CSC 424 Final Project
FIFA 2017: What does player rating mean?

Jennifer Piane, Ji Qi, Vishnu Vardhan, Xiaochang Liu, Zheheng Mao

Abstract
Rating for Soccer players come from a variety of sources and are often debated. In this paper, our team identified a regression model to determine if Ratings can be predicted from other statistics. We support this model by applying Correspondence Analysis to assess the relationship between Rating and Freekick_Accuracy and applying Linear Discriminant Analysis to classify Rating into a defined range. Additionally, we applied Principal Component Analysis to determine the relationships between variables in our dataset.

Introduction
In soccer, player ratings are determined by a sports journalist's or game expert's assessment without agreement among experts or use a specific formula.These values can have an impact on players salaries and by extension club budgets, player and team popularity which extends to television ratings and advertising revenue and even Kit sales. If player ratings are meaningful, there should be some common influencers driving them. For this reason, our team decided to explore the factors that impact the value of ratings. We used the FIFA 2017 dataset on Kaggle.com, which was scraped from www.fifaindex.com and is used in Video Games like EASports FIFA17. The recent development Optical Player Tracking software, software that uses video analysis to track player position on the field over time and calculate metrics, for Soccer matches has made a wide range of new statistics available for the sport. Our team's goal was to find out what factors into expert player ratings and determine the relationships between other statistics and answer the research question: Can player ratings given by journalists and experts be predicted from other statistics? Our team employed Principal Component Analysis, Linear Discriminant Analysis, Correspondence Analysis, Regression and other methods to answer this question.

Related Work
Arndt & Brefeld evaluated data from five seasons of the Bundesliga club in the German Football League System with the goal of predicting the performance of individual players in coming soccer matches in terms of post-match player ratings as provided by sports journalists using Ridge Regression and Support Vector Regression (SVR). The feature selection method, Recursive Feature Elimination (RFE), was chosen, in part, because it written to avoid taking a greedy approach like Forward or Backward Selection. In this implementation of RFE, features that had been removed may be added back into the model in later iterations. (Arndt & Brefeld, 2016) Features included metrics about the player and the opponent in the coming match. Arndt & Brefeld assess their model quality was assessed on Mean Absolute Error and found that Support Vector Regression produces the lowest value. (Arndt & Brefeld, 2016)

Louzada & Ara built an expert system, iSport, for identifying soccer talent in up and coming athletes. iSport used 6 tests to assess athletic ability and technical skill; the *Mor and Christian pass test* that simulates shooting into a goal, the *five cone dribbling test*, the *kick after pass test*, the *1000m on a track test* which is an indirect measure of oxygen usage, the *cyclic speed of 20m test* which measures the time to complete a cyclic course, and the anaerobic power test (RAST) which calculates the athlete's Relative Power during an interval course. (Louzada & Ara, 2016). The system evaluated these features by creating 3 new metrics, *Physical Score*, *Technical Score* and *General Score,* from evaluation of the 6 tests with PCA and Factor Analysis using the Varimax Rotation. Physical Score and *Technical Score* were the results of PCA using the correlation matrix of the variables from the 3 (respective) relevant tests. General Score used Factor Analysis of the Principal Components of all 6 variables, again from the correlation matrix, applying the varimax rotation. The first factor consisted of the 3 physical tests and the *five cone dribbling test*, while the two remaining variables fell into the second factor. These two factors explained 67% of the variance. (Louzada & Ara, 2016)

Abdullah et al analyzed performance of players on two teams in the Malaysian Super League to identify other players who are performing well, using Performance Parameters recorded during games by human annotators. A range of variables were recorded that included metrics like *shots*, *headings* and *fouls* but also metrics like *chasing loose balls* and *distribution of passes.* They applied PCA with the Varimax rotation on standardized variables for purpose reducing dimensionality. In their model, Abdullah et al used the Kaiser Criterion to determine which components to keep, those with eigenvalues > 1 settling on 8 components that explained ~82% of the variance. (Abdullah et al, 2016) In a following study, Abdullah et al collected anthropometric and fitness data which included sitting and standing height, body fat measurements, upper arm circumference, VO max during 20m shuttle run and 30m linear sprint speed. Previously defined tests like the 505 agility test, Vertec testing device for vertical jump height, dribbling under time pressure and others for total of 26 variables from both novice and elite players. Abdullah et al used PCA with the Varimax rotation to find the relationships between variables on Z-scale standardized variables, selecting 7 components to explain ~70% of the variance. (Abdullah et al, 2017)

Lago-Peñas et al used data from 380 games in the 2008-2009 Season of the Spanish Soccer League to predict Wins, Losses and Draws with Linear Discriminant Analysis from a set of 15 variables that included shots, assists, corners, crosses, fouls committed and received, venue and others. Lago-Peñas et al were able to identify the variables more relevant to prediction and correctly classify matches as Wins, Losses and Draws ~55% of the time. (Lago-Peñas, 2010)

Exploratory Analysis of Data

Data

For this project, the FIFA 2017 dataset on Kaggle.com, which was scraped from www.fifaindex.com was used. This data is used in Video Games like EASports FIFA17.
Dataset: https://www.kaggle.com/artimous/complete-fifa-2017-player-dataset-global

This dataset contains both metric and categorical variables, listed in Appendix A, Table 1, the Independent Variables table. Appendix A, Table 2 lists all variables and their data types and ranges. The variable rating, will be our dependent variable. This is not the typical 1-10 rating that journalists would assign a player, but, appears to be the result similar expert judgements. This variable will serve as the “Expert Rating” in our research question.

Data Cleaning

Some variables were not useful, such as the Kit variables, which were images of uniforms. Other variables were redundant, such as Age and Date of Birth (in this case Age was usable). Contract_Expiry was interesting, but, more useful when formatted as a metric variable than as a date and was therefore converted to No_of_Years_for_Contract_Expiry. In order to use categorical variables in the regression model, Dummy Variables were created for Work_Rate and Preferred_Foot. The height and weight variables were all in the same units in the original dataset, but, the units were included and had to be stripped out. Missing values were handled by setting the value to the mean.

Table 3 in Appendix A lists the variables in final, cleaned dataset used in our analysis.

PCA/FA

PCA was evaluated on this dataset in terms of identifying underlying relationships among attributes. Scaling was required because there are variables assess skills on scale of 1-100 and variables like Age, Height & Weight that are in different units and on different scales.

Varimax and Promax Rotations

Two versions of PCA were applied comparing different types of rotations. When comparing rotations, it was helpful to consider an orthogonal (Varimax) and an oblique (Promax) rotation. Since our data is highly correlated, the oblique rotation is may be better suited than the orthogonal rotation. In fact, the Promax rotation was helpful in making components more visible. With the Varimax rotation, some of variables that made up the components had noticeably higher magnitudes than others. Running the Promax rotation on the same data, those high magnitude variables would often be the only variables that make up that component. Figure 5 shows the fourth component with both the Varimax and Promax rotations. It’s easy to see how much better the relevant variables stand out in Promax.

