Marissa Ellingson, Sophie Giacobbe, Jason Nero, Kristine Wiggins

DSCI 478 Capstone Group Project

Dr. Emily King

3 May 2023

Effect of Timeouts and Other Correlates on NFL Pressure Kicks

Introduction & Background

American football is an incredibly popular sport and often has fans on the edge of their seats as the clock runs down in the fourth quarter. Close games are frequently decided in the last minutes when the losing team has just managed to drive into field goal range and is lining up to kick. If the kick is made, the offensive team might add a win to their record, or at least send the game into overtime for another chance at glory. At this point, the defending team may choose to deploy psychological strategies to influence the kick's outcome. "Icing the kicker" is a strategy in which the defending team calls a timeout just before the kicker begins the play. The intention behind this timeout is to throw off the kicker, forcing him to reprepare or give him time to "contemplate the negative outcomes if he fails to score (i.e., rumination)" (Goldschmeid & Cafri, 2010). If the kicker is affected enough by this timeout call, the idea is that he will miss the kick, handing the defensive team the ball and possibly the win. There are other aspects that a defensive team can potentially rely on, such as high winds or cold temperatures. This paper uses archival NFL data to analyze and quantify the effect of icing and other circumstances on the outcome of a last-minute field goal attempt. The models used to examine the effect and other circumstances are K-nearest neighbors, logistic regression, mixed effect logistic regression, and random forest. The models created will also be used to attempt predictions on a sample of the data.

This paper intends to expand on the findings of Nadav Goldschmeid and Guy Cafri in 2010. These men sought to quantify the effectiveness of icing to induce a negative outcome for the kicking team with their sports psychological research. Published their findings in their paper *Pressure Kicks in the NFL: An Archival Exploration Into the Deployment of Timeouts and Other Environmental Correlates*. Goldschmeid and Cafri would draw on NFL archival data for only the 2002-2008 National Football League seasons. Using a mixed-effect hierarchical linear model, "it was found that icing was successful in reducing scoring while other environmental factors such as experience, game location or game score were not associated with conversion success" (Goldschmeid & Cafri, 2010). Based on their findings, only about 64% of iced kickers made the kick, while about 84% of non-iced kickers made the kick, indicating that icing the kicker can reduce a kicker's chance of scoring on a decisive play. Lastly, Goldschmeid and Cafri determined that factors such as score, location, and kicker experience were not associated with altering outcomes (Goldschmeid & Cafri, 2010).

Data

The dataset was retrieved using *nflfastR*, an R package that quickly scrapes NFL data from multiple sources. Initially, the dataset consisted of every play of every game from the 2000-2020 seasons. The dataset was then filtered to remove any observations and variables unrelated to field goal attempts. Once complete, 46 possible predictor variables remained with one response (field_goal_result) and 22,803 kick observations to train and execute the models. After addressing NA values, three subsets were generated from the dataset to narrow the scope to specifically pressure kicks. For this paper, pressure kicks are considered field goal attempts occurring at or under five minutes remaining in the game. The three subsets contain plays occurring under one, three, and five minutes remaining. Please see Appendix A, *Variable Descriptions*, for the complete list of variables and their descriptions selected for this project.

Methodology

As previously mentioned, this paper intends to explore the impact and effectiveness of calling a timeout at the start of a decisive kick play by leveraging inferential and predictive models and algorithms to identify significant determinants of the outcome when the kick does occur. Additionally, the paper will aim to craft predictive solutions to determine the success of future icing attempts on decisive plays. The process started with exploring the data using correlation maps and Factor Analysis of Mixed Data (FAMD). Though these provided cursory insight into the possible relationships between the continuous variables (correlation map) and quantitative and qualitative variables (FAMD), such as the difference in scores and seconds remaining in the game, the process continued toward more complex models that could offer potentially greater insight. See Appendix B, *FAMD Plots*, and Appendix C, *Correlation Plots*, for the outputs of these models.

The first model implemented was K-nearest neighbors (KNN). This simple non-parametric supervised learning classification model uses the observations 'closest' in space to determine the outcome of an individual data point (IBM, 2021). This model type is useful due to its simplicity and ability to handle high-dimensional data. We used this model in anticipation that pressure kicks under similar circumstances lead to similar results. Multiple KNN models were fit using different subsets of the predictor variables and different values of k, selecting the two strongest models for discussion.

