I. Research Problem

The purpose of this study was to utilize data analytics as means to classify NFL offensive play types. This research more specifically examines the application of logistical regression analysis utilizing open source software in comparing two NFL teams’ play selections between the 2000 and 2012 seasons. Data were collected on two NFL teams’ first, second, and third down play selections in the fourth quarter over 13 seasons. Plays were identified as run or pass. The software R was used to analyze each team’s likelihood of running or passing on a given down.

Given the post-Moneyball sport environment, the use of analytics in the front office is well known (Lewis, 2004). However, the use of data to inform coaching decisions on the field of play is gaining recognition as well. If provided to coaches in a useable, informative format, they can use this type of analysis to assist in determining defensive or offensive strategies and tactics. This research contains timely information on how data analysis can provide insights into team tendencies. Coaches seek an edge in understanding not only their own team’s strengths, weaknesses, and tendencies, but also those of their opponents.

This article is likely to be useful to sport analytics personnel as well as to coaches. Other sport managers who oversee personnel and/or operations will have an interest in how sport analytics can benefit not only the business functions in the front office, but also the on-field performance of the team. Most franchises have analytics staff. Translating the data analyses conducted by sport analysts into a format that is useful to coaches is an essential step in making this process useful. The use of analytics is spreading across all sports and to amateur and intercollegiate play as well. The use of open source software to analyze data is a valuable asset to stakeholders in a wide variety of sport settings.
II. Issues

Data analytics are used in a variety of capacities in sport organizations. “Big data” is the collection of vast amounts of data. Sport organizations have access to “big data,” which can provide them with improved efficiency and effectiveness, ultimately leading to a competitive advantage. There is no specific size that determines how big “big data” has to be. It only has to be big enough to answer questions and guide decisions. This growth in available data requires new ways to transform it into useful information, and ultimately into competitive advantage and improved team performance. An available tool that can transform raw data into useful knowledge is the interactive statistical software R, which is an open-sourced computing and graphics platform that is available as free software. In sport, “big data” is valuable only in its usefulness as a competitive advantage.

As “big data” becomes more available and more usable in informing strategic decisions, the mining of relevant data is also being used to guide decisions in sport. Data mining involves looking for useful information in large datasets. Mining involves the collection of data from readily available sources. Aspects of data mining include classifying, clustering, trend and deviation analyses, and dependency modeling. The data collected via mining undergoes statistical analysis. Analytical applications essentially involve data input and output. In other words, sport analysts mine raw data, analyze it, and produce useful output. While data input is important, the focus of sport organizations is in the application of data output, the analysis of the data. The output must be presented in a format that can be used in decision-making processes.

Analytics can enhance productivity and efficiency in sport, whether evaluating prospective talent, individualizing customer service, or informing managerial and coaching decisions. Gaining a competitive advantage through sport analytics begins with analysts analyzing pertinent data. Next, that analysis must be translated into a format that can be used by managers and coaches in developing and implementing decision-making strategies. Managers and coaches must find value in using analytics to inform their respective decisions. That is likely to occur if evidence shows that the application of data analysis results in a competitive advantage.

Pursuing competitive advantage is a fundamental principle in sport. Data mining and subsequent predictive analyses in sport might yield a competitive advantage. There are many sport settings, both in the front office and on the field, where more efficient and effective decision making systems, powered by data mining and predictive analyses, may enhance performance. This study explores an on-field application of these techniques within the NFL, but the principles can apply across many sports and settings.
III. Summary

The purpose of this paper is to employ open source software to classify offensive plays by two NFL teams. Specifically, the Cleveland Browns and the Pittsburgh Steelers, a long-running rivalry in the American Football Conference (AFC), were selected for review. During the 13-year period of this study, the Browns were an unsuccessful team as measured by wins (71). In contrast, the Steelers were extremely successful, not only winning 64 (135–71) more games than the Browns but also the 2006 and 2009 Super Bowl championships. However, in many respects, these teams, separated by 135 miles, are similar. Both teams play in the same conference, thereby sharing many opponents. They both have similar “blue collar” cultures, and represent similar-sized markets.

IV. Analysis

The logistic regression used in this study calculates the logistical odds of something occurring. However, in practical terms, once odds are calculated, they can yield a more practical outcome, known as probability. Probability is the percentage likelihood of an event occurring or not occurring, for example, whether a play will be a run or pass. The range of probability is from 0% to 100%.

This analytical application was 66.4% accurate for Cleveland Browns. The model classified 8,090 play types correctly and 4,097 incorrectly. The application for the Pittsburgh Steelers was 66.9% accurate. The model classified 9,442 play types correctly and 4,681 incorrectly. This article also examines a sample play for each team, and a third play is used to compare teams side by side. Pittsburgh’s play was against Baltimore Ravens’ defense and resulted in a pass. It was the first play of the drive in the 14th minute of the fourth quarter. Pittsburgh was trailing Baltimore by a score of 10 to 13. Additionally, the Steelers only had a single timeout left versus the Ravens’ three. Specifically, it was first-and-10 on the Pittsburgh 30-yard line. Cleveland’s play was against Cincinnati’s defense and resulted in a rush. It was the second play of the drive in the sixth minute of the first quarter. Cleveland was trailing Cincinnati by a score of 0 to 7. Both teams had three timeouts, and it was second-and-3 on the Cleveland 23-yard line. A third example served as a direct comparison between the teams. In this situation, and ones like it, the gap between the teams can be large. In this case, the result is a classification of a Pittsburgh pass versus a Cleveland rush. Pittsburgh had a 50.1% probability of a pass while Cleveland had a 38.8%. So, why does this matter?

