8 Golden Goal Analytics

"It took me seventeen years and 114 days to become an overnight success."

—Lionel Messi

The sport we call soccer is recognized around the world as "Futbol." Although there are many different versions of the early beginnings of soccer, written text has been found referring to a similar game in China more than three thousand years ago. The game went by the name of "Tsu Chu" and involved kicking a leather ball into an empty hole. Other soccer historians note that, while this game was being played in the Eastern hemisphere, Native American tribes in the Western hemisphere were playing a kicking game called "pasuckuakohowog," which translates to "they gather to play ball with the foot." Researchers have also found traces of the game in Japan, Egypt, and Greece prior to its becoming popular in Europe and the Americas. In fact, kicking games may have been used to prepare warriors for battle.

Around 600-1600 AD versions of a kicking game were recorded in the Americas. Mesoamerican civilizations formed teams and set up baskets around a designated area with the objective of kicking a rubber ball into them. Ultimately, modern soccer originated in England, although among royalty it was frowned upon as a sport. English royalty were even known to imprison players of the game because of the ruckus it caused and its violent nature, which eventually led to its banning.

In spite of these events, rules for play were codified at the University of Cambridge in 1848 and were known as the Cambridge Rules. The sport grew and began to be played in schools and universities all over the country. Official soccer rules, known as the *Laws of the Game*, were drawn up by Cobb Morley, now recognized as the "father of soccer." These rules were eventually accepted by the Football Association in 1863 and have withstood the test of time. The first governing body of soccer was the Football Association, as the game was originally called association football. With time, the long name was shortened to "soccer."

The first official Football Association Cup was played in 1872. Leagues began to emerge in the late 1800s and early 1900s. The governing body of association football in Europe was the Union of European Football Association (UEFA) which was initiated in 1971. A few decades later the popularity of soccer spread to the United States and, in 1996, the organization of Major League Soccer (MLS) was developed.

In addition to league playoffs, countries participate in an international playoff every four years, the World Cup of the Fédération Internationale de Football Association (FIFA), the most watched sporting event worldwide. Common names of soccer legends to be well acquainted with include the likes of Pele, Diego Maradona, Ronaldo, Ronaldinho, David Beckham, Zinedine Zidane, Luis Enrique, Roberto Carlos, and future legends of the game, Lionel Messi and Cristiano Ronaldo.

The sport of soccer is played with eleven players on each team. It is a game in which hands are not allowed to touch the ball, with the exception of the goalie. The objective is to score as many goals as possible against the opposing team, and have the higher score by the end of the game. There is a ninety-minute clock, split in two halves, with halftime in the middle. After a fifteen-minute halftime, teams switch sides. The common goal is to score as many goals as possible and prevent the other team from scoring.

A team is considered to have scored a goal when the soccer ball crosses the goal line. In soccer, there is a sudden-death overtime that involves additional time for each team to score a goal. The winning goal in overtime is called the "golden goal." If neither team has scored in the two additional fifteen-minute overtime periods, penalty kicks take place to decide the game. Penalty shots also occur when a player commits a foul within the area of his own penalty box. Consequences of foul play include receiving a yellow or red card. A yellow card from the referee represents a warning. Two yellow cards are the equivalent of a red penalty card. A red card is given to a player when he has done something significantly wrong. The player is then required to leave the field immediately, and is suspended for the following game.

Equipment utilized in modern day soccer includes a pressurized spherical soccer ball, the inner layer of which is a latex bladder, and the outer layer stitched along the edges. The soccer ball's official dimensions are twenty-two centimeters in diameter and between sixty-eight to seventy centimeters in circumference.

A FIFA-approved soccer ball is the highest quality ball, as grueling tests of water absorption, air retention, air flight, and shape retention have been performed to make sure it is match-ready. The ball weighs between 410 and 450 grams.

Dimensions of the official soccer field are typically 115 yards in length, by 74 yards in width. The field consists of several landmarks including the goal line, halfway line, center circle, center spot, penalty box (eighteen-yard box), penalty spot, penalty arc, goal box (six-yard box), corner arc, and the technical area. Finally, soccer uniforms are fairly simple and consist of a jersey with the name and number of the player on the back, shorts, shin guards, and soccer-specific cleats. The goalie requires additional padding and gloves with which to catch the ball.

Although soccer technically consists of eleven positions, there are four distinct categories in which they can be placed. The first category is the goalie, second are the defenders (full backs), third includes midfielders (half-backs), and the fourth category is that of forwards (strikers).

The eleven positions can be shifted a bit, meaning that based on the formation, there may be more or fewer defenders, midfielders, or forwards on the field. There are two very popular formations common in today's game. There is the 4-4-2 formation, which consists of four defenders, four midfielders, two forwards, and the goalie. Then there is a 4-3-3 formation that has four positions as defenders, three as midfielders, three forwards, and of course the goalie.