The next model type considered was logistic regression, a classic model used to make binary predictions (IBM, 2021). Since the response variable is either "made" or "missed," it makes sense to use logistic regression. Many models containing subsets of predictor variables were fit for each of the three subsets of data looking for significant observations using single variable predictors. In doing so, this paper hoped to identify key variables for predictive models of pressure kick outcomes.

Next, the implementation of mixed effects logistic models sought to expand upon the results from the logistic regression models. The reason for this model and the implementation are similar to that of basic linear regression. The primary difference is that the mixed effects models include a random intercept for each kicker in the dataset. Including this aspect accounts for more variation and focuses the model on the effects of the other variables rather than differences between kickers and their abilities.

The final model type was random forest, a machine-learning algorithm that uses multiple decision trees to determine one result. The flexibility in this model is valuable when using complex data like the dataset considered here (IBM, 2021). For the random forest implementation, models were fit for each of the three data subsets. Each step implemented 200 simple trees, providing decent accuracy without excessive demand on system resources. Lastly, a Boruta algorithm was leveraged for all data subsets to verify variable importance within the context of the random forest model. The resulting models identify critical variables and predict the outcome of a pressure kick and how the inclusion of an attempt by the defense to 'ice the kicker' will influence the outcome.

Results & Discussion

Analyzing the results from the K-nearest-neighbors implementation shows decent predictive accuracy. Classifying a kick as either "made" or "missed," the paper, as previously mentioned, experimented using different subsets of the data and values of k. The two strongest-performing models using this method both had an accuracy of around 80%. The first model used a value of k=13 and 31 predictor variables, leading to an accuracy of 82.27%. The second most accurate model used a value of k=21 and a subset of 10 input variables with an accuracy of 83.18%.

Model 1

	MADE	MISSED
Predicted Made	1893	390
Predicted Missed	23	24

82.27% Accuracy

Model 2

	MADE	MISSED
Predicted Made	1923	376
Predicted Missed	15	16

83.18% Accuracy

Both KNN models are better at classifying whether a kick was made than predicting a missed one, only predicting a missed kick correctly at best 5.8% of the time. While correctly predicting a made kick 99.23% of the time, making this method not very useful in real life.

For the logistic regression analysis, the first model was if a timeout was called before the kick, or "timeout" in the dataset. This model gave the highest accuracy out of any of the logistic regression models across all three time restraints with 80.52% with one minute remaining, 68.36% with three minutes remaining, and 72.77% with five minutes remaining.

One Minute Remaining | MADE | MISSED
Three Minutes Remaining | MADE | MISSED
Five Minutes Remaining | MADE | MISSED

One Minute Remaining

	MADE	MISSED
Predicted	205	62
True Value	206	61

99.63% Accuracy

Three Minutes Remaining

	MADE	MISSED
Predicted	170	380
True Value	448	102

49.45% Accuracy

Five Minutes Remaining

	MADE	MISSED
Predicted	319	476
True Value	680	106

54.07% Accuracy

The second model also had only one predictor variable: the number of timeouts the other team (the defensive team) had remaining. The thought here is that if the other team had more timeouts remaining, the kicker would feel nervous that they might call a timeout, causing them to have a lower success rate. In the one-minute time restraint, this had the highest accuracy of any model, with 99.63% accuracy. However, the accuracy did not carry across the other time restraints, with 49.45% accuracy with three minutes remaining and 54.07% accuracy with five minutes remaining.

The third model used the possession team as the predictor variable. In the one-minute and five-minute time restraints, only the Baltimore Ravens were flagged as significant predictors. In the three-minute time restraint, the Baltimore Ravens, Buffalo Bills, and Detroit Lions were considered significant. These models had decent accuracy, with 60.30%, 72.91%, and 66.66% in the one-, three-, and five-minute time restraints, respectively.

One Minute Remaining

	MADE	MISSED
Predicted	100	167
True Value	206	61

60.30% Accuracy

Three Minutes Remaining

	MADE	MISSED
Predicted	299	251
True Value	448	102

72.91% Accuracy

Five Minutes Remaining

	MADE	MISSED
Predicted	418	368
True Value	680	106

66.66% Accuracy

The four and final model consisted of only one predictor used seconds remaining. This model had decent accuracy for the three- and five-minute time restraints (79.27% and 77.99%) but not for the one-minute restraint (50.56%).

