Project Report

Determinants of Average Points per Game in the NBA

Team: Harsimran Aneja, Ajjan Inthiran, Harry Yao
Professor: Dr. Matthias Schonlau

Section 1: Introduction

The goal of every basketball player is to win the game by scoring more points. We aim to explore the factors behind a player's point-scoring efficacy through a regression model. This data set consists of real-world observations, which includes a variety of factors and specific circumstances. The large number of observations should offer a good view of the behaviour of the data, eliminating the weight of outliers and unrepresentative data points. The natural discrepancies and normalcy of the data serve as a strong foundation to demonstrate regression techniques.

This data set has not been widely analyzed in academic literature but serves as an interesting regression exercise. This analysis can be applied in a multitude of real-life situations. Particularly, NBA decision-making, talent acquisition, fantasy basketball, sports betting, and more.

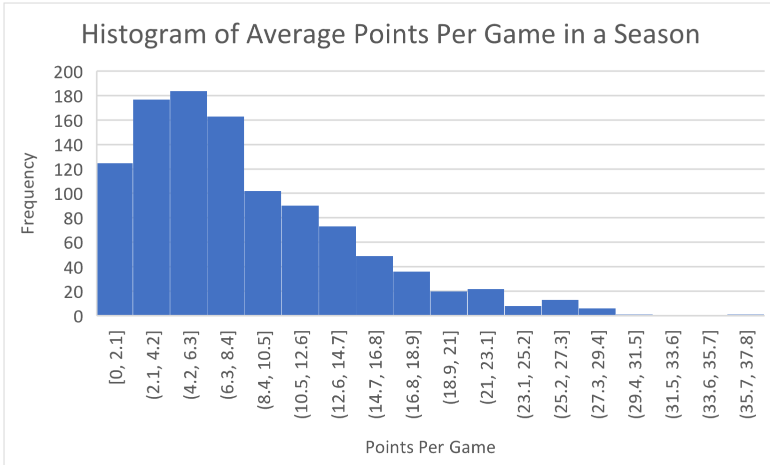
Section 1.b: X-Variables

We regress the response y-variable (average points scored per game) on the explanatory x-variables: age, height, weight, games played, assists, rebounds, draft year, draft round, draft number, true shooting percentage, and usage rate (total of 11 factors). These x-variables are attributes that describe the player over a season. We created indicator variables for draft round, draft number, and draft year, as they are categorical and binary; this results in 94 x-variables in the original regression model. We aim to discover how the average points per game (PPG) is associated with these variables.

The X-variables age, height, and weight represent physical attributes of a player. Games played is the total number of games a player played that season. Assists represent the times a player made a pass prior to a score. Rebounds represent a retrieval of the ball after a missed shot. True shooting percentage measures the player's efficacy when shooting the ball, the formula (Appendix B.1) considers free throws, and 2- and 3-point field goals. This x-variable is closely associated with y, as a player who makes a greater percentage of shots is likely to score more points in a game if he has sufficient chances to attempt shots. Usage rate measures the percentage of game plays a player was involved in during a game; the formula (Appendix B.2) accounts for the individual's and team's field goal attempts, free throw attempts, turnovers, and minutes played. The independent variables can impact the response variable in a variety of ways. For example, older players have more experience but may lose their edge as they age pass their athletic prime. These preliminary speculations will be explored by the regression analysis.

Section 1.c: The Data

The dataset contains observations from the 2017-18 and 2018-19 National Basketball Association (NBA) seasons. The data contains 1070 observations over the course of the two seasons, which is sufficient to illustrate the association the x-variables have to average PPG. Each observation in the data set corresponds to the mean PPG for a player over one season. This data set was from an online resource (link included in Works Cited). The data was originally obtained from the NBA statistics record. The league carefully records statistics of each game. In a sports league of this magnitude that spurs opportunities in sports betting, entertainment, fantasy leagues, and extensive network coverage, as well as performance being a major determinant of the players' salaries, the precision of the data is critical as it reflects the competence of the league. Due to the business of the NBA and the accuracy necessary for recording statistics, we believe in the quality of the data.



The histogram of the data represents frequency distribution for ranges of points. Each bar accounts for a range of 2.1 points. This shows that the observations are roughly normally distributed.
Note: additional scatter plots are included in Appendix D.

Age is not strongly correlated to average PPG (Appendix D.1). However, variation drops past age 35, likely due to the fact that players become more injury-prone/less athletic. At this point, teams value their veteran leadership and knowledge. Age will likely be insignificant in the regression model.

The true shooting percentage scatter plot demonstrates a positive correlation to average PPG (Appendix D.4). True shooting percentage is a measure of the player's efficacy when shooting the ball, so a higher value indicates a more efficient scorer. Thus, the scatter plot demonstrates that players with a high true shooting percentage are also high scorers (20+ PPG).

The usage rate scatter plot shows a strong positive correlation to average PPG (Appendix D.5), as usage rate indicates a player's involvement in games. The more involved the player is, the more opportunities he will get to score, which may explain why high usage rate players generally have higher PPG.

Section 2: Original Regression Model

2.a: Original Regression Output (can be found in Appendix C.1)
By analyzing the data, we gain an understanding of the impact of each of these x-variables on the response variable. The original regression model includes an extensive list of x-variables and a thorough and thorough data set. The initial analysis provides a comprehensive understanding of the response variable and which explanatory variables are most impactful.

The observations marked with asterisks represent significant values of the data (low p-values). A p-value close to zero indicates that the x-variable is more significant. The x-variables with higher significance include weight, rebounds, assists, games played, usage percentage, and true shooting percentage.

The output also shows many insignificant x-variables: some indicator variables for draft years and draft numbers. Logically, the efficacy of a player to score points are likely not too closely connected with the year they were drafted or the round in which they were drafted. If there is some perceived association, it is likely the individual circumstances of a particular year or a coincidence, not a significant element to the process over time. The ability of a player to score points is logically closer associated with his genetics (height, speed, strength), and training (skill, dexterity). Thus, it is no surprise that many of the indicator variables for draft year and draft round are not very significant. We expect these to be removed during model selection, as they do not contribute much to the regression's ability to explain the variation in y.

Near the top of the regression output (Appendix C.1) is the warning "2 not defined because of singularities". This is because these x-variables are linearly dependent to other x-variables, resulting in overdetermination of the model. We will remove these x-variables (draft numbers 9 and 60) for the final regression if they are not removed during model selection (they likely will). This is similar to the removal of a baseline category when creating indicator variables to remove dependence between variables without losing descriptiveness. VIF will be called again after model selection to analyze the remaining x-variables to search for additional collinearity.

From the output, we can conclude that the original regression model is quite good already. The R-squared value is 0.8374, meaning 83.74% of the variation in y is explained by the x-variables. That means the original model already does quite well to model the behaviour of the y-variable. We will expect this value to decrease slightly after model selection, as R-squared necessarily decreases with the number of x-variables. We are more interested in comparing the adjusted R-squared values between the original and the final regression models; if this metric increases in the final regression, we can be confident that the model selection yielded a satisfactory result.

2.b: <u>Diagnostic plots for original regression model</u> (can be found in Appendix D.7)
The QQ-plot output shows that the assumption of normality holds well for these data. The majority of the data points (in the middle) follow the dotted line, meaning it is close to linear with a slope of 1. The tails deviate from the dotted line, but that is expected of a real-world data set.

The Residuals vs. Fitted Values plot indicates non-constant variance, as it exhibits a clear funnel pattern. We will deal with this in the final interpretation section by using robust standard error.

The Studentized residuals are the same as standardized residuals (to unit variance), but are fitted ignoring the current observation for each point. This solves the problem that the variance of estimated residuals is not constant. From the lecture: "broadly speaking, absolute Studentized residual values over 3 are cause for concern, and absolute Studentized residual values over 4 are criminally high". In our original data, we see just 3 observations with Studentized residuals over 4, these will be removed in the final data set for discussion. Due to the size of our sample, we will allow the values between 3 and 4, as we are bound to see some amount of variation with such a large number of observations. These data points will be addressed, and removed if necessary, in the process of creating the final regression model.

The Leverage vs. Index plot plots the leverage of each observation by index (Appendix D.7). This plot identifies outliers in x-space. The plot has one point with noticeably high leverage compared to the rest of the observations. Upon further investigation, this point has leverage of 1. This player played just 2 games in the season (Appendix C.3), and did not score many points in either of them, making his x-variable values unusual in comparison to the other observations. This observation has a moderate studentized residual, so it will not affect the regression much. However, due to the fact that this player's career was barely existent in the NBA, he is not representative of the average player (or the regression), so this observation will be removed in the final regression model if the leverage of the observation is still high.