The number of fullbacks, midfielders, and forwards may deviate based on the strategy of play, according to whether the coach wants to incorporate a more offensive or defensive style of play. It is important to understand formations, as most often they are indicators of a team's style of play, whether offensive, balanced, or defensive. A team that has an offensive style of play is always trying to find a way to score and is on the attack. Typical formations of teams that play using a more offensive style include the 4-4-2, 4-3-3, 4-2-4, 3-4-3, and 3-5-2. On the other hand, a team whose goalie may not be strong or a team facing an opposing team whose offense is extremely strong may be better off using a defensive style of play. As a result, the team may implement a 4-5-1, 5-3-2 or the versatile 4-4-2 formation. See the soccer glossary (page 227) for additional information about the game.

Literature on physical and psychological measures for professional soccer players reveals very interesting findings. Rampinini et al. (2007) examined whether a repeated-sprint ability (RSA) protocol, which consisted of maximal sprints lasting from one to seven seconds with interspersed short recovery period. They found RSA to be a more accurate assessment of anaerobic power than a single sprint assessment of match performance for soccer players. The assessment was quantified by a video-computerized match analysis image recognition system called ProZone. The RSA measure was found to be a more precise measure of match performance due to its ability to simulate real-time soccer matches. Another study examining anthropometrics and age characteristics of professional soccer players found no significant differences in match performance by age. The Rampinini et al. (2007) study suggests that today's goalies, central defenders, and center forwards are older and taller than players in those positions forty years ago. Additionally, this study found that midfielders and wing players displayed lower BMI and reciprocal ponderal index (a measure of leanness) compared with players in positions centrally located on the soccer field.

In 2010, a protocol simulating soccer games was developed and termed the ball-sport endurance and sprint test, nicknamed the BEAST90 protocol by Williams et al. (2010). This assessment measures movement patterns and physical demands replicating those that occur in a real-life soccer match. Amateur soccer players were tested to assess the validity and reliability of this protocol. Validity and reliability were established by comparing the values obtained from two separate assessments (with a gap of seven days in between) of fifteen amateur soccer players on distance, movement, peak heart rate, and oxygen uptake. The values in the two trials were comparable and did not significantly differ, confirming the reliability and validity of the BEAST90 protocol. This protocol is recommended for assessment of cardiorespiratory endurance and anaerobic power typically evidenced in a ninety-minute soccer match.

The literature is scarce pertaining to psychological assessments for soccer. Filaire et al. (2001) examined seventeen male professional soccer players' mood states. The findings revealed that professional soccer players who played well displayed the positive attributes of the iceberg profile and performed optimally during their soccer matches. Additionally, they found a significant decrease in vigor, paired with increased levels of tension and depression in professional soccer players who performed poorly.

Some scientists have found that exhibiting a negative attitude and an absence of strong psychological skills is associated with a predisposition to injury, as well as delayed injury recovery in athletes (Woods et al. 2003). It is important to know that soccer players have a tendency, because of the nature of sport, to suffer disproportionately from pulled muscles in the thigh area (twenty-three percent more), knee injuries (seventeen percent more), and ankle injuries (also seventeen percent more) compared to athletes in other sports.

Another study examined the relationship between lower body muscles and anaerobic power in soccer players (Robineau et al. 2012). Researchers specifically assessed the quadriceps and hamstring muscle groups as they relate to sprint speed, squat jump, and countermovement jump height reach. They found that over the length of a soccer match, squat jump height and sprint speed were significantly reduced, while no changes were found in countermovement jump. This implies that the squat jump may be a more relevant measure of anaerobic power for soccer.

Chamari et al. (2008) examined lower body explosiveness utilized the squat jump and countermovement jump with a modification that included five consecutive strides prior to, and post, the actual jump. They called the five consecutive strides jump test the "5JT" and implemented it with fifteen elite soccer players. They found that the 5JT had a significant positive correlation with both the standard squat jump measure and countermovement jump.

Chamari et al. (2008) suggested that the 5JT could also be used as an assessment of anaerobic power. Additionally, these researchers noted that distance covered during a soccer game significantly differed between player positions, with midfielders and forwards covering approximately twelve kilometers compared to defenders who typically cover about ten kilometers. Although research on performance measures of soccer players is more abundant than for other sports, there is still a long way to go.

Speed, agility, and anaerobic power are typically assessed for all soccer positions. The UEFA training ground and FIFA websites are great resources often utilized in the MLS soccer combine, as well as by other teams seeking to measure and assess soccer performance. But it must be recognized that measures and assessments should be designed for specific player positions. Taking a closer look at the role of each position, information can be extracted and utilized to develop more relevant and predictive assessments of soccer performance.