One Minute Remaining

	MADE	MISSED
Predicted	74	193
True Value	206	61

50.56% Accuracy

Three Minutes Remaining

	MADE	MISSED
Predicted	334	216
True Value	448	102

79.27% Accuracy

Five Minutes Remaining

	MADE	MISSED
Predicted	507	279
True Value	680	106

77.99% Accuracy

After making a model with every predictor variable, the ones deemed significant (seconds remaining, drive, defending team remaining timeouts, the difference in score, probability of not scoring, probability of the other team scoring, and win probability) were used to make the next model. The model had accuracies of 54.31%, 64.73%, and 65.65% in the one-, three-, and five-time restraints, respectively.

One Minute Remaining

	MADE	MISSED
Predicted	84	182
True Value	206	61

54.31% Accuracy

Three Minutes Remaining

	MADE	MISSED
Predicted	254	295
True Value	448	102

64.73% Accuracy

Five Minutes Remaining

	MADE	MISSED
Predicted	410	374
True Value	680	106

65.65% Accuracy

Using this model, the significant predictors were taken out to create another model. This new model used the probability of not scoring and the probability of the other team scoring. The model's results had accuracies of 64.42%, 75.27%, and 75.19% in the one-, three-, and five-time restraints, respectively.

One Minute Remaining

	MADE	MISSED
Predicted	111	156
True Value	206	61

64.42% Accuracy

Three Minutes Remaining

	MADE	MISSED
Predicted	312	238
True Value	448	102

75.27% Accuracy

Five Minutes Remaining

	MADE	MISSED
Predicted	485	301
True Value	680	106

75.19% Accuracy

In the mixed effects regression models, the distance of the kick was the most prominently significant variable, but temperature and wind also had noticeable effects. The significance of kick distance was especially evident in field goals attempted in the last three minutes of the game. Other variables, such as score differential and remaining timeouts available for the defensive team, had little to no effect on the outcome of a kick in these models. The model containing only kick distance had an accuracy of 79%, 81.8%, and 84% for one, three, and five minutes remaining, respectively. The best model for each time subset had 79.4%, 81.64%, and 84.86% accuracy. These models included the number of seconds remaining in the game, the distance of the kick, the number of timeouts remaining for the defensive team, and the difference in score.

With the random forest implementation, the factor selection highlights similar results to the mixed effect models, with kicking distance being ranked with high importance to the relationship of the model with the response 'field_goal_result.' Factors associated with probability also ranked high in importance. However, this paper stresses concern over their importance to the outcome. Such factors may use the response variable to calculate their value, explaining their highly correlated relationship, lessening their impact on the result unless the kicker is aware of such probabilities going into the pressure kick. Unfortunately, the web scrape package does not provide equations that calculate these variables. not provided by the web scraper package.

On the predictive side of the random forest implementation, based on the selected variables and the current time remaining, for plays with or without a timeout, shows promise in predicting the outcome of a decisive kick with about 88% accuracy on the test data. The accuracy did not profoundly change between minutes remaining subsets. If this paper had access to the mathematics behind each variable, the accuracy would likely improve and factors once significant, now insignificant, could be weeded out, revealing the proper factors that influence the outcome.

Conclusion

By taking the time to expand upon the work of Goldschmeid and Cafri in their paper *Pressure Kicks in the NFL: An Archival Exploration Into the Deployment of Timeouts and Other Environmental Correlates*, their insights into the effect of the psychological tactic known as ‘icing’ on the effect of a pressure kick outcome is confirmed while offering new approaches to the subject via factor significance and prediction modeling. Based upon mixed effect modeling, factors such as distance of the kick, temperature, and wind speed are significant, while confirming that score and location remain insignificant as the number of seasons included in the dataset grows.

A logistic regression model using the one-minute restraint subset had the highest accuracy of any model, with 99.63% accuracy, with the number of timeouts remaining for the defensive team being the single predictor variable. This model highlights the significant relationship between the field goal kick success, and the defense calling a timeout, hinting at ‘icing’ continuing to influence the outcome of pressure kicks effectively when available to the defensive team. Looking at the complete usage of predictor variables in the dataset, the random forest model ranked the highest, about 88%, on the test data set across all time restraint subsets. The logistic and random forest models are a reminder that model selection is contextual and is not a one size fits all scenario. It is worth noting that due to the limitations of the R package, we could not determine the calculations associated with specific factors, which limited our ability to dig further into the potential impact of these factors.

