Sports Analytics and Data Science

Winning the Game with Methods and Models

THOMAS W. MILLER
Faculty Director of Northwestern University’s Predictive Analytics Program
Sports Analytics
and Data Science

Winning the Game with Methods and Models

Thomas W. Miller
# Contents

Preface v  
Figures ix  
Tables xi  
Exhibits xiii  

1 Understanding Sports Markets 1  
2 Assessing Players 23  
3 Ranking Teams 37  
4 Predicting Scores 49  
5 Making Game-Day Decisions 61  
6 Crafting a Message 69  
7 Promoting Brands and Products 101  
8 Growing Revenues 119  
9 Managing Finances 133
10 Playing What-if Games 147
11 Working with Sports Data 169
12 Competing on Analytics 193

A Data Science Methods 197
A.1 Mathematical Programming 200
A.2 Classical and Bayesian Statistics 203
A.3 Regression and Classification 206
A.4 Data Mining and Machine Learning 215
A.5 Text and Sentiment Analysis 217
A.6 Time Series, Sales Forecasting, and Market Response Models 226
A.7 Social Network Analysis 230
A.8 Data Visualization 234
A.9 Data Science: The Eclectic Discipline 240

B Professional Leagues and Teams 255

Data Science Glossary 261

Baseball Glossary 279

Bibliography 299

Index 329
“Sometimes you win, sometimes you lose, sometimes it rains.”

—TIM ROBBINS AS EBBY CALVIN LALOOSH IN Bull Durham (1988)

Businesses attract customers, politicians persuade voters, websites cajole visitors, and sports teams draw fans. Whatever the goal or target, data and models rule the day.

This book is about building winning teams and successful sports businesses. Winning and success are more likely when decisions are guided by data and models. Sports analytics is a source of competitive advantage.

This book provides an accessible guide to sports analytics. It is written for anyone who needs to know about sports analytics, including players, managers, owners, and fans. It is also a resource for analysts, data scientists, and programmers. The book views sports analytics in the context of data science, a discipline that blends business savvy, information technology, and modeling techniques.

To use analytics effectively in sports, we must first understand sports—the industry, the business, and what happens on the fields and courts of play. We need to know how to work with data—identifying data sources, gathering data, organizing and preparing them for analysis. We also need to know how to build models from data. Data do not speak for themselves. Useful predictions do not arise out of thin air. It is our job to learn from data and build models that work.
The best way to learn about sports analytics and data science is through examples. We provide a ready resource and reference guide for modeling techniques. We show programmers how to solve real world problems by building on a foundation of trustworthy methods and code.

The truth about what we do is in the programs we write. The code is there for everyone to see and for some to debug. Data sets and computer programs are available from the website for the Modeling Techniques series at http://www.ftpress.com/miller/. There is also a GitHub site at https://github.com/mtpa/.

When working on sports problems, some things are more easily accomplished with R, others with Python. And there are times when it is good to offer solutions in both languages, checking one against the other.

One of the things that distinguishes this book from others in the area of sports analytics is the range of data sources and topics discussed. Many researchers focus on numerical performance data for teams and players. We take a broader view of sports analytics—the view of data science. There are text data as well as numeric data. And with the growth of the World Wide Web, the sources of data are plentiful. Much can be learned from public domain sources through crawling and scraping the web and utilizing application programming interfaces (APIs).

I learn from my consulting work with professional sports organizations. Research Publishers LLC with its ToutBay division promotes what can be called “data science as a service.” Academic research and models can take us only so far. Eventually, to make a difference, we need to implement our ideas and models, sharing them with one another.

Many have influenced my intellectual development over the years. There were those good thinkers and good people, teachers and mentors for whom I will be forever grateful. Sadly, no longer with us are Gerald Hahn Hinkle in philosophy and Allan Lake Rice in languages at Ursinus College, and Herbert Feigl in philosophy at the University of Minnesota. I am also most thankful to David J. Weiss in psychometrics at the University of Minnesota and Kelly Eakin in economics, formerly at the University of Oregon.
My academic home is the Northwestern University School of Professional Studies. Courses in sports research methods and quantitative analysis, marketing analytics, database systems and data preparation, web and network data science, web information retrieval and real-time analytics, and data visualization provide inspiration for this book. Thanks to the many students and fellow faculty from whom I have learned. And thanks to colleagues and staff who administer excellent graduate programs, including the Master of Science in Predictive Analytics, Master of Arts in Sports Administration, Master of Science in Information Systems, and the Advanced Certificate in Data Science.

Lorena Martin reviewed this book and provided valuable feedback while she authored a companion volume on sports performance measurement and analytics (Martin 2016). Adam Grossman and Tom Robinson provided valuable feedback about coverage of topics in sports business management. Roy Sanford provided advice on statistics. Amy Hendrickson of TExnology Inc. applied her craft, making words, tables, and figures look beautiful in print—another victory for open source. Candice Bradley served dual roles as a reviewer and copyeditor for all books in the Modeling Techniques series. And Andy Beaster helped in preparing this book for final production. I am grateful for their guidance and encouragement.

Thanks go to my editor, Jeanne Glasser Levine, and publisher, Pearson/FT Press, for making this book possible. Any writing issues, errors, or items of unfinished business, of course, are my responsibility alone.

My good friend Brittney and her daughter Janiya keep me company when time permits. And my son Daniel is there for me in good times and bad, a friend for life. My greatest debt is to them because they believe in me.

Thomas W. Miller
Glendale, California
October 2015
This page intentionally left blank
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.2</td>
<td>Mets’ Away and Yankees’ Home Data (Offense and Defense)</td>
<td>154</td>
</tr>
<tr>
<td>10.3</td>
<td>Balanced Game-day Simulation (Offense and Defense)</td>
<td>155</td>
</tr>
<tr>
<td>10.4</td>
<td>Actual and Theoretical Runs-scored Distributions</td>
<td>157</td>
</tr>
<tr>
<td>10.5</td>
<td>Poisson Model for Mets vs. Yankees at Yankee Stadium</td>
<td>159</td>
</tr>
<tr>
<td>10.6</td>
<td>Negative Binomial Model for Mets vs. Yankees at Yankee Stadium</td>
<td>160</td>
</tr>
<tr>
<td>10.7</td>
<td>Probability of Home Team Winning (Negative Binomial Model)</td>
<td>162</td>
</tr>
<tr>
<td>10.8</td>
<td>Strategic Modeling Techniques in Sports</td>
<td>164</td>
</tr>
<tr>
<td>11.1</td>
<td>Software Stack for a Document Search and Selection System</td>
<td>173</td>
</tr>
<tr>
<td>11.2</td>
<td>The Information Supply Chain of Professional Team Sports</td>
<td>174</td>
</tr>
<tr>
<td>11.3</td>
<td>Automated Data Acquisition by Crawling, Scraping, and Parsing</td>
<td>177</td>
</tr>
<tr>
<td>11.4</td>
<td>Automated Data Acquisition with an API</td>
<td>179</td>
</tr>
<tr>
<td>11.5</td>
<td>Gathering and Organizing Data for Analysis</td>
<td>180</td>
</tr>
<tr>
<td>A.1</td>
<td>Mathematical Programming Modeling Methods</td>
<td>201</td>
</tr>
<tr>
<td>A.2</td>
<td>Evaluating the Predictive Accuracy of a Binary Classifier</td>
<td>212</td>
</tr>
<tr>
<td>A.3</td>
<td>Linguistic Foundations of Text Analytics</td>
<td>218</td>
</tr>
<tr>
<td>A.4</td>
<td>Creating a Terms-by-Documents Matrix</td>
<td>221</td>
</tr>
<tr>
<td>A.5</td>
<td>Data and Plots for the Anscombe Quartet</td>
<td>235</td>
</tr>
<tr>
<td>A.6</td>
<td>Visualizing Many Games Across a Season: Differential Runs Plot</td>
<td>236</td>
</tr>
<tr>
<td>A.7</td>
<td>Moving Fraction Plot for Basketball</td>
<td>237</td>
</tr>
<tr>
<td>A.8</td>
<td>Visualizing Basketball Play-by-Play Data</td>
<td>239</td>
</tr>
<tr>
<td>A.9</td>
<td>Data Science: The Eclectic Discipline</td>
<td>241</td>
</tr>
</tbody>
</table>
Tables

1.1 Sports and Recreation Activities in the United States ........................................... 3
1.2 MLB Team Valuation and Finances (March 2015) .................................................. 5
1.3 NBA Team Valuation and Finances (January 2015) ................................................. 6
1.4 NFL Team Valuation and Finances (August 2014) .................................................. 7
1.5 World Soccer Team Valuation and Finances (May 2015) ....................................... 8
2.1 Levels of Measurement .......................................................................................... 29
3.1 NBA Team Records (2014–2015 Season) ................................................................ 39
5.1 Twenty-five States of a Baseball Half-Inning ........................................................... 63
6.1 Dissimilarity Matrix for Entertainment Events and Activities ................................. 71
6.2 Consumer Preference Data for Dodger Stadium Seating ........................................ 76
7.1 Bobbleheads and Dodger Dogs ............................................................................. 103
7.2 Regression of Attendance on Month, Day of Week, and Promotion ......................... 110
9.1 Discounted Cash Flow Analysis of a Player Contract ............................................. 139
9.2 Would you like to buy the Brooklyn Nets? ............................................................. 141
10.1 New York Mets’ Early Season Games in 2007 ....................................................... 149
10.2 New York Yankees’ Early Season Games in 2007 .................................................. 150
A.1 Three Generalized Linear Models .......................................................................... 209
A.2 Social Network Data: MLB Player Transactions ................................................... 233
B.1 Women’s National Basketball Association (WNBA) .............................................. 255
B.2 Major League Baseball (MLB) .............................................................................. 256
B.3 Major League Soccer (MLS) ................................................................................ 257
B.4 National Basketball Association (NBA) ................................................................. 258
B.5 National Football League (NFL) ........................................................................... 259
B.6 National Hockey League (NHL) ............................................................................. 260
This page intentionally left blank
# Exhibits

1.1 MLB, NBA, and NFL Player Salaries (R) .................................................. 16  
1.2 Payroll and Performance in Major League Baseball (R) .................. 18  
1.3 Making a Perceptual Map of Sports (R) .............................................. 19  
3.1 Assessing Team Strength by Unidimensional Scaling (R) ............ 43  
6.1 Mapping Entertainment Events and Activities (R) ......................... 83  
6.2 Mapping Entertainment Events and Activities (Python) ................. 86  
6.3 Preferences for Sporting Events—Conjoint Analysis (R) .............. 88  
6.4 Preferences for Sporting Events—Conjoint Analysis (Python) ...... 99  
7.1 Shaking Our Bobbleheads Yes and No (R) ...................................... 113  
7.2 Shaking Our Bobbleheads Yes and No (Python) ............................. 116  
10.1 Team Winning Probabilities by Simulation (R) ............................ 167  
10.2 Team Winning Probabilities by Simulation (Python) .................... 168  
11.1 Simple One-Site Web Crawler and Scraper (Python) .................... 186  
11.2 Gathering Opinion Data from Twitter: Football Injuries (Python) . 189  
A.1 Programming the Anscombe Quartet (Python) ............................... 242  
A.2 Programming the Anscombe Quartet (R) ......................................... 244  
A.3 Making Differential Runs Plots for Baseball (R) ............................ 245  
A.4 Moving Fraction Plot: A Basketball Example (R) ......................... 246  
A.5 Visualizing Basketball Games (R) .................................................. 248  
A.6 Seeing Data Science as an Eclectic Discipline (R) ......................... 252
“Those of you on the floor at the end of the game, I’m proud of you. You played your guts out. I’m only going to say this one time. All of you have the weekend. Think about whether or not you want to be on this team under the following condition: What I say when it comes to this basketball team is the law, absolutely and without discussion.”