PCA with Varimax

Figure 3 shows the scree plot used to identify the components selected for PCA with the Varimax rotation. The knee was identified at the 5th component with ~78% of the variance explained. The resulting components selected were Performance, Defence Performance, Physical Performance, Reaction, Explosiveness.

PCA with Promax

Figure 7 shows the output of the PCA method in R and Figure 4 shows the scree plot. The components were considered in terms of both the scree plot and the amount of total variance.

Achieving a cumulative variance of 95% requires using 19 principal components. This is an improvement over 40 variables, but, further reduction in dimensionality is preferred. A cumulative variance of 86% only requires 8 principal components. These 8 components identified are described in Table 4, they are Offence, Defence, Strength, Yoga Class, Jumping, Contract_Expiry, Weak_foot and Motion.

Factor Analysis

Using Factor Analysis to obtain 4 Factors is expected to eliminate the components with a single variable. In the factors created are described in Table 5 in Appendix A. The new factors include interesting changes. The most notable change is that while PC3 has contained only the strength variable, Factor3 contains stats relevant to Goalkeeping, including Reactions and Jumping. The downside is that the output, in Figure 5 shows we've only explained 77% of the variance.

CA

We have several categorical variables such as Preffered_Foot, Preffered_Position & Work Rate of players in the data set. As player abilities vary based on position they play in or work rate category of individual player, we have decided to use these categorical variables for grouping observations rather than as categories or ordinals. Example, comparison of two players who play at positions "Center Forward" & "Second Striker" is not fair and bound to give wrong results. So we grouped observation based on "Preffered_Position" and performed further analysis.

We have chosen two features of data, **Rating & Freekick_Accuracy** of player to perform Correspondence analysis due to their correlation with each other which was observed while performing Linear regression. In order to perform Correspondence analysis, we have divided Overall Rating of player & Free kick Accuracy of all observation into 4 levels (as they are metric variables) as Minimum to 1st Quartile, 1st Quartile to Median, Median to 3rd Quartile & 3rd Quartile to Maximum value and these levels have been labelled and value of each observation has been mapped with one of labels to be used for classification & prediction.
Ordinal level labels generated by mapping **Rating & Freekick_Accuracy** with each observation have been made as new data frame and converted into table format for contingency values, Figure 8 in Appendix A.

Thus derived table of Rating level & Freekick_Accuracy levels has been observed and called "ca" function to generate correspondence analysis and summary of analysis is as in Figure 9a & 9b in Appendix A. Results of correspondence analysis were well interpreted and found that Freekick_Accuracy of the players has strong & positive association with Rating of the players. Rating levels or labels of players grouped with same level or labels of Freekick_Accuracy showing the correlation.

LDA

As we have chosen Rating of Individual player as dependent variable. We wanted to see how new observations will be classified to overall rating based other metric variables of the player and also to visualize the rating trend across various levels. LDA technique has been used to classify different levels of ratings and measure accuracy of prediction of rating for new player or observation.

In order to apply LDA technique over Rating variable, we have converted Rating metric variable into 4 ordinal levels of labels such as Minimum to 1st Quartile as label “45-66”, 1st Quartile to 3rd Quartile as “67-71” & 3rd Quartile to Maximum as “72-94”. We grouped observations based on “Preffered_Position”, removed Rating metric variable from individual observations.

We have randomly chosen 100 observations to apply LDA technique on the filtered data set and results of LDA technique have been interpreted through visualizations.

LDA learning function plot showed Players with ratings “45 to 66” were concentrated in one specific area of plot and same with Player ratings at levels of “67 to 71” & “72 to 94”. Ordinal level plotting of LDA function has less overlapping of individual groups showing good accuracy in classifying new observations.

Post visual analysis of LDA output, we have randomly sampled 100 instances of Train data and 100 instances of Test data to perform cross validation. And these train and test instances have been passed to LDA function to classify Rating labels. Output of LDA classified labels have been compared to actual labels of observations and the results are in Appendix A, Figure 10a & 10b.

Train data had accuracy of 89% in classifying observations to Rating labels. Test data had accurately classified 78% of observation to their actual Rating category or level. Also cross verified findings of application of LDA technique by building Hierarchical cluster analysis. Results from CA had shown same findings as that of LDA technique.

Model Selection

Assumptions

Normality was verified using histogram plots. When examining variables for normality, we didn’t find any variables that both needed and could be transformed without a negative side effect. Figure 11, in Appendix A includes some example histogram plots of independent variables.

Correlations

Correlation matrix, in Appendix A, Figure 1, shows that most of independent variable are relatively related with dependent variable. Looking at both the correlation matrix and Figure 2, which shows a sample of the Pearson Correlation values for Rating and some important

variables, we can find some evidence of Multicollinearity was an issue, specifically high correlation values among variables the Defence variables. This is supports the our PCA results.

Selection

The model regression model was fitted using Backward Selection, choosing the variable that causes the least increase In SSR. This method removed two variables: Long_Pass and Volleys; the final model had a P-value of whole model(goodness of fit) that was significant , F-value =1776.31. R-square =0.7522.Each variable in the model are significant at the 0.1 level.

Final equation

```
Rating =
      21.04189+Age*0.10988+Ball_Control*0.09101+Dribbling*-
      0.01189+Aggression*-
      0.01111+Reactions*0.49125+Attacking_Position*-
      0.08666+Interceptions*-
      0.02375+Vision*0.04072+Composure*0.04827+Crossing*0.02215+Short_P
      ass*0.04168+Stamina*-0.00845+Strength*0.09194+Heading*-
      0.01188+Shot_Power*0.01993+Finishing*0.01653+Long_Shots*-
      0.01287+Curve*0.01556+Freekick_Accuracy*-0.01365+Penalties*-
      0.02518+Preffered_FootLeft*0.29657
```

Validation

A random sampling method was used create total numbers of observation are 17588. 12312 of them are used as training data, 5276 are as testing data. The result of validation test has shown that the RMSE=3.56 and the MAE=2.080, the accuracy of the testing data is 86.2%.

Residuals

Residuals were analyzed by checking residuals vs. each independent variables and residuals vs. predicted values. The plot of the studentized residuals, appendix A, figure 12, was linear. The plots of residuals of individual variables show that with a few exceptions they ranged between +3 and -3 and evenly dispersed. This is can indication that our Model is appropriate for the data.

Analysis of Results

Conclusions

In our analysis, we were able to find a regression equation that would predict Rating. We were able to use LDA to predict rating when divided into categories. We were able to find correspondence between Freekick_Accuracy and Rating. Finally, we were able to find underlying relationships and build relevant PC components. The data suggests that Rating values, even though they are a judgement call, have other variables influencing them.

Limitations

There were some limitations to the models developed based on this dataset. There are male players and female players, but no attribute of gender. Player ratings assigned based on a

player's success in their respective position. In our analysis, we didn't include the influence of positions. This will lead to low accuracy for predicting. Some team members are short of soccer domain knowledge. Improving domain knowledge could lead to stronger models. The Rating used in the dataset was created from an unknown source. This model still needs to be applied to Ratings scraped from sports journals to best address the research questions..