Additional limitations and recommendations that the group would like to address are that this project runs on archival data and is limited to data provided by NFL.com. Alternative factors to the negative outcome of a decisive play, such as pre-existing mental or physical stressors, could not be adequately accounted for; thus, this project and its conclusions should be considered a chapter in a larger narrative toward identifying and controlling for hidden factors. Suggestions for expansion include but are not limited to pre- and post-kick interviews, behavioral and physical assessments or surveys, and gameplay footage analysis. In doing so, layers of insignificant factors can be stripped away to reveal what factors contribute to inducing a negative outcome to a pressure kick.

References

Competitor-Cutter. “K-Nearest Neighbors Algorithm with Examples in R (Simply Explained Knn).” Medium, Towards Data Science, 30 Dec. 2018, https://towardsdatascience.com/k-nearest-neighbors-algorithm-with-examples-in-r-simply-explained-knn-1f2c88da405c.

Goldschmeid, N., & Cafri, G. (September, 2010). *Pressure kicks in the NFL: An archival exploration into the deployment...* https://www.researchgate.net/profile/Nadav-Goldschmied/publication/289361244_Pressure_Kicks_in_the_NFL_An_Archival_Exploration_Into_the_Deployment_of_Timeouts_and_Other_Environmental_Correlates/links/572d8a0908aeb1c73d11be19/Pressure-Kicks-in-the-NFL-An-Archival-Exploration-Into-the-Deployment-of-Timeouts-and-Other-Environmental-Correlates.pdf

IBM (2021), *What is KNN?* IBM, Corporation. https://www.ibm.com/topics/knn

IBM (2021), *What is Logistic Regression?* IBM, Corporation. https://www.ibm.com/topics/logistic-regression

IBM (2021), *What is Random Forest?* IBM, Corporation. https://www.ibm.com/topics/knn

Kassambara, et al. “FAMD - Factor Analysis of Mixed Data in R: Essentials.” STHDA, Statistical Tools for High-Throughput Data Analysis, 24 Sept. 2017, http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/115-famd-factor-analysis-of-mixed-data-in-r-essentials/.

Lyubomirsky, S., Tucker, K.L., Caldwell, N.D., & Berg, K. (1999). Why ruminators are poor problem solvers: Clues from the phenomenology of dysphoric rumination. Journal of Personality and Social Psychology, 77, 1041–1060.

Appendix

A. Variable Descriptions

Variable	Description
play_id	Numeric play id that when used with game_id and drive provides the unique identifier for a single play.
game_id	Ten digit identifier for NFL game.
home_team	String abbreviation for the home team.
away_team	String abbreviation for the away team.
season_type	'REG' or 'POST' indicating if the game belongs to regular or post season.
week	Season week
posteam	String abbreviation for the team with possession.
posteam_type	String indicating whether the posteam team is home or away.
quarter_seconds_remaining	Numeric seconds remaining in the quarter.
half_second_remaining	Numeric seconds remaining in the half.
drive	Numeric drive number in the game.
sp	Binary indicator for whether or not a score occurred on the play.
qtr	Quarter of the game (5 is overtime).

time	Time at start of play provided in string format as minutes:seconds remaining in the quarter.
play_type	String indicating the type of play: pass (includes sacks), run (includes scrambles), punt, field_goal, kickoff, extra_point, qb_kneel, qb_spike, no_play (timeouts and penalties), and missing for rows indicating end of play.
field_goal_result	String indicator for result of field goal attempt: made, missed, or blocked.
kick_distance	Numeric distance in yards for kickoffs, field goals, and punts.
home_timeouts_remaining	Numeric timeouts remaining in the half for the home team.
away_timeouts_remaining	Numeric timeouts remaining in the half for the away team.
timeout	Binary indicator for whether or not a timeout was called by either team
timeout_team	String abbreviation for which team called the timeout.
posteam_timeouts_remaining	Number of timeouts remaining for the possession team.
defteam_timeouts_remaining	Number of timeouts remaining for the team on defense.
posteam_score	Score the posteam at the start of the play.
defteam_score	Score the defteam at the start of the play.
score_differential	Score differential between the posteam and defteam at the start of the play.