We will discuss the diagnostic plots in greater detail in the context of the final regression model, as that is more relevant for interpretation.

Section 3: Model Selection (using AIC)

3.a: Backward Selection
Since our original regression model had some x-variables that were heavily correlated (see VIF output in Appendix C.2), forward and backward selection strategies will yield different models. So, since backward selection was mentioned in the class forum to more likely yield the best result, we chose to use backward selection to choose our model.

We chose the Akaike's Information Criterion (AIC) for model selection. The AIC is defined with 2 summand terms. The first is the negative Gaussian log-likelihood up to a constant, and the second is a linear term in k (the number of variables in the sub-model) to penalize a model for having too many coefficients. Finally, we choose the model with the lowest AIC score. Since our original regression model had many x-variables, the AIC second term will help reduce the number of unnecessary variables.

3.b: Regression Output (regression and backward selection output can be found in Appendix C.4 & C.5)
Through backward selection, we are left with the final regression model with 15 x-variables.

We removed all of the draft number x-variables except draft number 1. Firstly, most of them were extremely insignificant. Secondly, the best players are selected in the first draft, so draft 1 is expected to have the greatest association with a player's ability to score points. We expect a player drafted in the first round to score a greater mean PPG than a player drafted in any of the subsequent rounds (baseline).

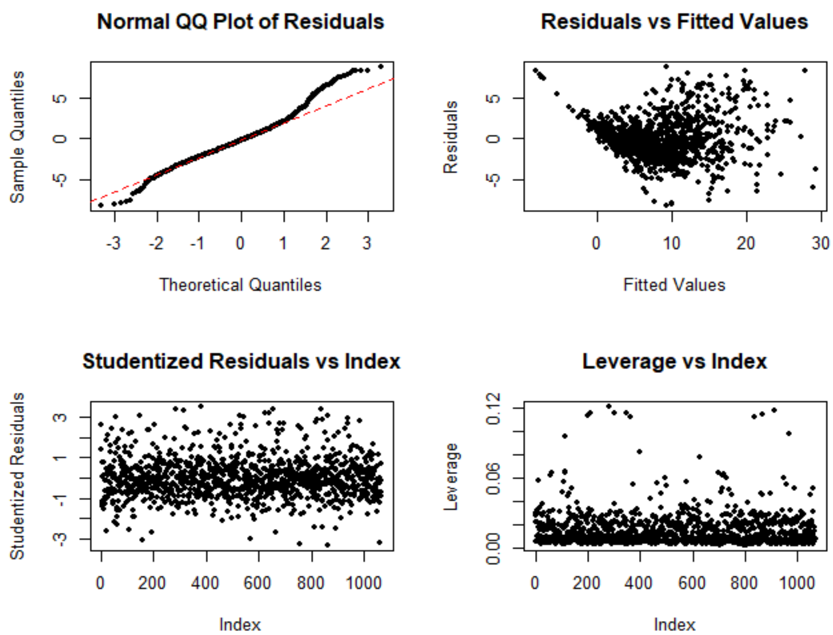
Model selection removed many draft year indicators. Just 7 draft year indicators remain in the final regression model, most of which are significant. These years may have had a particularly strong or weak class of new players. Players drafted these years showed a particularly strong association to an increase or decrease in average PPG later in their careers in the two seasons under consideration.

Section 4: Final regression

4.a : Final Regression Output (can be found in Appendix C.6)
Note: observations with absolute Studentized residuals > 4 observations were removed.

Compared to the original regression, R-squared has fallen slightly from 0.8374 to 0.8225 (Appendix C.6). This is expected, as more x-variables always explain more of the variation in y, even if they are insignificant. Thus, with model selection removing over half the x-variables from the original model, such a small decrease in R-squared is a good sign that our final regression model is not only much leaner than the original but also retains much of the efficacy in terms of explaining the y-variable of this data set.

Adjusted R-squared is a metric that also aims to measure the model's ability to explain the variation in y. However, the advantage of Adjusted R-squared is that, unlike R-squared, it does not necessarily increase with the number of x-variables. Thus, seeing that the final regression model has a greater adj. R-squared value of 0.823 to the original's 0.8226 indicates that the final model is indeed better than the original, despite having fewer variables. This gives us confidence that the redundant (multicollinearity) and insignificant variables that were removed manually and during model selection were justified removals, and that the resulting final regression model is indeed a better set of x-variables to explain the variation in y. Real-world data will never perfectly fit the regression due to individual circumstances of each observation. So, seeing an R-squared value over 0.8 is an indication that the x-variables explain the variation in y well.



The QQ-plot shows that the assumption of normality holds well in our data set. The red dotted line shows a slope of 1 against the theoretical quantiles, meaning if the sample data falls along that dotted line, we can safely say that the data is normally distributed. By the QQ-plot plotted from the original data set, the center fits nearly perfectly along that dotted line, meaning the standardized quantiles of the data set are indeed normally distributed N(0,1). We care most about the center, as that is where the vast majority of observations lie on a normal distribution. The sample quantiles deviate a bit from the straight line around the tails, which is to be expected in a real-world data set. The deviation from normality is slightly more noticeable on the right tail than the left tail; this may be because the number of points scored in a basketball game is technically unbounded above, while it is bounded below at 0. Thus, we may see some outliers of high-scoring players during some games, but there is less room to have outliers on the low end (capped at zero points below). Aside from testing the normality assumption, the QQ-plot can also reveal outliers in y (large residuals) and mean shifts. Our data does not have outliers in y, as no points deviate significantly from the rest of the data and the dotted line. The data also has no mean shifts, as there is no jump (or several jumps) in the middle of the data.

The Residuals vs. Fitted Values plot shows the residuals of each observation from the regression line. This diagnostic plot checks the constant variance assumption, which means we ideally want to see no big patterns in this plot. However, the Residuals vs. Fitted Values plot based on our original data set exhibits a funnel pattern, which indicates that the residuals generally increase with y (average points scored per game). This violates a regression assumption, so we must fix this for our final interpretation.

Due to non-constant variance, we considered transformations on y and robust standard errors. (se) We will use robust se in the final discussion, as transforming y detracts from the interpretation. For example, the log of average points scored per game makes little sense. Additionally, a transformation on y does not guarantee constant variance is satisfied (e.g., log may be too strong, while 1/y may not be strong enough). Transformations also change the significance of certain variables. For reference, the output of the transformations attempted are included in Appendix A.1 to A.3. In search of constant variance, none of these transformations resulted in a desirable Residuals vs. Fitted Values plot. Case in point, transformations are messy, imperfect, and detract from interpretation. Thus, we will employ robust standard errors to handle the non-constant variance in this data set, as they rely on fewer assumptions and are robust to heteroskedasticity, but do not change the model.

The VIF output on the final regression model (Appendix C.8) shows that, after model selection, none of the remaining x-variables in the optimal model are collinear. This can be seen in the fact that the VIF values are quite low, with the highest VIF being 2.27, which is well below the guideline value of concern of 10. Thus, there is no need to remove any variables or create new ones to handle any linear dependence. The VIF output on the original regression model (Appendix C.2) says that the original model had several highly correlated x-variables, with VIF values much greater than 10. As expected, since most of those x-variables were also insignificant to the regression model (as per the regression output), they were removed in the model selection process.

