EPL Survivor Analysis

Divine Amanluoghe

2026-01-09

Introduction & Objectives

This report analyzes the competitive dynamics of the nine Premier League clubs that successfully maintained their top-flight status continuously between the 2013/14 and 2023/24 seasons.The dataset effectively filters for 'survivorship,' excluding clubs that faced relegation.

The objective is to determine what drives performance variation within this surviving cohort. Specifically, the analysis investigates three competing hypotheses:

Financial Dominance: Whether the 'Big 6' rely on raw wage expenditure to secure points compared to the mid-table survivors (Everton, West Ham, Crystal Palace).

Managerial Stability: Whether keeping a manager for longer tenures yields a 'stability dividend' in points.

Squad Experience: Whether investing in older, more experienced players (higher Experience Ratio) provides a competitive edge over younger squads.

The analysis employs OLS regression with Fixed Effects to isolate the impact of 'Club Identity' versus marginal resource allocation

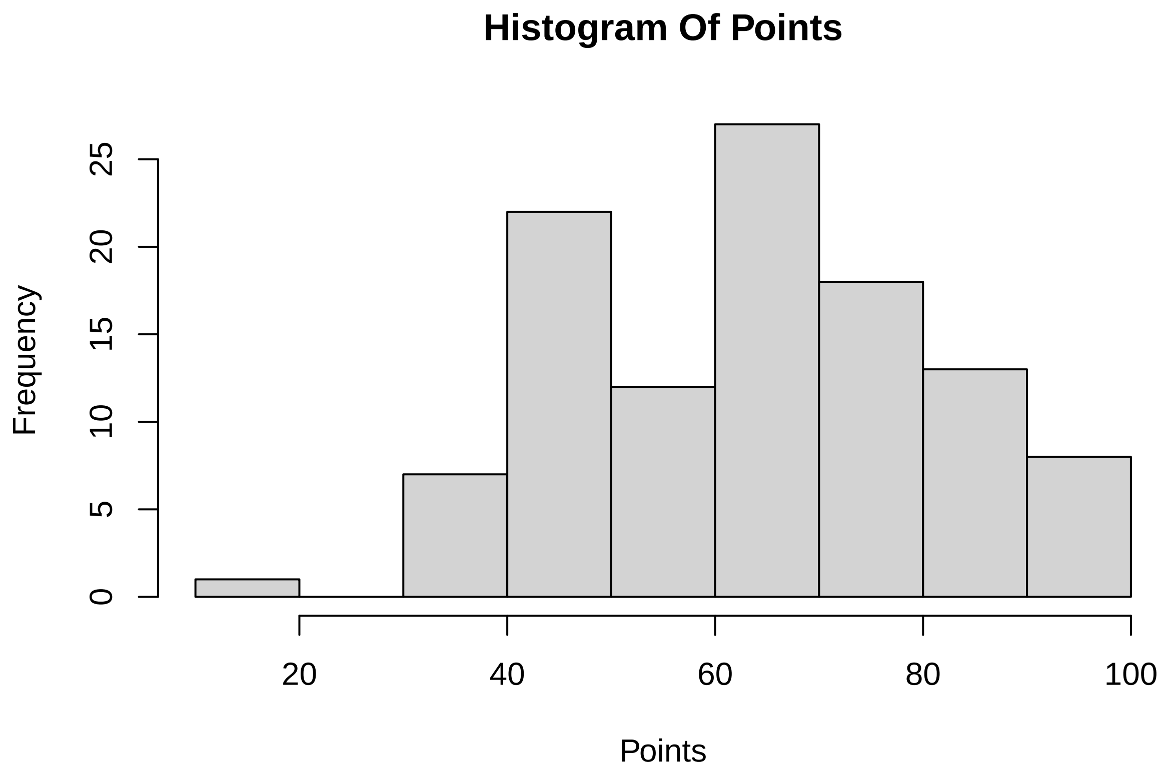
```r
main = read.csv("~/Desktop/Ball/Analysis/Master_File.csv")
head(main)
```

```
##           Team.Name Points Wage_Ratio Experience_Ratio Manager_Tenure Season
## 1           Arsenal     79  1.0515367        1.0308615           6181   2013
## 2           Everton     72  0.7544490        1.0041752             59   2013
## 3         Liverpool     84  0.9358047        0.8757246            458   2013
## 4 Manchester United     64  1.3886928        1.2878935             63   2013
## 5         Tottenham     69  0.7795104        0.7469898            426   2013
## 6           Chelsea     82  1.5117299        1.3978547             91   2013
```

Univariate Analysis

Histogram and Summary Of Points

```r
hist(main$Points,
     main ="Histogram Of Points",
     xlab = "Points",
     ylab = "Frequency")
```

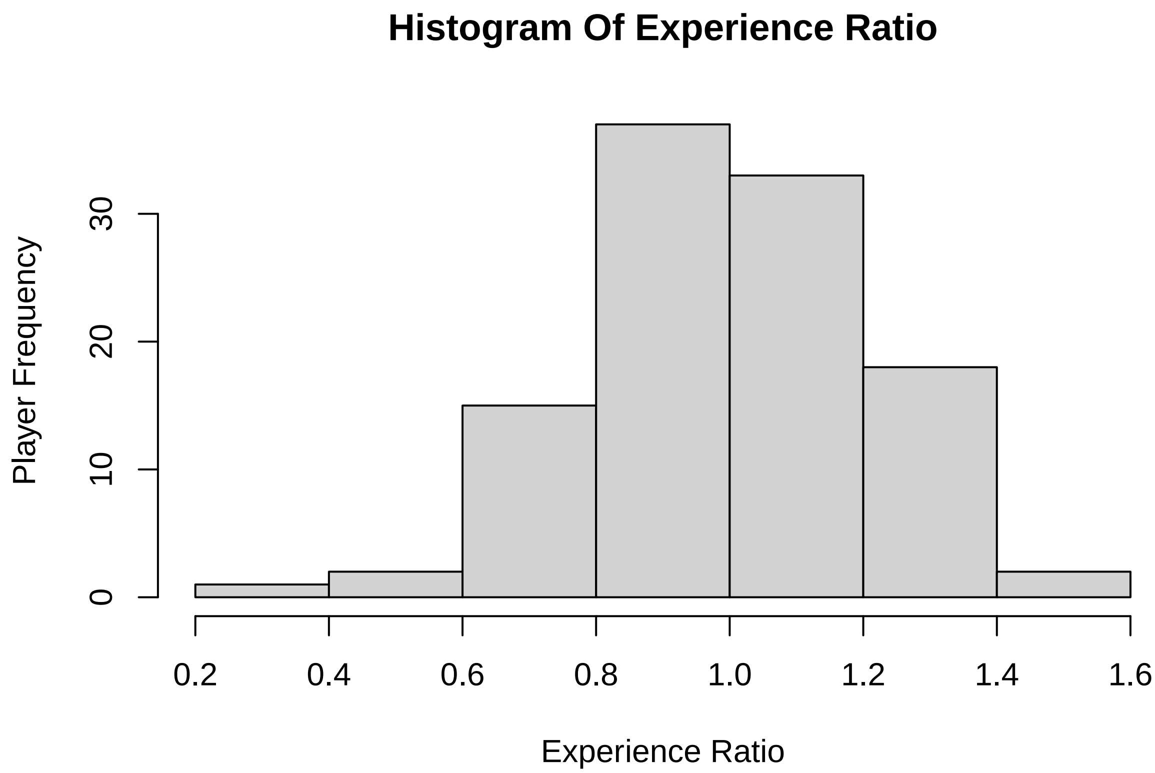


Histogram and Summary of Wage Ratio

```
hist(main$Wage_Ratio,
     main ="Histogram Of Wage Ratio",
     xlab = "Wage Ratio",
     ylab = "Frequency")
```

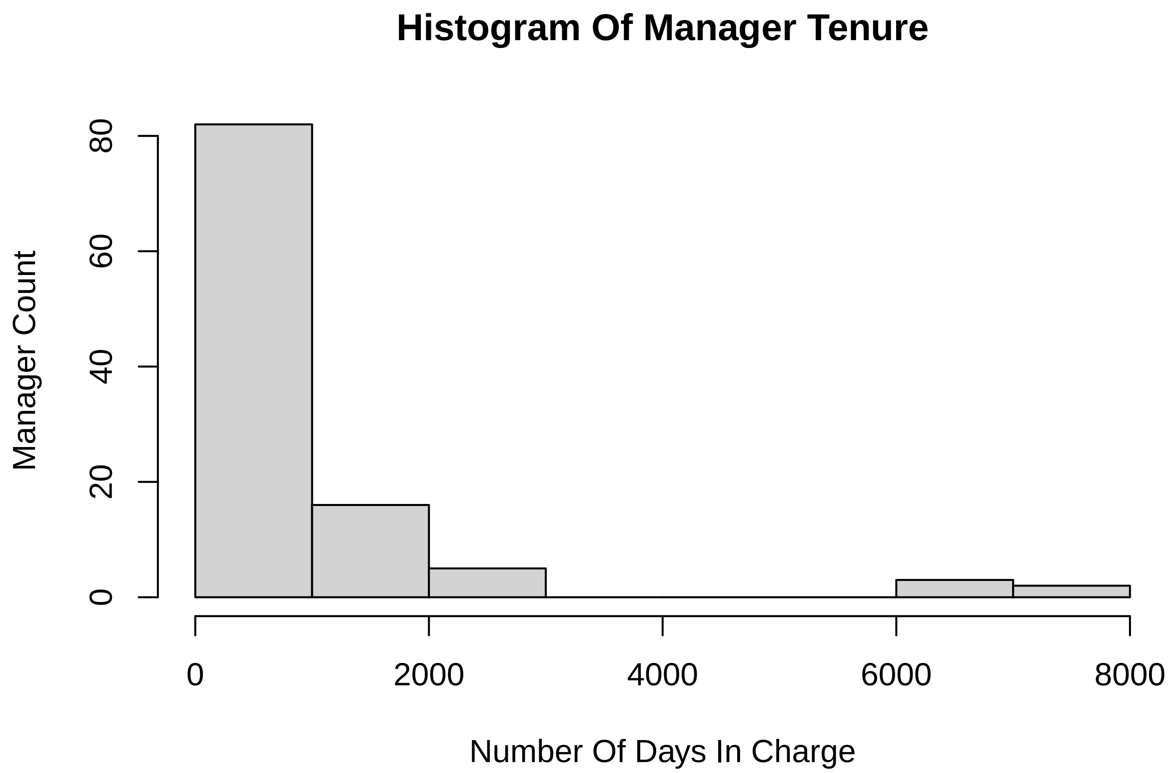
Histogram of Experience Index

```
hist(main$Experience_Ratio,
     main ="Histogram Of Experience Ratio",
     xlab="Experience Ratio",
     ylab="Player Frequency")
```



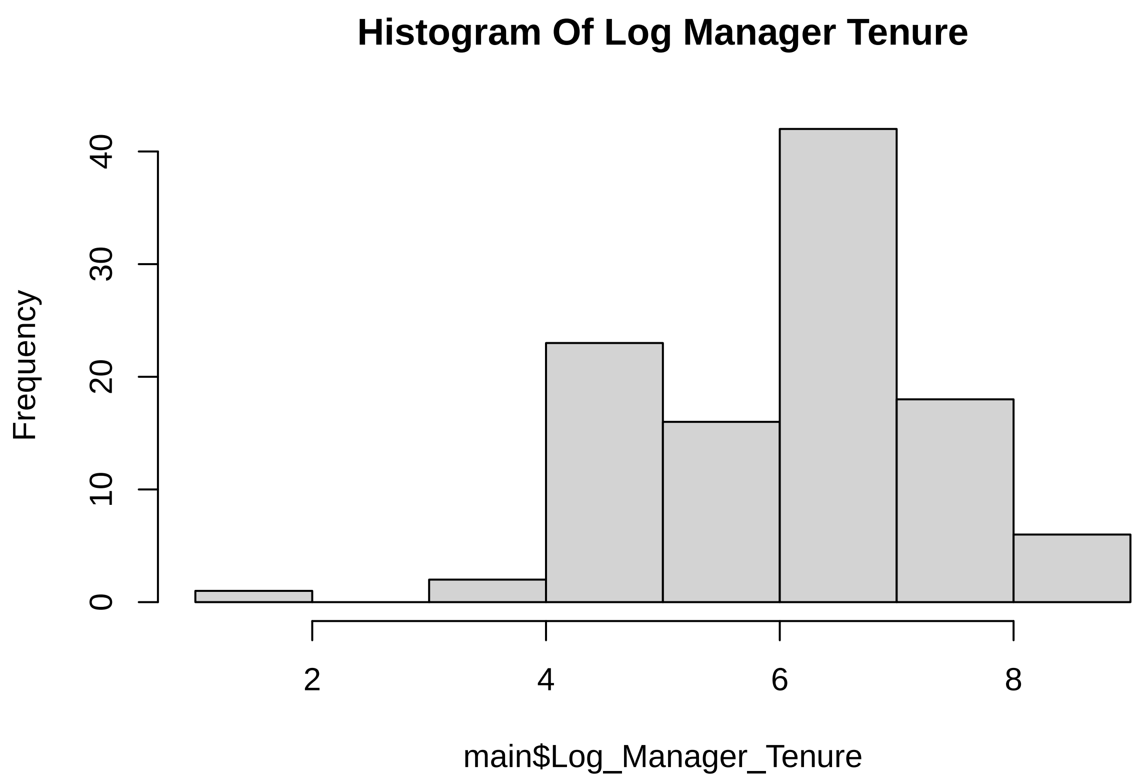
Histogram Of Manager Tenure

```r
hist(main$Manager_Tenure,
      main ="Histogram Of Manager Tenure",
     xlab="Number Of Days In Charge",
     ylab="Manager Count")
```