posteam_score_post	Score for the posteam at the end of the play.
defteam_score_post	Score for the defteam at the end of the play.
score_differential_post	Score differential between the posteam and defteam at the end of the play.
no_score_prob	Predicted probability of no score occurring for the rest of the half based on the expected points model.
opp_fg_prob	Predicted probability of the defteam scoring a FG next.
wp	Estimated win probabiity for the posteam given the current situation at the start of the given play.
def_wp	Estimated win probability for the defteam.
home_wp	Estimated win probability for the home team.
away_wp	Estimated win probability for the away team.
kicker_player_name	String name for the kicker on FG or kickoff.
kicker_player_id	Unique identifier for the kicker on FG or kickoff.
season	4 digit number indicating to which season the game belongs to.
play_clock	Time on the playclock when the ball was snapped.
location	Either 'Home' o 'Neutral' indicating if the home team played at home or at a neutral site.
result	Equals home_score - away_score and means the game outcome from the perspective of the home team.

roof	One of 'dome', 'outdoors', 'closed', 'open' indicating the roof status of the stadium the game was played in. (Source: Pro-Football-Reference)
surface	What type of ground the game was played on. (Source: Pro-Football-Reference)
temp	The temperature at the stadium only for 'roof' = 'outdoors' or 'open'.(Source: Pro-Football-Reference)
wind	The speed of the wind in miles/hour only for 'roof' = 'outdoors' or 'open'. (Source: Pro-Football-Reference)

Table A1: Descriptions of Variables as outlined by https://www.nflfastr.com/articles/field_descriptions.html