4.b: <u>Interpretation (with Robust SE)</u>

```
t test of coefficients:

                           Estimate Std. Error t value  Pr(>|t|)    
(Intercept)              -4.2063556  1.1686168 -3.5994 0.0003338 ***
age                       0.0242284  0.0192742  1.2570 0.2090186    
player_weight            -0.0473888  0.0098310 -4.8204 1.644e-06 ***
games_played              0.0422945  0.0038521 10.9795 < 2.2e-16 ***
average_rebounds_per_game 0.8059504  0.0551325 14.6184 < 2.2e-16 ***
average_assists_per_game  0.8819780  0.0868506 10.1551 < 2.2e-16 ***
usage_percentage         44.0509935  2.8864947 15.2611 < 2.2e-16 ***
true_shooting_percentage  3.8592375  1.0912945  3.5364 0.0004233 ***
draft_year_2001          -1.7463242  0.6432573 -2.7148 0.0067399 ** 
draft_year_2009           0.9262993  0.4695523  1.9727 0.0487882 *  
draft_year_2010          -0.3611358  0.3900569 -0.9259 0.3547344    
draft_year_2011           1.1308508  0.4209925  2.6862 0.0073419 ** 
draft_year_2012           0.8737449  0.4563654  1.9146 0.0558183 .  
draft_year_2014           0.7798575  0.3367205  2.3160 0.0207479 *  
draft_year_2016          -0.5181542  0.2937284 -1.7641 0.0780126 .  
draft_number_1            1.5164794  0.5922776  2.5604 0.0105934 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The intercept is -4.206, which represents the mean PPG for a player with no x values (all x-variables 0). This average PPG is impossible, so this model does not work for players who are outside the league, such as young players, without sufficient age, weight, or game statistics. This is expected, as most models have acceptable ranges for their x-variables. Any NBA player would have acceptable x-variable input values, as they are over 18 years old, likely relatively tall, and have values for their game statistics.

Each of the estimates of beta (coefficients) show the change in mean PPG for a unit increase in the corresponding x-variable, holding all other x-variables constant. Age has coefficient 0.024, meaning, in relation to other x-variables, players increase their mean PPG by 0.024 for each additional year of age. However, a p-value of 0.209 indicates that age is not significant. This is likely due to the fact that, with age comes experience, but also physical deterioration. So, age could work for or against a player's mean PPG, and likely cannot be well-modelled linearly. Thus, age is not great at explaining the y-variable and is insignificant in the regression model.

All else constant, the interpretation for weight is that for each additional pound a player gains, he will score 0.047 fewer PPG on average. Each additional game a player plays in a season is associated with an average of 0.042 more PPG. Interpretations for average rebounds and assists per game represent an additional rebound and assist respectively represent a 0.806 and 0.882 increase in average PPG.

Usage percentage has a high coefficient and is very significant. This means there is a great association between a player's overall involvement in games and how many points he scores per game over average. All else constant, each additional usage percentage is associated with a 44.051 increase in mean PPG. True shooting percentage, also significant, is associated with a 3.858 increase in mean PPG for each additional unit. Logically, players who are better at shooting the ball are likely to score more.

Some draft years are associated with an increase in mean PPG (2009, 2011, 2012, 2014), and others are associated with a decrease (2001, 2010, 2016). These years, for some reason, generally had better or worse than average rookies. All the other years do not show enough association to be included in the model. Additionally, players drafted in the first round (perceived as the best rookies that year) score an average of 1.516 PPG more than players drafted in any later draft; this positive association corroborates the fact that the best players are often drafted first.

Section 5: Discussion & Conclusion

The purpose of our regression analysis is to explain the variation in mean PPG. Through our analysis of the x-variables, we found that some are significant and present a strong correlation to the dependent variable (average points scored per game in a season). Other x-variables are not so significant, for a variety of reasons (discussed above). Model selection removed a large number of unwanted x-variables. The final regression model had just 15 x-variables, in comparison to the 94 in the original model. The reduction in x-variables makes the resulting regression model more interpretable, allowing us to pinpoint the contribution of each factor to explaining mean PPG. Additionally, to help with interpretation and model stability, we removed points of high leverage and high Studentized residual and used robust standard errors to handle the non-constant variance. The final model is also better at explaining the y-variable, with an increase in the adjusted R-squared over the original model. We arrived at a model with a high adjusted R-squared (0.823), meaning it explains a great portion of the variation in the response variable mean PPG.

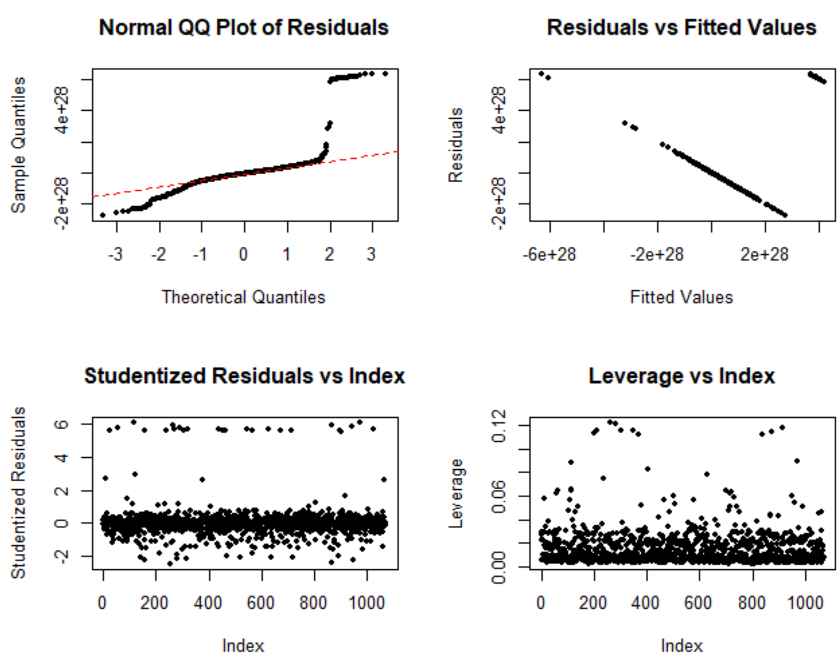
Works Cited

Link to the data source: https://www.kaggle.com/justinas/nba-players-data
2 seasons under consideration: 2017-2018 and 2018-2019 (1070 observations)

Appendix A: trial runs with different transformations on y

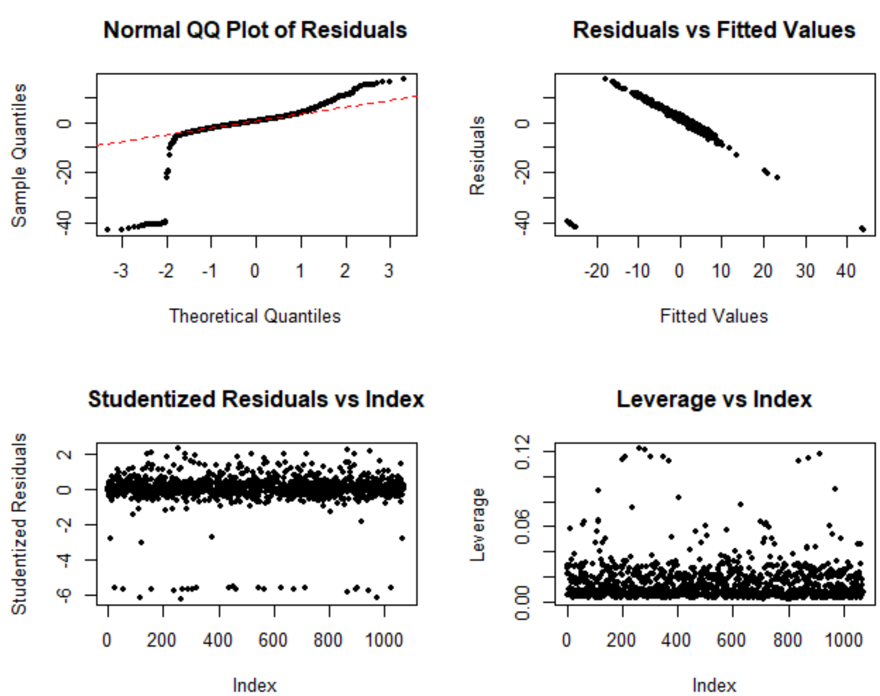
A.1: 1/y transformation on y

```
Call:
lm(formula = (1/average_points_per_game) ~ age + player_weight +
    games_played + average_rebounds_per_game + average_assists_per_game +
    usage_percentage + true_shooting_percentage + draft_year_2001 +
    draft_year_2009 + draft_year_2010 + draft_year_2011 + draft_year_2012 +
    draft_year_2014 + draft_year_2016 + draft_number_1, data = dataset_copy2)

Residuals:
       Min         1Q     Median         3Q        Max
-2.749e+28 -3.538e+27 -5.773e+26  1.975e+27  6.330e+28

