

Final Research Question 2

Research question :

Is the fight outcome associated with the number of submission attempts made by a fighter?

- **Outcome Variable:** Fight outcome (binary variable: Win or Loss)
- **Independent Variable:** Primary fight style (nominal variable)

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
vforcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr    1.3.1
v purrr    1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()   masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become non-conflicting
```

```
library(dplyr)
library(broom)
library(car) # For VIF calculation
```

Loading required package: carData

Attaching package: 'car'

The following object is masked from 'package:dplyr':

```
recode
```

The following object is masked from 'package:purrr':

```
some
```

Step 1: Data Loading and Preparation

```
# Load dataset
ufc_data <- read_csv('ufc-master.csv', show_col_types = FALSE)
```

Warning: One or more parsing issues, call `problems()` on your data frame for details,
e.g.:

```
dat <- vroom(...)
problems(dat)
```

```
# Glimpse of data structure
glimpse(ufc_data %>% select(BlueReachCms, RedReachCms, BlueAvgSigStrLanded,
                             RedAvgSigStrLanded, WeightClass, Winner,
                             RedAvgSubAtt, BlueAvgSubAtt, TotalFightTimeSecs))
```

Rows: 6,478

Columns: 9

```
$ BlueReachCms      <dbl> 172.72, 165.10, 205.74, 172.72, 190.50, 187.96, 16~
$ RedReachCms       <dbl> 177.80, 167.64, 198.12, 170.18, 187.96, 185.42, 17~
$ BlueAvgSigStrLanded <dbl> 2.72, 3.71, 3.16, 3.70, 3.47, 3.17, 5.38, 7.66, 2.~
$ RedAvgSigStrLanded <dbl> 3.99, 5.24, 5.82, 4.04, 5.98, 3.85, 4.06, 5.43, 3.~
$ WeightClass        <chr> "Flyweight", "Women's Flyweight", "Light Heavyweig~
$ Winner              <chr> "Red", "Red", "Blue", "Blue", "Blue", "Red", "Red"~
$ RedAvgSubAtt       <dbl> 0.4, 0.8, 0.0, 0.3, 0.1, 0.8, 0.0, 0.6, 0.6, 0.9, ~
$ BlueAvgSubAtt      <dbl> 0.5, 0.4, 0.4, 0.6, 1.1, 0.3, 0.5, 0.4, 1.5, 0.5, ~
$ TotalFightTimeSecs <dbl> 1500, 1500, 900, 748, 268, 900, 900, 523, 359, 900~
```

```
# Check for missing values
missing_values <- ufc_data %>%
  summarise(across(everything(), ~sum(is.na(.)))) %>%
  pivot_longer(cols = everything(), names_to = "Variable", values_to = "Missing_Count")

print("Missing values in each column:")
```

```

[1] "Missing values in each column:"
```

```

print(missing_values)
```

```

# A tibble: 118 x 2
  Variable      Missing_Count
  <chr>          <int>
1 RedFighter        0
2 BlueFighter       0
3 RedOdds         219
4 BlueOdds         219
5 RedExpectedValue 219
6 BlueExpectedValue 219
7 Date             0
8 Location          0
9 Country           0
10 Winner            0
# i 108 more rows
```

```

# Data cleaning
ufc_data <- ufc_data %>%
  mutate(
    Outcome = ifelse(Winner == "Red", 1, 0), # Binary outcome: 1 for Red win, 0 for Blue win
    WeightClass = as.factor(WeightClass),
    TotalRedSubAttempts = RedAvgSubAtt,           # Red's submission attempts
    TotalBlueSubAttempts = BlueAvgSubAtt
  ) %>%
  filter(
    !is.na(Outcome),
    !is.na(TotalRedSubAttempts),
    !is.na(TotalBlueSubAttempts),
    !is.na(BlueReachCms),
    !is.na(RedReachCms),
    !is.na(BlueAvgSigStrLanded),
    !is.na(RedAvgSigStrLanded),
    !is.na(TotalFightTimeSecs),
    !is.na(WeightClass)
  ) %>%
  mutate(
    LogRedSubAttempts = log1p(TotalRedSubAttempts),
    LogBlueSubAttempts = log1p(TotalBlueSubAttempts),
    LogBlueReach = log1p(BlueReachCms),

```