—Gene Hackman as Coach Norman Dale in Hoosiers (1986)

In applying the laws of economics to professional sports, we must consider the nature of sports and the motives of owners. Professional sports are different from other forms of business.

There are sellers and buyers of sports entertainment. The sellers are the players and teams within the leagues of professional sports. The buyers are consumers of sports, many of whom never go to games in person but who watch sports on television, listen to the radio, and buy sports team paraphernalia.

Sports compete with other forms of entertainment for people’s time and money. And various sports compete with one another, especially when their seasons overlap. Sports teams produce entertainment content that is distributed through the media. Sports teams license their brand names and logos to other organizations, including sports apparel manufacturers.
Sports teams are not independent businesses competing with one another. While players and teams compete on the fields and courts of play, they cooperate with one another as members of leagues. The core product of sports is the sporting contest, a joint product of two or more players or two or more teams.

Fifty-four sports and recreation activities, shown in table 1.1, are tracked by the National Sporting Goods Association (2015), which serves the sporting goods industry. In recent years, participation in baseball, basketball, football, and tennis has declined, while participation in soccer has increased. There has been growth in individual recreational sports, such as skateboarding and snowboarding. Of course, levels of participation in sports are not necessarily an indicator of levels of interest in sports as entertainment.

Sports businesses produce entertainment products by cooperating with one another. While it is illegal for businesses in most industries to collude in setting output and prices, sports leagues engage in cooperative output and pricing as a standard part of their business model. The number of games, indeed the entire schedule of games in a sport, is determined by the league. In fact, aspects of professional sports are granted monopoly power by the federal government in the United States.

When developing a model for a typical business or firm, an economist would assume profit maximization as a motive. But for a professional sports team, an owner’s motives may not be so easily understood. While one owner may operate his or her team for profit year by year, another may seek to maximize wins or overall utility. Another may look for capital appreciation—buying, then selling after a few years. Lacking knowledge of owners’ motives, it is difficult to predict what they will do.

Gaining market share and becoming the dominant player is a goal of firms in many industries. Not so in the business of professional sports. If one team were assured of victory in almost all of its contests, interest in those contests could wane. A team benefits by winning more often than losing, but winning all the time may be less beneficial than winning most of the time. Professional sports leagues claim to be seeking competitive balance, although there are dominant teams in many leagues.
### Table 1.1. Sports and Recreation Activities in the United States

<table>
<thead>
<tr>
<th>Activity</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic Exercising</td>
<td>Ice/Figure Skating</td>
</tr>
<tr>
<td>Archery (Target)</td>
<td>In-Line Roller Skating</td>
</tr>
<tr>
<td>Backpack/Wilderness Camping</td>
<td>Kayaking</td>
</tr>
<tr>
<td>Baseball</td>
<td>Lacrosse</td>
</tr>
<tr>
<td>Basketball</td>
<td>Martial Arts/MMA/Tae Kwon Do</td>
</tr>
<tr>
<td>Bicycle Riding</td>
<td>Mountain Biking (Off Road)</td>
</tr>
<tr>
<td>Billiards/Pool</td>
<td>Muzzleloading</td>
</tr>
<tr>
<td>Boating (Motor/Power)</td>
<td>Paintball Games</td>
</tr>
<tr>
<td>Bowling</td>
<td>Running/Jogging</td>
</tr>
<tr>
<td>Boxing</td>
<td>Scuba Diving (Open Water)</td>
</tr>
<tr>
<td>Camping (Vacation/Overnight)</td>
<td>Skateboarding</td>
</tr>
<tr>
<td>Canoeing</td>
<td>Skiing (Alpine)</td>
</tr>
<tr>
<td>Cheerleading</td>
<td>Skiing (Cross Country)</td>
</tr>
<tr>
<td>Dart Throwing</td>
<td>Snowboarding</td>
</tr>
<tr>
<td>Exercise Walking</td>
<td>Soccer</td>
</tr>
<tr>
<td>Exercising with Equipment</td>
<td>Softball</td>
</tr>
<tr>
<td>Fishing (Fresh Water)</td>
<td>Swimming</td>
</tr>
<tr>
<td>Fishing (Salt Water)</td>
<td>Table Tennis/Ping Pong</td>
</tr>
<tr>
<td>Football (Flag)</td>
<td>Target Shooting (Airgun)</td>
</tr>
<tr>
<td>Football (Tackle)</td>
<td>Target Shooting (Live Ammunition)</td>
</tr>
<tr>
<td>Football (Touch)</td>
<td>Tennis</td>
</tr>
<tr>
<td>Golf</td>
<td>Volleyball</td>
</tr>
<tr>
<td>Gymnastics</td>
<td>Water Skiing</td>
</tr>
<tr>
<td>Hiking</td>
<td>Weight Lifting</td>
</tr>
<tr>
<td>Hockey (Ice)</td>
<td>Work Out at Club/Gym/Fitness Studio</td>
</tr>
<tr>
<td>Hunting with Bow &amp; Arrow</td>
<td>Wrestling</td>
</tr>
<tr>
<td>Hunting with Firearms</td>
<td>Yoga</td>
</tr>
</tbody>
</table>
Sports is big business as shown by valuations and finances of the major professional sports in the United States and worldwide. Data from *Forbes* for Major League Baseball (MLB), the National Basketball Association (NBA), the National Football League (NFL), and worldwide soccer teams are shown in tables 1.2 through 1.5.

Professional sports teams most certainly compete with one another in the labor market, and labor in the form of star players is in short supply. Some argue that salary caps are necessary to preserve competitive balance. Salary caps also help teams in limiting expenditures on players.

Most professional sports in the United States have salary caps. The 2015 salary cap for NFL teams, with fifty-three player rosters, is set at $143.28 million (Patra 2015). Most teams have payrolls at or near the cap, making the average salary of an NFL player about $2.7 million. One player on an NFL team may be designated as a *franchise player*, restricting that player from entering free agency. The league sets minimum salaries for franchise players. For example, a franchise quarterback has a minimum salary of $18.544 million in 2015. The highest annual salary among NFL players is $22 million for Aaron Rodgers, Green Bay Packers quarterback (spotrac 2015c). The minimum annual salary is $420 thousand.

NBA teams have a $70 million salary cap for the 2015–16 season, with penalties for teams going over the cap. Maximum player salaries are based on a percentage of cap and years of service. For example, LeBron James, with ten years of experience, would have a maximum salary of $23 million (Mahoney 2015). New Orleans Pelicans Anthony Davis’ average salary of $29 million is the highest among NBA players (spotrac 2015b). Team rosters include fifteen players under contract, with as many as thirteen available to play in any particular game. The minimum annual salary is $428,498.