V. Discussion/Implications

The results of this study demonstrate that open source software can be effectively used in the analysis of situation-specific play selection in the NFL. It also provides evidence that logistic regression has promise as a classification system for opposing NFL offenses. The analysis of large amounts of data regarding play selection can inform decisions by head coaches, defensive coordinators, position coaches, opposing coaches, and even player personnel managers. Head coaches,
defensive coordinators, and position coaches can better develop a game plan and prepare athletes for opponents’ anticipated play selection. Conversely, offensive coaches might choose to alter their documented play selection pattern simply by being informed of it. Player personnel managers may identify and pursue talent specifically based on an analysis of situation-specific play selection, both by their own team and by their opponents.

This research demonstrated that data analytics can add value in the decision-making process in sport settings, in this case, the NFL. It also demonstrated that people are critical in the effective implementation of sport analytics. Slaton (2013) noted that “…the entire sports organization, from the lowliest assistant coach and marketing employee to the most senior leader needs to adopt the analytics philosophy if it is to be truly effective” (p. 1). Whether on the field or in the front office, coaches and managers are central in the collection of data, in the analysis of that data, and in the application of that data analysis.

If it is presented in a useful format, quantitative analysis provides valuable information that sport managers and coaches can use to inform their decisions. There is ample data available to support both on-the-field and front-office decisions in sport. It is incumbent on sport managers to obtain, analyze, and utilize appropriate data within their organization.

In the coming years, the United States will need up to 190,000 more expert analysts and an additional 1.5 million managers, including sport managers, who are prepared to use data to inform decisions (Manyika et al, 2011). The application of analytics in sport is no longer the niche pursuit of a few visionaries such as the Oakland A’s Billy Beane, the Houston Rockets’ Daryl Morey, or the Kraft Group’s Jessica Gelman, but rather it is a mainstream practice that sport organizations have embraced (Slaton, 2013). For example, over 2,700 interested researchers, analysts, and sport managers annually attend the Sloan Conference on the latest developments in sport analytics.

As evidence of the successful application of sport analytics builds, a new style of sport manager is evolving in sport organizations (Fry & Ohlmann, 2012). These new sport managers and coaches are proficient in the application of data analytics. The organizations they work for are stimulating the expansion of data mining and predictive analysis in sport settings. The proper utilization of data management, predictive analyses, and data-driven decision making by sport managers occur in pursuit of a competitive advantage. Coaches and managers seeking a competitive advantage can employ the data generated regarding opponent tendencies in their decisions on player personnel, formations, and play calls in these specific game situations.

This study supports the contention that the analysis of data in sport can be most useful in its application by qualified sport leaders. The increasing use of sport analytics, while informing decisions, does not negate the value of human insights and actions, or the need to translate quantitative data into useful information for
decision makers. Sport managers, informed by data analytics, must ultimately be the ones who employ these tools. Analysts are crucial in collecting appropriate data and in analyzing that data accordingly. Analysts and managers must be able to translate data that is applicable in the decision-making process. Effective coaches and sport managers in the future will need to be able to apply analytics to strategic decisions. For example, in this study, it is in the hands of head coaches, offensive and defensive coordinators, position coaches, and player personnel managers to determine the best way to effectively use the data analyses in the decision-making process. Additionally, their intended use of the data, and the questions they ask, can inform future analyses.

This study provided evidence of situation-specific tendencies for two teams in the NFL. The study was delimited to only the Cleveland Browns and Pittsburgh Steelers. All other teams and other professional sport leagues, while not included in this study, would likely have similar applications. This analytical approach could be easily applied to any team, in any sport, within any situation-specific parameters. Overtime plays were not analyzed in this study, but could be analyzed to inform decisions in that specific situation. Likewise, fourth-down plays were excluded, but could be reviewed in a separate analysis. In explicit game situations, such as time remaining or score, first-down plays, plays in the “red zone,” special team plays could be similarly analyzed. An analysis of the effectiveness of man-to-man or zone defenses, based upon points allowed or opponent shooting percentage, could inform coaching decisions in basketball.

To better inform the application needs of coaches and sport managers, analytical tools and procedures can be tailored to specific situations and directed toward specific questions. Changes in coaching staff, player personnel, and even opponents’ style of play occur regularly in the NFL. Analyses incorporating algorithms that address these factors can be developed. For example, when Tim Tebow replaced Kyle Orton as the Denver Broncos quarterback, the number of rushes increased significantly. Games with inclement weather may often have more rushes. So, adding an attribute for weather conditions may improve the analyses’ accuracy. Coaches, offensive coordinators, and audible-calling quarterbacks, such as Peyton Manning, may behave consistently in similar game situations despite the team. Thus, using data on a specific decision maker (e.g. quarterback) rather than as a team data set may improve the predictive accuracy.

Ultimately, analytics must be connected to the goals and performance of the sport organization. The data collected must be guided by organizational goals and directed toward gaining a competitive advantage. Identifying possible influencing factors and connecting the data collection and analyses to the questions that need to be answered will refine the process. The application of sport analytics will result in success only if the data analyses are translated into a useful format to inform coaches’ and sport managers’ decision making.