINK PURCHASES

Coca-Cola	Sprite	Pepsi
Diet Coke	Coca-Cola	Coca-Cola
Pepsi	Diet Coke	Coca-Cola
Diet Coke	Coca-Cola	Coca-Cola
Coca-Cola	Diet Coke	Pepsi
Coca-Cola	Coca-Cola	Dr. Pepper
Dr. Pepper	Sprite	Coca-Cola
Diet Coke	Pepsi	Diet Coke
Pepsi	Coca-Cola	Pepsi
Pepsi	Coca-Cola	Pepsi
Coca-Cola	Coca-Cola	Pepsi
Dr. Pepper	Pepsi	Pepsi
Sprite	Coca-Cola	Coca-Cola
Coca-Cola	Sprite	Dr. Pepper
Diet Coke	Dr. Pepper	Pepsi
Coca-Cola	Pepsi	Sprite
Coca-Cola	Diet Coke	

TABLE 2.4 - FREQUENCY DISTRIBUTION OF SOFT DRINK PURCHASES

Soft Drink	Frequency
Coca-Cola	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	5
Total	50

- The frequency distribution summarizes information about the popularity of the five soft drinks:
  - Coca-Cola is the leader, Pepsi is second, Diet Coke is third, and Sprite and Dr. Pepper are tied for fourth.

RELATIVE FREQUENCY AND PERCENT FREQUENCY DISTRIBUTIONS

- Relative frequency and percent frequency distributions
  - **Relative frequency distribution:** It is a tabular summary of data showing the relative frequency for each bin.
  - **Percent frequency distribution:** Summarizes the percent frequency of the data for each bin.
    - Used to provide estimates of the relative likelihoods of different values of a random variable.

TABLE 2.5 - RELATIVE FREQUENCY AND PERCENT FREQUENCY DISTRIBUTIONS OF SOFT DRINK PURCHASES

Soft Drink	Relative Frequency	Percent Frequency (%)
Coca-Cola	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	0.10	10
Total	1.00	100

FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA

- Frequency distributions for quantitative data
  - Three steps necessary to define the classes for a frequency distribution with quantitative data:

    1. Determine the number of nonoverlapping bins.
    2. Determine the width of each bin.
    3. Determine the bin limits.

CREATING DISTRIBUTIONS FROM DATA

Table 2.6 - Year-End Audit Times (Days)

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

Table 2.7 - Frequency, Relative Frequency, and Percent Frequency Distributions for the Audit Time Data

Audit Times (days)	Frequency	Relative Frequency	Percent Frequency
10–14	4	0.20	20
15–19	8	0.40	40
20–24	5	0.25	25
25–29	2	0.10	10
30–34	1	0.05	5

FIGURE 2.10 - USING EXCEL TO GENERATE A FREQUENCY DISTRIBUTION FOR AUDIT TIMES DATA

	A	B	C	D
1	Year-End Audit Times (in Days)			
2	12	14	19	18
3	15	15	18	17
4	20	27	22	23
5	22	21	33	28
6	14	18	16	13
7				
8				
9	Bin	Frequency		
10	14	=FREQUENCY(A2:D6,A10:A14)		
11	19	=FREQUENCY(A2:D6,A10:A14)		
12	24	=FREQUENCY(A2:D6,A10:A14)		
13	29	=FREQUENCY(A2:D6,A10:A14)		
14	34	=FREQUENCY(A2:D6,A10:A14)		

	A	B	C	D
1	Year-End Audit Times (in Days)			
2	12	14	19	18
3	15	15	18	17
4	20	27	22	23
5	22	21	33	28
6	14	18	16	13
7				
8				
9	Bin	Frequency		
10	14	4		
11	19	8		
12	24	5		
13	29	2		
14	34	1		

HISTOGRAM

- **Histogram**: A common graphical presentation of quantitative data
  - Constructed by placing the variable of interest on the horizontal axis and the selected frequency measure (absolute frequency, relative frequency, or percent frequency) on the vertical axis.
  - The frequency measure of each class is shown by drawing a rectangle whose base is determined by the class limits on the horizontal axis and whose height is the corresponding frequency measure.