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.139e+28  4.455e+27   7.045 3.33e-12 ***
age                       1.290e+26  9.105e+25   1.416   0.1569
player_weight             3.953e+25  4.089e+25   0.967   0.3339
games_played             -2.458e+25  1.598e+25  -1.538   0.1243
average_rebounds_per_game  1.189e+26  2.042e+26   0.582   0.5606
average_assists_per_game -5.935e+26  2.751e+26  -2.157   0.0312 *
usage_percentage          5.736e+27  7.025e+27   0.817   0.4144
true_shooting_percentage -6.774e+28  2.943e+27 -23.016  < 2e-16 ***
draft_year_2001          -2.537e+27  3.809e+27  -0.666   0.5055
draft_year_2009           1.665e+27  1.838e+27   0.906   0.3653
draft_year_2010          -2.355e+26  1.797e+27  -0.131   0.8957
draft_year_2011          -9.980e+25  1.496e+27  -0.067   0.9468
draft_year_2012          -2.888e+26  1.591e+27  -0.182   0.8560
draft_year_2014           3.605e+26  1.461e+27   0.247   0.8051
draft_year_2016          -5.005e+26  1.290e+27  -0.388   0.6980
draft_number_1            1.050e+27  2.341e+27   0.448   0.6539
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.1e+28 on 1054 degrees of freedom
Multiple R-squared:  0.3801,	Adjusted R-squared:  0.3713
F-statistic: 43.08 on 15 and 1054 DF,  p-value: < 2.2e-16
```



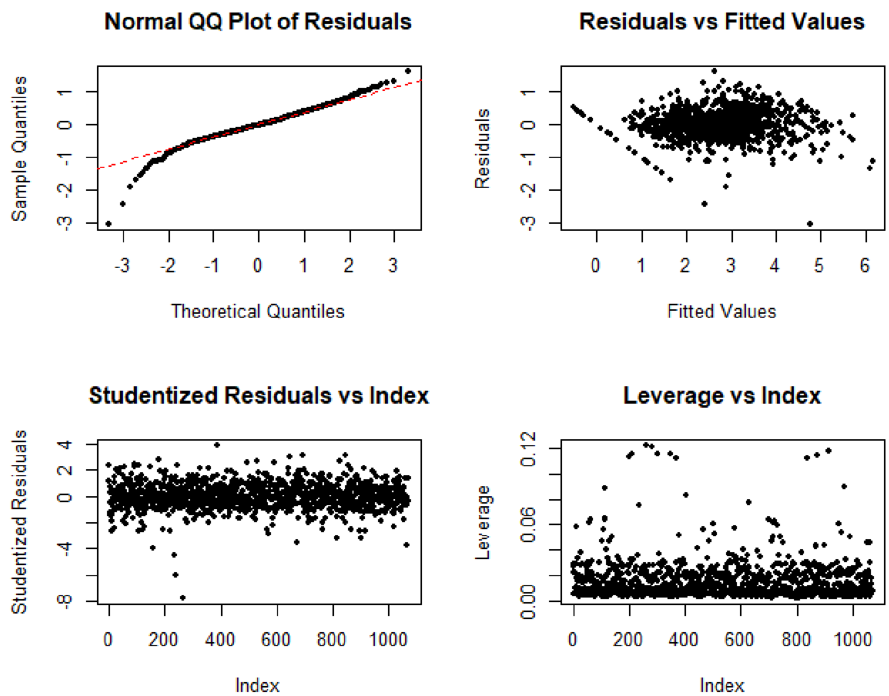
A.2: log transformation on y

```
Call:
lm(formula = log(average_points_per_game) ~ age + player_weight + 
    games_played + average_rebounds_per_game + average_assists_per_game + 
    usage_percentage + true_shooting_percentage + draft_year_2001 + 
    draft_year_2009 + draft_year_2010 + draft_year_2011 + draft_year_2012 + 
    draft_year_2014 + draft_year_2016 + draft_number_1, data = dataset_copy2)

Residuals:
    Min      1Q  Median      3Q     Max 
-42.815  -1.330   0.437   2.412  17.113 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -20.82906    2.98163  -6.986 5.01e-12 ***
age                        -0.07870    0.06094  -1.292  0.19681    
player_weight              -0.03590    0.02737  -1.312  0.18993    
games_played                0.02648    0.01069   2.476  0.01343 *  
average_rebounds_per_game   0.03220    0.13670   0.236  0.81384    
average_assists_per_game    0.49201    0.18412   2.672  0.00765 ** 
usage_percentage            0.02624    4.70167   0.006  0.99555    
true_shooting_percentage   46.53695    1.96984  23.625  < 2e-16 ***
draft_year_2001             1.49444    2.54946   0.586  0.55788    
draft_year_2009            -1.09599    1.23024  -0.891  0.37320    
draft_year_2010             0.03519    1.20259   0.029  0.97666    
draft_year_2011             0.19740    1.00096   0.197  0.84370    
draft_year_2012             0.26244    1.06480   0.246  0.80537    
draft_year_2014            -0.14058    0.97779  -0.144  0.88571    
draft_year_2016             0.28620    0.86312   0.332  0.74027    
draft_number_1             -0.74543    1.56657  -0.476  0.63429    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.365 on 1054 degrees of freedom
Multiple R-squared:  0.4148,	Adjusted R-squared:  0.4065 
F-statistic:  49.8 on 15 and 1054 DF,  p-value: < 2.2e-16
```



A.3: square root transformation on y

```
Call:
lm(formula = sqrt(average_points_per_game) ~ age + player_weight +
    games_played + average_rebounds_per_game + average_assists_per_game +
    usage_percentage + true_shooting_percentage + draft_year_2001 +
    draft_year_2009 + draft_year_2010 + draft_year_2011 + draft_year_2012 +
    draft_year_2014 + draft_year_2016 + draft_number_1, data = dataset_copy2)

Residuals:
     Min       1Q   Median       3Q      Max
-3.01973 -0.23877 -0.00648  0.26549  1.62289

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.3968335  0.1730528   2.293 0.022036 *
age                       0.0058789  0.0035367   1.662 0.096757 .
player_weight            -0.0106606  0.0015884  -6.711 3.14e-11 ***
games_played              0.0104330  0.0006207  16.810  < 2e-16 ***
average_rebounds_per_game 0.1453563  0.0079337  18.321  < 2e-16 ***
average_assists_per_game  0.1496218  0.0106865  14.001  < 2e-16 ***
usage_percentage          5.7836868  0.2728835  21.195  < 2e-16 ***
true_shooting_percentage  1.5771529  0.1143292  13.795  < 2e-16 ***
draft_year_2001          -0.2669010  0.1479701  -1.804 0.071556 .
draft_year_2009           0.0749880  0.0714025   1.050 0.293860
draft_year_2010          -0.1025854  0.0697981  -1.470 0.141929
draft_year_2011           0.1997230  0.0580954   3.438 0.000609 ***
draft_year_2012           0.1243884  0.0618009   2.013 0.044397 *
draft_year_2014           0.1443094  0.0567505   2.543 0.011137 *
draft_year_2016          -0.0657348  0.0500953  -1.312 0.189740
draft_number_1            0.0924086  0.0909234   1.016 0.309703
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4275 on 1054 degrees of freedom
Multiple R-squared:  0.8399,	Adjusted R-squared:  0.8376
F-statistic: 368.5 on 15 and 1054 DF,  p-value: < 2.2e-16
```



Appendix B: Formulas

B.1: True Shooting Percentage

$$TS\% = \frac{PTS}{2(FGA + (0.44 \times FTA))}$$

B.2: Usage Rate

$$USG\% = 100 * \frac{(FGA + 0.44 * FTA + TOV) * \left(\frac{TmMP}{5}\right)}{MP * (TmFGA + 0.44 * TmFTA + TmTOV)}$$

Appendix C: Additional Data and Output

C.1: <u>Original Regression Model Output</u>

```
Call:
lm(formula = average_points_per_game ~ ., data = dataset_copy)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.0386  -1.4858  -0.0776   1.3162   9.5576 