Logarithm Transformation

```
main$Log_Manager_Tenure=log(main$Manager_Tenure)
hist(main$Log_Manager_Tenure,main ="Histogram Of Log Manager Tenure")
```



Summary Statistics Table

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.4.3
```

```
my_vars <- c("Points", "Wage_Ratio", "Manager_Tenure", "Experience_Ratio")

get_summary_stats <- function(x) {
  c(
    Mean     = mean(x, na.rm = TRUE),
    Median   = median(x, na.rm = TRUE),
    SD       = sd(x, na.rm = TRUE),
    Min      = min(x, na.rm = TRUE),
    Max      = max(x, na.rm = TRUE),
    Skewness = skewness(x, na.rm = TRUE)
  )
}


stats_matrix <- sapply(main[my_vars], get_summary_stats)

final_table <- t(stats_matrix)
final_table <- round(final_table, 2)

print(final_table)
```

```
##                     Mean Median      SD   Min     Max Skewness
## Points             64.33     66   17.24 10.00  100.00    -0.10
## Wage_Ratio          1.00      1    0.35  0.28    1.73     0.21
## Manager_Tenure    936.73    445 1475.66  6.00 7642.00     3.16
## Experience_Ratio    1.00      1    0.21  0.28    1.59    -0.31
```

The dataset consists of 9 teams which have survived relagtion observed over 2013/2014– 2024/2025 season. Preliminary univariate analysis revealed significant skewness in key variables:

The variable ***Manager Tenure** revealed a strong right-skewed distribution. This was confirmed by the descriptive statistics in the summary statistics table, which showed a high skewness value of 3.16. Indicating that long-serving managers are rare outliers Consequently, a Log transformation was applied to normalize the variable for regression analysis.

Wage Ratio displayed a bimodal distribution, likely reflecting the financial stratification between 'Top 6' clubs and Mid-table Survivors(Everton, Crystal Palace, West Ham). However, as the skewness was low (0.21), the variable was preserved in its raw form to maintain interpretability.

Experience Ratio was found to be approximately normal (-0.31) indicating that squad age profiles are relatively consistent across both the 'Big 6' and the 'Survivor 3' and required no transformation.

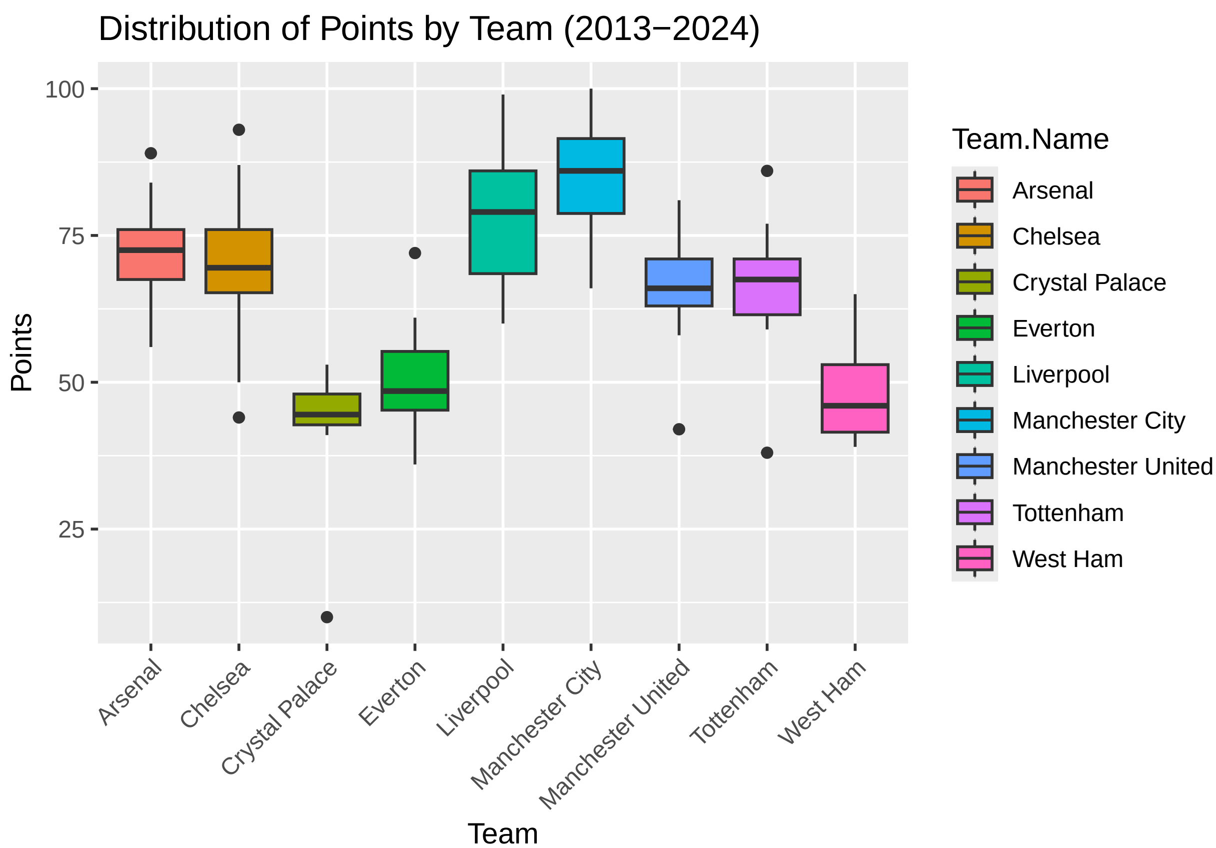
Bivariate Analysis

Boxplot of Points grouped by Teams

```r
library(ggplot2)
```

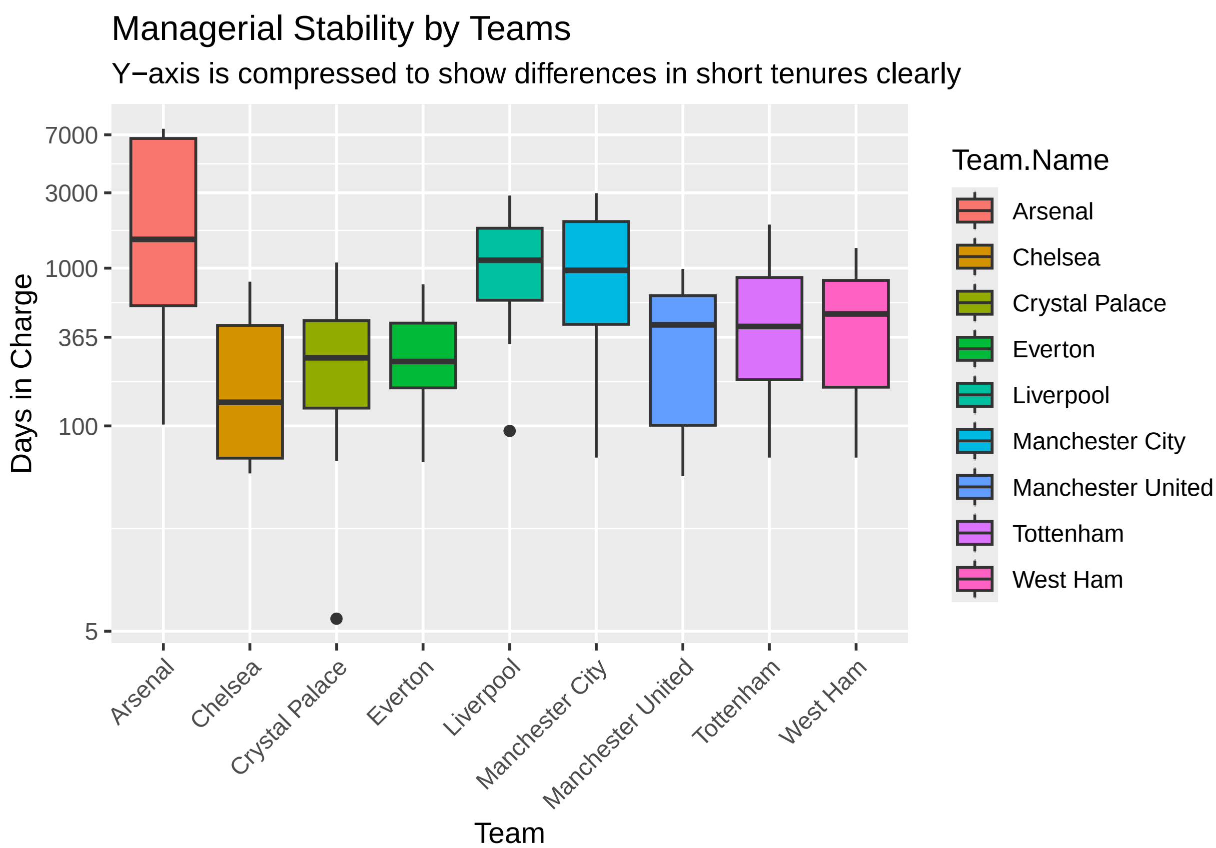
```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:e1071':
##
##     element
```

```r
ggplot(main, aes(x = Team.Name, y = Points, fill = Team.Name)) +
  geom_boxplot() +
  labs(
    title = "Distribution of Points by Team (2013-2024)",
    x = "Team",
    y = "Points"
  )+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