Future Work

This model gives us a good starting point for assessing the predictors of Rating. Going forward, we can further study Rating based on factors related to Position and build custom models for each position. Our original dataset included both, National_Position and Club_Position, which can be used to split the dataset into more focused subsets. We can use Principal Component Regression (PCR), integrating the components identified into a regression model. We can also add additional Rating dependent variables from a variety sources, identifying the predictors for based on publication or journalist.

REFERENCES

1. Abdullah, M. R., Maliki, A. B., Musa, R. M., Kosni, N. A., Juahir, H., & Mohamed, S. B. (2017). Identification and Comparative Analysis of Essential Performance Indicators in Two Levels of Soccer Expertise. *International Journal on Advanced Science, Engineering and Information Technology, 7*(1), 305. doi:10.18517/ijaseit.7.1.1150
2. Abdullah, M. R., Musa, R. M., Maliki, A. B., Kosni, N. A., & Aziz, M. A. (2016). The application of Principal Component Analysis to identify essential performance parameters in outfield soccer players. *Research Journal of Applied Sciences,* 1199-1205.
3. Arndt, C., & Brefeld, U. (2016). Predicting the future performance of soccer players. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 9*(5), 373-382. doi:10.1002/sam.11321
4. Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010, June). Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League. Retrieved November 05, 2017, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761743/
5. Louzada, F., Maiorano, A. C., & Ara, A. (2016). ISports: A web-oriented expert system for talent identification in soccer. *Expert Systems with Applications, 44*, 400-412. doi:10.1016/j.eswa.2015.09.007
6. Player Stats Database - FIFA 18 - FIFA Index. (n.d.). Retrieved November 04, 2017, from https://www.fifaindex.com/
7. Agarwal, S. (2017, April 13). Retrieved November 04, 2017, from https://www.kaggle.com/artimous/complete-fifa-2017-player-dataset-global

Appendices

Appendix A - Figures and Tables

	Category	Metric Variables	Categorical Variables	Date	Not Used
1	Basic Attributes	Height, Weight, Age	National_Position, Club_Position, Preferred_Foot, Preffered_Position, Work_Rate	Club_Joining, Contract_Expiry, Birth_Date	Name, Nationality, Club_Kit, National_Kit, Club
2	Ball Skill	Ball_Control, Dribbling, Curve, Skill_Moves			
3	Defence	Marking, Sliding_Tackle, Standing_Tackle, Heading, Interceptions			
4	Metal	Aggression, Vision, Attacking_Position, Composure			
5	Physical	Acceleration, Speed, Stamina, Strength, Balance, Agility, Jumping, Reactions			
6	Passing	Crossing, Short_Pass, Long_Pass			
7	Shooting	Shot_Power, Finishing, Long_Shots, Freekick_Accuracy, Penalties, Volleys			
8	Goalkeeping	GK_Positioning, GK_Diving, GK_Kicking, GK_Handling, GK_Reflexes			

Table 1: Independent Variables in original data set

Variable	Data Type	Range
Name	String	
Nationality	String	
National_Position	String	Category
National_Kit	URL to Image	
Club	String	
Club_Position	String	Category
Club_Kit	URL to Image	
Club_Joining	Date	
Contract_Expiry	Date	
Rating	Metric	0-100
Height	Metric	Unbounded
Weight	Metric	Unbounded
Preffered_Foot	String	Category
Birth_Date	Date	
Age	Metric	Unbounded
Preffered_Position	String	Category
Work_Rate	String	Category
Weak_foot	Metric	1-5
Skill_Moves	Metric	1-5
Ball_Control	Metric	0-100
Dribbling	Metric	0-100
Marking	Metric	0-100
Sliding_Tackle	Metric	0-100
Standing_Tackle	Metric	0-100
Aggression	Metric	0-100
Reactions	Metric	0-100
Attacking_Position	Metric	0-100
Interceptions	Metric	0-100
Vision	Metric	0-100
Composure	Metric	0-100
Crossing	Metric	0-100

Short_Pass	Metric	0-100
Long_Pass	Metric	0-100
Acceleration	Metric	0-100
Speed	Metric	0-100
Stamina	Metric	0-100
Strength	Metric	0-100
Balance	Metric	0-100
Agility	Metric	0-100
Jumping	Metric	0-100
Heading	Metric	0-100
Shot_Power	Metric	0-100
Finishing	Metric	0-100
Long_Shots	Metric	0-100
Curve	Metric	0-100
Freekick_Accuracy	Metric	0-100
Penalties	Metric	0-100
Volleys	Metric	0-100
GK_Positioning	Metric	0-100
GK_Diving	Metric	0-100
GK_Kicking	Metric	0-100
GK_Handling	Metric	0-100
GK_Reflexes	Metric	0-100