The position of defender typically requires greater anaerobic power capacity rather than aerobic capacity. Defenders also tend to be lower in technical skill, but have greater tackling ability in comparison to midfielders and strikers. As mentioned previously, players in center positions, particularly center defenders, are taller than right and left defenders. Athletes playing the sweeper position are usually smaller and quicker. Anaerobic power, agility, and lower body muscular strength should be assessed for defenders, utilizing the assessments designated in chapter 2. Additionally, the psychological makeup of a defender has been recognized to include high levels of motivation, persistence, and resiliency. Relevant psychological assessments include SMS-6, MMPI, and the 16PF.

A midfielder, regardless of whether designated to play in a more offensive or defensive style, should possess extremely high cardiorespiratory endurance and anaerobic power. A midfielder may have some advantage if he is tall, such as in winning headers. Conversely, a short midfielder may be quicker and more agile in eluding defenders. The midfielder position also requires the ability to see the game, spot teammates, make accurate passes, and set up plays. Of all soccer positions, midfielders are considered to control the pace of the game. They must be versatile enough to transition from defense to offense, pass the ball to the striker, and take the shot themselves. Assessments of cardiorespiratory endurance, as well as anaerobic power are strongly recommended. Additionally, the BEAST90 protocol is recommended for assessment and training, as it encompasses both aerobic and anaerobic capacity characteristics. Lower body muscular endurance and power assessments are also recommended for players in this position. The midfielder position requires that the athlete have high levels of vigor, the ability to think under pressure, create plays, as well as possess quick reflexes and decision-making ability. Thus, it is recommended that midfielders be assessed on the iceberg profile, the CSAI-2R, the Wonderlic, and the IAT to assess confidence and anxiety levels, cognitive ability, and reaction time reflexes.

Forwards get much of the credit for making goals. There are tall forwards and short forwards. The position of a forward requires quick decision making skills and the ability to score goals. Thus, forwards should have high levels of anaerobic power, extremely high shooting accuracy, great vision, anticipatory skills, and killer instincts. They must have the ability to handle pressure and be able to finish the play by scoring the goal. Measures of anaerobic power such as jump tests, RSA, and agility are strongly recommended for this position.

As for assessing the skill and technique needed for the scoring ability that is crucial to the position of a striker, incorporating a game-like simulation similar to that used by the NBA (for spot-up shooting and shooting off the dribble) should be implemented. For instance, for a forward it might be useful to incorporate drills and assessments that include dribbling and passing defenders, shooting from each side of the field, and scoring. Drills such as these will help to better assess skills and predict future in-game performance for the forward position. Also assessments of vision and reaction time using the IAT are recommended.

The SOQ and the TSCI are great assessments of confidence that are known to be reflective of handling high-pressured situations. The MMPI, TAT, and 16PF may be used to assess player personality, but there is yet to be a diagnostic that quantifies the "killer instinct."

A goalie requires great anticipatory skill, eye-hand coordination, flexibility, and anaerobic power. Recommended associations include the IAT, back scratch test, trunk rotation test, and the static squat jump. Additionally, a goalie should possess certain psychological traits specific to this position. The goalkeeper's position is very solitary. Thus, the player should display some degree of independence which may be assessed using the 16PF or the MMPI.

At the end of the soccer match, goalies have been said to feel one of two ways: like a winner, or like a failure. Goalies must have extreme concentration, thus an ability to get into the zone and remain in a state of flow should be assessed using either the Flow Questionnaire or the Flow State Scale. Finally, when it comes to being scored on, the goalie should have what is called a "short-term memory," so as not to affect future performance.

Combining the assessment of physical and psychological constructs with analytics, a new and better way of strategizing and predicting performance in soccer can be developed. Until recently, soccer analytics have been silent or dormant. A book about soccer, *The Numbers Game* (Anderson and Sally 2013), reveals that soccer analysts have been examining events when players have possession of the ball. Shockingly, 99 percent of the time players do not have the ball, and 98.5 percent of the time they run without it. The typical soccer player has possession of the ball for an average of 53.4 seconds throughout the entire course of a soccer match (Anderson and Sally 2013).

Sports analytics revolution has been at the forefront of sports science and technology, utilized by players, coaches, and team management to optimize performance and, consequently, the chances of winning. To illustrate this point, Peter Vermes, head coach of MLS team Sporting Kansas City is known to utilize sports data science to optimize performance and prevent injury in his players. He was quoted saying, "You're looking at what type of training sessions, or even exercises, that you may do with the players that are either extremely useful for your model of play, or actually detrimental.... No matter how you play, if you understand the science of it, you can train your team in the model of your play" (Schaerlaeckens 2015).