Coefficients: (2 not defined because of singularities)
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             -0.566335   3.225029  -0.176 0.860640    
age                      0.017161   0.040967   0.419 0.675386    
player_height           -0.010644   0.018450  -0.577 0.564131    
player_weight           -0.047427   0.013291  -3.568 0.000377 ***
games_played             0.041206   0.004027  10.231  < 2e-16 ***
average_rebounds_per_game  0.822417   0.052740  15.594  < 2e-16 ***
average_assists_per_game   0.922344   0.070243  13.131  < 2e-16 ***
usage_percentage        36.327020   1.722279  21.092  < 2e-16 ***
true_shooting_percentage   3.155227   0.699593   4.510 7.26e-06 ***
draft_year_1998         -0.102229   2.997668  -0.034 0.972802    
draft_year_1999         -0.617877   3.342719  -0.185 0.853391    
draft_year_2000         -1.962030   3.263587  -0.601 0.547854    
draft_year_2001         -3.610325   2.793286  -1.293 0.196488    
draft_year_2002         -2.530442   3.232873  -0.783 0.433979    
draft_year_2003         -0.065145   2.742881  -0.024 0.981056    
draft_year_2004         -0.905240   2.697326  -0.336 0.737239    
draft_year_2005         -0.949853   2.673561  -0.355 0.722459    
draft_year_2006          0.095780   2.701087   0.035 0.971720    
draft_year_2007          0.150372   2.654834   0.057 0.954843    
draft_year_2008         -0.123662   2.631014  -0.047 0.962521    
draft_year_2009          0.609111   2.630940   0.232 0.816961    
draft_year_2010         -1.054267   2.628117  -0.401 0.688398    
draft_year_2011          0.698044   2.571039   0.272 0.786062    
draft_year_2012          0.470446   2.615360   0.180 0.857286    
draft_year_2013         -0.138320   2.610059  -0.053 0.957747    
draft_year_2014          0.120753   2.605650   0.046 0.963046    
draft_year_2015         -0.680688   2.602265  -0.262 0.793704    
draft_year_2016         -1.102557   2.602970  -0.424 0.671968    
draft_year_2017         -0.769908   2.603317  -0.296 0.767490    
draft_year_2018         -0.795561   2.618566  -0.304 0.761333    
draft_round_1            0.473999   2.643445   0.179 0.857731    
draft_round_2            0.232362   2.810282   0.083 0.934121    
draft_number_1           2.382601   0.728482   3.271 0.001110 ** 
draft_number_10          1.293425   0.757163   1.708 0.087906 .  
draft_number_11          1.157892   0.767082   1.509 0.131500    
draft_number_12          0.476490   0.787659   0.605 0.545356    
draft_number_13          2.106542   0.804287   2.619 0.008951 ** 
draft_number_14          0.622733   0.849920   0.733 0.463919    
draft_number_15          2.063820   0.797559   2.588 0.009806 ** 
draft_number_16         -0.250694   0.786939  -0.319 0.750122    
draft_number_17          0.720506   0.817383   0.881 0.378275    
draft_number_18          0.020735   0.756044   0.027 0.978126    
draft_number_19          1.855166   0.815506   2.275 0.023129 *  
draft_number_2           1.602793   0.706414   2.269 0.023490 *  
draft_number_20         -0.134220   0.815777  -0.165 0.869347    
draft_number_21         -1.286455   0.839581  -1.532 0.125782    
draft_number_22         -0.456364   0.817914  -0.558 0.576998    
draft_number_23          0.256874   0.753504   0.341 0.733247    
```

```
draft_number_24          0.666540   0.806509   0.826 0.408750    
draft_number_25          0.701505   0.872525   0.804 0.421595    
draft_number_26         -0.325617   0.840197  -0.388 0.698434    
draft_number_27          0.706537   0.801241   0.882 0.378099    
draft_number_28          1.190141   0.890693   1.336 0.181795    
draft_number_29         -2.249820   0.995895  -2.259 0.024097 *  
draft_number_3           2.506637   0.724271   3.461 0.000562 ***
draft_number_30          1.138997   0.927716   1.228 0.219838    
draft_number_31          2.252897   1.410215   1.598 0.110464    
draft_number_32          0.256112   1.395070   0.184 0.854378    
draft_number_33         -0.246174   1.361426  -0.181 0.856546    
draft_number_34         -0.082997   1.328278  -0.062 0.950189    
draft_number_35         -2.233882   1.377860  -1.621 0.105282    
draft_number_36          1.505102   1.477273   1.019 0.308531    
draft_number_37          1.174085   1.423588   0.825 0.409722    
draft_number_38          0.306292   1.388172   0.221 0.825416    
draft_number_39          0.358023   1.420017   0.252 0.800997    
draft_number_4           0.537625   0.708334   0.759 0.448036    
draft_number_40          0.813871   1.360054   0.598 0.549704    
draft_number_41         -0.772567   1.443964  -0.535 0.592749    
draft_number_42         -0.852256   1.423015  -0.599 0.549372    
draft_number_43         -0.037446   1.483888  -0.025 0.979873    
draft_number_44          0.619488   1.513609   0.409 0.682424    
draft_number_45          1.678391   1.390111   1.207 0.227578    
draft_number_46          0.720692   1.399050   0.515 0.606579    
draft_number_47         -0.146738   1.441835  -0.102 0.918959    
draft_number_48          0.063467   1.519499   0.042 0.966692    
draft_number_49         -3.282020   1.890975  -1.736 0.082945 .  
draft_number_5          -0.375002   0.683142  -0.549 0.583173    
draft_number_50          0.509934   1.624574   0.314 0.753673    
draft_number_51          0.760140   1.776071   0.428 0.668753    
draft_number_52          0.230428   1.560294   0.148 0.882624    
draft_number_53          0.153424   1.873194   0.082 0.934739    
draft_number_54         -0.047581   1.862492  -0.026 0.979624    
draft_number_55          0.015648   1.505407   0.010 0.991709    
draft_number_56         -1.054667   1.512660  -0.697 0.485826    
draft_number_57         -1.418982   2.044332  -0.694 0.487780    
draft_number_58          0.373773   1.718743   0.217 0.827888    
draft_number_59          3.973054   2.157763   1.841 0.065882 .  
draft_number_6           1.630309   0.798254   2.042 0.041385 *  
draft_number_60                NA         NA      NA       NA    
draft_number_7           2.264143   0.712760   3.177 0.001537 ** 
draft_number_8           0.405081   0.757946   0.534 0.593154    
draft_number_9                 NA         NA      NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.537 on 980 degrees of freedom
Multiple R-squared:  0.8374,	Adjusted R-squared:  0.8226 
F-statistic: 56.71 on 89 and 980 DF,  p-value: < 2.2e-16
```

C.2: VIF on the initial model

(with draft_number_9 & draft_number_60 removed due to exact collinearity)

```
                    age             player_height             player_weight 
               4.921664                  4.073947                  3.598354 
 average_assists_per_game          usage_percentage  true_shooting_percentage 
               2.502420                  1.542770                  1.239688 
        draft_year_2000           draft_year_2001           draft_year_2002 
               3.303778                 10.819524                  3.241886 
        draft_year_2005           draft_year_2006           draft_year_2007 
              27.117892                 17.867219                 32.966389 
        draft_year_2010           draft_year_2011           draft_year_2012 
              42.316328                 59.082333                 53.540099 
        draft_year_2015           draft_year_2016           draft_year_2017 
              74.288627                 85.026014                 94.607961 
          draft_round_2            draft_number_1           draft_number_10 
             227.101160                  2.013325                  1.833973 
        draft_number_13           draft_number_14           draft_number_15 
               1.681583                  1.550843                  1.653568 
        draft_number_18           draft_number_19            draft_number_2 
               1.828554                  1.628673                  2.040734 
        draft_number_22           draft_number_23           draft_number_24 
               1.638308                  1.816291                  1.690891 
        draft_number_27           draft_number_28           draft_number_29 
               1.668871                  1.583048                  1.375321 
        draft_number_31           draft_number_32           draft_number_33 
               3.061230                  3.292304                  3.698503 
        draft_number_36           draft_number_37           draft_number_38 
               2.692497                  3.119565                  3.259826 
        draft_number_40           draft_number_41           draft_number_42 
               3.691048                  2.891280                  3.117053 
        draft_number_45           draft_number_46           draft_number_47 
               3.268939                  3.311117                  2.882759 
         draft_number_5           draft_number_50           draft_number_51 
               2.045926                  2.040882                  1.953246 
        draft_number_54           draft_number_55           draft_number_56 
               1.612478                  2.448828                  2.472482 
        draft_number_59            draft_number_6            draft_number_7 
               1.444203                  1.656451                  2.077560 


           games_played average_rebounds_per_game 
               1.929331                  2.840806 
        draft_year_1998           draft_year_1999 
               5.564213                  3.465935 
        draft_year_2003           draft_year_2004 
              15.012456                 21.098055 
        draft_year_2008           draft_year_2009 
              45.380130                 42.407279 
        draft_year_2013           draft_year_2014 
              62.760998                 64.406674 
        draft_year_2018             draft_round_1 
              52.710556                289.408302 
        draft_number_11           draft_number_12 
               1.794414                  1.706020 
        draft_number_16           draft_number_17 
               1.702901                  1.636180 
        draft_number_20           draft_number_21 
               1.629756                  1.619901 
        draft_number_25           draft_number_26 
               1.519124                  1.622279 
         draft_number_3           draft_number_30 
               1.990113                  1.455922 
        draft_number_34           draft_number_35 
               4.054534                  3.788331 
        draft_number_39            draft_number_4 
               3.103932                  1.977741 
        draft_number_43           draft_number_44 
               2.716667                  2.475584 
        draft_number_48           draft_number_49 
               2.494890                  1.662174 
        draft_number_52           draft_number_53 
               2.256965                  1.631062 
        draft_number_57           draft_number_58 
               1.942709                  1.829188 
         draft_number_8 
               2.008953 
```

C.3: High leverage point in original data (Observation and r output for leverage)

Index	age	player_height	player_weight	draft_year	draft_round	draft_number	gp	pts	reb	ast	usg_pct	ts_pct
307	30	195.58	92.98636	2011	Undrafted	Undrafted	2	0	1	0	0.136	0