Table 2: All variables available

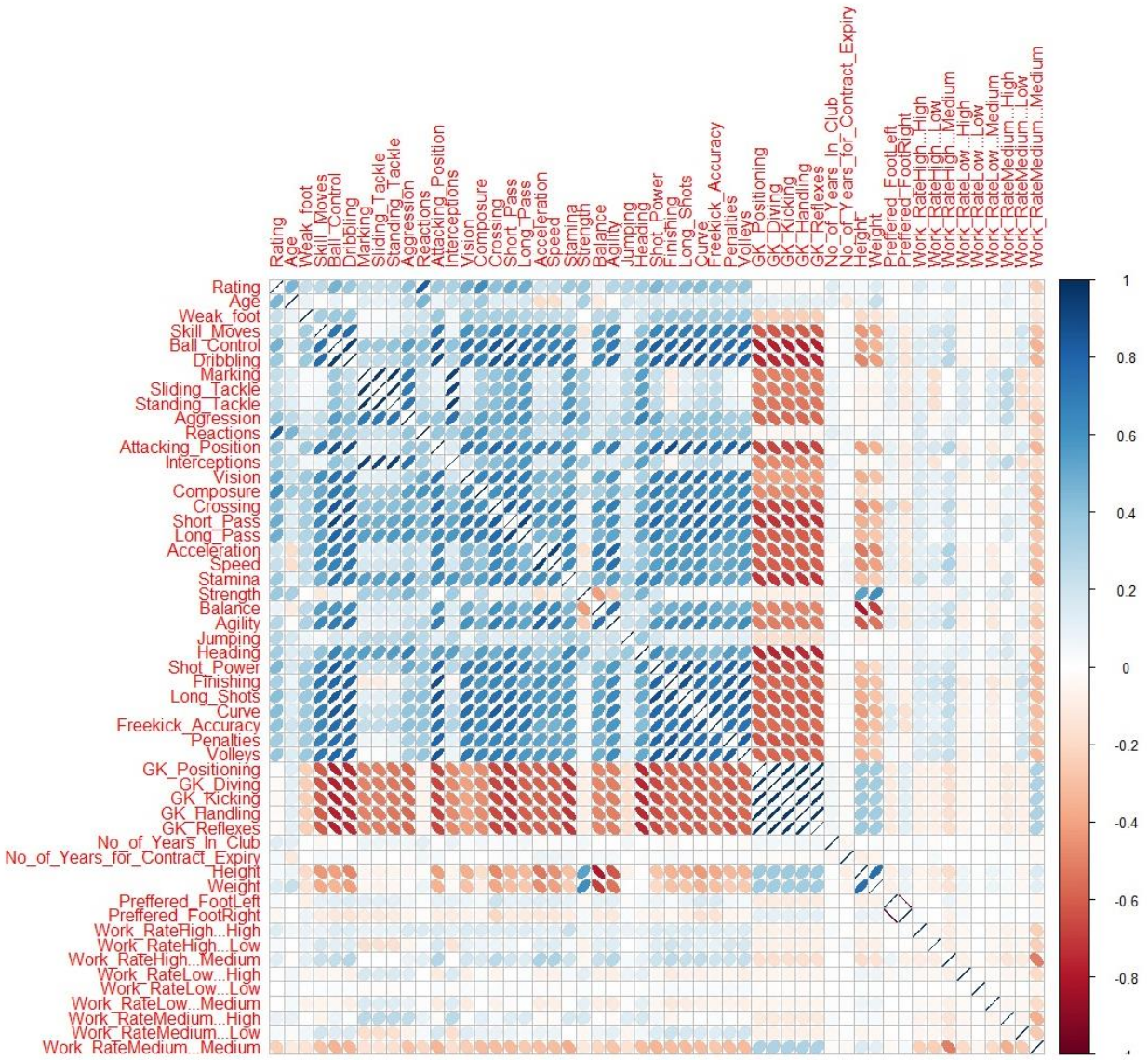

Figure 1: Correlation matrix.

Correlations

		Rating	Age	Weak_foot	Skill_Moves	Ball_Control	Dribbling	Marking
Pearson Correlation	Rating	1.000	.458	.226	.252	.463	.369	.237
	Age	.458	1.000	.086	-.016	.083	.005	.131
	Weak_foot	.226	.086	1.000	.337	.367	.363	.027
	Skill_Moves	.252	-.016	.337	1.000	.727	.763	.033
	Ball_Control	.463	.083	.367	.727	1.000	.931	.355
	Dribbling	.369	.005	.363	.763	.931	1.000	.228
	Marking	.237	.131	.027	.033	.355	.228	1.000
	Sliding_Tackle	.215	.097	.026	.043	.357	.243	.960
	Standing_Tackle	.249	.117	.044	.071	.392	.270	.960

Figure 2: Pearson Correlation between Rating and some key variables

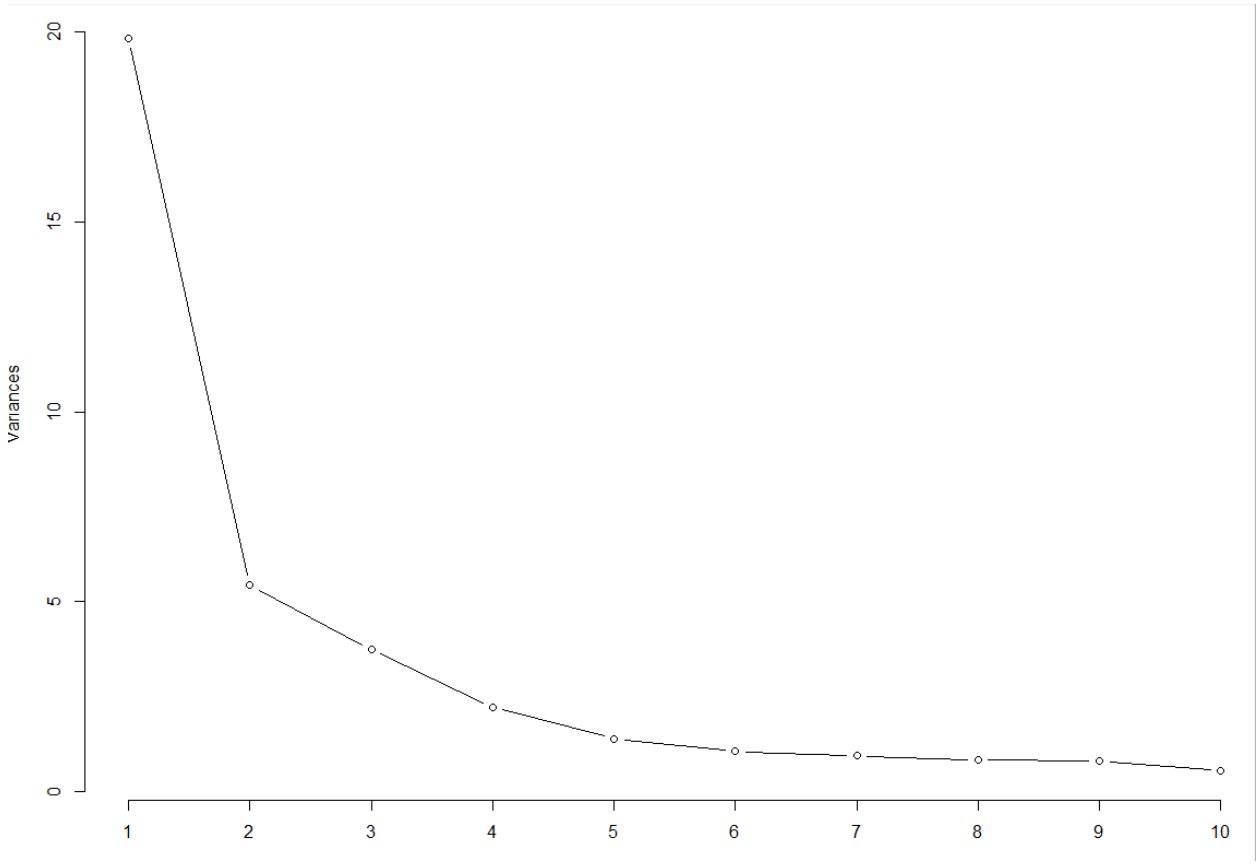

Figure 3: PCA with Varimax Rotation Scree Plot

Rating	Shot_Power
Age	Finishing
Weak_foot	Long_Shots
Skill_moves	Curve
Ball_Control	Freekick_Accuracy
Dribbling	Penalties
Marking	Volleys
Sliding_Tackle	GK_Positioning
Standing_Tackle	GK_Diving
Aggression	GK_Kicking
Reactions	GK_Heading
Attacking_Position	GK_Reflexes
Interceptions	No_of_Years_for_Contract_Expiry
Vision	Height
Composure	Weight
Crossing	Preferred_FootLeft
Short_Pass	Preferred_FootRight
Long_Pass	Work_RateHigh...High
Acceleration	Work_RateHigh...Medium
Speed	Work_RateHigh...Low
Stamina	Work_RateMedium...High
Strength	Work_RateMedium...Medium
Balance	Work_RateMedium...Low
Agility	Work_RateLow...High
Jumping	Work_RateLow...Medium
Heading	Work_RateLow...Low

Table 3: Final variables in the clean data set.