B. FAMD Plots

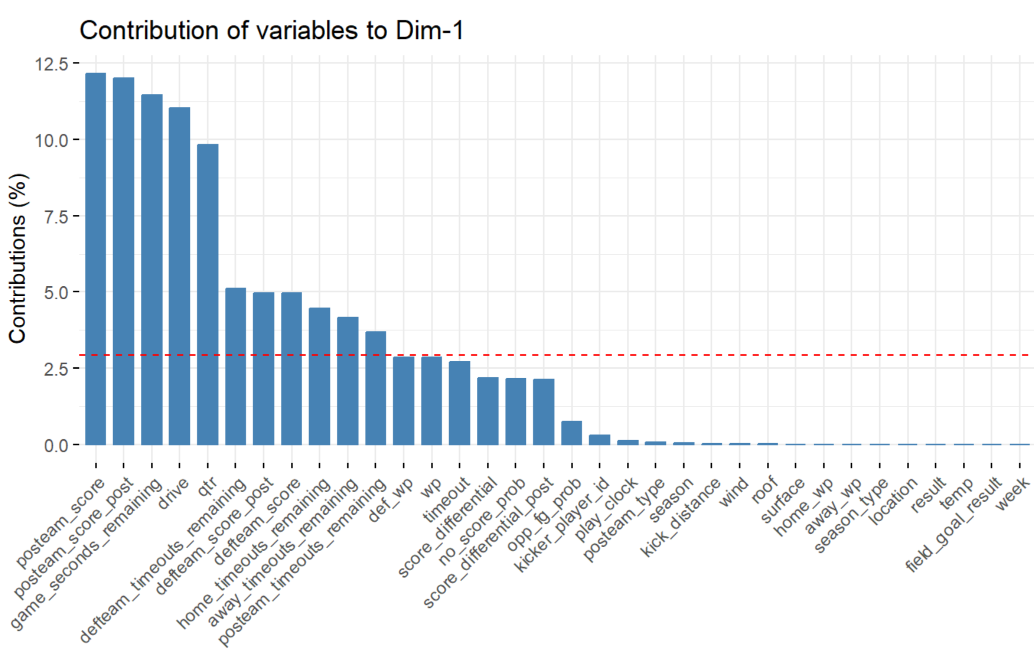

Figure B1: Variable Contributions to FAMD Dimension 1

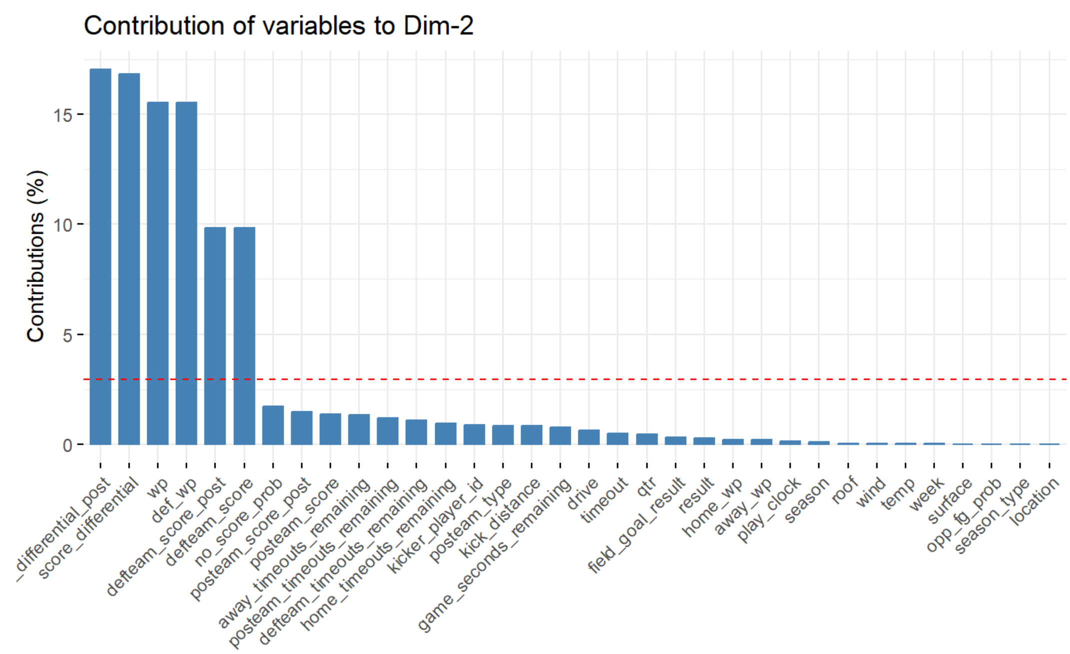

Figure B2: Variable Contributions to FAMD Dimension 2

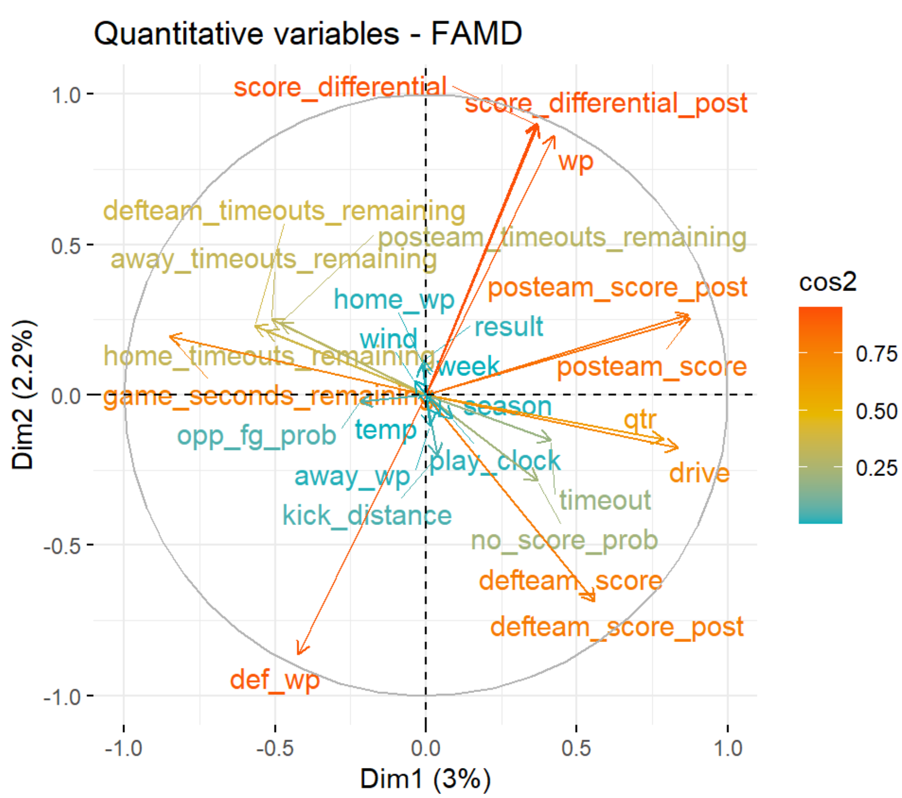