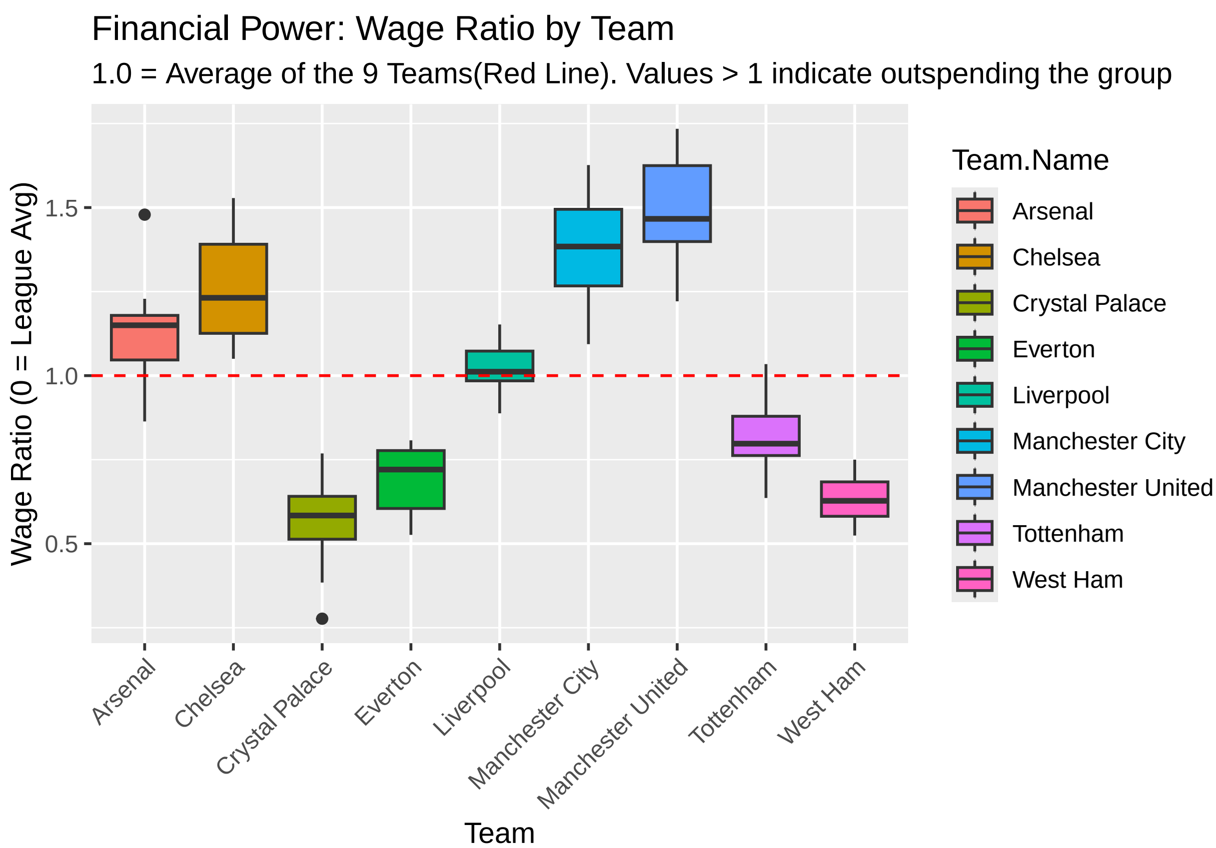
Boxplot of Manager Tenure grouped by Teams

```
ggplot(main, aes(x = Team.Name, y = Manager_Tenure, fill = Team.Name)) +
  geom_boxplot() +
  scale_y_log10(breaks = c(5,100, 365, 1000, 3000, 7000)) +
  labs(
    title = "Managerial Stability by Teams",
    subtitle = "Y-axis is compressed to show differences in short tenures clearly",
    x = "Team",
    y = "Days in Charge"
  ) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



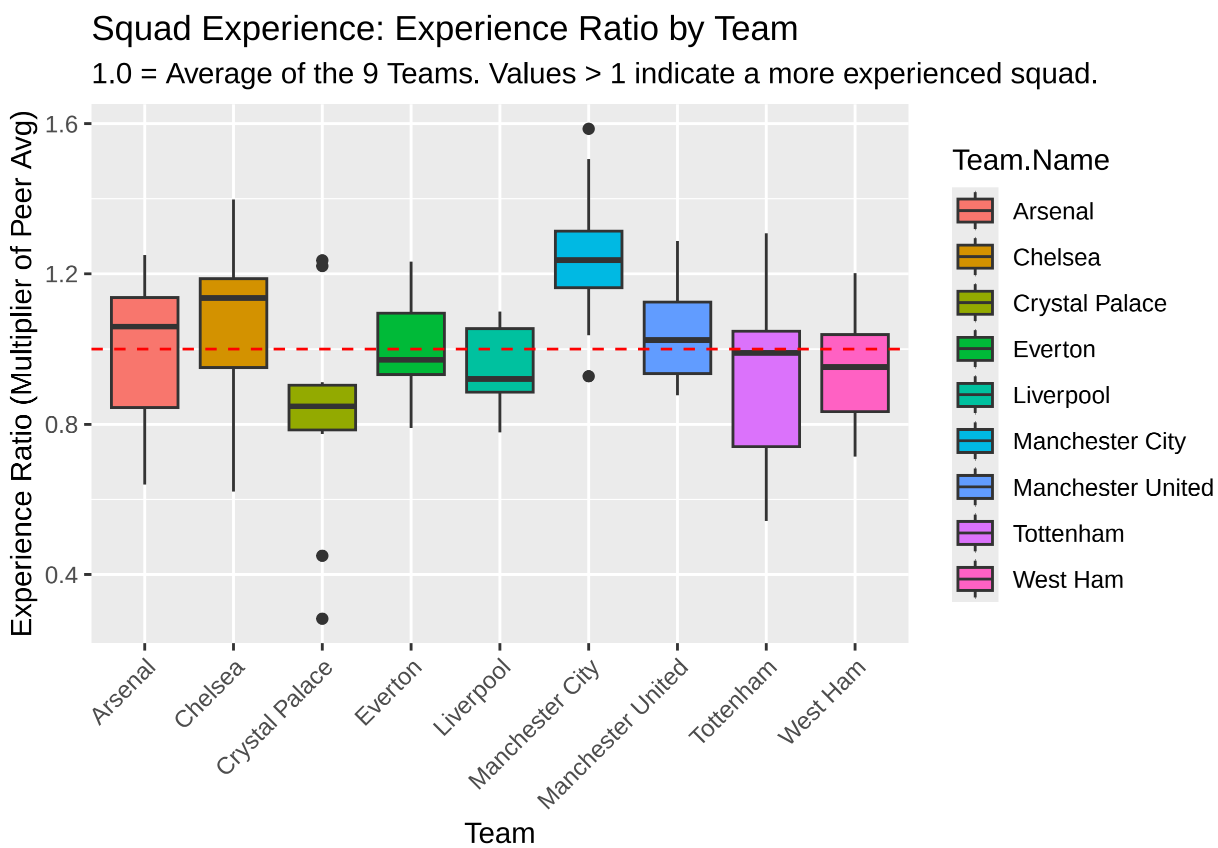
Boxplot of Wage Ratio By Teams

```
ggplot(main, aes(x = Team.Name, y = Wage_Ratio, fill = Team.Name)) +
  geom_boxplot() +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red",) +
  labs(
    title = "Financial Power: Wage Ratio by Team",
    subtitle = "1.0 = Average of the 9 Teams(Red Line). Values > 1 indicate outspending the group",,
    x = "Team",
    y = "Wage Ratio (0 = League Avg)"
  )+
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```



Boxplot of Experience Ratio By Teams

```
ggplot(main, aes(x = Team.Name, y = Experience_Ratio, fill = Team.Name)) +
  geom_boxplot() +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red",) +
  labs(
    title = "Squad Experience: Experience Ratio by Team",
       subtitle = "1.0 = Average of the 9 Teams. Values > 1 indicate a more experienced squad.",
       y = "Experience Ratio (Multiplier of Peer Avg)",
       x = "Team") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
```

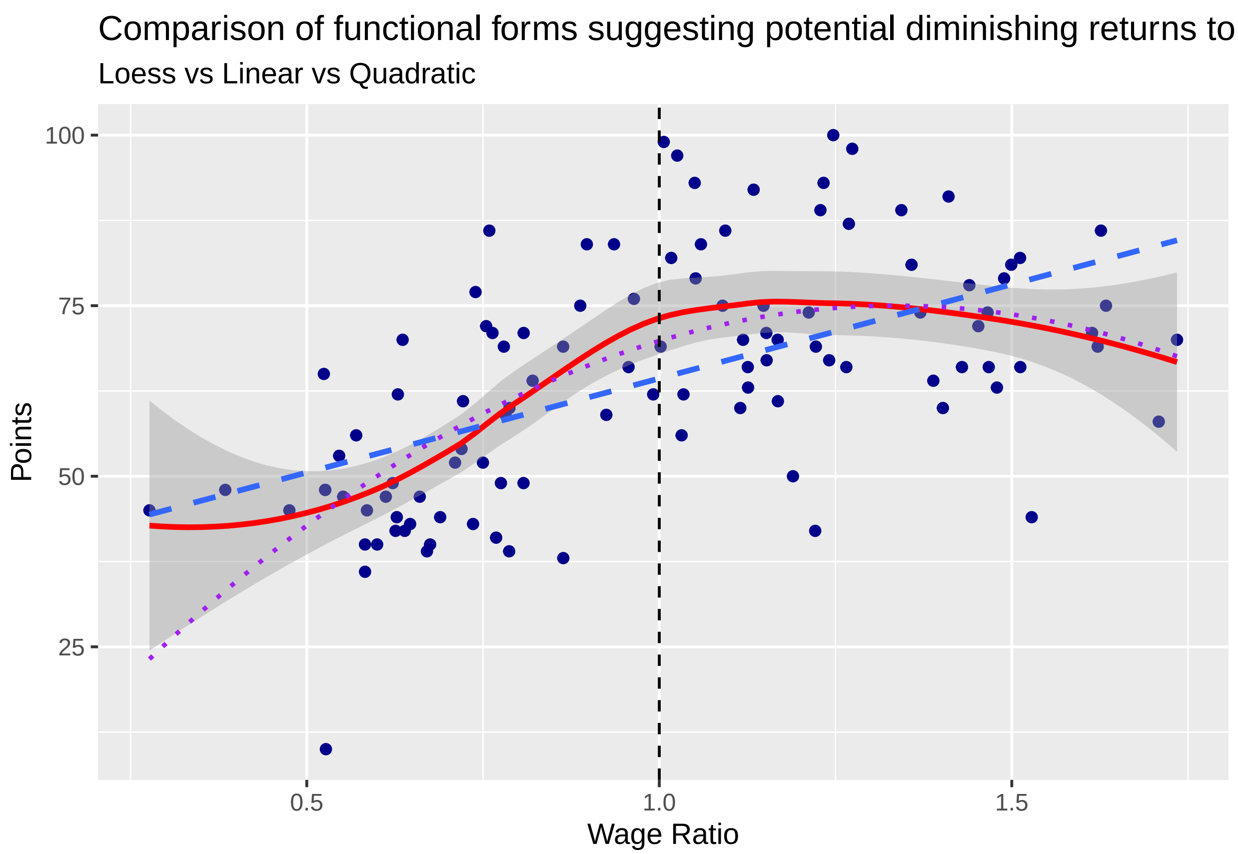


1.)Points vs. Financial Power (The "Diminishing Returns" Check)

```
ggplot(main, aes(x = Wage_Ratio, y = Points)) +
  geom_point( color = "darkblue") +  # The dots
  geom_smooth(method = "loess", color = "red", size = 1) +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, linetype = 2) +
  geom_smooth(method = "lm", formula = y ~ x + I(x^2), se = FALSE, color ="purple", linetype = 3, size =
  geom_vline(xintercept = 1.0, linetype = "dashed", color = "black") +
  labs(title = "Comparison of functional forms suggesting potential diminishing returns to wage spending
       subtitle = "Loess vs Linear vs Quadratic",
       x = "Wage Ratio",
       y = "Points")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

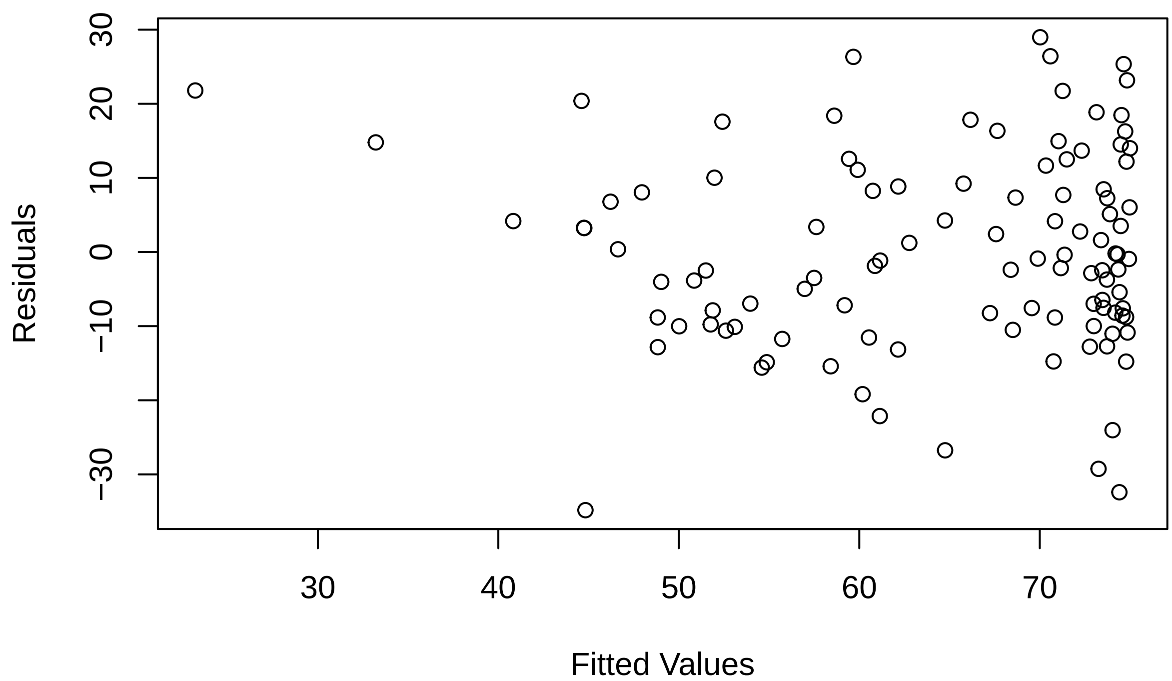