	Label	Variables
PC1	Offence	Long_shots, Volleys, Curve, Finishing, Shot_Power, Freekick_Accuracy, Penalties, Vision, Attacking_Position, Dribbling, Ball_Control
PC2	Defence	Interceptions, Standing_Tackle, Sliding_Tackle, Marking, Aggression
PC3	Strength	
PC4	Yoga Class	Height, Weight, Strength, Balance
PC5	Jumping	
PC6	Contract Expiry	
PC7	Week Foot	
PC8	Motion	Speed, Acceleration

Table 4: PCA with Promax Rotation Components

	Name	Variables
Factor1	Offence	Long_shots, Volleys, Curve, Finishing, Shot_Power, Freekick_Accuracy, Penalties, Vision, Attacking_Position, Dribbling, Ball_Control, Composure, Reactions, Skill_Moves
Factor2	Defence	Interceptions, Standing_Tackle, Sliding_Tackle, Marking, Aggression
Factor3	Goalkeeping	GK_Reflexes, GK_Handling, GK_Kicking, GK_Diving, GK_Positioning, Jumping, Reactions
Factor4	Yoga Class	Height, Weight, Strength, Balance, Agility

Table 5: Factor Analysis Factors

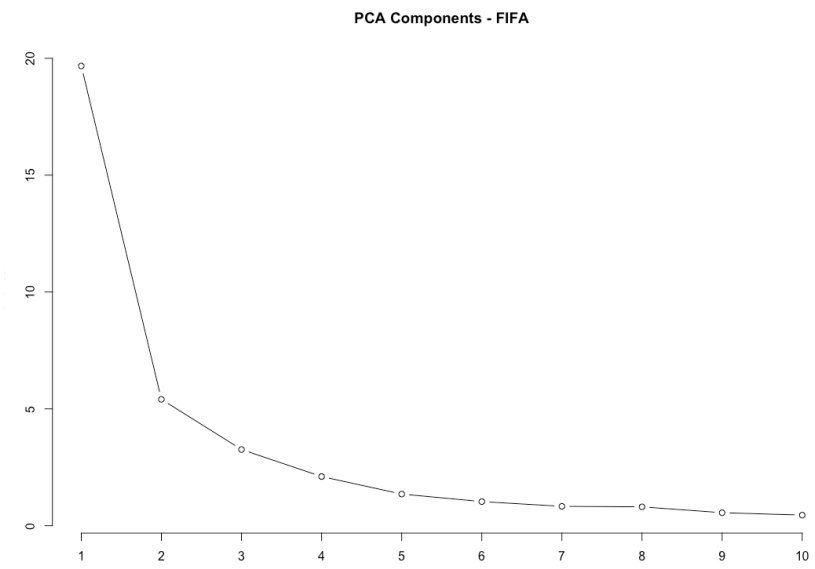

Figure 4: PCA with Promax Rotation Scree Plot

	Factor1	Factor2	Factor3	Factor4
SS loadings	15.051	7.068	4.685	4.029
Proportion Var	0.376	0.177	0.117	0.101
Cumulative Var	0.376	0.553	0.670	0.771