FIGURE 2.11 - HISTOGRAM FOR THE AUDIT TIME DATA

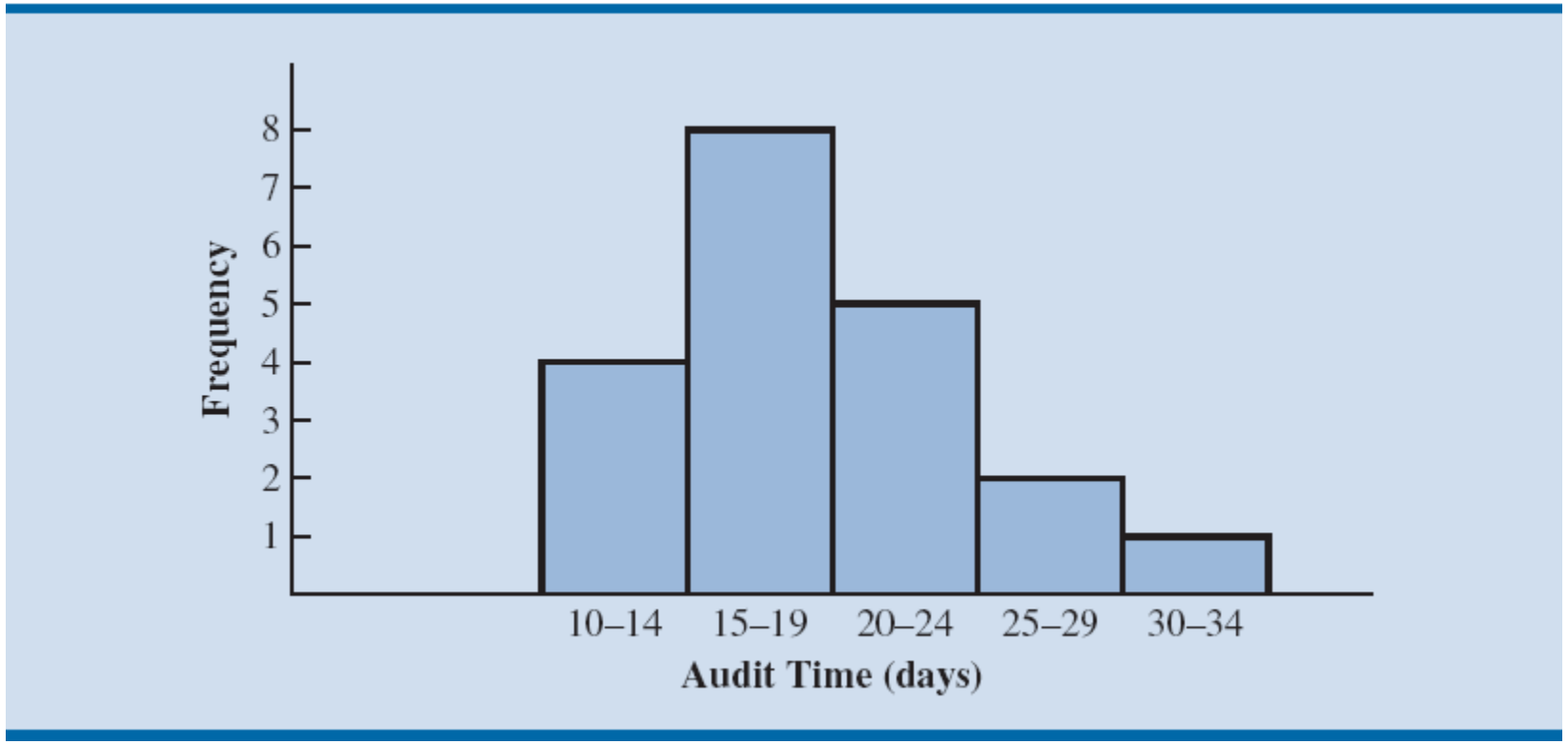


FIGURE 2.12 - CREATING A HISTOGRAM FOR THE