Major League Baseball (MLB) has a “luxury tax” for teams with payrolls in excess of $189 million. There is a regular-player roster of twenty-five or twenty-six players for double-header days/night. A forty-man roster includes players under contract and eligible to play. Between September 1 and the end of the regular season the roster is expanded to forty players. The roster drops back to twenty-five players for the playoffs. The minimum MLB annual salary is $505,700 in 2015. The highest MLB annual salary is $31 million for Miguel Cabrera of the Detroit Tigers (spotrac 2015a).
### Table 1.2. MLB Team Valuation and Finances (March 2015)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Current Value ($ Millions)</th>
<th>One-Year Change in Value (Percentage)</th>
<th>Debt/Value (Percentage)</th>
<th>Revenue ($ Millions)</th>
<th>Operating Income ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York Yankees</td>
<td>3,200</td>
<td>28</td>
<td>0</td>
<td>508</td>
<td>8.1</td>
</tr>
<tr>
<td>2</td>
<td>Los Angeles Dodgers</td>
<td>2,400</td>
<td>20</td>
<td>17</td>
<td>403</td>
<td>-12.2</td>
</tr>
<tr>
<td>3</td>
<td>Boston Red Sox</td>
<td>2,100</td>
<td>40</td>
<td>0</td>
<td>370</td>
<td>49.2</td>
</tr>
<tr>
<td>4</td>
<td>San Francisco Giants</td>
<td>2,000</td>
<td>100</td>
<td>4</td>
<td>387</td>
<td>68.4</td>
</tr>
<tr>
<td>5</td>
<td>Chicago Cubs</td>
<td>1,800</td>
<td>50</td>
<td>24</td>
<td>302</td>
<td>73.3</td>
</tr>
<tr>
<td>6</td>
<td>St Louis Cardinals</td>
<td>1,400</td>
<td>71</td>
<td>21</td>
<td>294</td>
<td>73.6</td>
</tr>
<tr>
<td>7</td>
<td>New York Mets</td>
<td>1,350</td>
<td>69</td>
<td>26</td>
<td>263</td>
<td>25.0</td>
</tr>
<tr>
<td>8</td>
<td>Los Angeles Angels</td>
<td>1,300</td>
<td>68</td>
<td>0</td>
<td>304</td>
<td>16.7</td>
</tr>
<tr>
<td>9</td>
<td>Chicago Cubs</td>
<td>1,280</td>
<td>83</td>
<td>27</td>
<td>287</td>
<td>41.4</td>
</tr>
<tr>
<td>10</td>
<td>Philadelphia Phillies</td>
<td>1,250</td>
<td>28</td>
<td>8</td>
<td>265</td>
<td>-39.0</td>
</tr>
<tr>
<td>11</td>
<td>Texas Rangers</td>
<td>1,220</td>
<td>48</td>
<td>13</td>
<td>266</td>
<td>3.5</td>
</tr>
<tr>
<td>12</td>
<td>Atlanta Braves</td>
<td>1,150</td>
<td>58</td>
<td>0</td>
<td>267</td>
<td>33.2</td>
</tr>
<tr>
<td>13</td>
<td>Detroit Tigers</td>
<td>1,125</td>
<td>65</td>
<td>15</td>
<td>254</td>
<td>-20.7</td>
</tr>
<tr>
<td>14</td>
<td>Seattle Mariners</td>
<td>1,100</td>
<td>55</td>
<td>0</td>
<td>250</td>
<td>26.4</td>
</tr>
<tr>
<td>15</td>
<td>Baltimore Orioles</td>
<td>1,000</td>
<td>61</td>
<td>15</td>
<td>245</td>
<td>31.4</td>
</tr>
<tr>
<td>16</td>
<td>Chicago White Sox</td>
<td>975</td>
<td>58</td>
<td>5</td>
<td>227</td>
<td>31.9</td>
</tr>
<tr>
<td>17</td>
<td>Pittsburgh Pirates</td>
<td>900</td>
<td>57</td>
<td>10</td>
<td>229</td>
<td>43.6</td>
</tr>
<tr>
<td>18</td>
<td>Minnesota Twins</td>
<td>895</td>
<td>48</td>
<td>25</td>
<td>223</td>
<td>21.3</td>
</tr>
<tr>
<td>19</td>
<td>San Diego Padres</td>
<td>890</td>
<td>45</td>
<td>22</td>
<td>224</td>
<td>35.0</td>
</tr>
<tr>
<td>20</td>
<td>Cincinnati Reds</td>
<td>885</td>
<td>48</td>
<td>6</td>
<td>227</td>
<td>2.2</td>
</tr>
<tr>
<td>21</td>
<td>Milwaukee Brewers</td>
<td>875</td>
<td>55</td>
<td>6</td>
<td>226</td>
<td>11.3</td>
</tr>
<tr>
<td>22</td>
<td>Toronto Blue Jays</td>
<td>870</td>
<td>43</td>
<td>0</td>
<td>227</td>
<td>-17.9</td>
</tr>
<tr>
<td>23</td>
<td>Colorado Rockies</td>
<td>855</td>
<td>49</td>
<td>7</td>
<td>214</td>
<td>12.6</td>
</tr>
<tr>
<td>24</td>
<td>Arizona Diamondbacks</td>
<td>840</td>
<td>44</td>
<td>17</td>
<td>211</td>
<td>-2.2</td>
</tr>
<tr>
<td>25</td>
<td>Cleveland Indians</td>
<td>825</td>
<td>45</td>
<td>9</td>
<td>207</td>
<td>8.9</td>
</tr>
<tr>
<td>26</td>
<td>Houston Astros</td>
<td>800</td>
<td>51</td>
<td>34</td>
<td>175</td>
<td>21.6</td>
</tr>
<tr>
<td>27</td>
<td>Oakland Athletics</td>
<td>725</td>
<td>46</td>
<td>8</td>
<td>202</td>
<td>20.8</td>
</tr>
<tr>
<td>28</td>
<td>Kansas City Royals</td>
<td>700</td>
<td>43</td>
<td>8</td>
<td>231</td>
<td>26.6</td>
</tr>
<tr>
<td>29</td>
<td>Miami Marlins</td>
<td>650</td>
<td>30</td>
<td>34</td>
<td>188</td>
<td>15.4</td>
</tr>
<tr>
<td>30</td>
<td>Tampa Bay Rays</td>
<td>625</td>
<td>29</td>
<td>22</td>
<td>188</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Table 1.3. NBA Team Valuation and Finances (January 2015)

<table>
<thead>
<tr>
<th>Team Rank</th>
<th>Team</th>
<th>Current Value ($ Millions)</th>
<th>One-Year Change in Value (Percentage)</th>
<th>Debt/Value (Percentage)</th>
<th>Operating Income ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Los Angeles Lakers</td>
<td>2,600</td>
<td>93</td>
<td>2</td>
<td>293</td>
</tr>
<tr>
<td>2</td>
<td>New York Knicks</td>
<td>2,500</td>
<td>79</td>
<td>0</td>
<td>278</td>
</tr>
<tr>
<td>3</td>
<td>Chicago Bulls</td>
<td>2,000</td>
<td>100</td>
<td>3</td>
<td>201</td>
</tr>
<tr>
<td>4</td>
<td>Boston Celtics</td>
<td>1,700</td>
<td>94</td>
<td>9</td>
<td>173</td>
</tr>
<tr>
<td>5</td>
<td>Los Angeles Clippers</td>
<td>1,600</td>
<td>178</td>
<td>0</td>
<td>146</td>
</tr>
<tr>
<td>6</td>
<td>Brooklyn Nets</td>
<td>1,500</td>
<td>92</td>
<td>19</td>
<td>212</td>
</tr>
<tr>
<td>7</td>
<td>Golden State Warriors</td>
<td>1,300</td>
<td>73</td>
<td>12</td>
<td>168</td>
</tr>
<tr>
<td>8</td>
<td>Houston Rockets</td>
<td>1,250</td>
<td>61</td>
<td>8</td>
<td>175</td>
</tr>
<tr>
<td>9</td>
<td>Miami Heat</td>
<td>1,175</td>
<td>53</td>
<td>8</td>
<td>188</td>
</tr>
<tr>
<td>10</td>
<td>Dallas Mavericks</td>
<td>1,150</td>
<td>50</td>
<td>17</td>
<td>168</td>
</tr>
<tr>
<td>11</td>
<td>San Antonio Spurs</td>
<td>1,000</td>
<td>52</td>
<td>8</td>
<td>172</td>
</tr>
<tr>
<td>12</td>
<td>Portland Trail Blazers</td>
<td>940</td>
<td>60</td>
<td>11</td>
<td>153</td>
</tr>
<tr>
<td>13</td>
<td>Oklahoma City Thunder</td>
<td>930</td>
<td>58</td>
<td>15</td>
<td>152</td>
</tr>
<tr>
<td>14</td>
<td>Toronto Raptors</td>
<td>920</td>
<td>77</td>
<td>16</td>
<td>151</td>
</tr>
<tr>
<td>15</td>
<td>Cleveland Cavaliers</td>
<td>915</td>
<td>78</td>
<td>22</td>
<td>149</td>
</tr>
<tr>
<td>16</td>
<td>Phoenix Suns</td>
<td>910</td>
<td>61</td>
<td>20</td>
<td>145</td>
</tr>
<tr>
<td>17</td>
<td>Washington Wizards</td>
<td>900</td>
<td>86</td>
<td>14</td>
<td>143</td>
</tr>
<tr>
<td>18</td>
<td>Orlando Magic</td>
<td>875</td>
<td>56</td>
<td>17</td>
<td>143</td>
</tr>
<tr>
<td>19</td>
<td>Denver Nuggets</td>
<td>855</td>
<td>73</td>
<td>1</td>
<td>136</td>
</tr>
<tr>
<td>20</td>
<td>Utah Jazz</td>
<td>850</td>
<td>62</td>
<td>6</td>
<td>142</td>
</tr>
<tr>
<td>21</td>
<td>Indiana Pacers</td>
<td>830</td>
<td>75</td>
<td>18</td>
<td>139</td>
</tr>
<tr>
<td>22</td>
<td>Atlanta Hawks</td>
<td>825</td>
<td>94</td>
<td>21</td>
<td>133</td>
</tr>
<tr>
<td>23</td>
<td>Detroit Pistons</td>
<td>810</td>
<td>80</td>
<td>23</td>
<td>144</td>
</tr>
<tr>
<td>24</td>
<td>Sacramento Kings</td>
<td>800</td>
<td>45</td>
<td>29</td>
<td>125</td>
</tr>
<tr>
<td>25</td>
<td>Memphis Grizzlies</td>
<td>750</td>
<td>66</td>
<td>23</td>
<td>135</td>
</tr>
<tr>
<td>26</td>
<td>Charlotte Hornets</td>
<td>725</td>
<td>77</td>
<td>21</td>
<td>130</td>
</tr>
<tr>
<td>27</td>
<td>Philadelphia 76ers</td>
<td>700</td>
<td>49</td>
<td>21</td>
<td>125</td>
</tr>
<tr>
<td>28</td>
<td>New Orleans Pelicans</td>
<td>650</td>
<td>55</td>
<td>19</td>
<td>131</td>
</tr>
<tr>
<td>29</td>
<td>Minnesota Timberwolves</td>
<td>625</td>
<td>45</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>30</td>
<td>Milwaukee Bucks</td>
<td>600</td>
<td>48</td>
<td>29</td>
<td>110</td>
</tr>
</tbody>
</table>