Figure 5: Summary of Factor Analysis Output

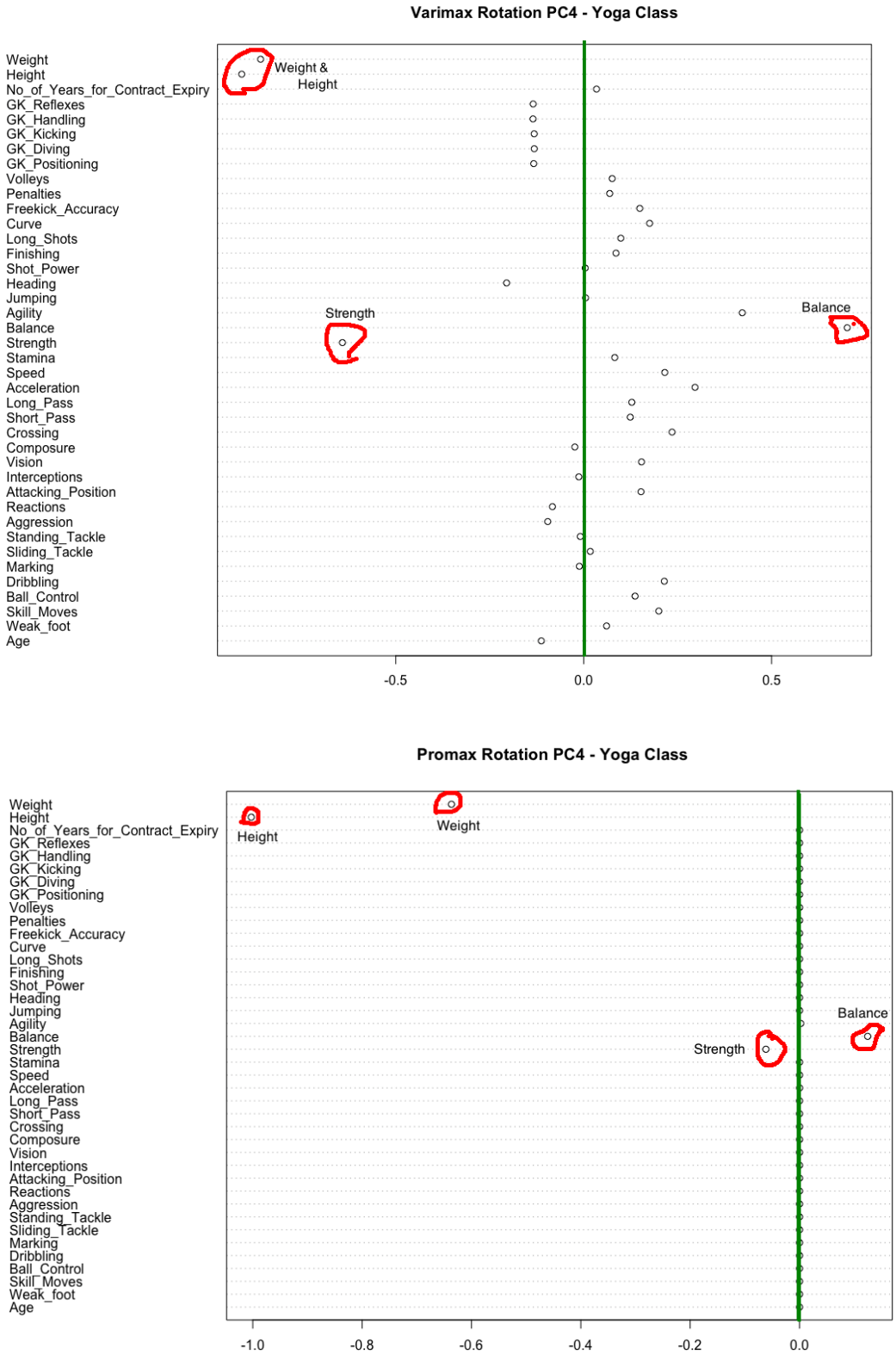

Figure 6: Varimax Rotation on PC4 (top) Promax Rotation on PC4 (bottom). No cutoff was used, Promax eliminated the noise.

```
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
Standard deviation     4.4352 2.3256 1.80585 1.45059 1.16364 1.01541 0.91039 0.89825 0.74593 0.67350
Proportion of Variance 0.4918 0.1352 0.08153 0.05261 0.03385 0.02578 0.02072 0.02017 0.01391 0.01134
Cumulative Proportion  0.4918 0.6270 0.70850 0.76111 0.79496 0.82074 0.84146 0.86163 0.87554 0.88688
                          PC11    PC12    PC13    PC14   PC15    PC16   PC17    PC18    PC19    PC20
Standard deviation     0.62466 0.59619 0.55373 0.54661 0.5177 0.50021 0.4939 0.47466 0.46321 0.45396
Proportion of Variance 0.00976 0.00889 0.00767 0.00747 0.0067 0.00626 0.0061 0.00563 0.00536 0.00515
Cumulative Proportion  0.89663 0.90552 0.91319 0.92066 0.9274 0.93361 0.9397 0.94534 0.95071 0.95586
                          PC21    PC22    PC23    PC24    PC25    PC26    PC27    PC28    PC29
Standard deviation     0.44484 0.42289 0.39386 0.37300 0.36738 0.36043 0.33339 0.30844 0.29978
Proportion of Variance 0.00495 0.00447 0.00388 0.00348 0.00337 0.00325 0.00278 0.00238 0.00225
Cumulative Proportion  0.96081 0.96528 0.96915 0.97263 0.97601 0.97925 0.98203 0.98441 0.98666
                          PC30    PC31    PC32    PC33    PC34   PC35    PC36    PC37   PC38    PC39
Standard deviation     0.28923 0.27114 0.26493 0.25368 0.21157 0.1997 0.19876 0.18238 0.1784 0.16553
Proportion of Variance 0.00209 0.00184 0.00175 0.00161 0.00112 0.0010 0.00099 0.00083 0.0008 0.00069
Cumulative Proportion  0.98875 0.99059 0.99234 0.99395 0.99507 0.9961 0.99705 0.99789 0.9987 0.99937
                          PC40
Standard deviation     0.15913
Proportion of Variance 0.00063
Cumulative Proportion  1.00000
```

Figure 7: Summary output of PCA on FIFA dataset with Promax Rotation

```
> CA1
           FAgroup
Ratinggroup 4-31 32-42 43-57 58-93
      45-62  578   223    42    20
      63-66  369   263   106    45
      67-71  379   281   194   121
      72-94  211   241   202   139
```