```

LogRedReach = log1p(RedReachCms),
LogBlueSigStr = log1p(BlueAvgSigStrLanded),
LogRedSigStr = log1p(RedAvgSigStrLanded),
LogFightTime = log1p(TotalFightTimeSecs)
)

# Check dimensions of the cleaned dataset
dim(ufc_data)

```

[1] 4895 128

Step 2: Exploratory Data Analysis (EDA)

```

# Summary statistics for numerical variables
summary_stats <- ufc_data %>%
  summarise(across(where(is.numeric), list(mean = mean, sd = sd, median = median,
                  min = min, max = max), na.rm = TRUE)) %>%
  pivot_longer(cols = everything(), names_to = c("Variable", ".value"), names_sep = "_")

Warning: There was 1 warning in `summarise()` .
i In argument: `across(...)` .
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `~.fns` through an anonymous function instead.

# Previously
across(a:b, mean, na.rm = TRUE)

# Now
across(a:b, ~mean(x, na.rm = TRUE))

print("Summary statistics for numerical variables:")

```

[1] "Summary statistics for numerical variables:"

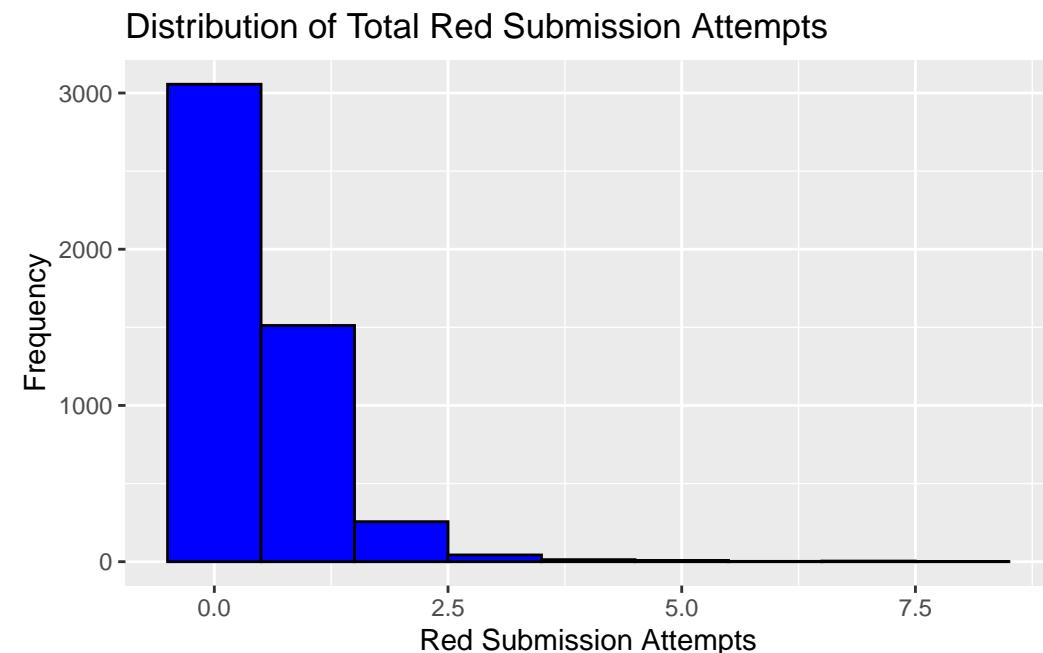
```