```
> hatvalues(lm_a)[307]
307
  1
```

C.4: Stepwise Model Path (Anova output of model selection)

```
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
average_points_per_game ~ age + player_height + player_weight +
    games_played + average_rebounds_per_game + average_assists_per_game +
    usage_percentage + true_shooting_percentage + draft_year_1998 +
    draft_year_1999 + draft_year_2000 + draft_year_2001 + draft_year_2002 +
    draft_year_2003 + draft_year_2004 + draft_year_2005 + draft_year_2006 +
    draft_year_2007 + draft_year_2008 + draft_year_2009 + draft_year_2010 +
    draft_year_2011 + draft_year_2012 + draft_year_2013 + draft_year_2014 +
    draft_year_2015 + draft_year_2016 + draft_year_2017 + draft_year_2018 +
    draft_round_1 + draft_round_2 + draft_number_1 + draft_number_10 +
    draft_number_11 + draft_number_12 + draft_number_13 + draft_number_14 +
    draft_number_15 + draft_number_16 + draft_number_17 + draft_number_18 +
    draft_number_19 + draft_number_2 + draft_number_20 + draft_number_21 +
    draft_number_22 + draft_number_23 + draft_number_24 + draft_number_25 +
    draft_number_26 + draft_number_27 + draft_number_28 + draft_number_29 +
    draft_number_3 + draft_number_30 + draft_number_31 + draft_number_32 +
    draft_number_33 + draft_number_34 + draft_number_35 + draft_number_36 +
    draft_number_37 + draft_number_38 + draft_number_39 + draft_number_4 +
    draft_number_40 + draft_number_41 + draft_number_42 + draft_number_43 +
    draft_number_44 + draft_number_45 + draft_number_46 + draft_number_47 +
    draft_number_48 + draft_number_49 + draft_number_5 + draft_number_50 +
    draft_number_51 + draft_number_52 + draft_number_53 + draft_number_54 +
    draft_number_55 + draft_number_56 + draft_number_57 + draft_number_58 +
    draft_number_59 + draft_number_6 + draft_number_60 + draft_number_7 +
    draft_number_8 + draft_number_9

Final Model:
average_points_per_game ~ age + player_weight + games_played +
    average_rebounds_per_game + average_assists_per_game + usage_percentage +
    true_shooting_percentage + draft_year_2001 + draft_year_2009 +
    draft_year_2010 + draft_year_2011 + draft_year_2012 + draft_year_2014 +
    draft_year_2016 + draft_number_1 + draft_number_10 + draft_number_11 +
    draft_number_13 + draft_number_15 + draft_number_19 + draft_number_2 +
    draft_number_21 + draft_number_29 + draft_number_3 + draft_number_31 +
    draft_number_35 + draft_number_41 + draft_number_42 + draft_number_45 +
    draft_number_49 + draft_number_56 + draft_number_57 + draft_number_59 +
    draft_number_6 + draft_number_7
```

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				980	6307.002	2078.182
2	- draft_number_9	0	0.000000e+00	980	6307.002	2078.182
3	- draft_number_60	0	0.000000e+00	980	6307.002	2078.182
4	- draft_number_55	1	6.953132e-04	981	6307.003	2076.182
5	- draft_year_2003	1	3.572755e-03	982	6307.007	2074.183
6	- draft_year_1998	1	4.618907e-03	983	6307.011	2072.183
7	- draft_number_18	1	6.256992e-03	984	6307.017	2070.185
8	- draft_number_54	1	7.690492e-03	985	6307.025	2068.186
9	- draft_number_43	1	6.294906e-03	986	6307.031	2066.187
10	- draft_year_2008	1	3.155921e-02	987	6307.063	2064.192
11	- draft_year_2013	1	1.589117e-02	988	6307.079	2062.195
12	- draft_number_48	1	3.075803e-02	989	6307.110	2060.200
13	- draft_number_53	1	5.809752e-02	990	6307.168	2058.210
14	- draft_number_34	1	1.033445e-01	991	6307.271	2056.228
15	- draft_number_47	1	1.327815e-01	992	6307.404	2054.250
16	- draft_round_2	1	1.300833e-01	993	6307.534	2052.272
17	- draft_number_20	1	2.510486e-01	994	6307.785	2050.315
18	- draft_number_33	1	3.572281e-01	995	6308.142	2048.375
19	- draft_year_1999	1	3.951915e-01	996	6308.537	2046.442
20	- draft_number_52	1	5.882015e-01	997	6309.126	2044.542
21	- draft_year_2006	1	6.101545e-01	998	6309.736	2042.646
22	- draft_number_16	1	6.994801e-01	999	6310.435	2040.764
23	- draft_number_26	1	7.705324e-01	1000	6311.206	2038.895
24	- draft_number_58	1	7.839689e-01	1001	6311.990	2037.028
25	- draft_number_32	1	1.097371e+00	1002	6313.087	2035.214
26	- draft_number_38	1	1.244900e+00	1003	6314.332	2033.425
27	- draft_number_50	1	1.402269e+00	1004	6315.734	2031.662
28	- draft_round_1	1	1.236329e+00	1005	6316.971	2029.872
29	- draft_number_5	1	7.644814e-01	1006	6317.735	2028.001
30	- draft_number_22	1	8.123646e-01	1007	6318.548	2026.139
31	- draft_number_39	1	1.166424e+00	1008	6319.714	2024.336
32	- player_height	1	1.671055e+00	1009	6321.385	2022.619
33	- draft_year_2007	1	1.977391e+00	1010	6323.362	2020.954
34	- draft_number_51	1	1.887202e+00	1011	6325.250	2019.273
35	- draft_number_44	1	2.379261e+00	1012	6327.629	2017.676
36	- draft_number_46	1	4.543642e+00	1013	6332.172	2016.444
37	- draft_number_23	1	4.304360e+00	1014	6336.477	2015.171
38	- draft_number_40	1	5.919725e+00	1015	6342.397	2014.170
39	- draft_number_12	1	6.579325e+00	1016	6348.976	2013.279
40	- draft_number_8	1	5.957335e+00	1017	6354.933	2012.283
41	- draft_year_2000	1	4.376583e+00	1018	6359.310	2011.019
42	- draft_number_14	1	5.898323e+00	1019	6365.208	2010.011
43	- draft_year_2015	1	5.363608e+00	1020	6370.572	2008.913
44	- draft_year_2018	1	3.900885e+00	1021	6374.473	2007.568
45	- draft_year_2017	1	3.003206e+00	1022	6377.476	2006.072
46	- draft_number_25	1	6.050433e+00	1023	6383.526	2005.086
47	- draft_number_27	1	7.142733e+00	1024	6390.669	2004.283
48	- draft_number_4	1	6.884797e+00	1025	6397.554	2003.435
49	- draft_number_24	1	7.530099e+00	1026	6405.084	2002.694
50	- draft_number_17	1	7.260028e+00	1027	6412.344	2001.906
51	- draft_number_37	1	7.452867e+00	1028	6419.797	2001.149
52	- draft_year_2004	1	8.624753e+00	1029	6428.422	2000.585
53	- draft_number_30	1	8.768638e+00	1030	6437.190	2000.044
54	- draft_year_2005	1	8.983094e+00	1031	6446.173	1999.536
55	- draft_number_36	1	6.841460e+00	1032	6453.015	1998.671
56	- draft_year_2002	1	9.351266e+00	1033	6462.366	1998.220
57	- draft_number_28	1	1.051585e+01	1034	6472.882	1997.960

C.5: Regression output for model created after Model Selection

```
Call:
lm(formula = average_points_per_game ~ age + player_weight + 
    games_played + average_rebounds_per_game + average_assists_per_game + 
    usage_percentage + true_shooting_percentage + draft_year_2001 + 
    draft_year_2009 + draft_year_2010 + draft_year_2011 + draft_year_2012 + 
    draft_year_2014 + draft_year_2016 + draft_number_1, data = dataset_copy)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.7563  -1.5769  -0.1806   1.2899   9.2748 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -3.321701   1.062983  -3.125  0.00183 ** 
age                        0.022954   0.021724   1.057  0.29093    
player_weight             -0.046656   0.009757  -4.782 1.98e-06 ***
games_played               0.043604   0.003812  11.437  < 2e-16 ***
average_rebounds_per_game  0.841763   0.048733  17.273  < 2e-16 ***
average_assists_per_game   0.965213   0.065642  14.704  < 2e-16 ***
usage_percentage          38.522481   1.676197  22.982  < 2e-16 ***
true_shooting_percentage   3.223862   0.702271   4.591 4.95e-06 ***
draft_year_2001           -1.791526   0.908912  -1.971  0.04898 *  
draft_year_2009            0.960847   0.438592   2.191  0.02869 *  
draft_year_2010           -0.342992   0.428737  -0.800  0.42389    
draft_year_2011            1.288542   0.356853   3.611  0.00032 ***
draft_year_2012            0.895818   0.379614   2.360  0.01847 *  
draft_year_2014            0.858698   0.348592   2.463  0.01392 *  
draft_year_2016           -0.431629   0.307712  -1.403  0.16100    
draft_number_1             1.588869   0.558500   2.845  0.00453 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.626 on 1054 degrees of freedom
Multiple R-squared:  0.8127,	Adjusted R-squared:   0.81 
F-statistic: 304.8 on 15 and 1054 DF,  p-value: < 2.2e-16
```

C.6: Regression output for Final Model