```
model_wage_quad = lm(Points ~Wage_Ratio + I(Wage_Ratio^2), data = main)
summary(model_wage_quad)
```

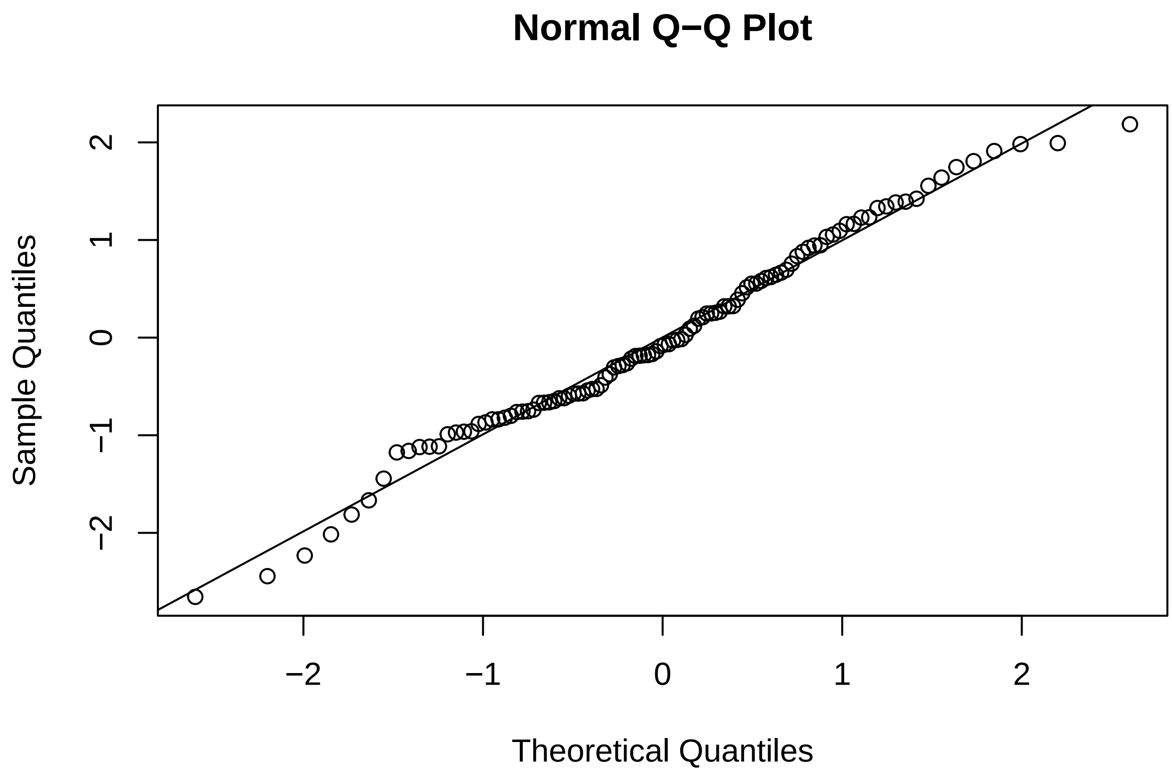
```
##
## Call:
## lm(formula = Points ~ Wage_Ratio + I(Wage_Ratio^2), data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.827  -8.833  -1.049   8.936  28.975
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -7.464     11.119  -0.671    0.503
## Wage_Ratio       123.628     22.880   5.403 4.11e-07 ***
## I(Wage_Ratio^2)  -46.333     10.889  -4.255 4.55e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.38 on 105 degrees of freedom
## Multiple R-squared:  0.4085, Adjusted R-squared:  0.3972
## F-statistic: 36.25 on 2 and 105 DF,  p-value: 1.067e-12
```

```
wageq.anova=aov(Points ~ Wage_Ratio + I(Wage_Ratio^2), data=main)
wageq.fitted = fitted.values(wageq.anova)
```

```r
wageq.residuals = residuals(wageq.anova)

# Residual Plot
plot(wageq.fitted,wageq.residuals, xlab = "Fitted Values", ylab = "Residuals")
```



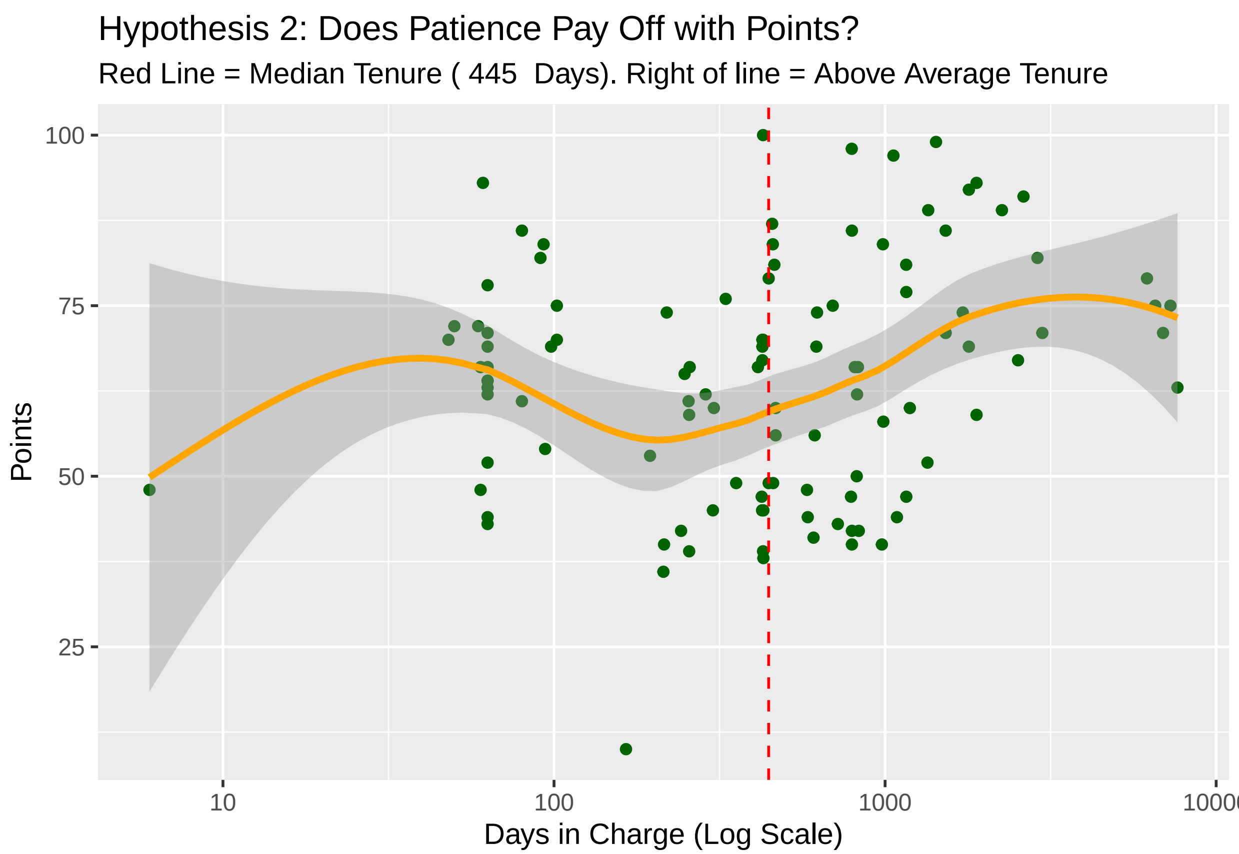

```r
#QQ Plot
wageq.stdres = rstandard(wageq.anova)
qqnorm(wageq.stdres)
qqline(wageq.stdres)
```



2.) Points vs. Managerial Stability (The "Patience" Check)

```
middle_tenure = median(main$Manager_Tenure)


ggplot(main,aes(x=Manager_Tenure, y= Points)) +
  geom_point(color = "darkgreen") +
  geom_smooth(method = "loess", color = "orange", size = 1.2) +
  geom_vline(xintercept = middle_tenure, linetype ="dashed", color="red")+
  scale_x_log10() + # Used log scale on X axis so short tenures aren't squashed
  labs(title = "Hypothesis 2: Does Patience Pay Off with Points?",
       subtitle =paste("Red Line = Median Tenure (", round(middle_tenure), " Days). Right of line = Abo
       x = "Days in Charge (Log Scale)",
       y = "Points")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

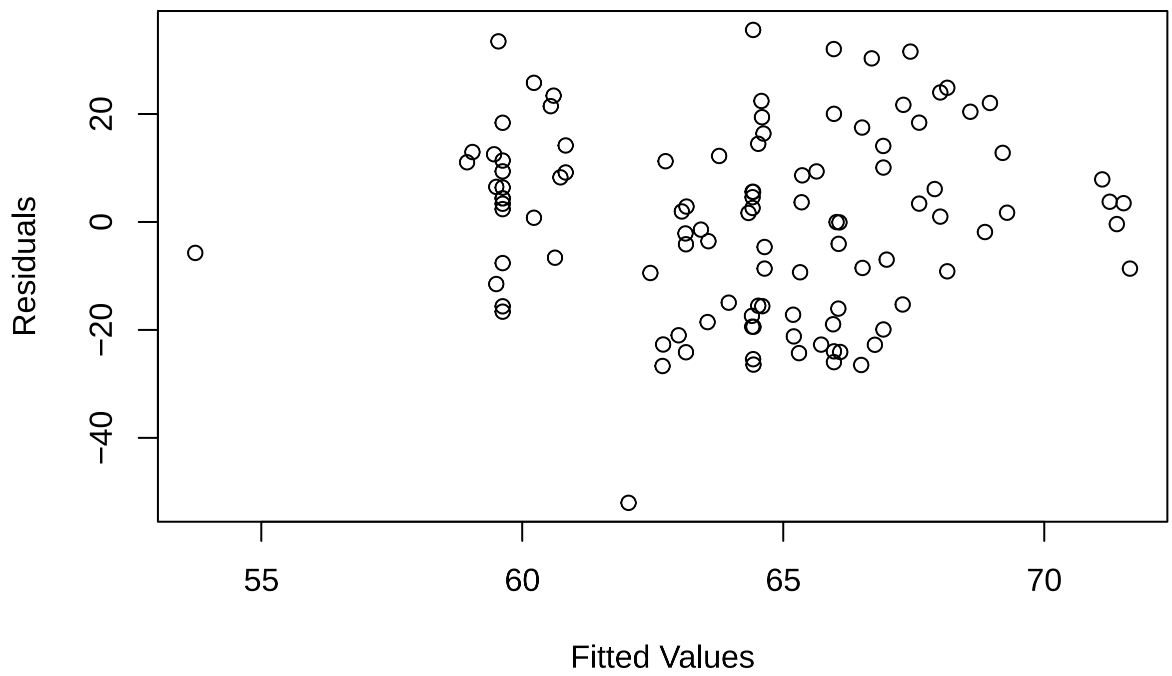



```
model_tenure = lm(Points ~Log_Manager_Tenure, data = main)
summary(model_tenure)
```

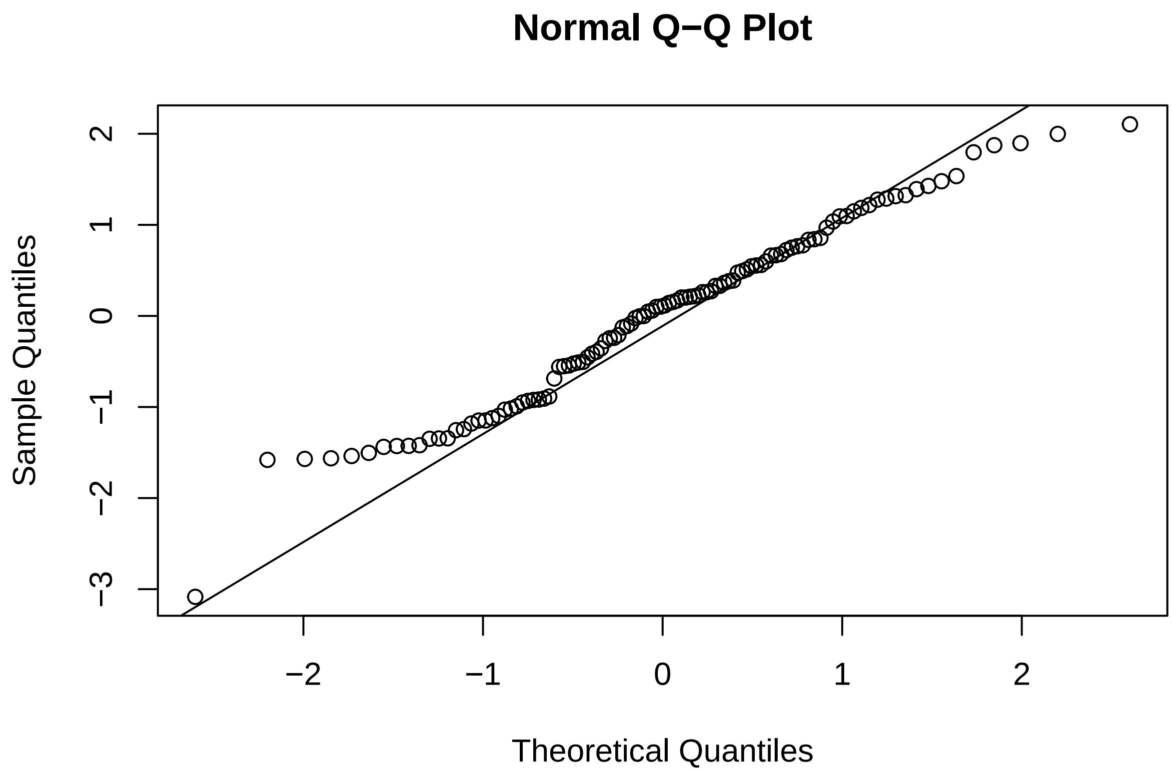
```
##
## Call:
## lm(formula = Points ~ Log_Manager_Tenure, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.035 -15.345   1.829  11.590  35.578
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            49.245      7.536   6.534 2.28e-09 ***
## Log_Manager_Tenure      2.505      1.221   2.051   0.0427 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.99 on 106 degrees of freedom
## Multiple R-squared:  0.03817,    Adjusted R-squared:  0.0291
## F-statistic: 4.207 on 1 and 106 DF,  p-value: 0.04273
```

```
tenure.anova=aov(Points~Log_Manager_Tenure, data=main)
tenure.fitted = fitted.values(tenure.anova)
tenure.residuals = residuals(tenure.anova)
```

```r
# Residual Plot
plot(tenure.fitted,tenure.residuals, xlab = "Fitted Values", ylab = "Residuals")
```