print(summary_stats)

```

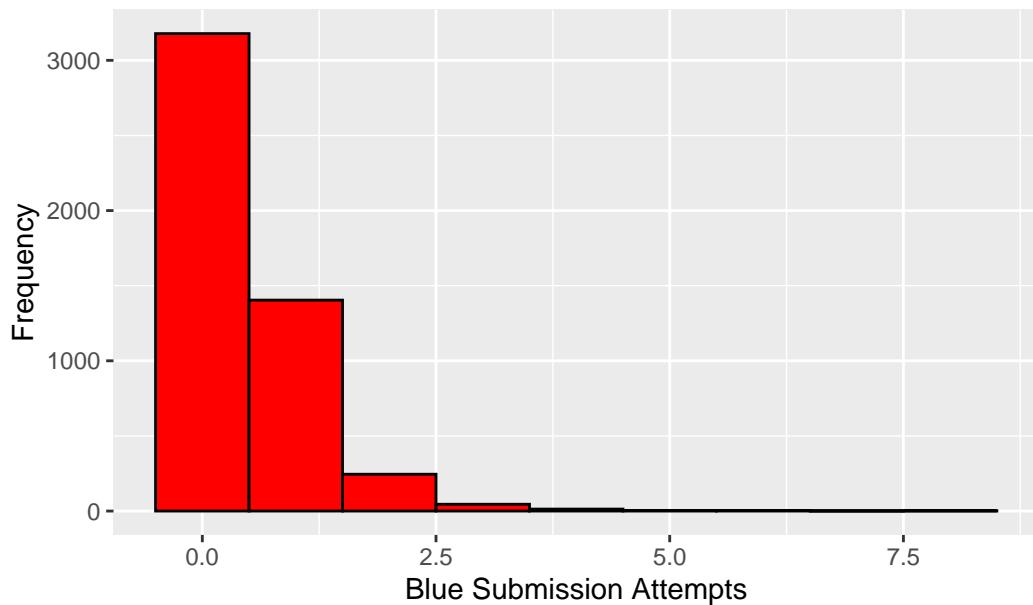
```
# A tibble: 111 x 6
  Variable      mean     sd median    min    max
  <chr>       <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
1 RedOdds     -116.    276.   -150   -2100    700
2 BlueOdds      60.5    251.    130   -1100   1150
3 RedExpectedValue 96.2    84.6    66.7    4.76    700
4 BlueExpectedValue 165.    136.    130    9.09   1150
5 NumberOfRounds  3.21    0.614    3      3      5
6 BlueCurrentLoseStreak 0.571    0.829    0      0      6
7 BlueCurrentWinStreak 1.13    1.46     1      0     12
8 BlueDraws      0.0247   0.163    0      0      2
9 BlueAvgSigStrLanded 20.9    20.6    13      0    154
10 BlueAvgSigStrPct  0.457    0.104   0.46     0      1
# i 101 more rows
```