```
Call:
lm(formula = average_points_per_game ~ age + player_weight + 
    games_played + average_rebounds_per_game + average_assists_per_game + 
    usage_percentage + true_shooting_percentage + draft_year_2001 + 
    draft_year_2009 + draft_year_2010 + draft_year_2011 + draft_year_2012 + 
    draft_year_2014 + draft_year_2016 + draft_number_1, data = dataset_copy)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.1386 -1.5975 -0.1833  1.2431  8.7396 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -4.206356   1.031932  -4.076 4.92e-05 ***
age                       0.024228   0.020987   1.154  0.24858    
player_weight            -0.047389   0.009413  -5.034 5.64e-07 ***
games_played              0.042294   0.003687  11.470  < 2e-16 ***
average_rebounds_per_game  0.805950   0.047196  17.077  < 2e-16 ***
average_assists_per_game  0.881978   0.064011  13.779  < 2e-16 ***
usage_percentage         44.050994   1.730713  25.453  < 2e-16 ***
true_shooting_percentage  3.859238   0.709972   5.436 6.78e-08 ***
draft_year_2001          -1.746324   0.876877  -1.992  0.04668 *  
draft_year_2009           0.926299   0.423130   2.189  0.02880 *  
draft_year_2010          -0.361136   0.413602  -0.873  0.38278    
draft_year_2011           1.130851   0.344712   3.281  0.00107 ** 
draft_year_2012           0.873745   0.366214   2.386  0.01721 *  
draft_year_2014           0.779858   0.336404   2.318  0.02063 *  
draft_year_2016          -0.518154   0.297035  -1.744  0.08138 .  
draft_number_1            1.516479   0.538832   2.814  0.00498 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.533 on 1051 degrees of freedom
Multiple R-squared:  0.8255,	Adjusted R-squared:  0.823 
F-statistic: 331.4 on 15 and 1051 DF,  p-value: < 2.2e-16
```

C.7: Anova output for final model

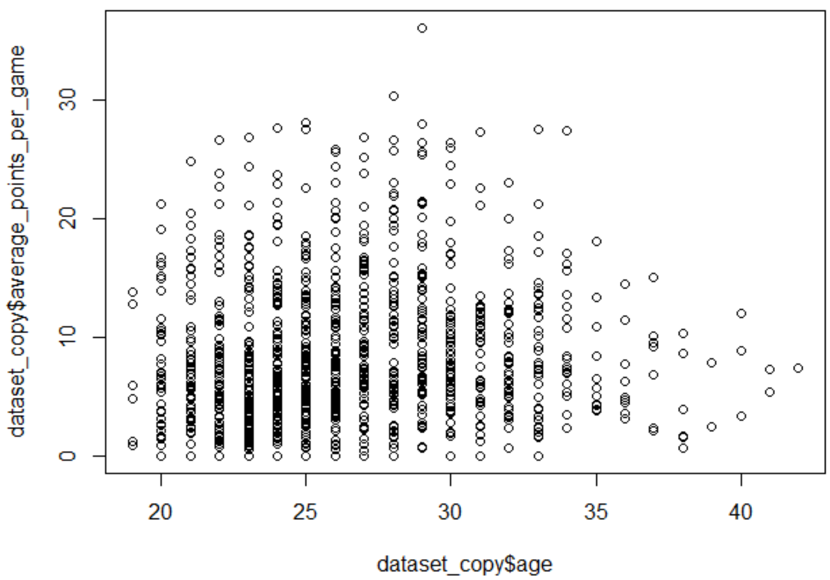
```
> anova(lm_final)
Analysis of Variance Table

Response: average_points_per_game
                            Df  Sum Sq Mean Sq   F value    Pr(>F)    
age                          1   351.8   351.8   54.8264 2.695e-13 ***
player_weight                1    26.2    26.2    4.0872  0.043461 *  
games_played                 1 11901.5 11901.5 1855.0615 < 2.2e-16 ***
average_rebounds_per_game    1  9460.4  9460.4 1474.5708 < 2.2e-16 ***
average_assists_per_game     1  5100.5  5100.5  795.0016 < 2.2e-16 ***
usage_percentage             1  4595.2  4595.2  716.2501 < 2.2e-16 ***
true_shooting_percentage     1   188.7   188.7   29.4111 7.263e-08 ***
draft_year_2001              1    37.6    37.6    5.8637  0.015624 *  
draft_year_2009              1    19.6    19.6    3.0558  0.080742 .  
draft_year_2010              1    11.3    11.3    1.7565  0.185346    
draft_year_2011              1    57.2    57.2    8.9173  0.002890 ** 
draft_year_2012              1    34.0    34.0    5.2970  0.021557 *  
draft_year_2014              1    40.4    40.4    6.3027  0.012205 *  
draft_year_2016              1    18.9    18.9    2.9416  0.086621 .  
draft_number_1               1    50.8    50.8    7.9207  0.004978 ** 
Residuals                 1051  6742.9     6.4                        
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

C.8: VIF output for final model

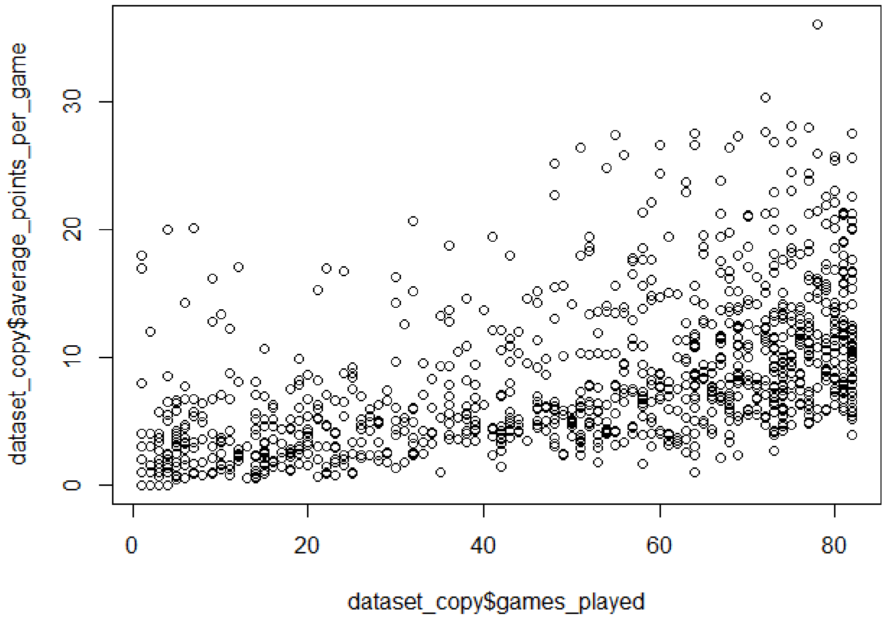
```
> vif(lm_final)
                 age            player_weight             games_played
            1.293091                 1.810252                 1.607941
    usage_percentage  true_shooting_percentage           draft_year_2001
            1.438056                 1.184642                 1.069539
     draft_year_2011          draft_year_2012           draft_year_2014
            1.065204                 1.052869                 1.076702

            average_rebounds_per_game  average_assists_per_game
                             2.270480                  2.077837
                      draft_year_2009           draft_year_2010
                             1.100197                  1.051205
                      draft_year_2016             draft_number_1
                             1.110383                  1.104857
```