For additional information about the physiology of soccer, see Bangsbo (1993) and Stølen et al. (2005). Taskin (2008) reviews physical assessments relevant to the sport. And Bangsbo and Peitersen (2000) and Anderson and Sally (2013) discuss soccer analytics and strategy.

Table 8.1 (page 146) shows UEFA Champions League tournament phase leading scorers. Table 8.2 (page 147) shows UEFA Champions League tournament phase leaders in assists. And table 8.3 (page 148) shows key performance measures in soccer and their abbreviations.

Exhibit 8.1 (page 151) shows an R program and listing for analyzing performance of the UEFA Champions League leaders in assists and goals scored. Figure 8.1 (page 149) is a boxplot that displays the number of assists by player position (UEFA). And figure 8.2 (page 149) is a boxplot that displays the number of goals scored by player position (UEFA).

Continuing with the exploration of these data, exhibit 8.2 (page 154) shows an R program and listing for additional analyses of performance of the UEFA Champions League leading soccer players. Figure 8.3 (page 150) is a boxplot detailing passes attempted by player position (UEFA). And figure 8.4 (page 150) is a boxplot detailing passes completed by player position (UEFA).

Table 8.1. *UEFA Champions League Tournament Phase Leading Scorers*

Player	Team	Position	Total Points	Playing Time (minutes)
Neymar	Barcelona	3	10	1026
Cristiano Ronaldo	Real Madrid	3	10	1065
Lionel Messi	Barcelona	3	10	1147
Luiz Adriano	Shakhtar Donetsk	3	9	628
Jackson Martínez	Porto	3	7	629
Thomas Müller	Bayern	2	7	777
Luis Suárez	Barcelona	3	7	827
Carlos Tévez	Juventus	3	7	1156
Sergio Agüero	Man. City	3	6	550
Karim Benzema	Real Madrid	3	6	664
Edinson Cavani	Paris	3	6	920
Robert Lewandowski	Bayern	3	6	932
Klaas-Jan Huntelaar	Schalke	3	5	663
Yacine Brahimi	Porto	2	5	682
Álvaro Morata	Juventus	3	5	744
Mario Mandžukić	Atlético	3	5	832
Ciro Immobile	Dortmund	3	4	404
Nani	Sporting CP	2	4	449
Mario Götze	Bayern	2	4	747
Adrián Ramos	Dortmund	3	3	98
Lukas Podolski	Arsenal	4	3	127
Vincent Aboubakar	Porto	3	3	271
Marco Reus	Dortmund	4	3	341
Franck Ribéry	Bayern	2	3	387
Aaron Ramsey	Arsenal	2	3	404
Seydou Doumbia	CSKA Moskva	3	3	405
Gervinho	AS Roma	3	3	434
Ricardo Quaresma	Porto	2	3	469
Danny Welbeck	Arsenal	3	3	500
Lasse Schöne	Ajax	3	3	501

Note: Position numbers from the traditional 4-4-2 formation are as follows:
2 = right fullback, 3, 5, and 6 = left fullback, 4 = defensive midfielder, 7 = right midfielder,
8 = central/attacking midfielder, 9 = striker, 10 = second/support striker, 11 = left midfielder.

Source. `http://www.uefa.com/uefachampionsleague/season=2015/statistics/round=2000548/players/type=topscorers/index.html`.

Table 8.2. *UEFA Champions League Tournament Phase Leaders in Assists*

Player	Team	Position	Total Assists	Playing Time (minutes)
Lionel Messi	Barcelona	3	6	1147
Andrés Iniesta	Barcelona	2	5	786
Bastian Schweinsteiger	Bayern	2	4	456
Cesc Fàbregas	Chelsea	2	4	696
Koke	Atlético	2	4	833
Dani Alves	Barcelona	1	4	961
Cristiano Ronaldo	Real Madrid	3	4	1065
Pajtim Kasami	Olympiacos	3	3	391
Łukasz Piszczek	Dortmund	1	3	392
Pedro Rodríguez	Barcelona	4	3	397
Eden Hazard	Chelsea	4	3	654
Fabian Frei	Basel	3	3	670
Yacine Brahimi	Porto	4	3	682
Karim Bellarabi	Leverkusen	2	3	712
Thomas Müller	Bayern	3	3	777
Hector Herrera	Porto	2	3	789
Gregory van der Wiel	Paris	2	3	812
Luis Suárez	Barcelona	3	3	827
Paul Pogba	Juventus	2	3	833
Juanfran	Atlético	2	3	930
Robert Lewandowski	Bayern	3	3	932
Toni Kroos	Real Madrid	3	3	968
Jérôme Boateng	Bayern	1	3	990
Carlos Tévez	Juventus	3	3	1156
Rafinha	Barcelona	2	2	114
Bernard	Shakhtar Donetsk	4	2	158
Ricardo Kishna	Ajax	3	2	241
Juan Manuel Iturbe	Roma	3	2	244
Arkadiusz Milik	Ajax	3	2	256
Jefferson	Sporting CP	1	2	270

Note: Position numbers from the traditional 4-4-2 formation are as follows:
2 = right fullback, 3, 5, and 6 = left fullback, 4 = defensive midfielder, 7 = right midfielder,
8 = central/attacking midfielder, 9 = striker, 10 = second/support striker, 11 = left midfielder.