## Table 1.4. NFL Team Valuation and Finances (August 2014)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Team</th>
<th>Current Value ($ Millions)</th>
<th>One-Year Change in Value (Percentage)</th>
<th>Debt/Value (Percentage)</th>
<th>Revenue ($ Millions)</th>
<th>Operating Income ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dallas Cowboys</td>
<td>3,200</td>
<td>39</td>
<td>6</td>
<td>560</td>
<td>245.7</td>
</tr>
<tr>
<td>2</td>
<td>New England Patriots</td>
<td>2,600</td>
<td>44</td>
<td>9</td>
<td>428</td>
<td>147.2</td>
</tr>
<tr>
<td>3</td>
<td>Washington Redskins</td>
<td>2,400</td>
<td>41</td>
<td>10</td>
<td>395</td>
<td>143.4</td>
</tr>
<tr>
<td>4</td>
<td>New York Giants</td>
<td>2,100</td>
<td>35</td>
<td>25</td>
<td>353</td>
<td>87.3</td>
</tr>
<tr>
<td>5</td>
<td>Houston Texans</td>
<td>1,850</td>
<td>28</td>
<td>11</td>
<td>339</td>
<td>102.8</td>
</tr>
<tr>
<td>6</td>
<td>New York Jets</td>
<td>1,800</td>
<td>30</td>
<td>33</td>
<td>333</td>
<td>79.5</td>
</tr>
<tr>
<td>7</td>
<td>Philadelphia Eagles</td>
<td>1,750</td>
<td>33</td>
<td>11</td>
<td>330</td>
<td>73.2</td>
</tr>
<tr>
<td>8</td>
<td>Chicago Bears</td>
<td>1,700</td>
<td>36</td>
<td>6</td>
<td>309</td>
<td>57.1</td>
</tr>
<tr>
<td>9</td>
<td>San Francisco 49ers</td>
<td>1,600</td>
<td>31</td>
<td>53</td>
<td>270</td>
<td>24.8</td>
</tr>
<tr>
<td>10</td>
<td>Baltimore Ravens</td>
<td>1,500</td>
<td>22</td>
<td>18</td>
<td>304</td>
<td>56.7</td>
</tr>
<tr>
<td>11</td>
<td>Denver Broncos</td>
<td>1,450</td>
<td>25</td>
<td>8</td>
<td>301</td>
<td>30.7</td>
</tr>
<tr>
<td>12</td>
<td>Indianapolis Colts</td>
<td>1,400</td>
<td>17</td>
<td>4</td>
<td>285</td>
<td>60.7</td>
</tr>
<tr>
<td>13</td>
<td>Green Bay Packers</td>
<td>1,375</td>
<td>16</td>
<td>1</td>
<td>299</td>
<td>25.6</td>
</tr>
<tr>
<td>14</td>
<td>Pittsburgh Steelers</td>
<td>1,350</td>
<td>21</td>
<td>15</td>
<td>287</td>
<td>52.4</td>
</tr>
<tr>
<td>15</td>
<td>Seattle Seahawks</td>
<td>1,330</td>
<td>23</td>
<td>9</td>
<td>288</td>
<td>27.3</td>
</tr>
<tr>
<td>16</td>
<td>Miami Dolphins</td>
<td>1,300</td>
<td>21</td>
<td>29</td>
<td>281</td>
<td>8.0</td>
</tr>
<tr>
<td>17</td>
<td>Carolina Panthers</td>
<td>1,250</td>
<td>18</td>
<td>5</td>
<td>283</td>
<td>55.6</td>
</tr>
<tr>
<td>18</td>
<td>Tampa Bay Buccaneers</td>
<td>1,225</td>
<td>15</td>
<td>15</td>
<td>275</td>
<td>46.4</td>
</tr>
<tr>
<td>19</td>
<td>Tennessee Titans</td>
<td>1,160</td>
<td>10</td>
<td>11</td>
<td>278</td>
<td>35.6</td>
</tr>
<tr>
<td>20</td>
<td>Minnesota Vikings</td>
<td>1,150</td>
<td>14</td>
<td>43</td>
<td>250</td>
<td>5.3</td>
</tr>
<tr>
<td>21</td>
<td>Atlanta Falcons</td>
<td>1,125</td>
<td>21</td>
<td>27</td>
<td>264</td>
<td>13.1</td>
</tr>
<tr>
<td>22</td>
<td>Cleveland Browns</td>
<td>1,120</td>
<td>11</td>
<td>18</td>
<td>276</td>
<td>35.0</td>
</tr>
<tr>
<td>23</td>
<td>New Orleans Saints</td>
<td>1,110</td>
<td>11</td>
<td>7</td>
<td>278</td>
<td>50.1</td>
</tr>
<tr>
<td>24</td>
<td>Kansas City Chiefs</td>
<td>1,100</td>
<td>9</td>
<td>6</td>
<td>260</td>
<td>10.0</td>
</tr>
<tr>
<td>25</td>
<td>Arizona Cardinals</td>
<td>1,000</td>
<td>4</td>
<td>15</td>
<td>266</td>
<td>42.8</td>
</tr>
<tr>
<td>26</td>
<td>San Diego Chargers</td>
<td>995</td>
<td>5</td>
<td>10</td>
<td>262</td>
<td>39.9</td>
</tr>
<tr>
<td>27</td>
<td>Cincinnati Bengals</td>
<td>990</td>
<td>7</td>
<td>10</td>
<td>258</td>
<td>11.9</td>
</tr>
<tr>
<td>28</td>
<td>Oakland Raiders</td>
<td>970</td>
<td>18</td>
<td>21</td>
<td>244</td>
<td>42.8</td>
</tr>
<tr>
<td>29</td>
<td>Jacksonville Jaguars</td>
<td>965</td>
<td>15</td>
<td>21</td>
<td>263</td>
<td>56.9</td>
</tr>
<tr>
<td>30</td>
<td>Detroit Lions</td>
<td>960</td>
<td>7</td>
<td>29</td>
<td>254</td>
<td>-15.9</td>
</tr>
<tr>
<td>31</td>
<td>Buffalo Bills</td>
<td>935</td>
<td>7</td>
<td>13</td>
<td>252</td>
<td>38.0</td>
</tr>
<tr>
<td>32</td>
<td>St Louis Rams</td>
<td>930</td>
<td>6</td>
<td>12</td>
<td>250</td>
<td>16.2</td>
</tr>
</tbody>
</table>

Table 1.5. World Soccer Team Valuation and Finances (May 2015)

<table>
<thead>
<tr>
<th>Team Rank</th>
<th>Team</th>
<th>Current Value ($ Millions)</th>
<th>One-Year Change in Value (Percentage)</th>
<th>Debt/Value (Percentage)</th>
<th>Revenue ($ Millions)</th>
<th>Operating Income ($ Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real Madrid</td>
<td>3,263</td>
<td>-5</td>
<td>4</td>
<td>746</td>
<td>170</td>
</tr>
<tr>
<td>2</td>
<td>Barcelona</td>
<td>3,163</td>
<td>-1</td>
<td>3</td>
<td>657</td>
<td>174</td>
</tr>
<tr>
<td>3</td>
<td>Manchester United</td>
<td>3,104</td>
<td>10</td>
<td>20</td>
<td>703</td>
<td>211</td>
</tr>
<tr>
<td>4</td>
<td>Bayern Munich</td>
<td>2,347</td>
<td>27</td>
<td>0</td>
<td>661</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>Manchester City</td>
<td>1,375</td>
<td>59</td>
<td>0</td>
<td>562</td>
<td>122</td>
</tr>
<tr>
<td>6</td>
<td>Chelsea</td>
<td>1,370</td>
<td>58</td>
<td>0</td>
<td>526</td>
<td>83</td>
</tr>
<tr>
<td>7</td>
<td>Arsenal</td>
<td>1,307</td>
<td>-2</td>
<td>30</td>
<td>487</td>
<td>101</td>
</tr>
<tr>
<td>8</td>
<td>Liverpool</td>
<td>982</td>
<td>42</td>
<td>10</td>
<td>415</td>
<td>86</td>
</tr>
<tr>
<td>9</td>
<td>Juventus</td>
<td>837</td>
<td>-2</td>
<td>9</td>
<td>379</td>
<td>50</td>
</tr>
<tr>
<td>10</td>
<td>AC Milan</td>
<td>775</td>
<td>-10</td>
<td>44</td>
<td>339</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>Borussia Dortmund</td>
<td>700</td>
<td>17</td>
<td>6</td>
<td>355</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>Paris Saint-Germain</td>
<td>634</td>
<td>53</td>
<td>0</td>
<td>643</td>
<td>-1</td>
</tr>
<tr>
<td>13</td>
<td>Tottenham Hotspur</td>
<td>600</td>
<td>17</td>
<td>9</td>
<td>293</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>Schalke 04</td>
<td>572</td>
<td>-1</td>
<td>0</td>
<td>290</td>
<td>57</td>
</tr>
<tr>
<td>15</td>
<td>Inter Milan</td>
<td>439</td>
<td>-9</td>
<td>56</td>
<td>222</td>
<td>-41</td>
</tr>
<tr>
<td>16</td>
<td>Atletico de Madrid</td>
<td>436</td>
<td>33</td>
<td>53</td>
<td>231</td>
<td>47</td>
</tr>
<tr>
<td>17</td>
<td>Napoli</td>
<td>353</td>
<td>19</td>
<td>0</td>
<td>224</td>
<td>43</td>
</tr>
<tr>
<td>18</td>
<td>Newcastle United</td>
<td>349</td>
<td>33</td>
<td>0</td>
<td>210</td>
<td>44</td>
</tr>
<tr>
<td>19</td>
<td>West Ham United</td>
<td>309</td>
<td>33</td>
<td>12</td>
<td>186</td>
<td>54</td>
</tr>
<tr>
<td>20</td>
<td>Galatasaray</td>
<td>294</td>
<td>-15</td>
<td>17</td>
<td>220</td>
<td>-37</td>
</tr>
</tbody>
</table>

Chapter 1. Understanding Sports Markets

Figure 1.1, a histogram lattice, shows how player salaries compare across the MLB, NBA, and NFL in August 2015. Player salary distributions are positively skewed. The mean salary across NFL players is around $1.7 million, but the median is $630 thousand. The mean salary across NBA players is around $5.1 million, with median salary $2.8 million. The mean salary across MLB players is around $4.1 million, with the median $1.1 million.

Do team expenditures on players buy success? This is a meaningful question to ask for leagues that have no salary caps. Szymanski (2015) reports studies showing that between 60 and 90 percent of the variability in U.K. soccer team positions may be explained by wages paid to players. Major League Baseball has a luxury tax in place of a salary cap, and team payrolls vary widely in size. The New York Yankees have been known for having the highest payrolls in baseball. Recently, the Los Angeles Dodgers have surpassed the Yankees with the highest player payroll—more than $257 million at the end of the 2014 season (Woody 2014).

Figure 1.2 shows baseball team salaries at the beginning of the 2014 season plotted against the percentage of games won across the regular season. Notice how teams that made the playoffs in 2014, labeled with team abbreviations, have a wide range of payrolls. While the biggest spenders in baseball are often among the set of teams going to the playoffs, the relationship between team payrolls and team performance is weak at best—less than 7 percent of the variability in win/loss percentages is explained by player payrolls.

The thesis of Michael Lewis’ Moneyball (2003) and what has become the ethos of sports analytics is that small-market baseball teams can win by spending their money wisely. Star players demand top salaries due as much to their celebrity status as to their skills. Players with high on-base percentages, overlooked by major-market teams, can be hired at much lower salaries than star players.

Teams, although associated with particular cities, can be known nationwide or worldwide. The media of television and the Internet provide opportunities for reaching consumers across the globe. A Super Bowl at the Rose Bowl in Pasadena, California or AT&T Stadium in Arlington, Texas may be attended by around 100 thousand fans (Alder 2015), while U.S. television audiences have grown to over 100 million (statista 2015).
Figure 1.1. MLB, NBA, and NFL Average Annual Salaries

Sources. spotrac (2015a, 2015b, 2015c).
Figure 1.2. MLB Team Payrolls and Win/Loss Performance (2014 Season)

See Appendix B, page 255, for team abbreviations and names.
Media revenues are important to successful sports teams. Other revenues come from business partnerships, sponsorships, advertising, and stadium naming rights. City governments understand well the power of sports to promote business. Locating sports arenas in cities can help to revitalize downtown areas, as demonstrated by the experience of the Oklahoma City Thunder. Indianapolis, Indiana promotes itself as a sports capital with the Colts and Pacers (Rein, Shields, and Grossman 2015).