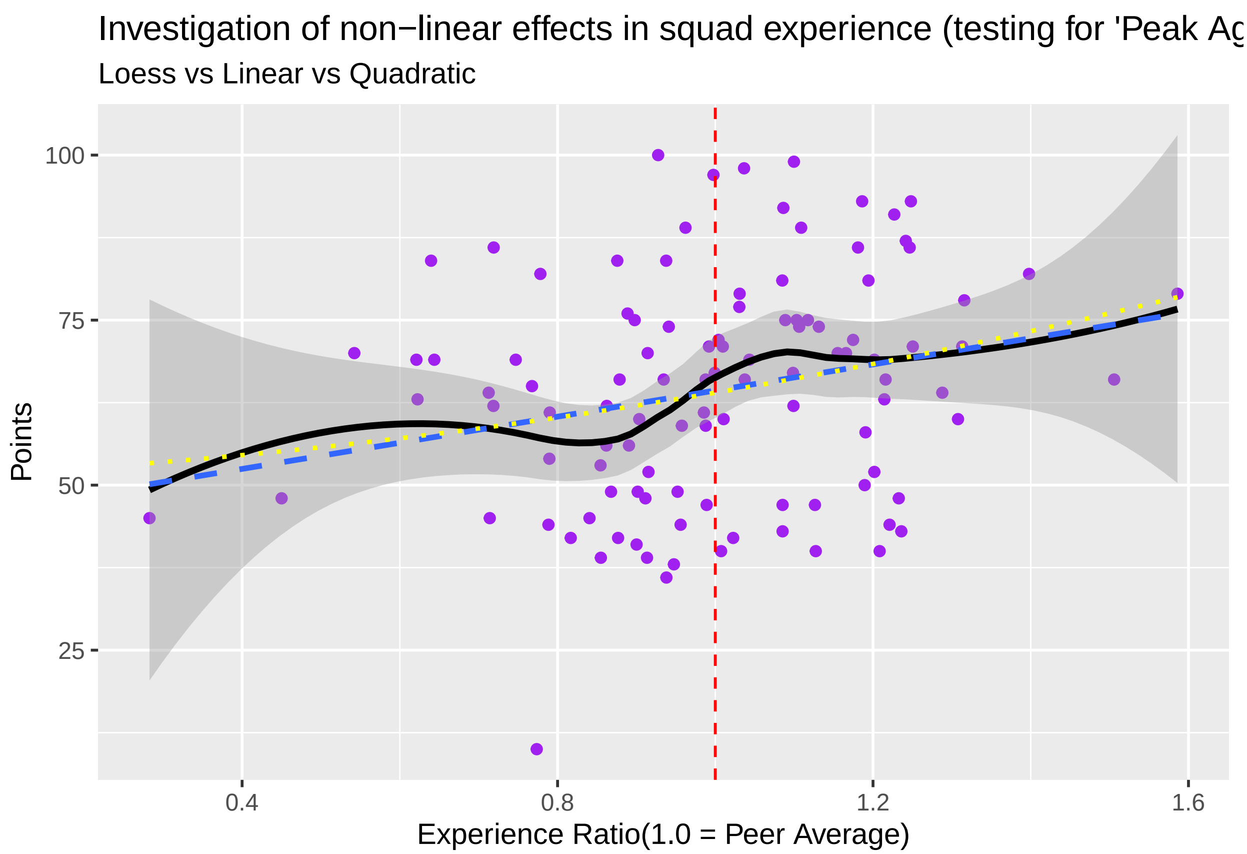

```r
#QQ Plot
tenure.stdres = rstandard(tenure.anova)
qqnorm(tenure.stdres)
qqline(tenure.stdres)
```



3.)Points Vs. Experience Ratio

```
ggplot(main,aes(x=Experience_Ratio, y= Points)) +
  geom_point(color = "purple") +
  geom_smooth(method = "loess", color = "black", size = 1.2) +
    geom_smooth(method = "lm", formula = y ~ x, se = FALSE, linetype = 2) +
  geom_smooth(method = "lm", formula = y ~ x + I(x^2), se = FALSE, color ="yellow", linetype = 3, size =
  geom_vline(xintercept = 1.0, linetype = "dashed", color = "red") +
  labs(title = "Investigation of non-linear effects in squad experience (testing for 'Peak Age' efficie
       x = "Experience Ratio(1.0 = Peer Average)",
       y = "Points")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

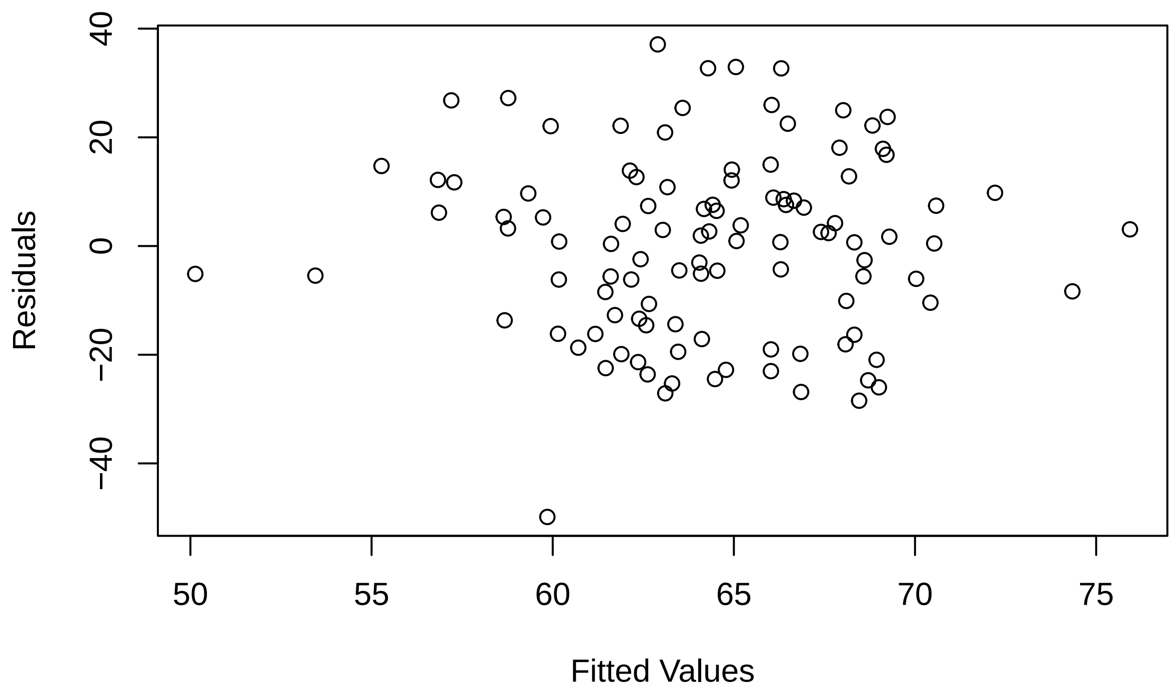



```r
model_exp = lm(Points ~ Experience_Ratio, data = main)
summary(model_exp)
```

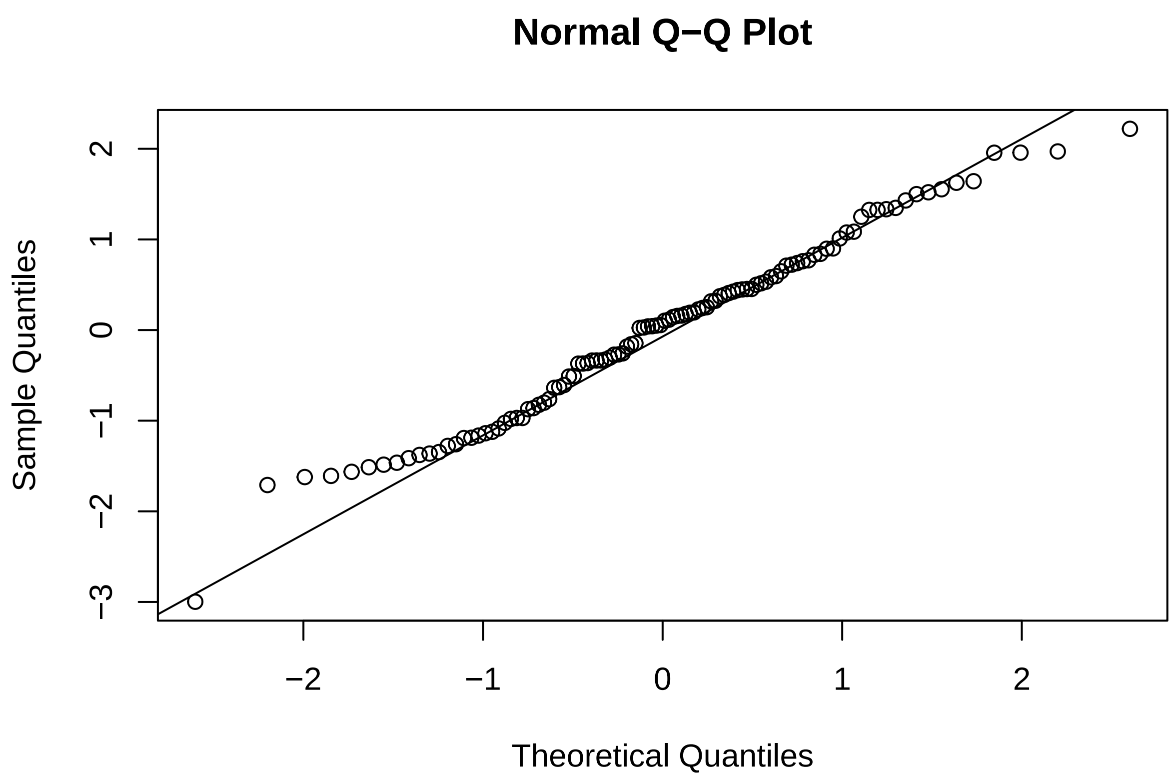
```
##
## Call:
## lm(formula = Points ~ Experience_Ratio, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.852 -13.458   1.318  11.054  37.100
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        44.541      7.832   5.687 1.15e-07 ***
## Experience_Ratio   19.793      7.663   2.583   0.0112 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.8 on 106 degrees of freedom
## Multiple R-squared:  0.0592, Adjusted R-squared:  0.05033
## F-statistic: 6.671 on 1 and 106 DF,  p-value: 0.01117
```

```r
exp.anova=aov(Points~Experience_Ratio, data=main)
exp.fitted = fitted.values(exp.anova)
exp.residuals = residuals(exp.anova)
```