Source. `http://www.uefa.com/uefachampionsleague/season=2015/statistics/round=2000548/players/type=assists/index.html`.

Table 8.3. *Soccer Performance Measures*

Performance Measure	Abbreviation
Assists	A
Average goals expected per shot	ExpG
Completed percentage	CMP%
Corner kicks	CK
Forward completed percentage	Forward Comp %
Forward pass percentage	Forward Pass %
Fouls committed	FC
Fouls suffered	FS
Game winning goals	GWG
Games played	GP
Games started	GS
Goal percentage	GP%
Goals	G
Goals above average	GAA
Goals against	GA
Goals for	GF
Minutes	Min
Offsides	OFF
Passes attempted	PA
Passes completed	PC
Penalty kick attempts	PKA
Penalty kick goals	PKG
Penalty kick saves	PKS
Saves	SV
Shots	SHT
Shots on goal	SOG
Shutouts	SO

Figure 8.1. *Number of Assists by Player Position (UEFA)*

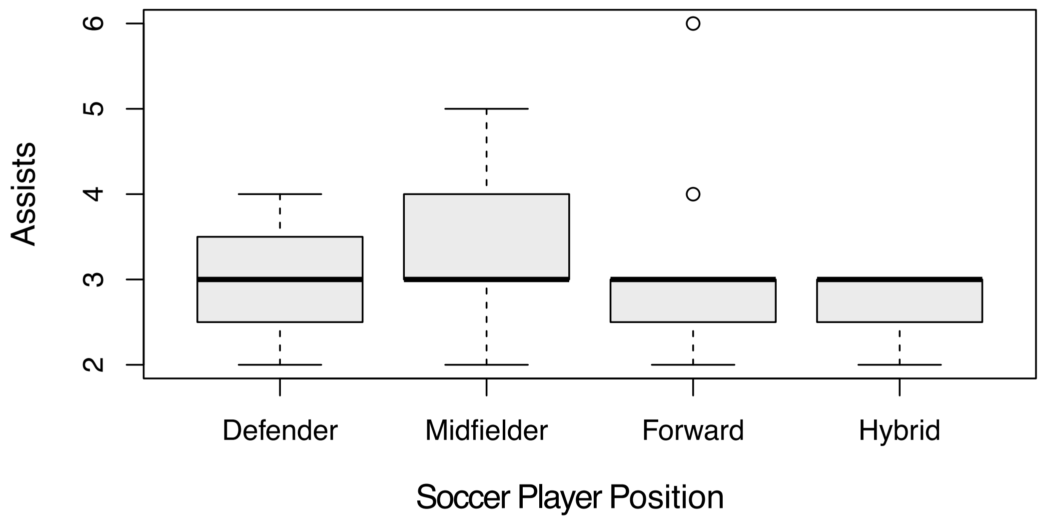

Figure 8.2. *Number of Goals Scored by Player (UEFA)*

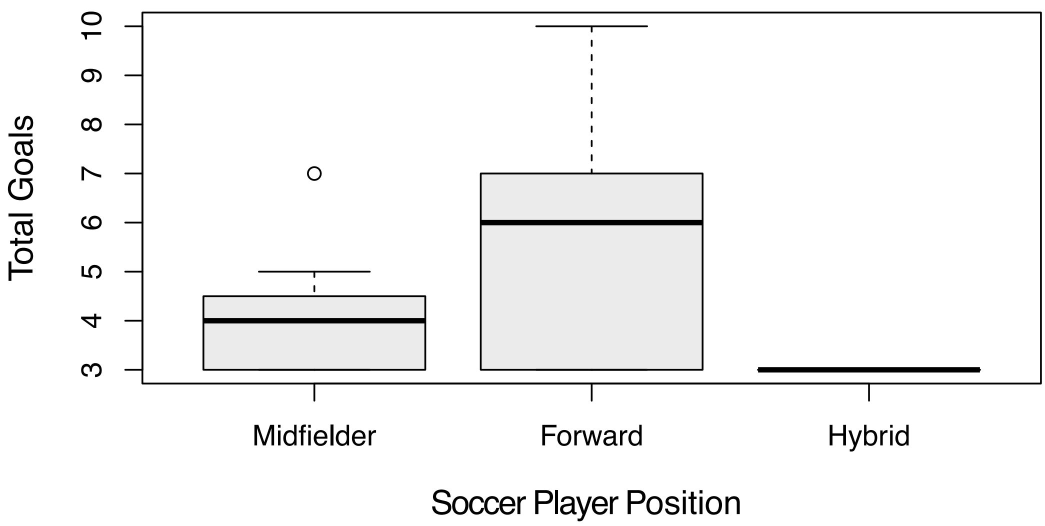

Figure 8.3. *Number of Passes Attempted by Player Position (UEFA)*

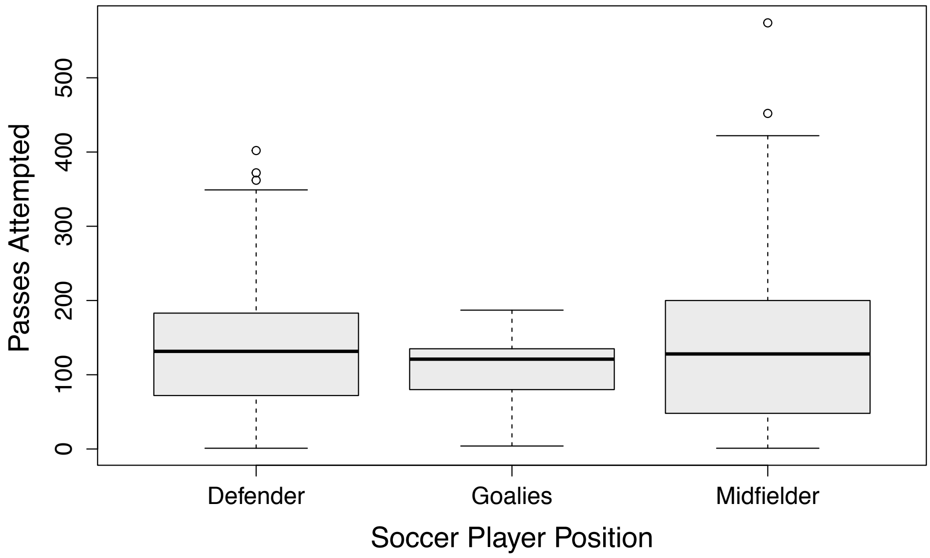

Figure 8.4. *Number of Passes Completed by Player Position (UEFA)*

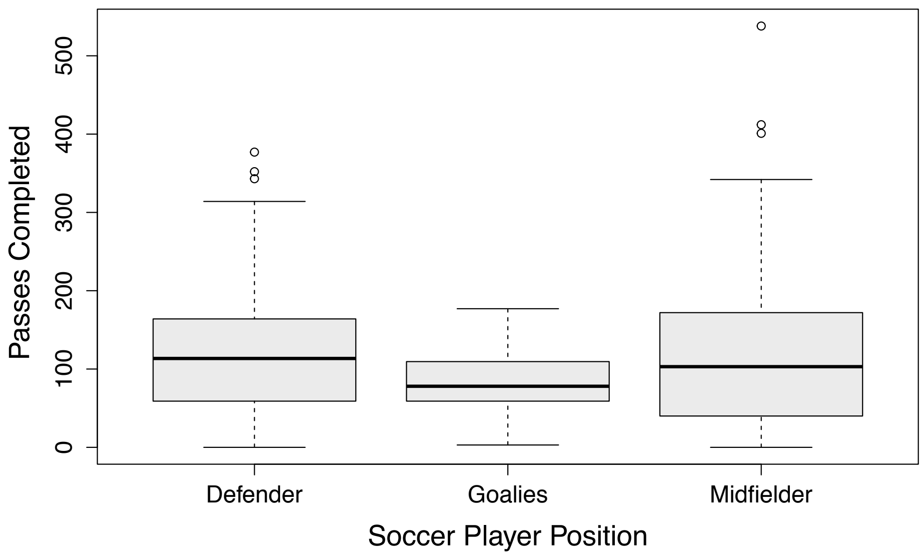

Exhibit 8.1. *Analyzing UEFA Assists and Goals Scored (R)*

```
# Analyzing UEFA  Assists and Goals Scored (R)

####################################
## Golden Goal Analytics R code   ##
####################################
# Reading in the data
setwd('/Users/Desktop/PerformanceMeasurementAnalytics/soccer')
SoccerAssists <- read.csv('goalassist.csv')
SoccerGoalsScored <- read.csv('goalscored.csv')

# Checking the structure of the data
str(SoccerAssists)
str(SoccerGoalsScored)

# Changing the positions variable to character
SoccerAssists$Position <- as.character(SoccerAssists$Position)
SoccerGoalsScored$Position <- as.character(SoccerGoalsScored$Position)

# Testing for a relationship between Time and Assists,
# since there is no relationship using this dataset,
# we can now ignore it in the regression model
AssistsTimeModel1 <- lm(Time ~ Position, data = SoccerAssists)

# Check summary for significance
summary(AssistsTimeModel1)

##################################################
##   UEFA Champions League Leaders in Assists   ##
##################################################