Figure 8: Contingency Table for Rating and Freekick_Accuracy

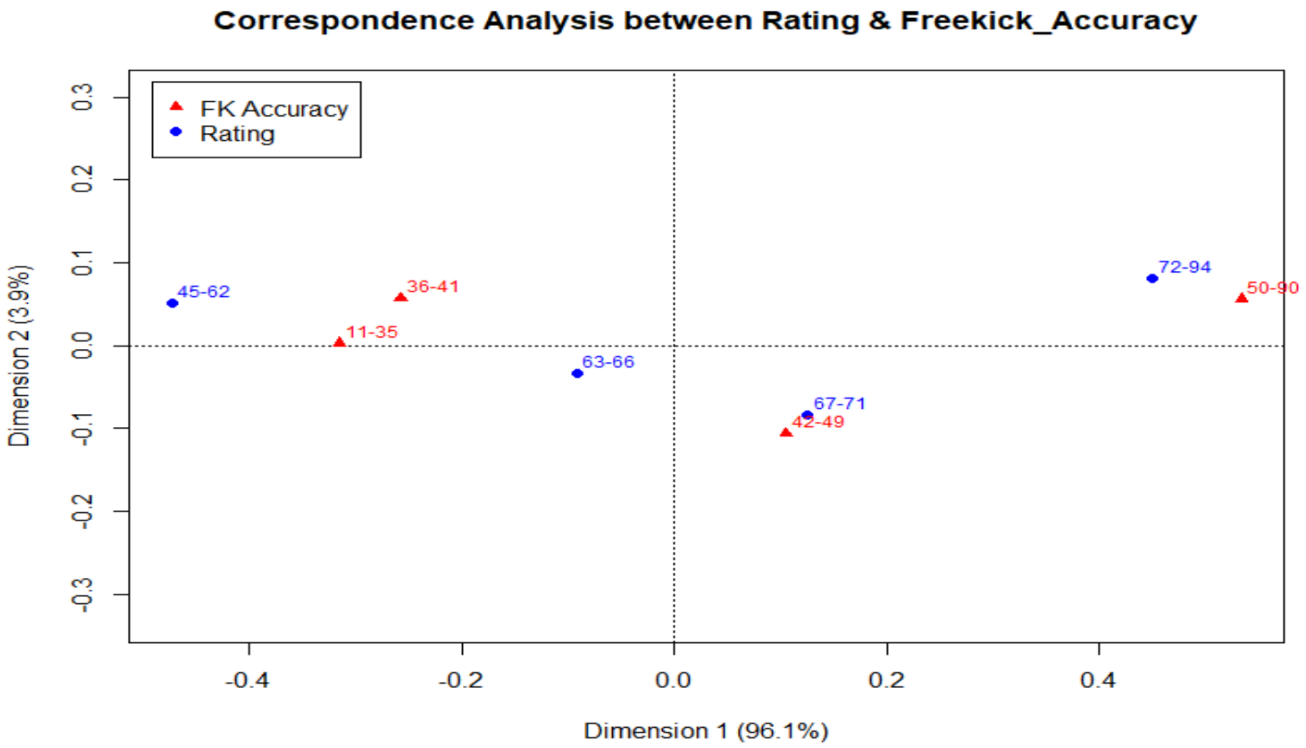

Figure 9a: Summary of Correspondence Analysis

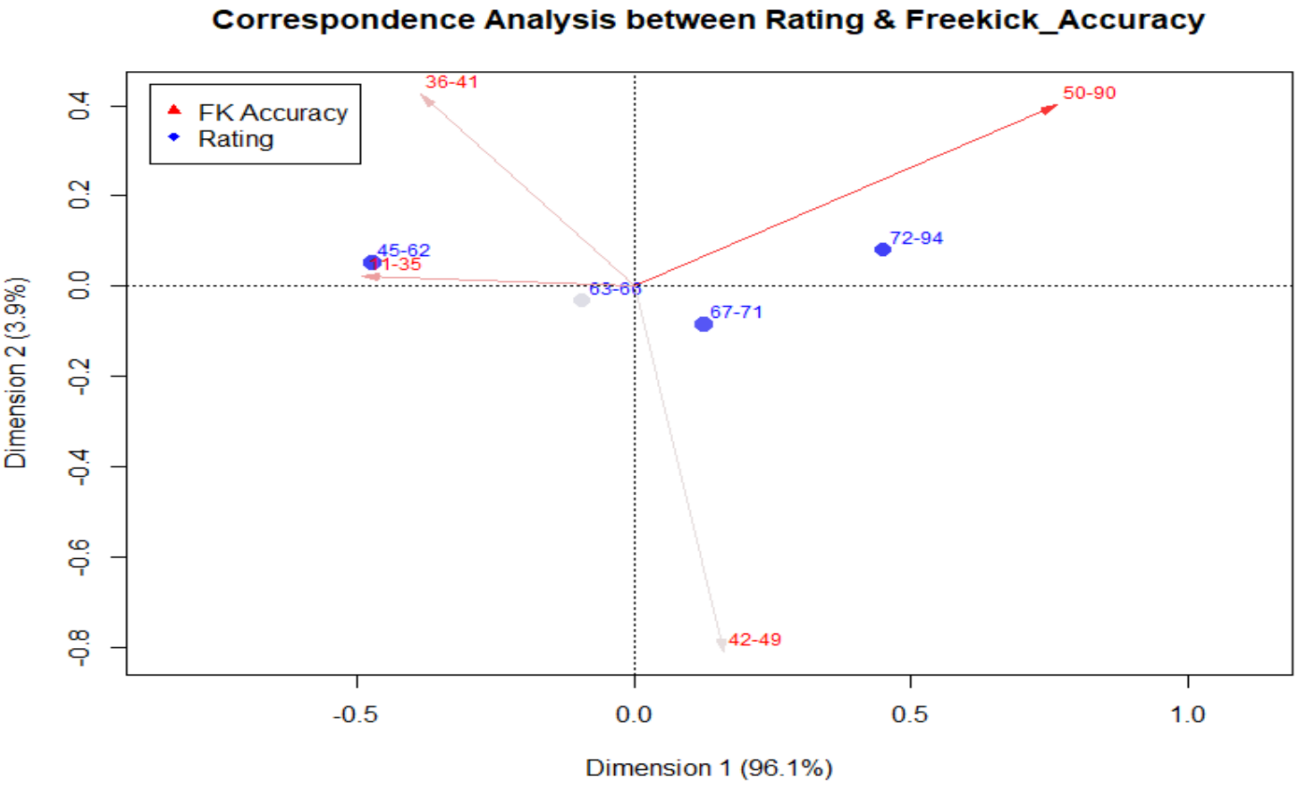

Figure 9b: Summary of Correspondence Analysis

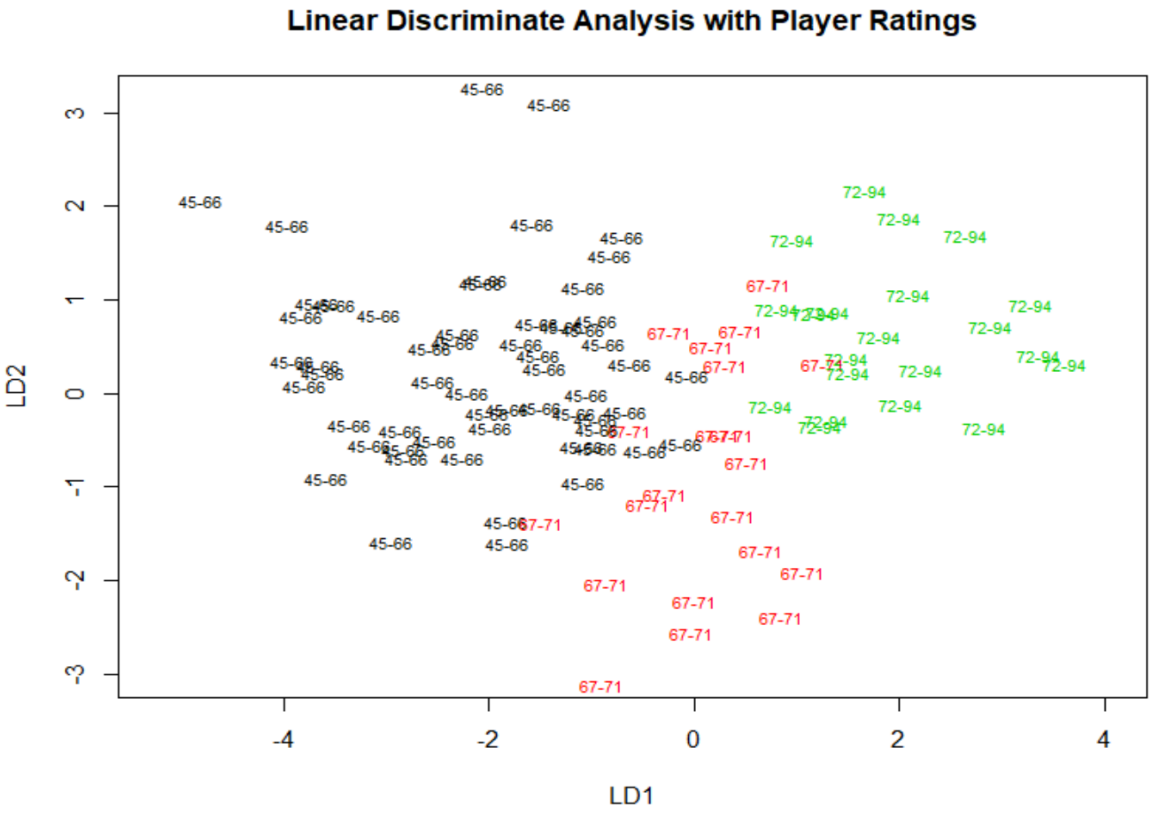

Figure 10a: LDA Plot.

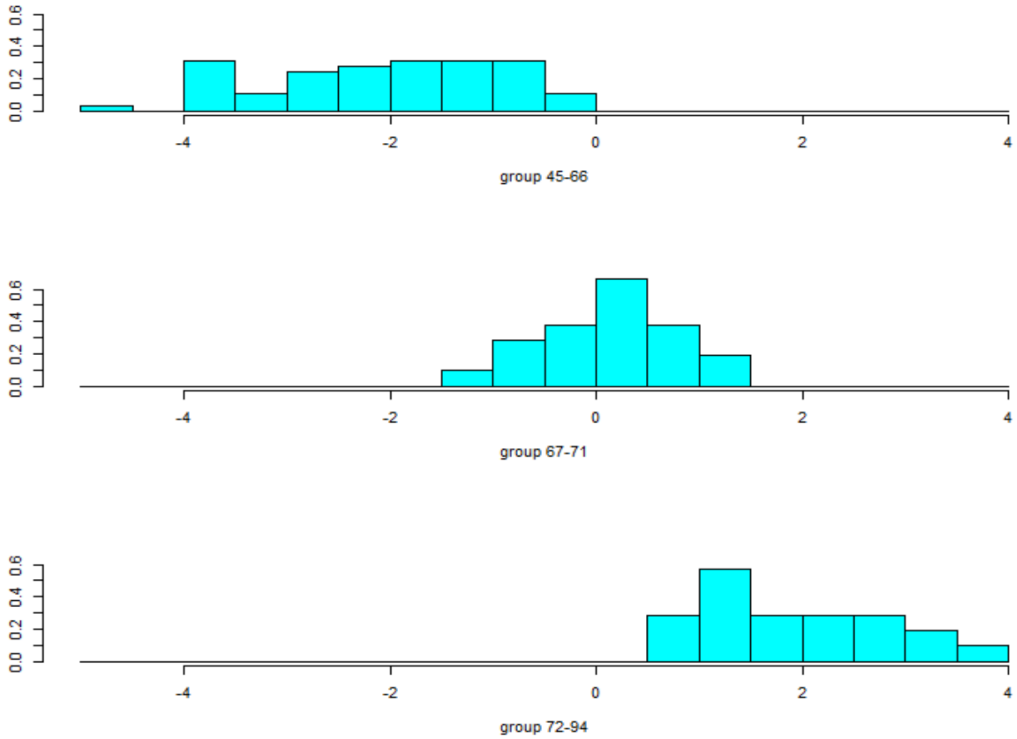

Figure 10b: LDA Class distinction Plot.