```r
# Residual Plot
plot(exp.fitted,exp.residuals, xlab = "Fitted Values", ylab = "Residuals")
```



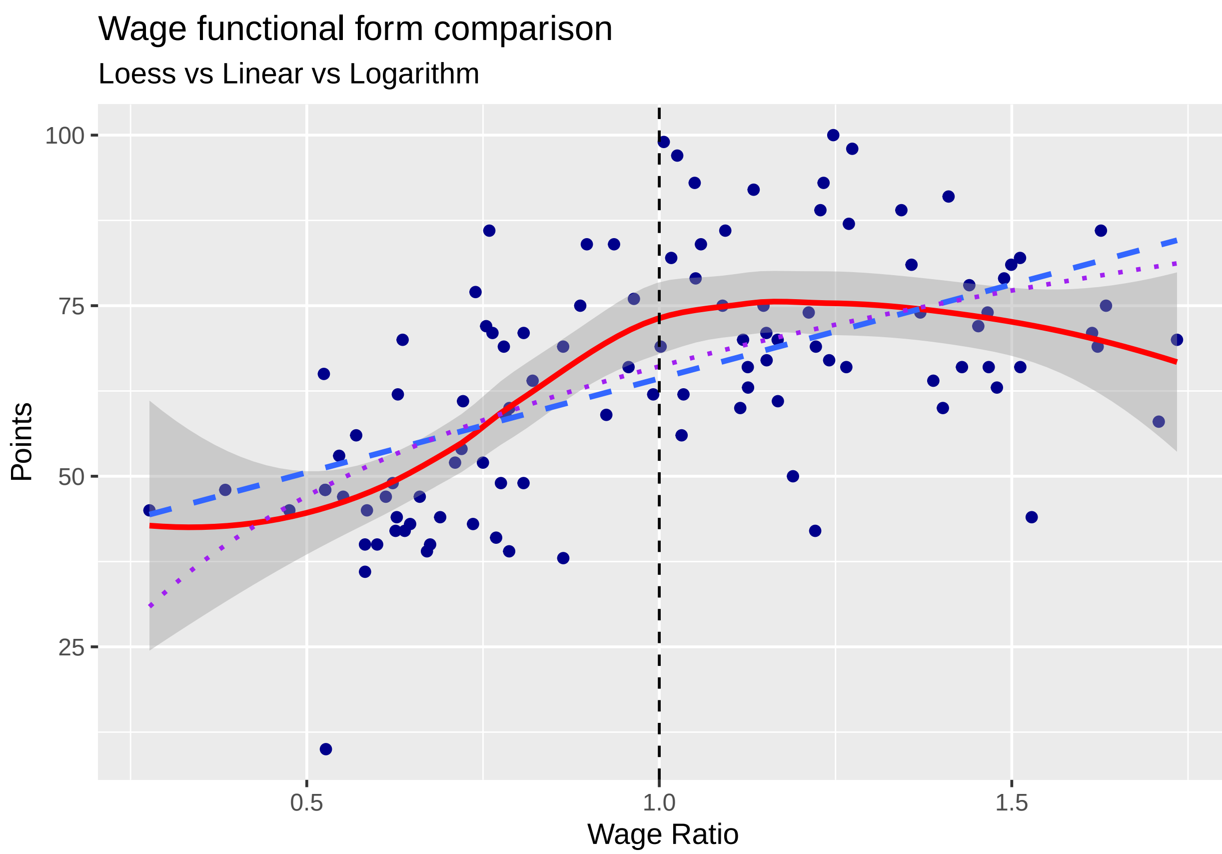

```r
#QQ Plot
exp.stdres = rstandard(exp.anova)
qqnorm(exp.stdres)
qqline(exp.stdres)
```



Bivariate log wage model

```
# 1C) Bivariate log wage model
ggplot(main, aes(x = Wage_Ratio, y = Points)) +
  geom_point( color = "darkblue") +  # The dots
  geom_smooth(method = "loess", color = "red", size = 1) +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE, linetype = 2) +
    geom_smooth(method="lm", formula = y ~ log(x), se=FALSE,color="purple",size =0.8, linetype=3)+
  geom_vline(xintercept = 1.0, linetype = "dashed", color = "black") +
  labs(title = "Wage functional form comparison",
       subtitle = "Loess vs Linear vs Logarithm",
       x = "Wage Ratio",
       y = "Points")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

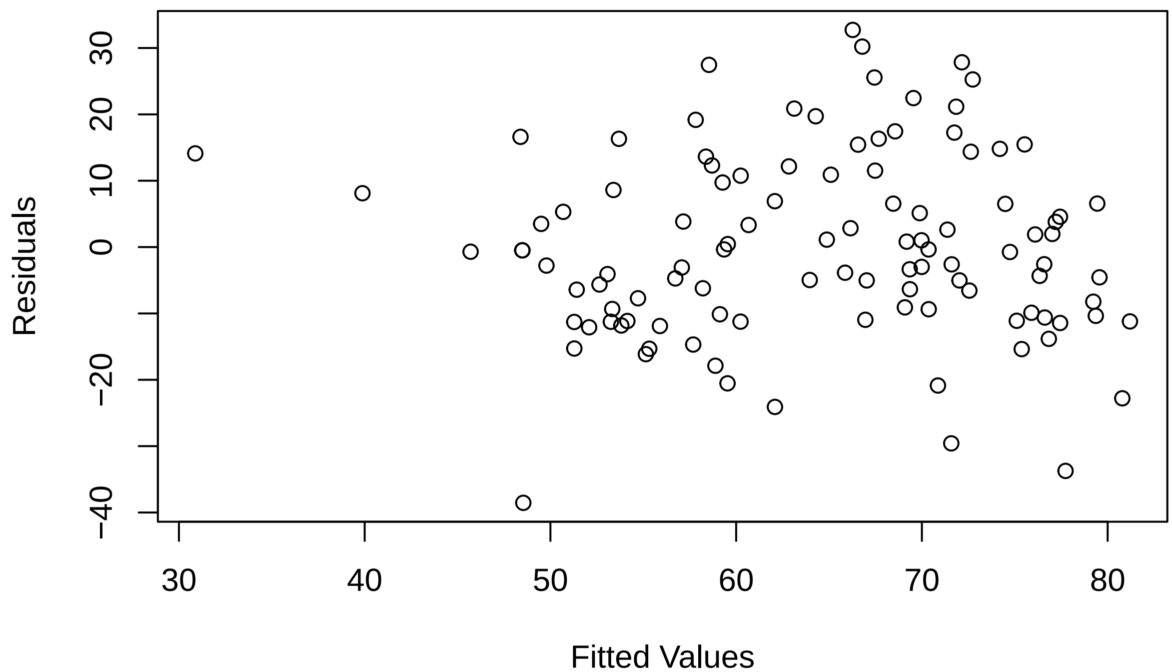



```r
model_wage_log = lm(Points ~log(Wage_Ratio ), data = main)
summary(model_wage_log)
```

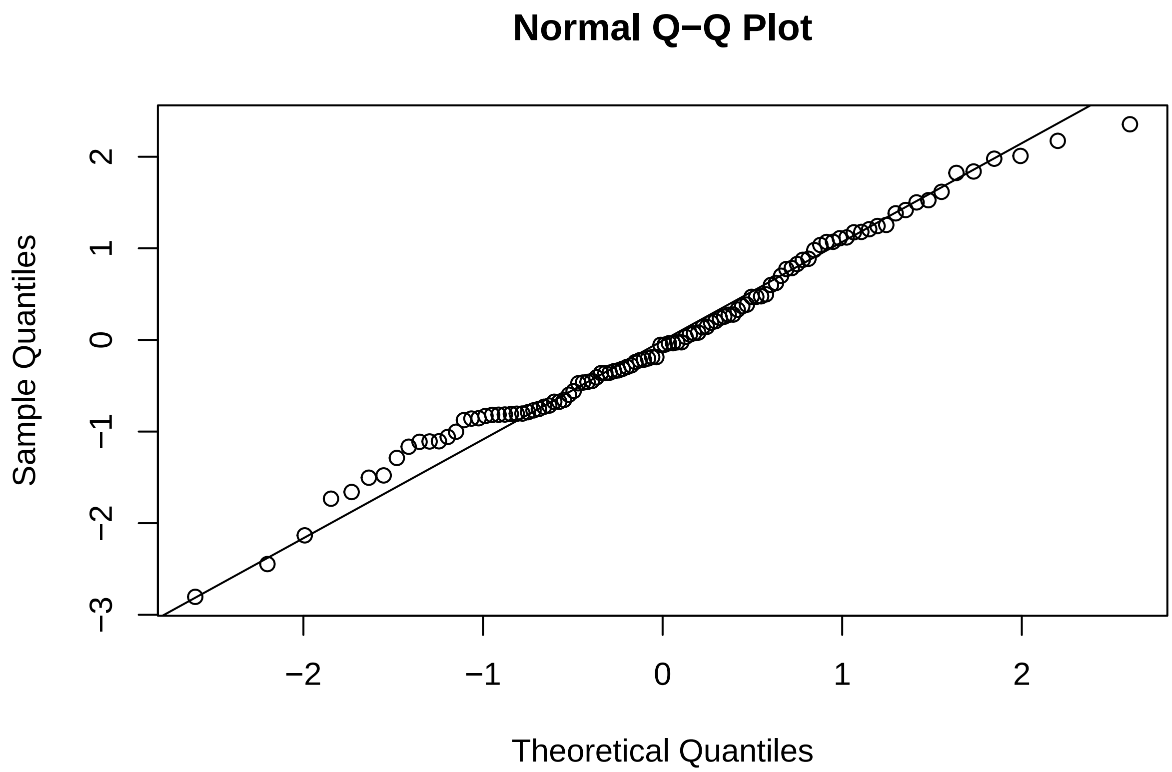
```
##
## Call:
## lm(formula = Points ~ log(Wage_Ratio), data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.541 -10.184  -0.722   9.986  32.725
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       66.102      1.364  48.454  < 2e-16 ***
## log(Wage_Ratio)   27.423      3.633   7.549 1.61e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.97 on 106 degrees of freedom
## Multiple R-squared:  0.3497, Adjusted R-squared:  0.3435
## F-statistic: 56.99 on 1 and 106 DF,  p-value: 1.607e-11
```

```r
wagelog.anova=aov(Points ~log(Wage_Ratio ), data=main)
wagelog.fitted = fitted.values(wagelog.anova)
wagelog.residuals = residuals(wagelog.anova)
```

```r
# Residual Plot
plot(wagelog.fitted,wagelog.residuals, xlab = "Fitted Values", ylab = "Residuals")
```




```r
#QQ Plot
wagelog.stdres = rstandard(wagelog.anova)
qqnorm(wagelog.stdres)
qqline(wagelog.stdres)
```



Visual inspection suggested a stratified league structure. Boxplots of points by team show distinct competitive tiers, with clubs like Manchester City consistently outperforming peers like Crystal Palace.

Initial bivariate scatterplots tested three hypotheses:

Diminishing Returns: A Loess curve fitted to Wages vs. Points suggested a potential non-linear relationship, warranting tests for quadratic effects.

Stability: The relationship between Tenure and Points appeared weak, with a low correlation visible in the scatterplot.

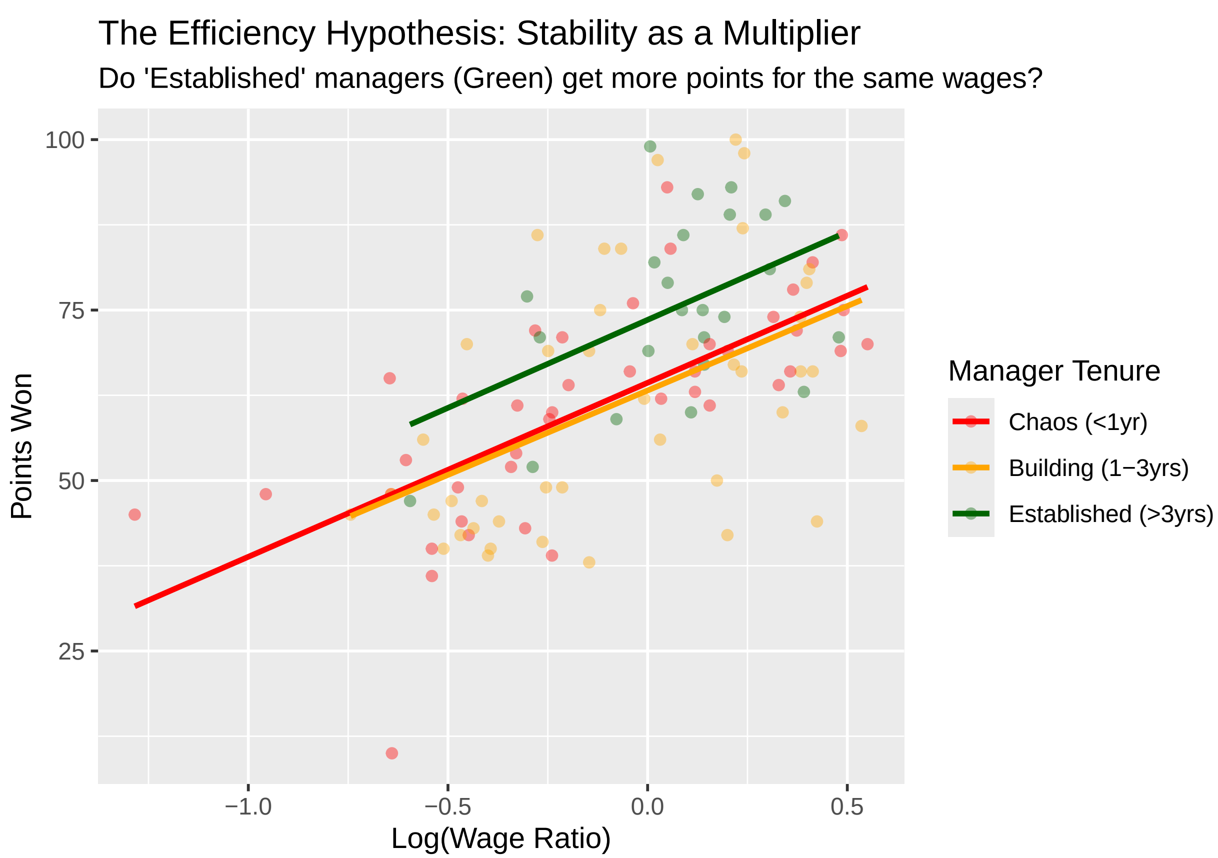
Experience: The 'Experience Efficiency' plot hinted that veteran squads might offer slightly better returns on wages, prompting an investigation into interaction effects."