# Assists by position - simple linear regression model
AssistsbyPositionLinearModel2 <- lm(Assists ~ Position, data = SoccerAssists)

# Check summary for significance
summary(AssistsbyPositionLinearModel2)

# Assists by position -  ANOVA
AssistsbyPositionANOVAModel3 <- aov(Assists ~ Position, data = SoccerAssists)

# Check summary for significance
summary(AssistsbyPositionANOVAModel3)

# Robust alternative - Kruskal Wallis use if data is not normally distributed
kruskal.test(Assists ~ factor(Position), data = SoccerAssists)

# Count data -> use a Poisson regression. Uses a log link to model data.
AssistsbyPositionsPoissonModel4 <- glm(Assists ~ Position,
                                       data = SoccerAssists, family = poisson)

# Check summary for significance
summary(AssistsbyPositionsPoissonModel4)
```

```
# Examining the boxplot below we see that
# there seem to be differences by player position.
# However they are not statistically significant.
# The boxplot does show a trend of
# greater number of assists by the midfielder position.

# Figure
# Creates a space to save the figure
pdf('SoccerAssists.pdf', height = 4, width = 4 * (1 + sqrt(5)) / 2)
# Which variables to use
with(SoccerAssists, boxplot(Assists ~ Position,
                       # Changes x and y labels
                       xlab = 'Soccer Player Position', ylab = 'Assists',
                       # Names under the boxplots
                       names = c('Defender', 'Midfielder', 'Forward', 'Hybrid'),
                       # Color - see color chart
                       col = 'gray92',
                       # Main title
                       main = 'Number of Assists by Player Position (UEFA)',
                       # Size of x and y labels
                       cex.lab = 1.2))
# Closes the figure space
dev.off()

########################################################
##   UEFA Champions League Leaders in Goals Scored    ##
########################################################

# Total by position - regression linear model (same as ANOVA)
SoccerGoalsScoredLinearModel1 <- lm(Total ~ Position, data = SoccerGoalsScored)
SoccerGoalsScoredANOVAModel2 <- aov(Total ~ Position, data = SoccerGoalsScored)

# Robust alternative - Kruskal Wallis since data is not normally distributed
kruskal.test(Total ~ factor(Position), data = SoccerGoalsScored)