Figure B3: Plot of Quantitative Variable Contributions to FAMD Dimensions

C. Correlation Plots

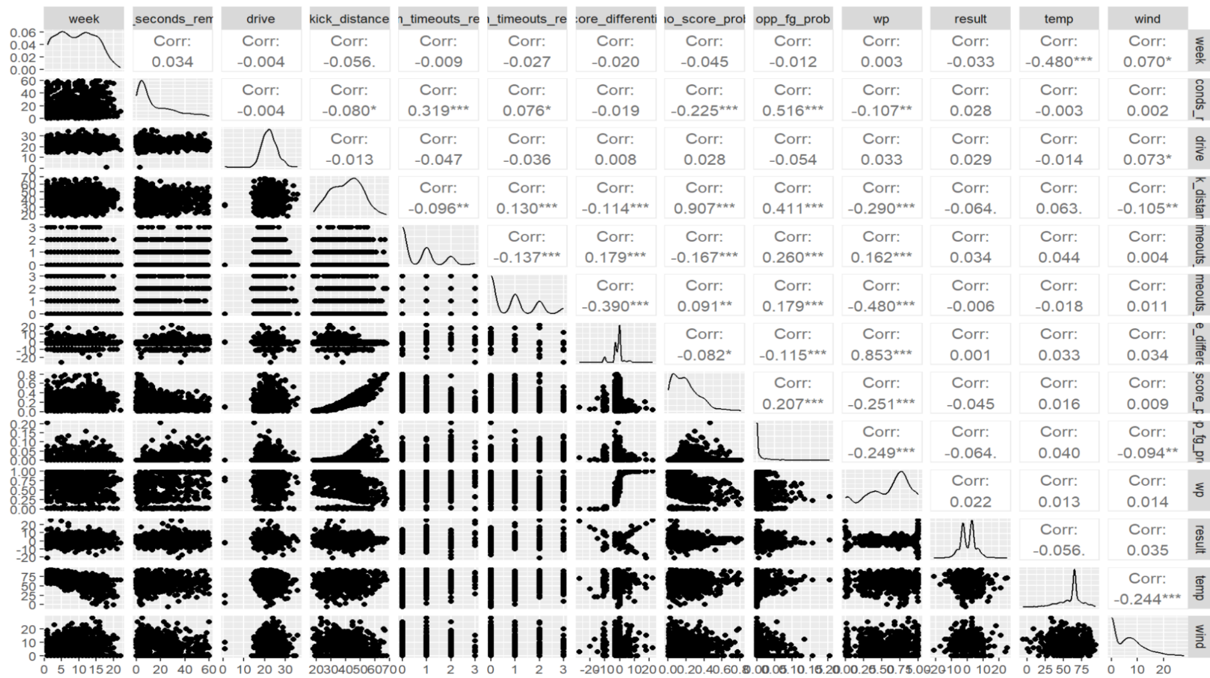

Figure C1: Correlation of Quantitative Variables in 1-Minute Data Subset