Interaction Plots

```
# 1. Binning Manager Stability (Tenure)
main$Stability_Status <- cut(main$Manager_Tenure,
                             breaks = c(-Inf, 365, 1095, Inf),
                             labels = c("Chaos (<1yr)", "Building (1-3yrs)", "Established (>3yrs)"))

# 2. Binning Experience (Young vs Veteran)
main$Squad_Maturity <- ifelse(main$Experience_Ratio > median(main$Experience_Ratio),
                              "Veteran Squad",
                              "Young Squad")

main$Squad_Maturity <- as.factor(main$Squad_Maturity)
main$Stability_Status <- as.factor(main$Stability_Status)
```

```r
ggplot(main, aes(x = log(Wage_Ratio), y=Points, color = Stability_Status))+
  geom_point(alpha=0.4)+
  geom_smooth(method = "lm", se = FALSE) +
  scale_color_manual(values = c("Chaos (<1yr)" = "red",
                                "Building (1-3yrs)" = "orange",
                                "Established (>3yrs)" = "darkgreen")) +

  labs(title = "The Efficiency Hypothesis: Stability as a Multiplier",
       subtitle = "Do 'Established' managers (Green) get more points for the same wages?",
       x = "Log(Wage Ratio)",
       y = "Points Won",
       color = "Manager Tenure")
```

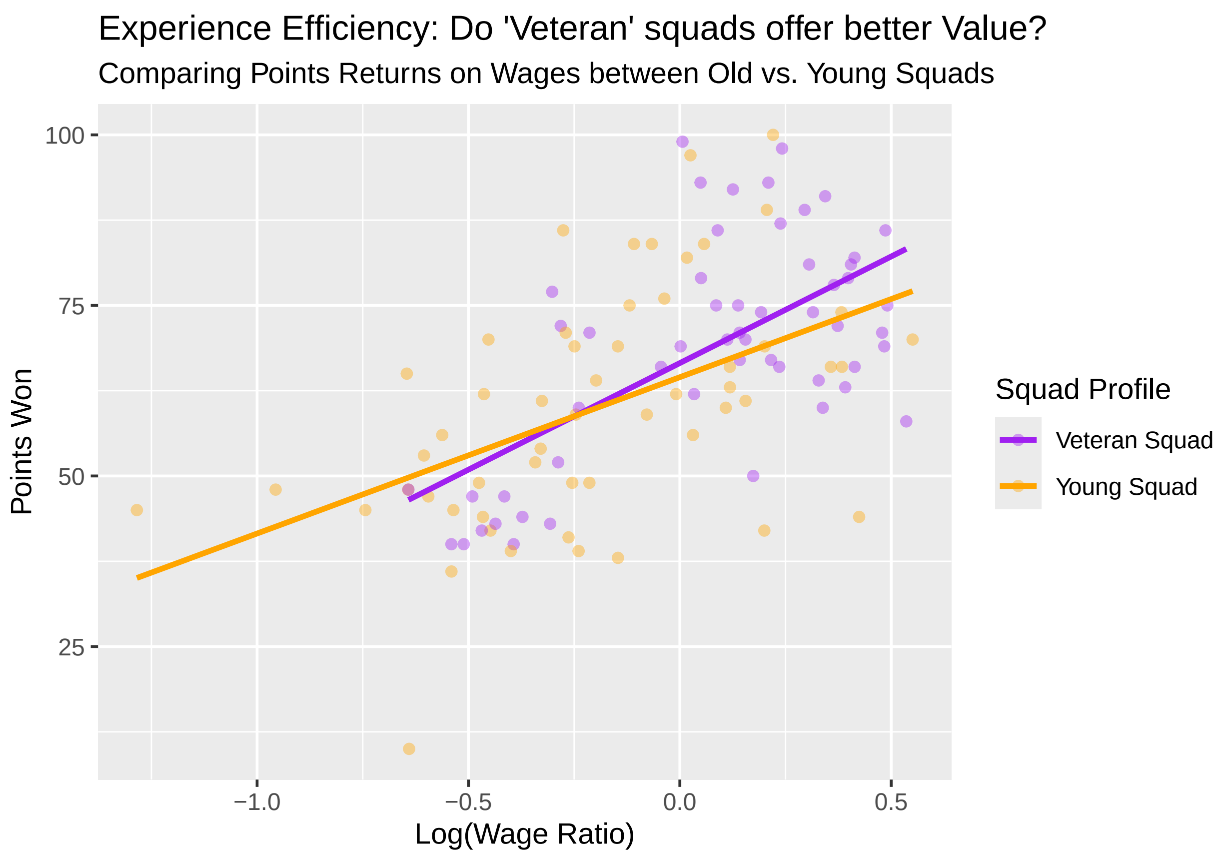
```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
ggplot(main, aes(x = log(Wage_Ratio), y = Points, color = Squad_Maturity)) +
  geom_point(alpha=0.4) +
  geom_smooth(method = "lm", se = FALSE,) +
  scale_color_manual(values = c("Veteran Squad" = "purple",
                                "Young Squad" = "orange")) +

  labs(title = "Experience Efficiency: Do 'Veteran' squads offer better Value?",
       subtitle = "Comparing Points Returns on Wages between Old vs. Young Squads",
       x = "Log(Wage Ratio)",
```

```
         y = "Points Won",
         color = "Squad Profile")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

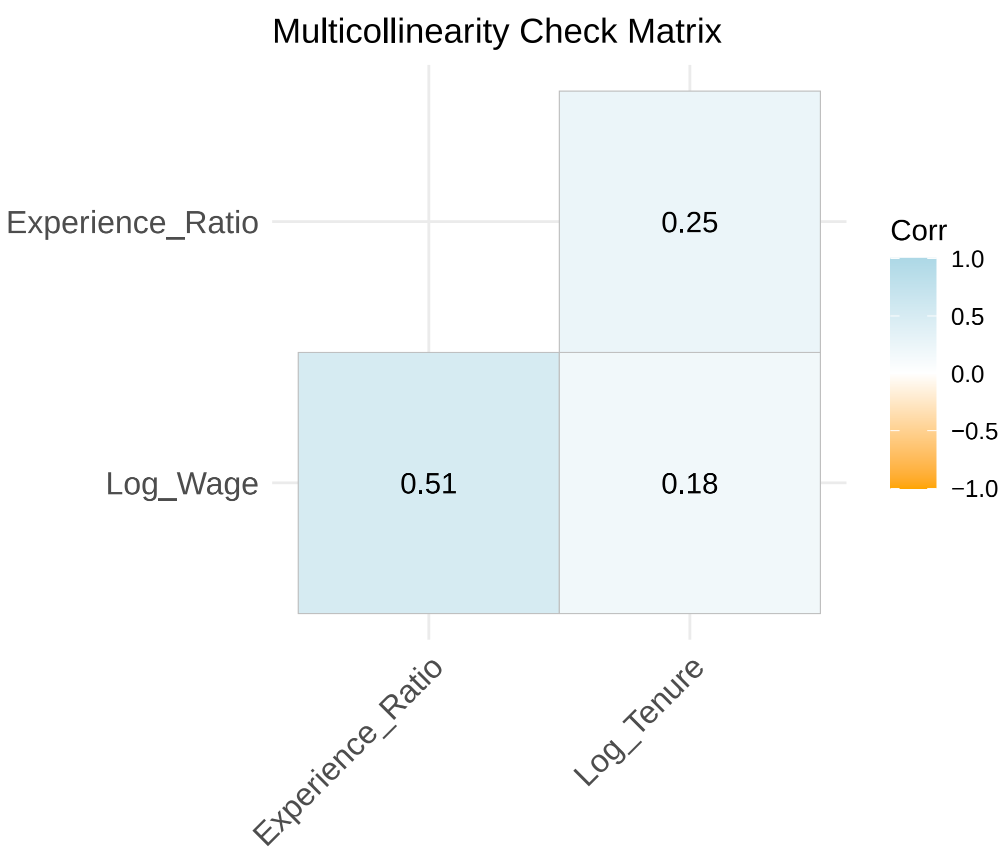



```
library(ggcorrplot)
```

```
## Warning: package 'ggcorrplot' was built under R version 4.4.3
```

```
main$Log_Wage <- log(main$Wage_Ratio)
main$Log_Tenure <- log(main$Manager_Tenure)

my_predictors <- main[, c("Log_Wage", "Experience_Ratio", "Log_Tenure")]

cor_matrix = cor(my_predictors)

ggcorrplot(cor_matrix,
           method = "square",
           type = "lower",
           lab = TRUE,
           colors = c("orange", "white", "lightblue"),
           title = "Multicollinearity Check Matrix")
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## i The deprecated feature was likely used in the ggcorrplot package.
##   Please report the issue at <https://github.com/kassambara/ggcorrplot/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Regression Models ***Simple Model***

```
base.lm=lm(Points~Log_Wage+Log_Tenure+Experience_Ratio+Team.Name,
            data =main)
summary(base.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage + Log_Tenure + Experience_Ratio +
##     Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.034  -5.673   0.508   5.933  20.946
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                80.3945     8.8934   9.040 1.72e-14 ***
## Log_Wage                   -7.5405     8.6858  -0.868  0.38749
## Log_Tenure                 -1.1761     0.9722  -1.210  0.22935
## Experience_Ratio            1.2677     6.9552   0.182  0.85576
## Team.NameChelsea           -4.2070     5.0905  -0.826  0.41061
## Team.NameCrystal Palace   -37.1649     7.6774  -4.841 4.94e-06 ***
## Team.NameEverton          -27.9430     6.5836  -4.244 5.06e-05 ***
## Team.NameLiverpool          5.4870     4.5901   1.195  0.23488
## Team.NameManchester City   13.0825     4.9410   2.648  0.00947 **
## Team.NameManchester United  -6.2024    5.2810  -1.174  0.24311
## Team.NameTottenham        -10.1185     5.4248  -1.865  0.06520 .
## Team.NameWest Ham         -29.5108     6.7910  -4.346 3.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.04 on 96 degrees of freedom
## Multiple R-squared:  0.6318, Adjusted R-squared:  0.5896
## F-statistic: 14.98 on 11 and 96 DF,  p-value: < 2.2e-16
```

Complex Models ****Quadratic Model***

```
complex.lm=lm(Points ~ Log_Wage + I(Log_Wage^2)+ Log_Tenure + I(Log_Tenure^2) + Experience_Ratio +Team.N
              data = main
)

summary(complex.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage + I(Log_Wage^2) + Log_Tenure +
##     I(Log_Tenure^2) + Experience_Ratio + Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.528  -5.901   0.848   5.445  20.590
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                99.3847    16.7778   5.924 5.14e-08 ***
## Log_Wage                  -12.8196    10.2708  -1.248 0.215074
## I(Log_Wage^2)              -8.2745     8.5145  -0.972 0.333643
## Log_Tenure                 -8.2434     5.6068  -1.470 0.144834
## I(Log_Tenure^2)             0.6107     0.4815   1.268 0.207828
## Experience_Ratio            0.7642     6.9643   0.110 0.912861
## Team.NameChelsea           -1.3613     5.4519  -0.250 0.803369
## Team.NameCrystal Palace   -36.0937     7.7838  -4.637 1.14e-05 ***
## Team.NameEverton          -27.0581     6.9924  -3.870 0.000201 ***
## Team.NameLiverpool          6.4108     4.7800   1.341 0.183094
## Team.NameManchester City   16.4635     5.4993   2.994 0.003521 **
## Team.NameManchester United  -1.4580    6.2754  -0.232 0.816786
## Team.NameTottenham         -9.6141     5.7960  -1.659 0.100504
## Team.NameWest Ham         -28.9275     7.0976  -4.076 9.59e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.03 on 94 degrees of freedom
## Multiple R-squared:   0.64,  Adjusted R-squared:  0.5902
## F-statistic: 12.85 on 13 and 94 DF,  p-value: 1.013e-15
```