# Count data -> use a Poisson regression. Uses a log link to model data.
SoccerGoalsScoredPoissonModel3 <- glm(Total ~ Position,
                                      data = SoccerGoalsScored, family = poisson)

# Examining the boxplot below we see
# that there seem to be differences by player position.
# However they are not statistically significant.
# The boxplot does show a trend of greater number
# of goals scored by the forward position.
# The lack of statistical significance could be due to the small sample size.
# These data are based on only the world leading scorers
# At that level midfielders also have a high scoring record.

# Figure
# Creates a space to save the figure
pdf('SoccerGoalsScored.pdf', height = 4, width = 4 * (1 + sqrt(5)) / 2)
# Which variables to use
```

```
with(SoccerGoalsScored, boxplot(Total ~ Position,
                       # Changes x and y labels
                       xlab = 'Soccer Player Position', ylab = 'Total Goals',
                       # Names under the boxplots
                       names = c('Midfielder', 'Forward', 'Hybrid'),
                       # Color - see color chart
                       col = 'gray92',
                       # Main title
                       main = 'Number of Goals Scored by Player (UEFA)',
                       # Size of x and y labels
                       cex.lab = 1.2))
# Closes the figure space
dev.off()

##############################################################
##  UEFA Champions League Performance by Player Position   ##
##############################################################

# Reading in the data
setwd('/Users/Desktop/PerformanceMeasurementAnalytics/soccer')
UEFAdefenders <- read.csv('UEFAdefense.csv')
UEFAgoalies <- read.csv('UEFAgoalies.csv')
UEFAmidfielders <- read.csv('UEFAmidfielder.csv')

# Combine the rows for the three files read in and named the new file uefa
uefa <- rbind(UEFAdefenders, UEFAgoalies, UEFAmidfielders)
uefa$Position <- c(rep('Defender', nrow(UEFAdefenders)),
                   rep('Goalie', nrow(UEFAgoalies)),
                   rep('Midfielder', nrow(UEFAmidfielders)))

# Get rid of rows with NA (missing values) across
uefa <- uefa[!is.na(uefa$PA), ]

# This presents a summary of your data.
summary(uefa)
```