Teams seek to build their brands, developing a positive reputation in the minds of consumers. Players, like fans, are attracted to teams with a reputation for hard work, courage, fair play, honesty, teamwork, and community service. The character of a team is often as important as its likelihood of winning. The Cubs are associated with Chicago, but Cub fans may be found from Maine to California. This is despite the fact that the Cubs have not won the World Series since 1908. Teams in U.S. professional sports vie to become “America’s team,” with fans across the land wearing their logo-embossed hats and jerseys.

The demand for sports and the feelings of sports consumers are not so easily understood. Fans can be fickle and fandom fleeting. Fans can be loyal to a sport, to a team, or to individual players. Multivariate methods can help us understand how sports consumers think by revealing relationships among products or brands.

Figure 1.3 provides an example, a perceptual map of seven sports. Along the horizontal dimension, we move from individual, non-contact sports on the left-hand side, to team sports with little contact, to team sports with contact on the right-hand side. The vertical dimension, less easily described, may be thought of as relating to the aerobic versus anaerobic nature of sports and to other characteristics such as physicality and skill. Sports such as tennis, soccer, and basketball entail aerobic exercise. These are endurance sports, while football is an example of a sport that involves both aerobic and anaerobic exercise, including intense exercise for short durations. Sports close together on the map have similarities. Baseball and golf, for example, involve special skills, such as precision in hitting a ball. Soccer and hockey involve almost continuous movement and getting a ball through the goal. Football and hockey have high physicality or player contact.
In many respects, professional sports teams are decidedly different from other businesses. They are in the public eye. They live and die in the media. And a substantial portion of their revenues come from media.

An overview of sports marketing is provided by Mullin, Hardy, and Sutton (2014). Rein, Kotler, and Shields (2006) and Carter (2011) discuss the convergence of entertainment and sports. Miller (2015a) reviews methods in marketing data science, including product positioning maps, market segmentation, target marketing, customer relationship management, and competitive analysis.

Sports also represents a laboratory for labor market research. Sports is one of the few industries in which job performance and compensation are public knowledge. Economic studies examine player performance measures and value of individual players to teams (Kahn 2000; Bradbury 2007). Miller (1991), Abrams (2010), and Lowenfish (2010) review baseball labor relations. And Early (2011) provides insight into labor and racial discrimination in professional sports.

Sports wagering markets have been studied extensively by economists because they provide public information about price, volume, and rates of return. Furthermore, sports betting opportunities have fixed beginning and ending times and published odds or point spreads, making them easier to study than many financial investment opportunities. As a result, sports wagering markets have become a virtual field laboratory for the study of market efficiency. Sauer (1998) provides a comprehensive review of the economics of wagering markets.

When management objectives can be defined clearly in mathematical terms, teams use mathematical programming methods—constrained optimization. Teams attempt to maximize revenue or minimize costs subject to known situational factors. There has been extensive work on league schedules, for which the league objective may be to have teams playing one another an equal number of times while minimizing total distance traveled between cities. Alternatively, league officials may seek home/away schedules, revenue sharing formulas, or draft lottery rules that maximize competitive balance. Briskorn (2008) reviews methods for scheduling sports competition, drawing on integer programming, combinatorics, and graph theory. Wright (2009) provides an overview of operations research in sport.
Extensive data about sports are in the public domain, readily available in newspapers and online sources. These data offer opportunities for predictive modeling and research. Throughout the book we also identify places to apply methods of operations research, including mathematical programming and simulation.

Exhibit 1.1 shows an R program for exploring distributions of player salaries across the MLB, NBA, and NFL. The program draws on software for statistical graphics from Sarkar (2008).

Exhibit 1.2 (page 18) shows an R program for examining the relationship between MLB payrolls and win-loss performance. The program draws on software for statistical graphics from Wickham and Chang (2014).

Exhibit 1.3 (page 19) shows an R program to obtain a perceptual map of seven sports, showing their relationships with one another. The program draws on modeling software for multidimensional scaling.
# MLB, NBA, and NFL Player Salaries (R)

library(lattice) # statistical graphics

# variables in contract data from spotrac.com (August 2015)
# player: player name (contract years)
# position: position on team
# team: team abbreviation
# teamsignedwith: team that signed the original contract
# age: age in years as of August 2015
# years: years as player in league
# contract: dollars in contract
# guaranteed: guaranteed dollars in contract
# guaranteedpct: percentage of contract dollars guaranteed
# salary: annual salary in dollars
# yearfreeagent: year player becomes free agent
#
# additional created variables
# salarymm: salary in millions
# leaguename: full league name
# league: league abbreviation

# read data for Major League Baseball
mlb_contract_data <- read.csv("mlb_player_salaries_2015.csv")
mlb_contract_data$leaguename <- rep("Major League Baseball",
 length = nrow(mlb_contract_data))
for (i in seq(along = mlb_contract_data$yearfreeagent))
  if (mlb_contract_data$yearfreeagent[i] == 0)
    mlb_contract_data$yearfreeagent[i] <- NA
for (i in seq(along = mlb_contract_data$age))
  if (mlb_contract_data$age[i] == 0)
    mlb_contract_data$age[i] <- NA
mlb_contract_data$salarymm <- mlb_contract_data$salary/1000000
mlb_contract_data$league <- rep("MLB", length = nrow(mlb_contract_data))
print(summary(mlb_contract_data))

# variables for plotting
mlb_data_plot <- mlb_contract_data[, c("salarymm","leaguename")]

nba_contract_data <- read.csv("nba_player_salaries_2015.csv")
nba_contract_data$leaguename <- rep("National Basketball Association",
 length = nrow(nba_contract_data))
for (i in seq(along = nba_contract_data$yearfreeagent))
  if (nba_contract_data$yearfreeagent[i] == 0)
    nba_contract_data$yearfreeagent[i] <- NA
for (i in seq(along = nba_contract_data$age))
  if (nba_contract_data$age[i] == 0)
    nba_contract_data$age[i] <- NA
nba_contract_data$salarymm <- nba_contract_data$salary/1000000
nba_contract_data$league <- rep("NBA", length = nrow(nba_contract_data))
print(summary(nba_contract_data))
# variables for plotting
nba_data_plot <- nba_contract_data[, c("salarymm","leaguename")]

nfl_contract_data <- read.csv("nfl_player_salaries_2015.csv")
nfl_contract_data$leaguename <- rep("National Football League",
    length = nrow(nfl_contract_data))
for (i in seq(along = nfl_contract_data$yearfreeagent))
    if (nfl_contract_data$yearfreeagent[i] == 0)
        nfl_contract_data$yearfreeagent[i] <- NA
for (i in seq(along = nfl_contract_data$age))
    if (nfl_contract_data$age[i] == 0)
        nfl_contract_data$age[i] <- NA
nfl_contract_data$salarymm <- nfl_contract_data$salary/1000000
nfl_contract_data$league <- rep("NFL", length = nrow(nfl_contract_data))
print(summary(nfl_contract_data))
# variables for plotting
nfl_data_plot <- nfl_contract_data[, c("salarymm","leaguename")]

# merge contract data with variables for plotting
plotting_data_frame <- rbind(mlb_data_plot, nba_data_plot, nfl_data_plot)

# generate the histogram lattice for comparing player salaries
# across the three leagues in this study
lattice_object <- histogram(~salarymm | leaguename, plotting_data_frame,
    type = "density", xlab = "Annual Salary ($ millions)", layout = c(1,3))

# print to file
pdf(file = "fig_understanding_markets_player_salaries.pdf",
    width = 8.5, height = 11)
print(lattice_object)
dev.off()
# Payroll and Performance in Major League Baseball (R)

library(ggplot2) # statistical graphics

# functions used with grid graphics to split the plotting region
# to set margins and to plot more than one ggplot object on one page/screen
vplayout <- function(x, y)
viewport(layout.pos.row=x, layout.pos.col=y)

# user-defined function to plot a ggplot object with margins
ggplot.print.with.margins <- function(ggplot.object.name,
left.margin.pct=10,
right.margin.pct=10,top.margin.pct=10,bottom.margin.pct=10)
{
# begin function for printing ggplot objects with margins
# margins expressed as percentages of total... use integers
grid.newpage()
pushViewport(viewport(layout=grid.layout(100,100)))
print(ggplot.object.name,
vp=vplayout((0 + top.margin.pct):(100 - bottom.margin.pct),
(0 + left.margin.pct):(100 - right.margin.pct)))
}

# read in payroll and performance data
# including annotation text for team abbreviations
mlb_data <- read.csv("mlb_payroll_performance_2014.csv")
mlb_data$millions <- mlb_data$payroll/1000000
mlb_data$winpercent <- mlb_data$wlpct * 100

cat("\nCorrelation between Payroll and Performance:\n")
with(mlb_data, print(cor(millions, winpercent)))

cat("\nProportion of win/loss percentage explained by payrolls:\n")
with(mlb_data, print(cor(millions, winpercent)^2))

pdf(file = "fig_understanding_markets_payroll_performance.pdf",
width = 5.5, height = 5.5)
ggplot_object <- ggplot(data = mlb_data,
aes(x = millions, y = winpercent)) +
geom_point(colour = "darkblue", size = 3) +
xlab("Team Payroll (Millions of Dollars)") +
ylab("Percentage of Games Won") +
geom_text(aes(label = textleft), size = 3, hjust = 1.3) +
geom_text(aes(label = textright), size = 3, hjust = -0.25)

ggplot.print.with.margins(ggplot_object, left.margin.pct = 5,
right.margin.pct = 5, top.margin.pct = 5, bottom.margin.pct = 5)

dev.off()
library(MASS)  # includes functions for multidimensional scaling
library(wordcloud)  # textplot utility to avoid overlapping text

USE_METRIC_MDS <- FALSE  # metric versus non-metric toggle

# utility function for converting a distance structure
# to a distance matrix as required for some routines and
# for printing of the complete matrix for visual inspection.
make.distance.matrix <- function(distance_structure) {
  n <- attr(distance_structure, "Size")
  full <- matrix(0,n,n)
  full[lower.tri(full)] <- distance_structure
  full+t(full)
}

# enter data into a distance structure as required for various
# distance-based routines. That is, we enter the upper triangle
# of the distance matrix as a single vector of distances
distance_structure <- as.single(c(9,11,10,5,14,4,15,6,12,13,16,1,18,2,20,7,3,19,17,8,21))

# provide a character vector of sports names
sport_names <- c("Baseball", "Basketball", "Football", "Soccer", "Tennis", "Hockey", "Golf")

attr(distance_structure, "Size") <- length(sport_names)  # set size attribute

# check to see that the distance structure has been entered correctly
# by converting the distance structure to a distance matrix
# using the utility function make.distance.matrix, which we had defined
distance_matrix <- unlist(make.distance.matrix(distance_structure))
cat("\n","Distance Matrix of Seven Sports","\n")
print(distance_matrix)

if (USE_METRIC_MDS) {
  # apply the metric multidimensional scaling algorithm and plot the map
  mds_solution <- cmdscale(distance_structure, k=2, eig=T)
}