Full Complex Model

```
full.lm=lm(Points ~ Log_Wage + I(Log_Wage^2)+ Log_Tenure + I(Log_Tenure^2) + Experience_Ratio + Log_Wag
              data = main
)

summary(full.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage + I(Log_Wage^2) + Log_Tenure +
##     I(Log_Tenure^2) + Experience_Ratio + Log_Wage * Experience_Ratio +
##     Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.580  -5.941   0.297   5.573  20.210
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                96.7051    17.2912   5.593 2.23e-07 ***
## Log_Wage                  -26.0687    22.2149  -1.173 0.243601
## I(Log_Wage^2)             -14.9269    13.0609  -1.143 0.256027
## Log_Tenure                 -7.7712     5.6667  -1.371 0.173561
## I(Log_Tenure^2)             0.5768     0.4855   1.188 0.237842
## Experience_Ratio            1.9574     7.2061   0.272 0.786511
## Team.NameChelsea           -0.9719     5.4983  -0.177 0.860078
## Team.NameCrystal Palace   -35.5431     7.8493  -4.528 1.76e-05 ***
## Team.NameEverton          -26.2874     7.1057  -3.700 0.000366 ***
## Team.NameLiverpool          6.3627     4.7945   1.327 0.187729
## Team.NameManchester City   16.1142     5.5397   2.909 0.004537 **
## Team.NameManchester United  -0.3128     6.5196  -0.048 0.961830
## Team.NameTottenham         -9.7916     5.8190  -1.683 0.095786 .
## Team.NameWest Ham         -28.3963     7.1620  -3.965 0.000144 ***
## Log_Wage:Experience_Ratio   12.1656    18.0728   0.673 0.502523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.07 on 93 degrees of freedom
## Multiple R-squared:  0.6417, Adjusted R-squared:  0.5878
## F-statistic:  11.9 on 14 and 93 DF,  p-value: 3.028e-15
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.4.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.4.3
```

```r
vif(full.lm)
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
##                                GVIF Df GVIF^(1/(2*Df))
## Log_Wage                  59.562052  1        7.717645
## I(Log_Wage^2)              6.132148  1        2.476317
## Log_Tenure                50.707080  1        7.120890
## I(Log_Tenure^2)           53.202423  1        7.293999
## Experience_Ratio           2.037101  1        1.427271
## Team.Name                 35.026013  8        1.248891
## Log_Wage:Experience_Ratio 36.474615  1        6.039422
```

Base (After VIF Check) Model

```r
main$Log_Tenure_Center <- main$Log_Tenure - mean(main$Log_Tenure)
main$Log_Wage_Center <- main$Log_Wage - mean(main$Log_Wage)
main$Exp_Ratio_Center <- main$Experience_Ratio - mean(main$Experience_Ratio)
base_vif.lm = lm(Points ~ Log_Wage_Center + Log_Tenure_Center +
                      Exp_Ratio_Center + Team.Name,
                      data = main)
summary(base_vif.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage_Center + Log_Tenure_Center + Exp_Ratio_Center +
##     Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.034  -5.673   0.508   5.933  20.946
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   75.0641     3.8286  19.606  < 2e-16 ***
## Log_Wage_Center               -7.5405     8.6858  -0.868  0.38749
## Log_Tenure_Center             -1.1761     0.9722  -1.210  0.22935
## Exp_Ratio_Center               1.2677     6.9552   0.182  0.85576
## Team.NameChelsea              -4.2070     5.0905  -0.826  0.41061
## Team.NameCrystal Palace      -37.1649     7.6774  -4.841 4.94e-06 ***
## Team.NameEverton             -27.9430     6.5836  -4.244 5.06e-05 ***
## Team.NameLiverpool             5.4870     4.5901   1.195  0.23488
## Team.NameManchester City      13.0825     4.9410   2.648  0.00947 **
## Team.NameManchester United    -6.2024     5.2810  -1.174  0.24311
## Team.NameTottenham           -10.1185     5.4248  -1.865  0.06520 .
## Team.NameWest Ham            -29.5108     6.7910  -4.346 3.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 11.04 on 96 degrees of freedom
## Multiple R-squared:  0.6318, Adjusted R-squared:  0.5896
## F-statistic: 14.98 on 11 and 96 DF,  p-value: < 2.2e-16
```

Complex (After VIF Check) Model

```
complex_vif.lm=lm(Points ~ Log_Wage_Center + I(Log_Wage_Center^2)+ Log_Tenure_Center + I(Log_Tenure_Cen

summary(complex_vif.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage_Center + I(Log_Wage_Center^2) +
##     Log_Tenure_Center + I(Log_Tenure_Center^2) + Exp_Ratio_Center +
##     Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.528  -5.901   0.848   5.445  20.590
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   73.4433     4.5005  16.319  < 2e-16 ***
## Log_Wage_Center              -11.7524     9.7281  -1.208 0.230043
## I(Log_Wage_Center^2)          -8.2745     8.5145  -0.972 0.333643
## Log_Tenure_Center             -0.8867     1.0147  -0.874 0.384470
## I(Log_Tenure_Center^2)         0.6107     0.4815   1.268 0.207828
## Exp_Ratio_Center               0.7642     6.9643   0.110 0.912861
## Team.NameChelsea              -1.3613     5.4519  -0.250 0.803369
## Team.NameCrystal Palace      -36.0937     7.7838  -4.637 1.14e-05 ***
## Team.NameEverton             -27.0581     6.9924  -3.870 0.000201 ***
## Team.NameLiverpool             6.4108     4.7800   1.341 0.183094
## Team.NameManchester City      16.4635     5.4993   2.994 0.003521 **
## Team.NameManchester United    -1.4580     6.2754  -0.232 0.816786
## Team.NameTottenham            -9.6141     5.7960  -1.659 0.100504
## Team.NameWest Ham            -28.9275     7.0976  -4.076 9.59e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.03 on 94 degrees of freedom
## Multiple R-squared:   0.64,  Adjusted R-squared:  0.5902
## F-statistic: 12.85 on 13 and 94 DF,  p-value: 1.013e-15
```

Full Complex (After VIF Check) Model

```
full_vif.lm=lm(Points ~ Log_Wage_Center + I(Log_Wage_Center^2)+ Log_Tenure_Center + I(Log_Tenure_Center
              data = main
)

summary(full_vif.lm)
```

```
##
```

```
## Call:
## lm(formula = Points ~ Log_Wage_Center + I(Log_Wage_Center^2) +
##     Log_Tenure_Center + I(Log_Tenure_Center^2) + Exp_Ratio_Center +
##     Log_Wage_Center * Exp_Ratio_Center + Team.Name, data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.580  -5.941   0.297   5.573  20.210
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       73.6148     4.5209  16.283  < 2e-16 ***
## Log_Wage_Center                  -11.9780     9.7623  -1.227 0.222933
## I(Log_Wage_Center^2)             -14.9269    13.0609  -1.143 0.256027
## Log_Tenure_Center                 -0.8224     1.0222  -0.805 0.423139
## I(Log_Tenure_Center^2)             0.5768     0.4855   1.188 0.237842
## Exp_Ratio_Center                   1.1729     7.0110   0.167 0.867506
## Team.NameChelsea                  -0.9719     5.4983  -0.177 0.860078
## Team.NameCrystal Palace          -35.5431     7.8493  -4.528 1.76e-05 ***
## Team.NameEverton                 -26.2874     7.1057  -3.700 0.000366 ***
## Team.NameLiverpool                 6.3627     4.7945   1.327 0.187729
## Team.NameManchester City          16.1142     5.5397   2.909 0.004537 **
## Team.NameManchester United        -0.3128     6.5196  -0.048 0.961830
## Team.NameTottenham                -9.7916     5.8190  -1.683 0.095786 .
## Team.NameWest Ham                -28.3963     7.1620  -3.965 0.000144 ***
## Log_Wage_Center:Exp_Ratio_Center  12.1656    18.0728   0.673 0.502523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.07 on 93 degrees of freedom
## Multiple R-squared:  0.6417, Adjusted R-squared:  0.5878
## F-statistic:  11.9 on 14 and 93 DF,  p-value: 3.028e-15
```

Anova Comparisons *Base VS Complex*

```r
anova(base_vif.lm,complex_vif.lm)
```

```
## Analysis of Variance Table
##
## Model 1: Points ~ Log_Wage_Center + Log_Tenure_Center + Exp_Ratio_Center +
##     Team.Name
## Model 2: Points ~ Log_Wage_Center + I(Log_Wage_Center^2) + Log_Tenure_Center +
##     I(Log_Tenure_Center^2) + Exp_Ratio_Center + Team.Name
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     96 11707
## 2     94 11446  2    260.28 1.0687 0.3476
```

Complex Vs Full

```r
anova(complex_vif.lm,full_vif.lm)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: Points ~ Log_Wage_Center + I(Log_Wage_Center^2) + Log_Tenure_Center +
##     I(Log_Tenure_Center^2) + Exp_Ratio_Center + Team.Name
## Model 2: Points ~ Log_Wage_Center + I(Log_Wage_Center^2) + Log_Tenure_Center +
##     I(Log_Tenure_Center^2) + Exp_Ratio_Center + Log_Wage_Center *
##     Exp_Ratio_Center + Team.Name
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     94 11446
## 2     93 11391  1      55.5 0.4531 0.5025
```

```
library(car)
durbinWatsonTest(base.lm)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.04389493      1.896253   0.686
##  Alternative hypothesis: rho != 0
```

Base vs Base without Teams

```
base_exteam.lm=lm(Points ~Log_Wage_Center + Log_Tenure_Center +
                      Exp_Ratio_Center, data = main)

summary(base_exteam.lm)
```

```
##
## Call:
## lm(formula = Points ~ Log_Wage_Center + Log_Tenure_Center + Exp_Ratio_Center,
##     data = main)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.250  -9.259  -0.774  10.158  31.697
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         64.333      1.339  48.030  < 2e-16 ***
## Log_Wage_Center     29.023      4.226   6.868  4.9e-10 ***
## Log_Tenure_Center    1.429      1.035   1.381    0.170
## Exp_Ratio_Center    -8.571      7.530  -1.138    0.258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.92 on 104 degrees of freedom
## Multiple R-squared:  0.3662, Adjusted R-squared:  0.3479
## F-statistic: 20.03 on 3 and 104 DF,  p-value: 2.538e-10
```

```
anova(base_vif.lm,base_exteam.lm)
```

```
## Analysis of Variance Table
##
## Model 1: Points ~ Log_Wage_Center + Log_Tenure_Center + Exp_Ratio_Center +
##     Team.Name
```

```
## Model 2: Points ~ Log_Wage_Center + Log_Tenure_Center + Exp_Ratio_Center
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     96 11707
## 2    104 20152 -8   -8444.9 8.6566 7.955e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To determine the functional form of the relationship, a series of Nested F-tests (ANOVA) were conducted to test for complexity:

Testing for Diminishing Returns: A quadratic model (including Log_Wage^2 and Log_Tenure^2 was compared to the Base Linear Model. The ANOVA returned a p-value of 0.3476, failing to reject the null hypothesis. This indicates that adding curvature did not significantly improve model fit.

Testing for Efficiency (Interactions): A Full Model adding a Wage * Experience interaction was tested against the complex model. The p-value was 0.5025. This suggests that experienced squads do not statistically convert wages into points more efficiently than younger squad

The complex terms were dropped in favor of the Base Linear Model.

Final Model Results The final model explains approximately 63% of the variance in points ($R^2 = 0.6318$). The results highlight the dominance of structural factors over marginal decisions:

The Dominance of Club Identity: A partial F-test confirmed that including 'Team' as a fixed effect was highly significant ($p < .001$).

Interpretation of Fixed Effects: Relative to the baseline (Arsenal), clubs face significant structural starting points. For example, Crystal Palace is predicted to finish approximately 37 points lower than Arsenal holding all else constant, while Manchester City holds a 13 point structural advantage.

The Insignificance of Wages: Once club fixed effects were controlled for, the Log_Wage variable became statistically insignificant ($p = 0.387$). This implies that high wages are a feature of big clubs as to prevent other clubs from poaching their players, not necessarily the direct cause of seasonal variation in points.

Diagnostics

Diagnostic checks confirmed the validity of the OLS assumptions. The Q-Q plot showed residuals tracking the normal line, and the Durbin-Watson test confirmed the independence of errors (no autocorrelation), validating the results The Durbin-Watson test indicated no significant autocorrelation ($\mathrm{DW} = 1.89, p > .05$), confirming that seasonal observations were statistically independent