AUDIT TIME DATA USING DATA ANALYSIS TOOLPAK IN EXCEL

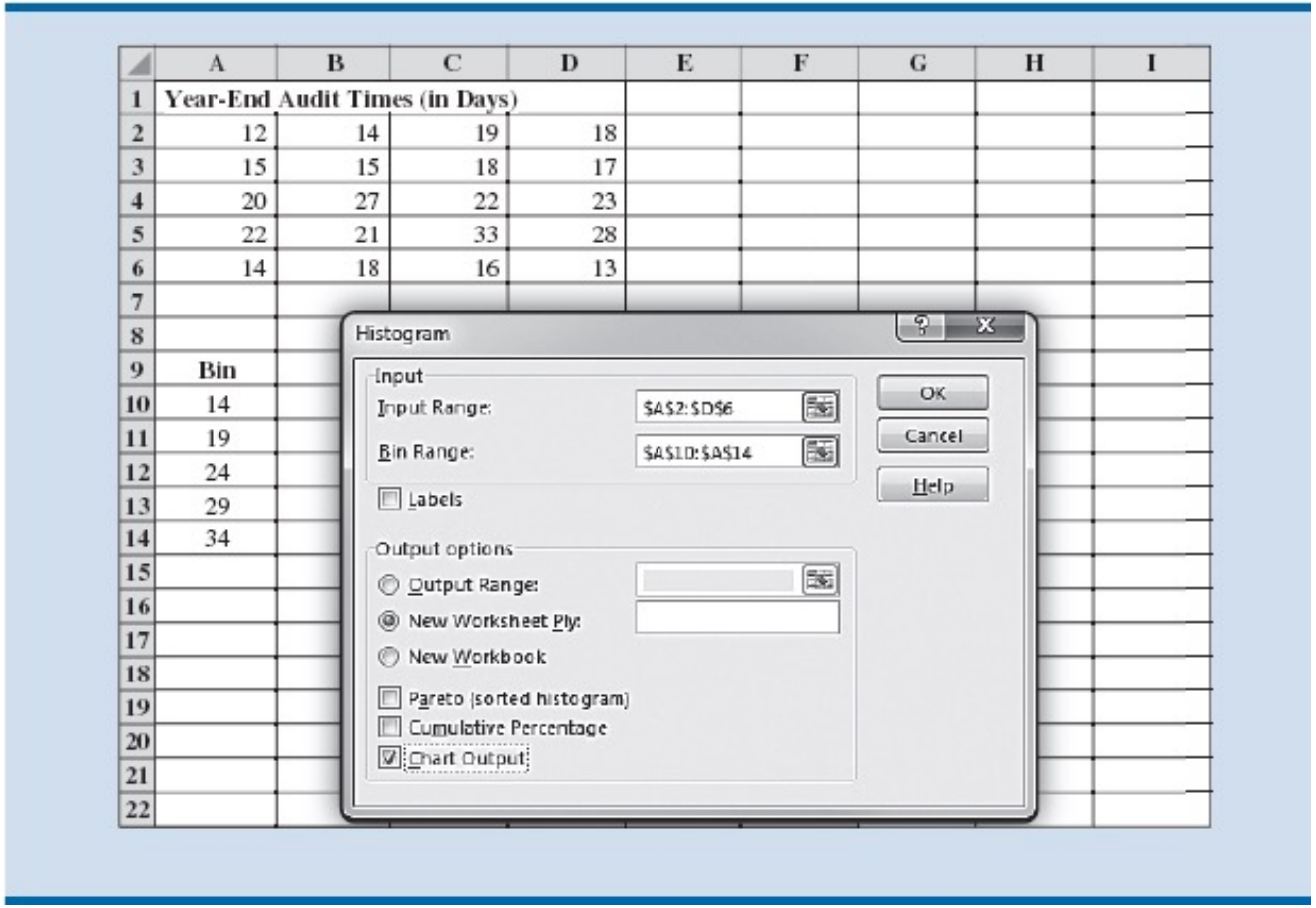
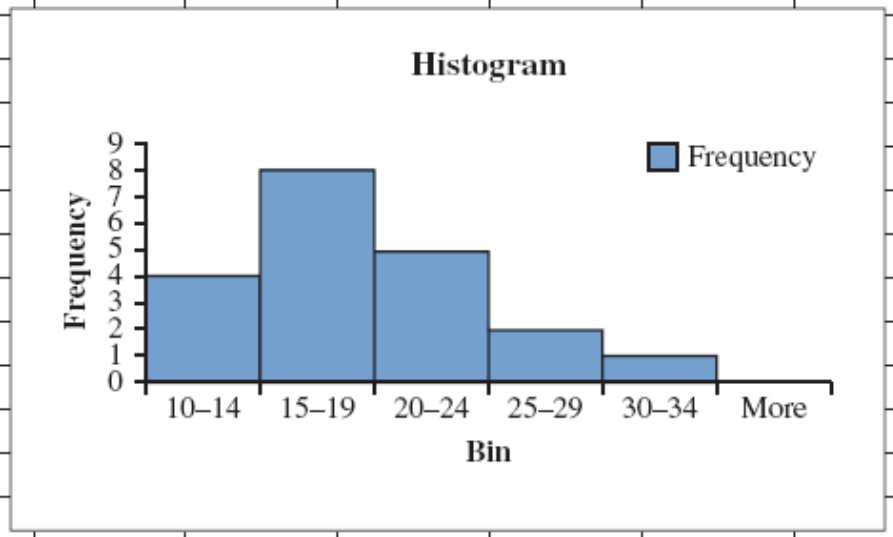