# apply the non-metric multidimensional scaling algorithm
# this is more appropriate for rank-order data
# and provides a more satisfactory solution here
if (!USE_METRIC_MDS) {
  mds_solution <- isoMDS(distance_matrix, k = 2, trace = FALSE)
}
First_Dimension <- mds_solution$points[,1]
Second_Dimension <- mds_solution$points[,2]

plot(First_Dimension, Second_Dimension, type = "n", cex = 1.5,
xlim = c(-15, 15), ylim = c(-15, 15)) # first page of pdf plots

plot(First_Dimension, Second_Dimension, type = "n", cex = 1.5,
xlim = c(-15, 15), ylim = c(-15, 15)) # second page of pdf plots

plot(First_Dimension, Second_Dimension, type = "n", cex = 1.5,
xlim = c(-15, 15), ylim = c(-15, 15)) # third page of pdf plots

plot(First_Dimension, Second_Dimension, type = "n", cex = 1.5,
xlim = c(-15, 15), ylim = c(-15, 15)) # fourth page of pdf plots

pdf(file = "plot_pretty_original_mds_seven_sports.pdf",
width=8.5, height=8.5) # opens pdf plotting device

pdf(file = "plot_nonmetric_mds_seven_sports.pdf",
width=8.5, height=8.5) # opens pdf plotting device
Chapter 1. Understanding Sports Markets

First_Dimension <- mds_solution$points[,1]  # no reflection
Second_Dimension <- mds_solution$points[,2]  # no reflection
# wordcloud utility for plotting with no overlapping text
textplot(x = First_Dimension,
y = Second_Dimension,
words = sport_names,
show.lines = FALSE,
xlim = c(-15, 15),  # extent of horizontal axis range
ylim = c(-15, 15),  # extent of vertical axis range
xaxt = "n",  # suppress tick marks
yaxt = "n",  # suppress tick marks
cex = 1.15,  # size of text points
mgp = c(0.85, 1, 0.85),  # position of axis labels
cex.lab = 1.5,  # magnification of axis label text
xlab = "",
ylab = "")
dev.off()  # closes the pdf plotting device

pdf(file = "fig_sports_perceptual_map.pdf",
width=8.5, height=8.5)  # opens pdf plotting device
# use par(mar = c(bottom, left, top, right)) to set up margins on the plot
par(mar=c(7.5, 7.5, 7.5, 5))
First_Dimension <- mds_solution$points[,1] * -1  # reflect horizontal
Second_Dimension <- mds_solution$points[,2]
# wordcloud utility for plotting with no overlapping text
textplot(x = First_Dimension,
y = Second_Dimension,
words = sport_names,
show.lines = FALSE,
xlim = c(-15, 15),  # extent of horizontal axis range
ylim = c(-15, 15),  # extent of vertical axis range
xaxt = "n",  # suppress tick marks
yaxt = "n",  # suppress tick marks
cex = 1.15,  # size of text points
mgp = c(0.85, 1, 0.85),  # position of axis labels
cex.lab = 1.5,  # magnification of axis label text
xlab = "First Dimension (Individual/Team, Degree of Contact)",
ylab = "Second Dimension (Anaerobic/Aerobic, Other)"
dev.off()  # closes the pdf plotting device
This page intentionally left blank
Index

A
adjacency matrix, 261
advertising research, 229
agent, 261
agent-based modeling, 261
Akaike information criterion (AIC), 54
algebraic modeling system, 145, 261
Apache Hadoop, see Hadoop
Apache Lucene, see Lucene
Apache Software Foundation, 261
Apache Spark, see Spark
API, see application programming interface
application programming interface, 175, 178, 179, 261
ARPANET, 175, 262
ASP, 262
association rules, 262
asynchronous focus group, see blog