Exhibit 8.2. *Analyzing UEFA Passes Attempted and Completed (R)*

```
# Analyzing UEFA Passes Attempted and Completed (R)

####################################
## Golden Goal Analytics R code   ##
####################################
# Reading in the data
setwd('/Users/Desktop/PerformanceMeasurementAnalytics/soccer')
SoccerAssists <- read.csv('goalassist.csv')
SoccerGoalsScored <- read.csv('goalscored.csv')

# Checking the structure of the data
str(SoccerAssists)
str(SoccerGoalsScored)

# Changing the positions variable to character
SoccerAssists$Position <- as.character(SoccerAssists$Position)
SoccerGoalsScored$Position <- as.character(SoccerGoalsScored$Position)

# Reading in the data
setwd('/Users/Desktop/PerformanceMeasurementAnalytics/soccer')
UEFAdefenders <- read.csv('UEFAdefense.csv')
UEFAgoalies <- read.csv('UEFAgoalies.csv')
UEFAmidfielders <- read.csv('UEFAmidfielder.csv')

# Combine the rows for the three files read in and named the new file uefa
uefa <- rbind(UEFAdefenders, UEFAgoalies, UEFAmidfielders)
uefa$Position <- c(rep('Defender', nrow(UEFAdefenders)),
                   rep('Goalie', nrow(UEFAgoalies)),
                   rep('Midfielder', nrow(UEFAmidfielders)))

# Get rid of rows with NA (missing values) across
uefa <- uefa[!is.na(uefa$PA), ]

# Presents a summary of your data
summary(uefa)

###################################################################
##  UEFA Champions League Passes Attempted by Player Position    ##
###################################################################

# Use a linear model to regress player position on pass attempts
PassAttemptsbyPositionLinearModel1 <- lm(PA ~ Position, data = uefa)

# If you do not want to include outliers then
# check the Median using tapply function because the
# mean will be skewed by outliers
tapply(uefa$PA, uefa$Position, median)
tapply(uefa$PA, uefa$Position, mean)

# Kruskal Wallis Test (a non-parametric equivalent of anova)
kruskal.test(PA ~ factor(Position), data = uefa)
```

```
# The following output is displayed
# Kruskal-Wallis rank sum test
# data:  PA by factor(Position)
# Kruskal-Wallis chi-squared = 2.7198, df = 2, p-value = 0.2567
# There was no significant difference by player position on
# passes attempted with the sample from this dataset.
# Examining the boxplot below we can tell that certain positions
# do attempt to pass more although the results
# were not statistically significant.

# Figure
# Creates a space to save the figure
pdf('UEFAPA.pdf', height = 4, width = 4 * (1 + sqrt(5)) / 2)
# Which variables to use
with(uefa, boxplot(PA ~ Position,
                   # Changes x and y labels
                   xlab = 'Position', ylab = 'Soccer Player Position',
                   # Color - see color chart
                   col = 'gray92',
                   # Main title
                   main = 'Number of Passes Attempted by Player Position (UEFA)',
                   # Size of x and y labels
                   cex.lab = 1.2))
# Closes the figure space
dev.off()


#################################################################
##  UEFA Champions League Passes Completed by Player Position  ##
#################################################################

# Use a linear model to regress player position on pass attempts
PassesCompletedbyPositionLinearModel1 <- lm(PC ~ Position, data = uefa)

# Get rid of rows with NA (missing values) across
uefa <- uefa[!is.na(uefa$PC), ]

# Use a linear model to regress player position on passes completed
PassAttemptsbyPositionLinearModel1 <- lm(PC ~ Position, data = uefa)

# If you do not want to include outliers then check the Median
# using tapply function because the mean will be skewed by outliers.
tapply(uefa$PC, uefa$Position, median)
tapply(uefa$PC, uefa$Position, mean)

# Kruskal Test
kruskal.test(PC ~ factor(Position), data = uefa)

# The following output is displayed
# Kruskal-Wallis rank sum test
# data:  PC by factor(Position)
# Kruskal-Wallis chi-squared = 6.5932, df = 2, p-value = 0.03701
```

```
# There is a significant difference by player position
# on passes completed p=0.03

# Figure
# Creates a space to save the figure
pdf('UEFApassescompleted.pdf',
    height = 4, width = 4 * (1 + sqrt(5)) / 2)
# Which variables to use
with(uefa, boxplot(PC ~ Position,
                   # Changes x and y labels
                   xlab = 'Soccer Player Position',
                   # Color - see color chart
                   col = 'gray92',
                   # Main title
                   main = 'Number of Passes Completed by Player Position (UEFA)',
                   # Size of x and y labels
                   cex.lab = 1.2))
# Closes the figure space
dev.off()
```