FIGURE 2.13 - COMPLETED HISTOGRAM FOR THE AUDIT TIME DATA USING DATA ANALYSIS TOOLPAK IN EXCEL

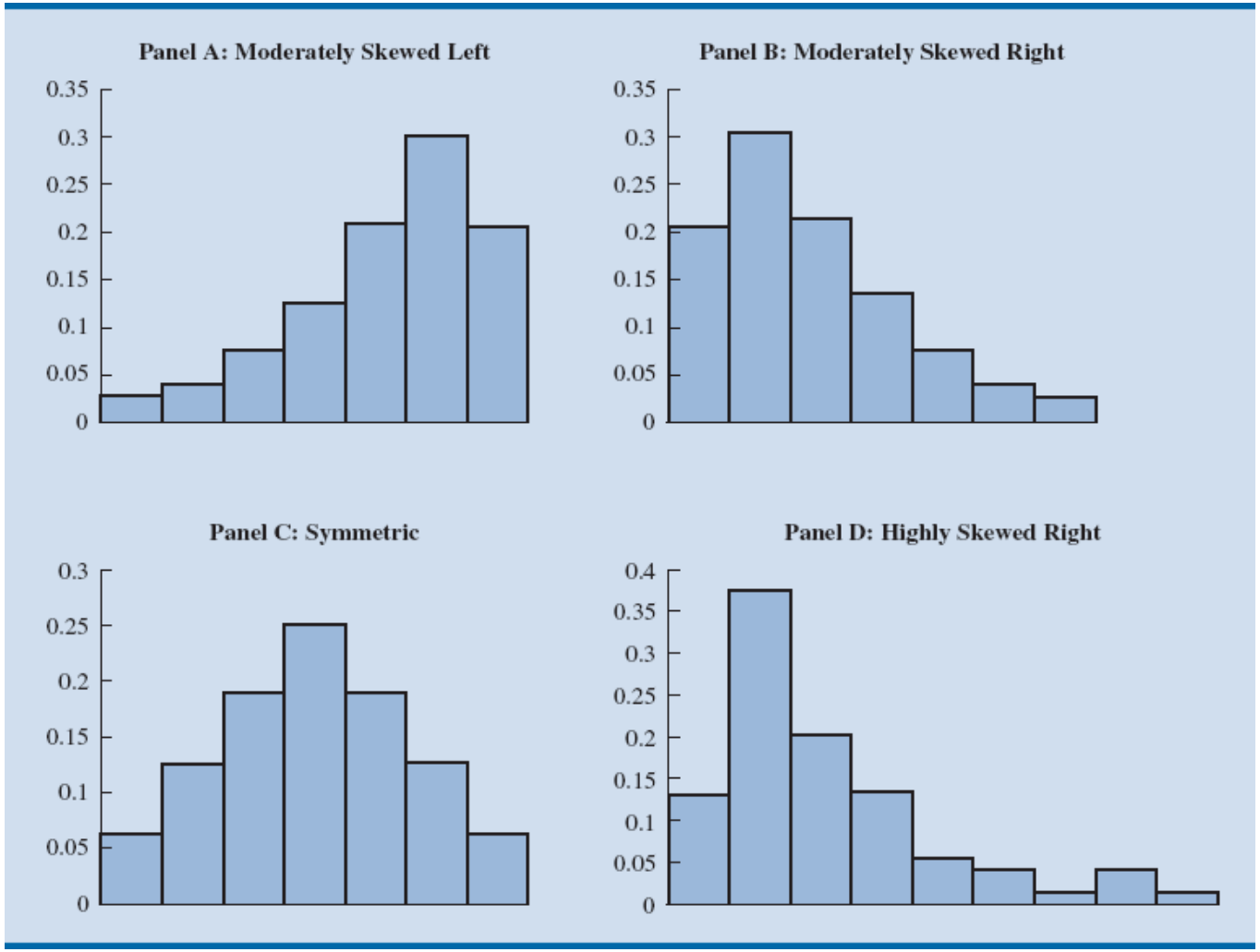
	A	B	C	D	E	F	G	H	I	J
1	*Bin*	*Frequency*								
2	10–14	4								
3	15–19	8								
4	20–24	5								
5	25–29	2								
6	30–34	1								
7	More	0								
8										
9										
10										
11										
12										
13										
14										



HISTOGRAM

- Histogram provides information about the shape, or form, of a distribution.
- **Skewness:** Lack of symmetry
  - Important characteristic of the shape of a distribution

FIGURE 2.14 - HISTOGRAMS SHOWING DISTRIBUTIONS WITH DIFFERENT LEVELS OF SKEWNESS



CUMULATIVE FREQUENCY DISTRIBUTIONS

- Cumulative Distributions
- **Cumulative frequency distribution**: A variation of the frequency distribution that provides another tabular summary of quantitative data.
  - Uses the number of classes, class widths, and class limits developed for the frequency distribution.
  - Shows the number of data items with values less than or equal to the upper class limit of each class.

TABLE 2.8 - CUMULATIVE FREQUENCY, CUMULATIVE RELATIVE FREQUENCY, AND CUMULATIVE PERCENT FREQUENCY DISTRIBUTIONS FOR THE AUDIT TIME DATA

Audit Time (days)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
Less than or equal to 14	4	0.20	20
Less than or equal to 19	12	0.60	60
Less than or equal to 24	17	0.85	85
Less than or equal to 29	19	0.95	95
Less than or equal to 34	20	1.00	100