B
bandwidth, 262
baseball term
“out”, 289
“play ball”, 291
“safe”, 293
“time”, 296
ahead in the count, 279
All Star Game, 279
American League, 279
around the horn, 279
at bats (AB), 279
away team, 279
bailing out, 279
balk, 279
ball, 280
ban on women, 280
base, 280
base coach, 280
base hit, 280
base on balls (BB), 280
base runner, 280
baserunner, 280
baserunning error, 280
bases loaded, 280
batter, 280
batter in the hole, 280
batter on deck, 280
batter’s box, 280
battery, 280
batting average (BA, AVG), 280
batting stance, 281
batting team, 281
behind in the count, 281
bench, 281
big leagues, 281
BIP (BPIP), 281
bloop single, 281
bunt, 281
call, 281
called game, 281
catcher, 281
catch looking, 281
catch off base, 281
catch stealing (CS), 281
center fielder, 282
Championship Series, 282
changeup, 282
check swing, 282
choking up, 282
chop single, 282
clean-up hitter, 282
closed batting stance, 282
closer, 282
clutch hitter, 282
coach, 282
command, 282
control, 282
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>cover the bases</td>
<td>282</td>
</tr>
<tr>
<td>crowd the plate</td>
<td>283</td>
</tr>
<tr>
<td>curveball (curve)</td>
<td>283</td>
</tr>
<tr>
<td>cut fastball (cutter)</td>
<td>283</td>
</tr>
<tr>
<td>cut-off position</td>
<td>283</td>
</tr>
<tr>
<td>dead ball</td>
<td>283</td>
</tr>
<tr>
<td>dead ball era</td>
<td>283</td>
</tr>
<tr>
<td>dead red hitter</td>
<td>283</td>
</tr>
<tr>
<td>defense</td>
<td>283</td>
</tr>
<tr>
<td>defensive indifference</td>
<td>283</td>
</tr>
<tr>
<td>designated hitter (DH)</td>
<td>283</td>
</tr>
<tr>
<td>diamond</td>
<td>283</td>
</tr>
<tr>
<td>dig in</td>
<td>284</td>
</tr>
<tr>
<td>Division Series</td>
<td>284</td>
</tr>
<tr>
<td>double (2B)</td>
<td>284</td>
</tr>
<tr>
<td>double play</td>
<td>284</td>
</tr>
<tr>
<td>double-header</td>
<td>284</td>
</tr>
<tr>
<td>double-switch</td>
<td>284</td>
</tr>
<tr>
<td>dugout</td>
<td>284</td>
</tr>
<tr>
<td>earned run average (ERA)</td>
<td>284</td>
</tr>
<tr>
<td>expected runs, 62–65</td>
<td>284</td>
</tr>
<tr>
<td>extra-base hit</td>
<td>284</td>
</tr>
<tr>
<td>fair ball</td>
<td>284</td>
</tr>
<tr>
<td>fair territory</td>
<td>284</td>
</tr>
<tr>
<td>fan</td>
<td>284</td>
</tr>
<tr>
<td>fantasy baseball</td>
<td>284</td>
</tr>
<tr>
<td>fastball</td>
<td>284</td>
</tr>
<tr>
<td>fielder</td>
<td>285</td>
</tr>
<tr>
<td>fielder’s choice</td>
<td>285</td>
</tr>
<tr>
<td>fielding error</td>
<td>285</td>
</tr>
<tr>
<td>first base</td>
<td>285</td>
</tr>
<tr>
<td>first baseman</td>
<td>285</td>
</tr>
<tr>
<td>five-tool player</td>
<td>285</td>
</tr>
<tr>
<td>fly ball</td>
<td>285</td>
</tr>
<tr>
<td>fly out</td>
<td>285</td>
</tr>
<tr>
<td>force out</td>
<td>285</td>
</tr>
<tr>
<td>forfeited game</td>
<td>285</td>
</tr>
<tr>
<td>foul ball</td>
<td>285</td>
</tr>
<tr>
<td>foul territory</td>
<td>285</td>
</tr>
<tr>
<td>foul tip</td>
<td>285</td>
</tr>
<tr>
<td>frame (a pitch)</td>
<td>285</td>
</tr>
<tr>
<td>free agent</td>
<td>285</td>
</tr>
<tr>
<td>full count</td>
<td>285</td>
</tr>
<tr>
<td>game (G)</td>
<td>285</td>
</tr>
<tr>
<td>grand slam</td>
<td>285</td>
</tr>
<tr>
<td>ground ball</td>
<td>285</td>
</tr>
<tr>
<td>ground out</td>
<td>286</td>
</tr>
<tr>
<td>ground-rule double</td>
<td>286</td>
</tr>
<tr>
<td>hit (H)</td>
<td>286</td>
</tr>
<tr>
<td>hit batsman (hit by pitch, HBP)</td>
<td>286</td>
</tr>
<tr>
<td>hit for the cycle</td>
<td>286</td>
</tr>
<tr>
<td>hit-and-run</td>
<td>286</td>
</tr>
<tr>
<td>hitter</td>
<td>286</td>
</tr>
<tr>
<td>hitter’s park</td>
<td>286</td>
</tr>
<tr>
<td>hitting for power</td>
<td>286</td>
</tr>
<tr>
<td>hitting slump</td>
<td>286</td>
</tr>
<tr>
<td>hitting streak</td>
<td>286</td>
</tr>
<tr>
<td>holding runner on base</td>
<td>286</td>
</tr>
<tr>
<td>home plate</td>
<td>286</td>
</tr>
<tr>
<td>home run (HR)</td>
<td>286</td>
</tr>
<tr>
<td>home team</td>
<td>287</td>
</tr>
<tr>
<td>illegal pitch</td>
<td>287</td>
</tr>
<tr>
<td>in the hole</td>
<td>287</td>
</tr>
<tr>
<td>infielder</td>
<td>287</td>
</tr>
<tr>
<td>inning</td>
<td>287</td>
</tr>
<tr>
<td>intentional base on balls</td>
<td>287</td>
</tr>
<tr>
<td>interference</td>
<td>287</td>
</tr>
<tr>
<td>keystone sack</td>
<td>287</td>
</tr>
<tr>
<td>knuckleball</td>
<td>287</td>
</tr>
<tr>
<td>lead-off hitter</td>
<td>287</td>
</tr>
<tr>
<td>leave the yard (go yard)</td>
<td>287</td>
</tr>
<tr>
<td>left fielder (LF)</td>
<td>287</td>
</tr>
<tr>
<td>left on base (LOB)</td>
<td>287</td>
</tr>
<tr>
<td>lefty</td>
<td>287</td>
</tr>
<tr>
<td>line drive</td>
<td>288</td>
</tr>
<tr>
<td>lineup</td>
<td>288</td>
</tr>
<tr>
<td>live ball</td>
<td>288</td>
</tr>
<tr>
<td>live ball era</td>
<td>288</td>
</tr>
<tr>
<td>making the turn</td>
<td>288</td>
</tr>
<tr>
<td>manager</td>
<td>288</td>
</tr>
<tr>
<td>manufactured run</td>
<td>193</td>
</tr>
<tr>
<td>men on base</td>
<td>288</td>
</tr>
<tr>
<td>Mendoza Line</td>
<td>288</td>
</tr>
<tr>
<td>middle infielder</td>
<td>288</td>
</tr>
<tr>
<td>middle reliever</td>
<td>288</td>
</tr>
<tr>
<td>National League</td>
<td>288</td>
</tr>
<tr>
<td>neighborhood play</td>
<td>288</td>
</tr>
<tr>
<td>no hitter</td>
<td>289</td>
</tr>
<tr>
<td>no-no</td>
<td>289</td>
</tr>
<tr>
<td>obstruction</td>
<td>289</td>
</tr>
<tr>
<td>offense</td>
<td>289</td>
</tr>
<tr>
<td>official scorer</td>
<td>289</td>
</tr>
<tr>
<td>on the field (team)</td>
<td>289</td>
</tr>
<tr>
<td>on-base percentage (OBP)</td>
<td>289</td>
</tr>
<tr>
<td>open batting stance</td>
<td>289</td>
</tr>
<tr>
<td>out</td>
<td>289</td>
</tr>
<tr>
<td>outfielder</td>
<td>289</td>
</tr>
<tr>
<td>overslide</td>
<td>290</td>
</tr>
<tr>
<td>pace of play</td>
<td>290</td>
</tr>
<tr>
<td>passed ball</td>
<td>290</td>
</tr>
<tr>
<td>PECOTA, 33, 34, 290</td>
<td>290</td>
</tr>
<tr>
<td>perfect game</td>
<td>290</td>
</tr>
<tr>
<td>pick off assignment</td>
<td>290</td>
</tr>
<tr>
<td>pick off play</td>
<td>290</td>
</tr>
<tr>
<td>pinch hitter</td>
<td>290</td>
</tr>
<tr>
<td>pinch runner</td>
<td>290</td>
</tr>
<tr>
<td>pitch</td>
<td>290</td>
</tr>
<tr>
<td>pitch count</td>
<td>290</td>
</tr>
<tr>
<td>Index</td>
<td>331</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>pitcher (P), 291</td>
<td>spin rate, 294</td>
</tr>
<tr>
<td>pitcher’s duel, 291</td>
<td>spitball, 295</td>
</tr>
<tr>
<td>pitcher’s park, 291</td>
<td>squeeze play, 295</td>
</tr>
<tr>
<td>pitcher’s plate, 291</td>
<td>starting pitcher, 295</td>
</tr>
<tr>
<td>pitching depth, 291</td>
<td>steal (stolen base, SB), 295</td>
</tr>
<tr>
<td>pitching from the stretch, 291</td>
<td>stepping in the bucket, 295</td>
</tr>
<tr>
<td>pitching mound, 291</td>
<td>strike, 295</td>
</tr>
<tr>
<td>pitching rotation, 291</td>
<td>strike zone, 295</td>
</tr>
<tr>
<td>pivot foot, 291</td>
<td>strikeout (K), 295</td>
</tr>
<tr>
<td>place hitter, 291</td>
<td>suspended game, 296</td>
</tr>
<tr>
<td>plate, 291</td>
<td>sweep, 296</td>
</tr>
<tr>
<td>plate appearance, 291</td>
<td>switch hitter, 296</td>
</tr>
<tr>
<td>platooning, 291</td>
<td>switch pitcher, 296</td>
</tr>
<tr>
<td>pop-up, 292</td>
<td>tag out, 296</td>
</tr>
<tr>
<td>position number, 292</td>
<td>tagging up, 296</td>
</tr>
<tr>
<td>position player, 292</td>
<td>take a lead (off base), 296</td>
</tr>
<tr>
<td>power hitter, 292</td>
<td>take a pitcher deep, 296</td>
</tr>
<tr>
<td>productive at bat, 292</td>
<td>Texas Leaguer, 296</td>
</tr>
<tr>
<td>pull hitter, 292</td>
<td>third base, 296</td>
</tr>
<tr>
<td>pull the string, 292</td>
<td>third baseman, 296</td>
</tr>
<tr>
<td>quick pitch, 292</td>
<td>three-bagger, 296</td>
</tr>
<tr>
<td>reaching for the fences, 292</td>
<td>throw, 296</td>
</tr>
<tr>
<td>regulation game, 292</td>
<td>throwing error, 296</td>
</tr>
<tr>
<td>relief pitcher, 292</td>
<td>tie game, 296</td>
</tr>
<tr>
<td>replay review, 292</td>
<td>tip a pitch, 296</td>
</tr>
<tr>
<td>retouch, 292</td>
<td>total bases (TB), 297</td>
</tr>
<tr>
<td>reverse curve, 292</td>
<td>triple (3B), 297</td>
</tr>
<tr>
<td>right fielder (RF), 292</td>
<td>triple crown, 297</td>
</tr>
<tr>
<td>RISP, 293</td>
<td>triple-play, 297</td>
</tr>
<tr>
<td>rounding the bases, 293</td>
<td>two-bagger, 297</td>
</tr>
<tr>
<td>run, 293</td>
<td>umpire, 297</td>
</tr>
<tr>
<td>run batted in (RBI), 293</td>
<td>umpire-in-chief, 297</td>
</tr>
<tr>
<td>run down, 293</td>
<td>up the middle, 297</td>
</tr>
<tr>
<td>runner, 293</td>
<td>up to bat (team), 297</td>
</tr>
<tr>
<td>sacrifice bunt, 293</td>
<td>visiting team, 297</td>
</tr>
<tr>
<td>sacrifice fly, 293</td>
<td>VORP, 297</td>
</tr>
<tr>
<td>scoring position, 293</td>
<td>walk, 297</td>
</tr>
<tr>
<td>screwball, 293</td>
<td>walk-off balk, 297</td>
</tr>
<tr>
<td>season, 293</td>
<td>walk-off hit, 297</td>
</tr>
<tr>
<td>second base, 293</td>
<td>walk-off home run, 298</td>
</tr>
<tr>
<td>second baseman, 293</td>
<td>WAR (WARP), 298</td>
</tr>
<tr>
<td>secondary lead, 293</td>
<td>WHIP, 298</td>
</tr>
<tr>
<td>semi-intentional walk, 293</td>
<td>Wild Card Game, 298</td>
</tr>
<tr>
<td>set position, 293</td>
<td>wild pitch, 298</td>
</tr>
<tr>
<td>shadow ball, 294</td>
<td>windup position, 298</td>
</tr>
<tr>
<td>shift, 294</td>
<td>World Series, 298</td>
</tr>
<tr>
<td>shine ball, 294</td>
<td>Bayesian statistics, 54, 144, 198, 199, 204, 205, 262</td>
</tr>
<tr>
<td>shortstop, 294</td>
<td>Bayes information criterion (BIC), 54</td>
</tr>
<tr>
<td>shutout, 294</td>
<td>Bayes’ theorem, 204</td>
</tr>
<tr>
<td>side-arm delivery, 294</td>
<td>hierarchical modeling, 79, 267</td>
</tr>
<tr>
<td>single (1B), 294</td>
<td>best-case/worst-case approach, 144, 262</td>
</tr>
<tr>
<td>slider, 294</td>
<td>betweenness centrality, 232, 262</td>
</tr>
<tr>
<td>slugging percentage (SLG), 294</td>
<td>big data, 262</td>
</tr>
<tr>
<td>small ball, 294</td>
<td></td>
</tr>
</tbody>
</table>
binary variable, 262
biologically-inspired methods, 216
black box model, 215
bootstrap method, 55, 57
bootstrap sampling, 262
bot, see crawler (web crawler)
boundary (of a network), 263
bps, 263
brand positioning, see marketing, brand positioning
break-even analysis, see cost-volume-profit analysis
brute force approach, 263
bulletin board, see blog
C
C, C++, C#, 172, 173, 263
CART, 263
cascading style sheet (CSS), 263
case study
Return of the Bobbleheads, 112
censoring, 112, 206, 207
chat room, 263
choice uncertainty, see decision analysis
classical statistics, 199, 203, 205, 263
null hypothesis, 203
power, 204
statistical significance, 203, 204
classification, 198, 211–213, 215, 263
predictive accuracy, 211–213
client, 263
client-server application, 263
closeness centrality, 232, 263
cloud computing, 183
cluster analysis, 71, 73, 122, 216, 263
coefficient of determination, 211
comma-delimited text (csv), 263
compile cycle, 264
complexity, of model, 213
conjoint analysis, 78, 79, 264
consumer heterogeneity, see market segmentation, consumer heterogeneity
consumer surplus, see pricing research, consumer surplus
content analysis, 264
continuous random variable, see random variable, continuous
cookie, 264
corpus, 264
cost-volume-profit analysis, 135–138
crawler, 175–178, 183, 264
cross-sectional data, 264
cross-validation, 55, 56, 214, 264
CSS, see cascading style sheet (CSS)
csv, see comma-delimited text (csv)
cumulative frequency distribution, see data visualization, cumulative frequency distribution
customer lifetime value, 125
CVP analysis, see cost-volume-profit analysis
Cython, 172
D
data mining, see machine learning
data organization, 227
data partitioning, 55
data preparation, 181
missing data, 182
data science, 197, 198
data visualization, 264
biplot, 81
box plot, 102, 104
cumulative frequency distribution, 239, 248
diagnostics, 109, 213
differential runs plot, 236
heat map, 161, 162
histogram, 10, 154, 157, 159, 160
lattice plot, 10, 105–108, 238
moving fraction plot, 237, 245, 246
Sankey diagram, 80
scatter plot, 11
spine chart, 72, 77, 78, 82, 88
strip plot, 105, 107
data-adaptive research, 52, 53
database system, 180, 181
MongoDB, 183, 270
MySQL, 270
non-relational, 183
NoSQL, 271
PostgreSQL, 183, 272
relational, 181, 183
decision analysis, 142, 143
decision tree, see decision analysis
decision uncertainty, see decision analysis
declarative language, 264
degree (of a network), 264
degree centrality, 265
degree distribution, 265
density (of a network), 265
descriptive statistics, 265
discounted cash flow analysis, 139–141
discount rate, 139
net present value (NPV), 139
payback period, 140
return on investment (ROI), 140
discrete random variable, see random variable, discrete
Document Object Model (DOM), 176, 265, 267
document store, 173, 177, 179, 180
DOM, see Document Object Model (DOM)