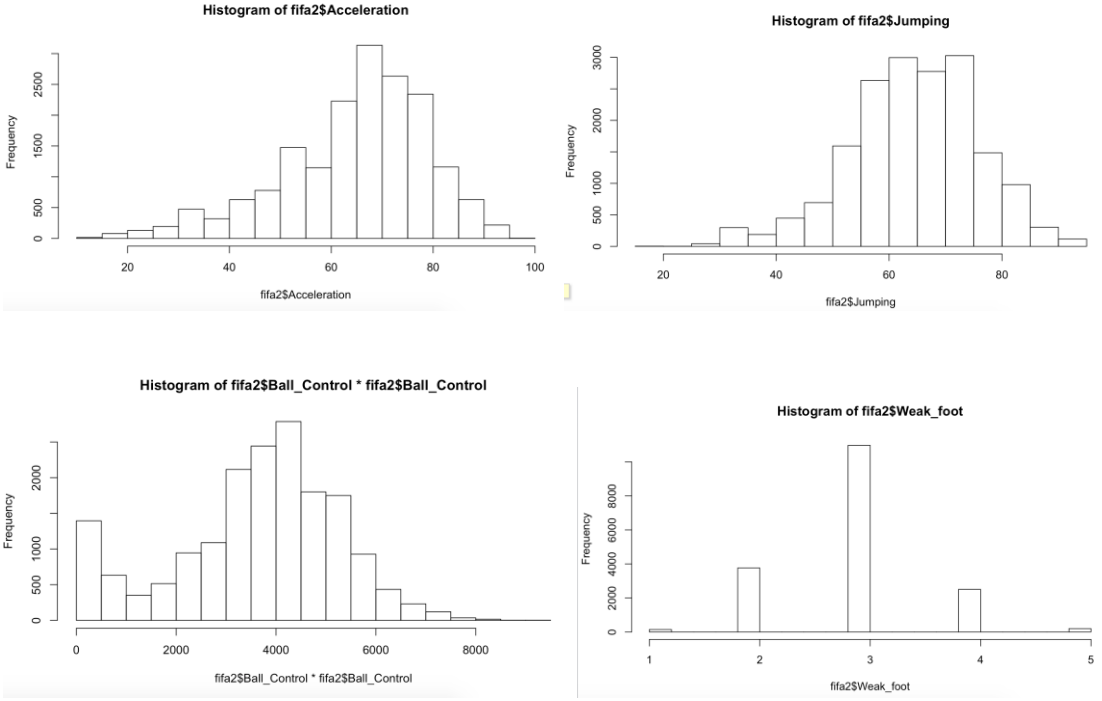

Figure 11: Examples of normal plots in our the cleaned dataset. Acceleration (top, left), Jumping (top, right), Ball_Control (bottom, left) Weak_foot(bottom, right).

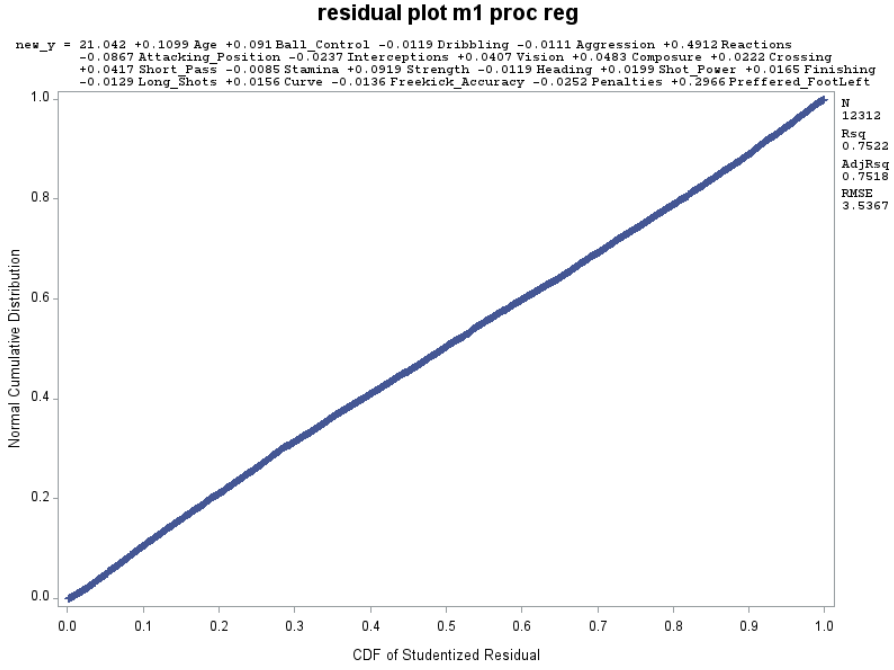

Figure 12: Residual analysis plot on regression model

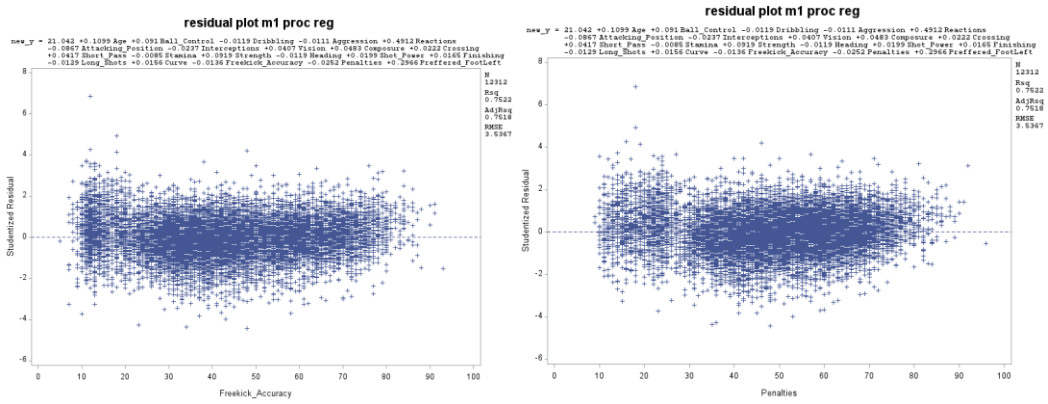

Figure 13: Residual plot for Freekick_accuracy (left) and Penalties (right)