Appendix D: Additional Graphs and Plots

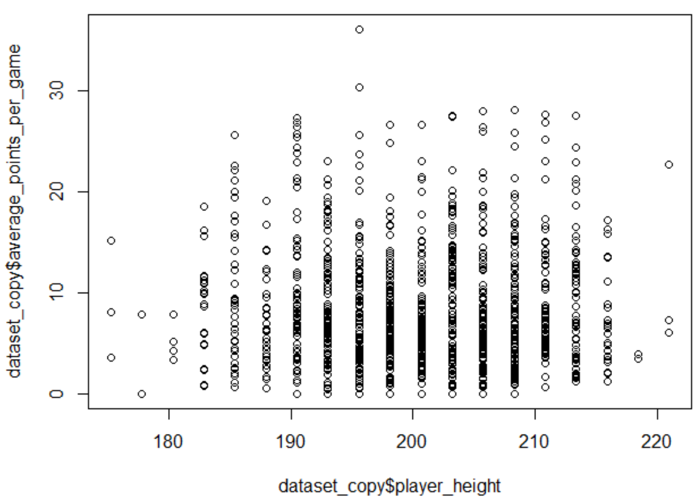
Appendix D.1: Age vs. Average Points per Game



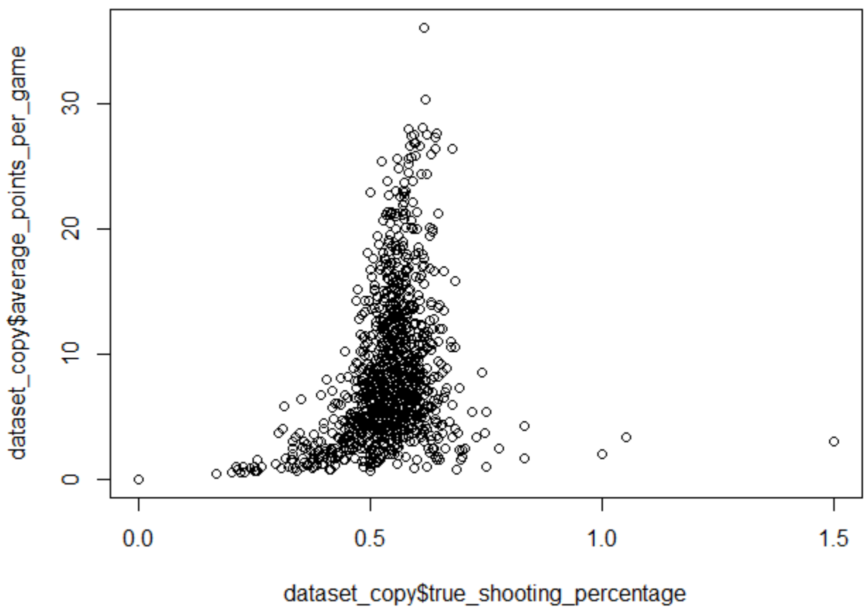
D.2: <u>Scatter Plot for Average Points per Game versus Games Played:</u>
Number of games played alone is not strongly correlated to average PPG. The number of games played per season indicates a player's health, dependability, and potentially a positive trajectory. In conjunction with other x-variables, the number of games played could indicate a potential for greater average PPG.



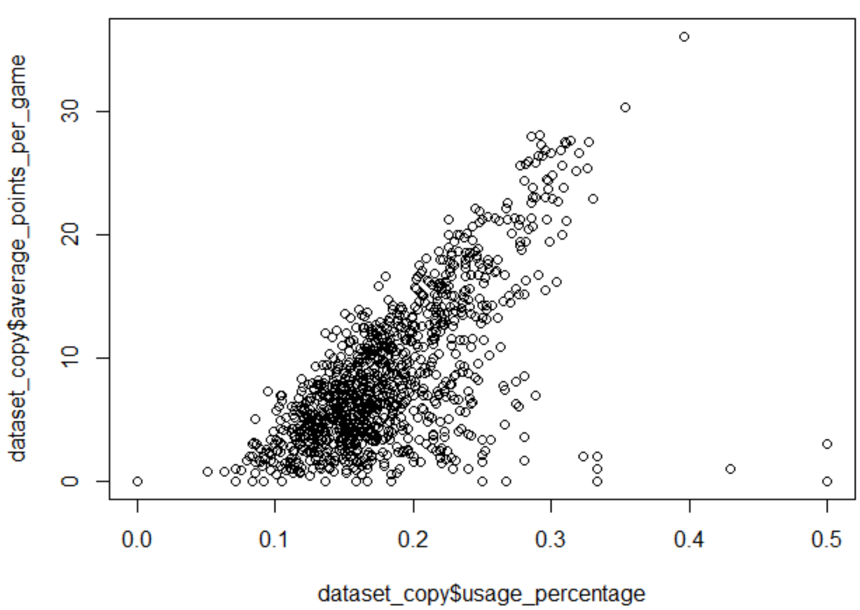
D.3: <u>Scatter Plot for Average Points per Game versus Player Height:</u>
Player height alone is not strongly correlated to average PPG. However, lower height is generally considered a weakness in the NBA, since that individual is likely less effective against taller opponents. Shorter players also have a harder time reaching the net.



D.4: Scatter Plot for Average Points per Game versus True Shooting Percentage

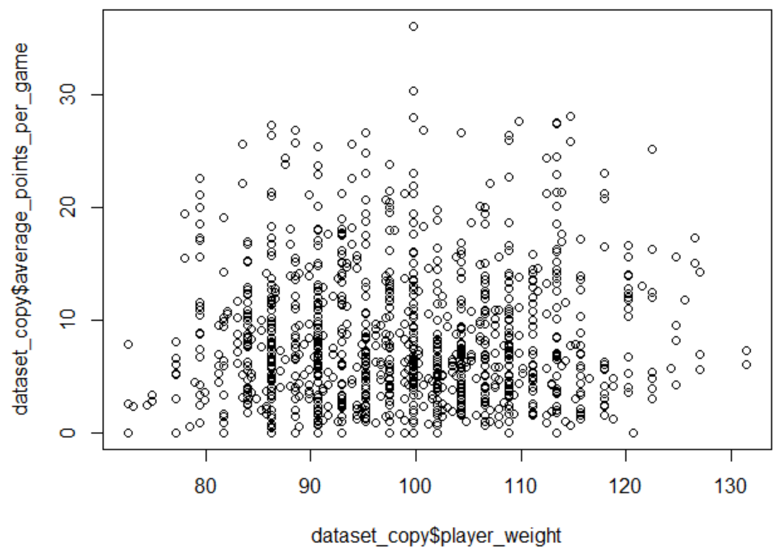


D.5: Scatter Plot for Average Points per Game versus Usage Percentage

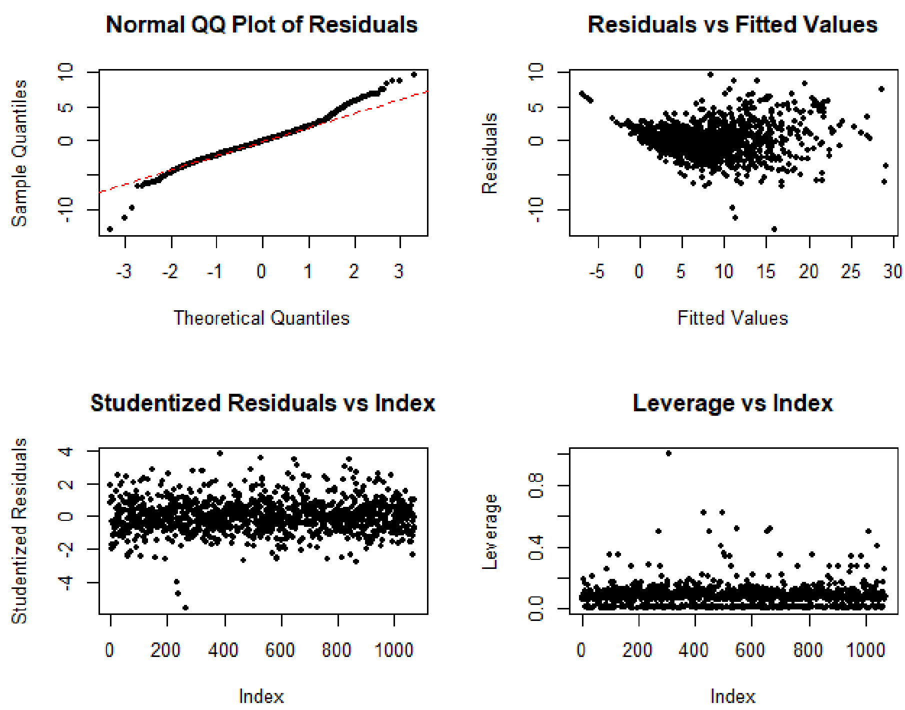


D.6: Scatter Plot for Average Points per Game versus Player Weight:

Player weight does not depict much correlation to average PPG. There are certain players that could be overweight/underweight depicting qualities of being out of shape, lack of strength, etc, which is why observations at the ends of the x-axis have low points per game. These values could be considered to be high leverage.



D.7: Diagnostic Plots for Original Regression Model



D.8: Diagnostic Plots for Model after Model Selection