E
e-mail, 265
eigenvector centrality, 232, 265
Elasticsearch, 173, 182, 183, 225, 265
Elo rating/ranking system, 40
emoji, see emoticon
emotion, 265
ethnography, 265
expected value, 265
experimental research, 265
explanatory model, 198, 266
explanatory variable, 51, 52, 124, 207, 266
exploratory data analysis, 102

F
factor analysis, 81, 266
fantasy sports, 35, 129, 130
forecasting, 229
Fortran, 172
frame, 266
ftp, 266
functional language, 266

G
game theory, 266
game-day strategy, see strategy, playing
General Inquirer, 223
generalized linear model, 214, 266
generative grammar, 266
genetic algorithm, 266
genetic algorithms, 216
Go (Golang), 172, 182, 266
graph theory, 267
grounded theory, 267

H
Hadoop, 267
heuristic, 216, 267
hierarchical modeling, see Bayesian statistics,
  hierarchical modeling
HTML, 175, 267
HTTP, 267

I
imputation, see multiple imputation
in-game strategy, see strategy, playing
indexing, 267
inferential statistics, 267
information retrieval, 182, 183, 220, 225, 267
injuries, 185, 189
integer knapsack problem, see mathematical
  programming, knapsack problem
integer programming, 267, see mathematical
  programming, integer programming
integrated development environment (IDE), 267
interaction effect, 213
Internet, 175, 267
Internet of Things, 170, 268
Internet Services Provider, 268
intranet, 268
investment analysis, see discounted cash flow
  analysis
IoT, see Internet of Things
IRC, 268
IT, 268

J
Java, 172, 173, 268
Java Virtual Machine, see Java
JavaScript, 268
JavaScript Object Notation (JSON), 179, 268
JPEG, 268
JSON, see JavaScript Object Notation (JSON)
JVM, see Java

K
kbps, 268
keyword, 268

L
LAMP, 268
leading indicator, 229
leave-one-out cross-validation, see cross-validation
levels of measurement, 269
linear model, 207, 214
linear predictor, 207
linear programming, 269, see mathematical
  programming, linear programming
linear regression, 269
Linux, 173, 183
listserv, 269
logging, see system logging
logistic regression, 42, 52, 211, 269
longitudinal data, 269
Lucene, 173, 225, 269
M

machine learning, 215, 216, 269
Major League Baseball, see MLB
Major League Soccer, see MLS
market basket analysis, 73
market response model, 111, 226–229
market segmentation, 79, 120–122
consumer heterogeneity, 127, 131
marketing
competitive analysis, 128, 129
mass marketing, 123
one-to-one marketing, 123
product positioning, 71, 72, 74
strategy, 128, 129
substitute products, 74
target marketing, 120
marketing mix model, 111, 228
Markov chain, 62, 269
transition probability, 62, 269
mass marketing, see marketing, mass marketing
mathematical programming, 127, 200, 269
integer programming, 200, 201, 267
knapsack problem, 41, 200, 201, 268
linear programming, 200, 269
mixed integer programming, 200, 269
stochastic programming, 134, 145, 275
measurement, 23–35, 269
accessible measure, 31
comprehensible measure, 31, 32
construct validity, 24, 27
content validity, 24, 224
convergent validity, 27
discriminant validity, 27
explicit measure, 31
face validity, 24, 224
levels of measurement, 28–29
multitrait-multimethod matrix, 24–26
predictive validity, 24
reliability, 24–26, 30, 274
tractable measure, 31
transparent measure, 31, 33
validity, 24, 26, 27, 31, 32, 277
MEG, see listserv
mixed integer programming, 269, see mathematical programming, mixed integer programming
MKP (multidimensional knapsack problem), see mathematical programming, knapsack problem
MLB, 4, 5, 9–11, 64, 65, 170, 201, 232, 233, 256, 288
MLS, 257
model, 270

model-dependent research, 52, 53
modem, 270
Moneyball, 9
MongoDB, see database system, MongoDB
morphology, 270
multi-fold cross-validation, see cross-validation
multi-level categorical variable, 270
multi-level modeling, see Bayesian statistics, hierarchical modeling
multidimensional knapsack problem, see mathematical programming, knapsack problem
multidimensional scaling, 71, 72, 81
Jaccard index, 73
perceptual map, 12, 13, 70–72, 241
similarity judgment, 71–73
multinomial variable, see multi-level categorical variable
multiple imputation, 270
multivariate methods, 81
MySQL, see database system, MySQL

N

nàïve Bayes model, 270
National Basketball Association, see NBA
National Football League, see NFL
National Hockey League, see NHL
natural language processing, see text analytics, natural language processing
NBA, 4, 6, 9, 10, 12, 38–40, 65–67, 170, 238, 239, 258
nearest-neighbor analysis, 73
net present value, see discounted cash flow analysis, net present value (NPV)
netiquette, 270
genre, 270
network science, 270
neural network, 270
NFL, 4, 7, 9, 10, 170, 259
NHL, 260
NLP, see text analytics, natural language processing
NoSQL, see database system, NoSQL
NPV, see discounted cash flow analysis, net present value (NPV)

O

object-oriented language, 271
observational research, 271
on-base percentage plus slugging (OPS), 289
one-to-one marketing, see marketing, one-to-one marketing
online community, 271, 277
operations research, 271
optimization, 216
organic search, 271
Orme, B. K., 131
outlier, 271
over-fitting, 213

P
PageRank, 271
paid search, 271
paired comparisons, 42
panel, 271
panel data, 271, see longitudinal data
parameter, 272
parametric models, 213
parser, 176, 177, 272
payback period, see discounted cash flow analysis
PECOTA, see baseball term, PECOTA
perceptual map, see multidimensional scaling, perceptual map
Perl, 272
PHP, 272
Poisson distribution, 272
Poisson regression, 206, 207, 210, 214, 272
population, 272
population distribution, 272
Porter five-forces model, 128, 129
post, 272
posterior distribution, 272
PostgreSQL, see database system, PostgreSQL
predictive analytics
definition, 51
predictive model, 198, 272
preference scaling, 81
price elasticity, see pricing research, price elasticity
price sensitivity, see pricing research, price sensitivity
pricing research, 126
consumer surplus, 128
price elasticity, 126
price sensitivity, 127
reference price, 127
three Cs of pricing, 126
willingness to pay, 75, 78, 79, 126, 127
principal component analysis, 81, 216, 272
prior distribution, 273
probability, 64
binomial distribution, 158
negative binomial distribution, 156, 158, 161
Poisson distribution, 156, 158, 161
procedural language, 273
product positioning, 81, see marketing, product positioning
psychographics, 273
Python, 171–173, 182, 273
Python package
json, 189
matplotlib, 86, 116, 242
numpy, 86, 99, 116, 168, 242
os, 186
pandas, 99, 116, 242
patsy, 99
scipy, 116, 168
scrapy, 186
sklearn, 86
statsmodels, 99, 116, 242
twitter, 189
Queuing theory, 273
R
R, 171–173, 182, 273
R package
car, 113
ggplot2, 18, 245, 246, 248
grid, 18, 245, 246, 248
lattice, 16, 113, 167
lubridate, 248
MASS, 83
plyr, 248
support.CEs, 88
R-squared, 211
random forests, 273
random network (random graph), 231, 273
random variable, 273
continuous, 273
discrete, 273
real-time analytics, 171, 172, 182
real-time focus group, see chat room
recency-frequency-monetary value (RFM) model, 126
reference price, see pricing research, reference price
regression, 52, 108–111, 119, 206, 213, 273
nonlinear regression, 214
robust methods, 214
time series regression, 228
regular expressions, 178, 273
regularized regression, 213
relational database, 274
reliability, see measurement, reliability
resampling, 274
response, 51
response variable, 124, 206
return on investment, see discounted cash flow analysis
robot, see crawler (web crawler)
ROI, see discounted cash flow analysis
root-mean-square error (RMSE), 211

S
sales forecasting, 226, 228
sampling, 274
  sampling variability, 204
sampling distribution, 274
sampling frame, 274
Scala, 172, 274
scatter plot, 274
scheduling, 216
scraper, 176, 177, 274
search, see information retrieval
semantic web, 274
semantics, 274
semi-supervised learning, 216, 274
sensitivity analysis, 145, 274
sentiment analysis, 222–225
shrinkage estimators, 213
similarity judgment, see multidimensional scaling, similarity judgment
simulated annealing, 274
simulation, 148, 153, 214, 275
  benchmark study, 214, 215
  game-day, 59, 151, 152
  what-if analysis, 198
small-world network, 231
smoothing methods, 213
  splines, 214
social network analysis, 230, 275
Solr, 225, 275
Spark, 172, 182, 275
spatial data, 275
spider, see crawler (web crawler)
SQL, 181, 275
statistic
  interval estimate, 203
  p-value, 203
  point estimate, 203
  test statistic, 203
statistical learning, see machine learning
stemming (word stemming), 275
stochastic process, 275
stochastic programming, see mathematical programming, stochastic programming
strategy
  playing, 61–68
strength of schedule, 38
structured query language, see SQL

T
target marketing, see marketing, target marketing
TCP/IP, 275
telnet, 276
term frequency-inverse document frequency, see TF-IDF
text analytics, 217–225
  content analysis, 223
  document annotation, 35
generative grammar, 217, 218
  latent Dirichlet allocation, 216
  latent semantic analysis, 216
  morphology, 218
natural language processing, 217, 224, 270
  semantics, 218
  stemming, 219
  syntax, 218
terms-by-documents matrix, 219, 221
text analysis, 276
text feature, 35
text measure, 35, 223–225
text summarization, 222
thematic analysis, 216, 223
text measure, 276
text mining, 276
text parser, see parser
TF-IDF, 220, 276
thread, of discussion, 276
time series, 276
time series analysis, 226–229
  ARIMA model, 227
  state space model, 228
time-value of money, 142
traditional research, 52
traditional statistics, 276
training-and-test regimen, 54, 55, 105, 108, 198, 199
transition probability, see Markov chain, transition probability
transitivity, 276
tree-structured model, 276
U
unidimensional scaling, 38, 40, 42, 43
unsupervised learning, 216, 220, 276
URL, 175, 277
Usenet, 277

V
validity, see measurement, validity
variable transformation, 213
virtual facility, 277
virtual private network (VPN), 277
VPN, see virtual private network

W
web, see World Wide Web
web board, see blog
web presence testing, 277
web scraper, see scraper
weblog, see blog
Wiki, 277
willingness to pay, see pricing research, willingness to pay
WNBA, 255
Women’s National Basketball Association, see WNBA
World Wide Web, 174, 175, 183, 277
WWW, see World Wide Web

X
XML, 277
XPath, 176, 277