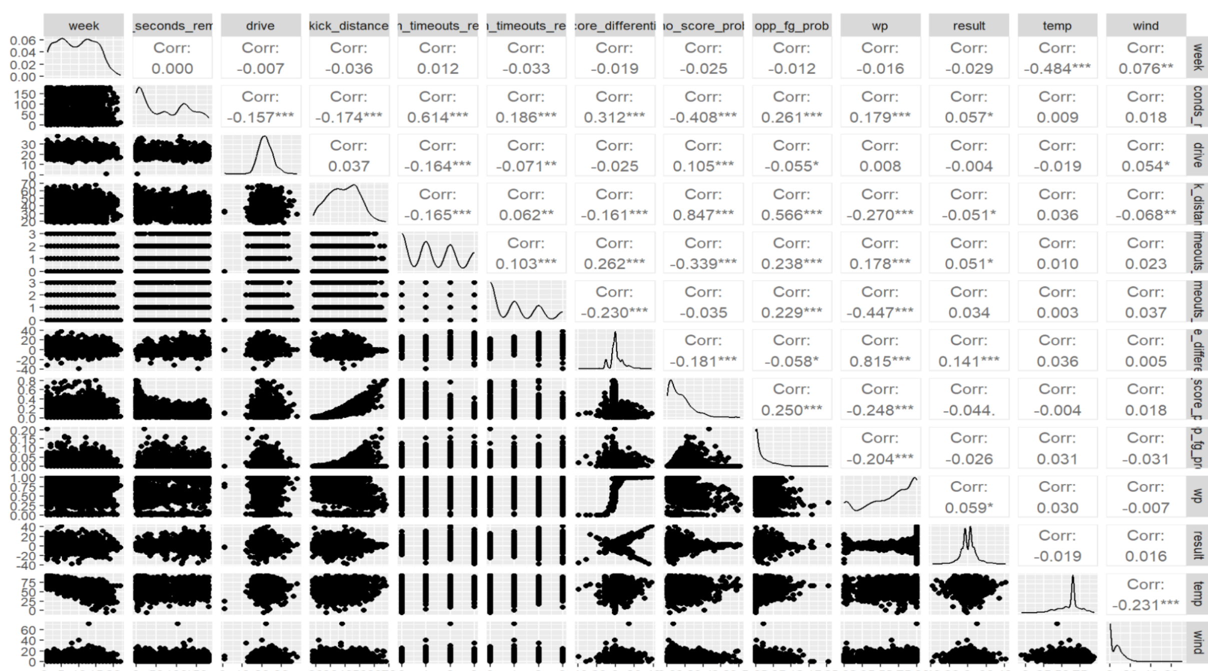

Figure C2: Correlation of Quantitative Variables in 3-Minute Data Subset

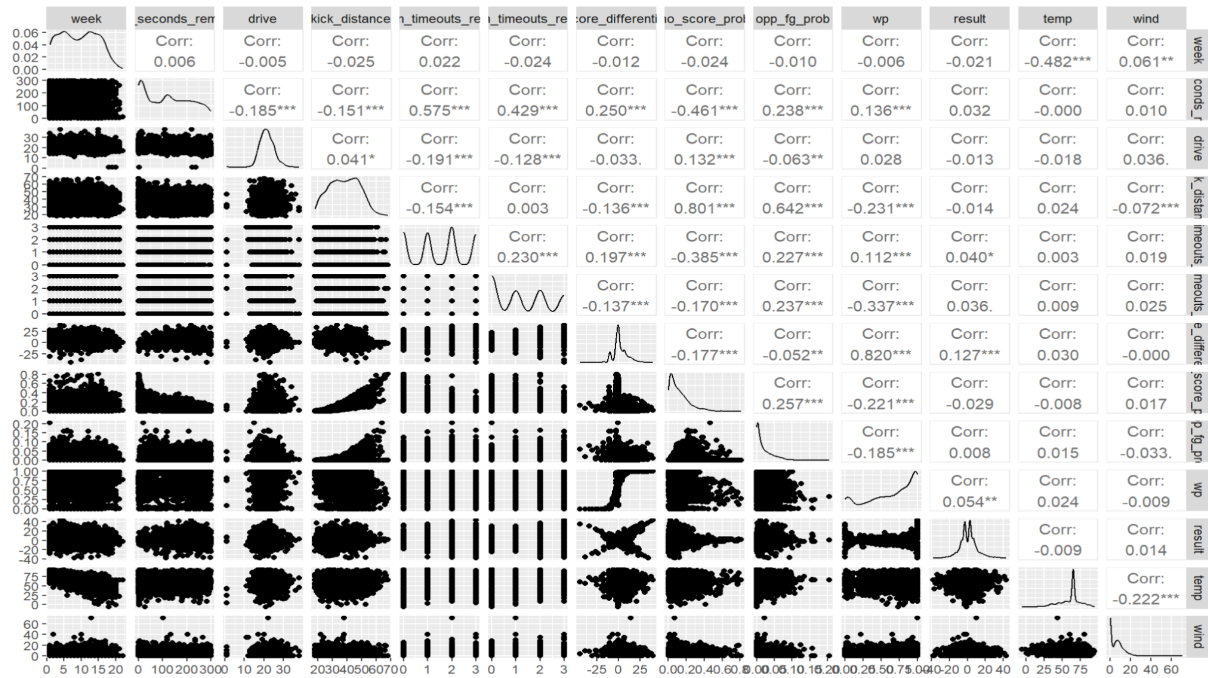

Figure C3: Correlation of Quantitative Variables in 5-Minute Data Subset