```
# Plot distributions of key numeric variables
ggplot(ufc_data, aes(x = TotalRedSubAttempts)) +
  geom_histogram(binwidth = 1, fill = 'blue', color = 'black') +
  labs(title = "Distribution of Total Red Submission Attempts", x = "Red Submission Attempts")
```



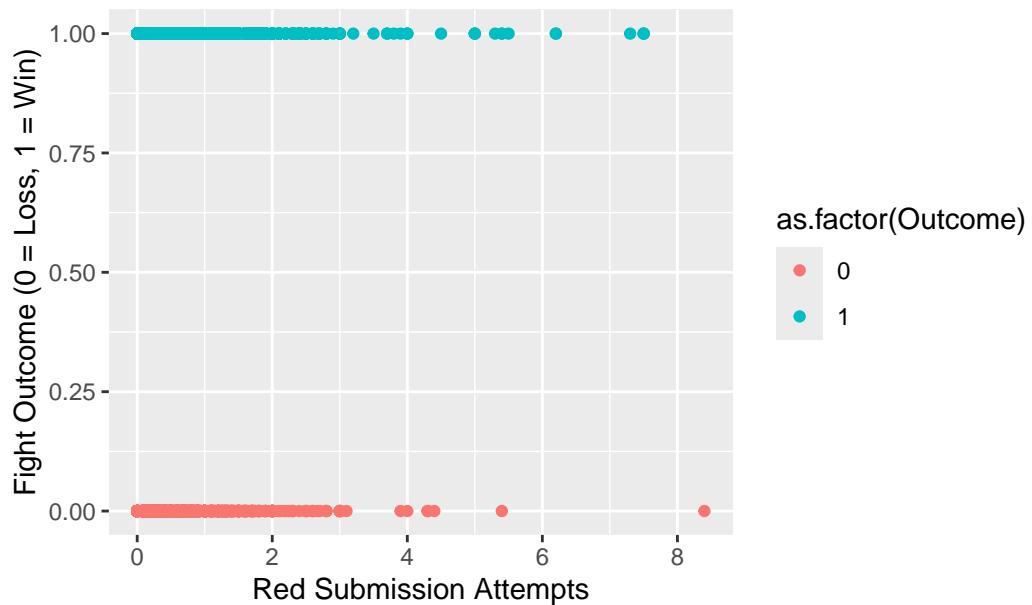
```
ggplot(ufc_data, aes(x = TotalBlueSubAttempts)) +
  geom_histogram(binwidth = 1, fill = 'red', color = 'black') +
  labs(title = "Distribution of Total Blue Submission Attempts", x = "Blue Submission Attempts")
```

Distribution of Total Blue Submission Attempts



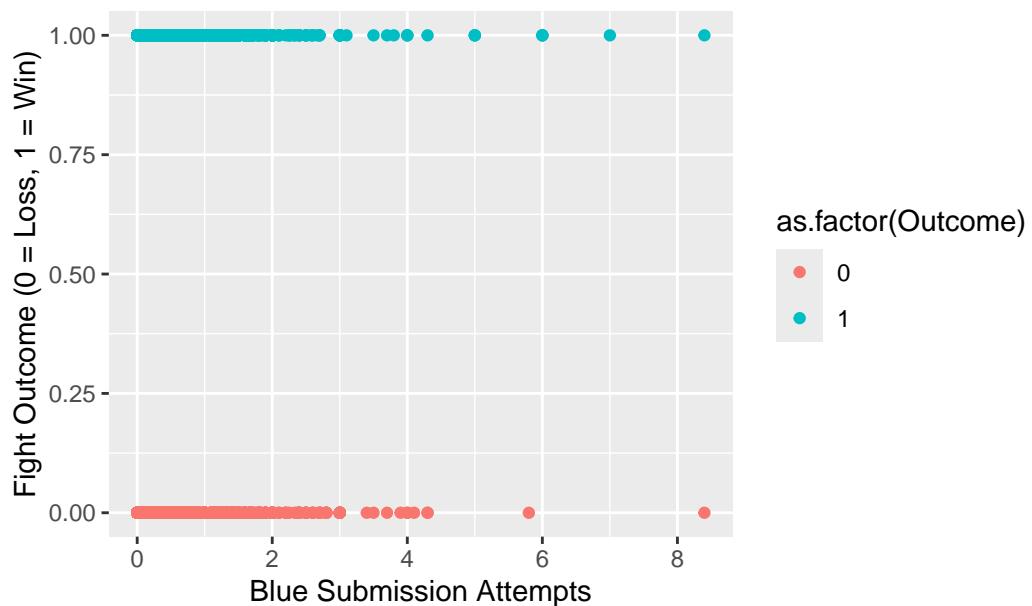
```
# Relationship between submission attempts and fight outcome
ggplot(ufc_data, aes(x = TotalRedSubAttempts, y = Outcome, color = as.factor(Outcome))) +
  geom_point() +
  labs(title = "Red Submission Attempts vs Fight Outcome", x = "Red Submission Attempts", y = "Fight Outcome")
```

Red Submission Attempts vs Fight Outcome



```
ggplot(ufc_data, aes(x = TotalBlueSubAttempts, y = Outcome, color = as.factor(Outcome))) +  
  geom_point() +  
  labs(title = "Blue Submission Attempts vs Fight Outcome", x = "Blue Submission Attempts", y = "Fight Outcome (0 = Loss, 1 = Win)", color = "as.factor(Outcome)")
```

Blue Submission Attempts vs Fight Outcome



Step 3: Logistic Regression Models

Initial Logistic Regression

```
sim_logistic_model <- glm(  
  Outcome ~  
    LogRedSubAttempts +  
    LogBlueSubAttempts +  
    LogBlueReach +  
    LogRedReach +  
    LogBlueSigStr +  
    LogRedSigStr +  
    LogFightTime +  
    WeightClass,  
  data = ufc_data,  
  family = binomial  
)  
  
# Model summary  
summary(sim_logistic_model)
```

Call:

```
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +  
  LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr +  
  LogFightTime + WeightClass, family = binomial, data = ufc_data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.351057	6.459545	-0.828	0.407447
LogRedSubAttempts	0.405680	0.089319	4.542	5.57e-06 ***
LogBlueSubAttempts	-0.292197	0.084704	-3.450	0.000561 ***
LogBlueReach	-1.439347	0.907894	-1.585	0.112883
LogRedReach	2.491817	0.903558	2.758	0.005819 **
LogBlueSigStr	-0.365003	0.048374	-7.545	4.51e-14 ***
LogRedSigStr	0.373853	0.050081	7.465	8.33e-14 ***
LogFightTime	0.035227	0.032697	1.077	0.281313
WeightClassCatch Weight	-0.057436	0.343726	-0.167	0.867293
WeightClassFeatherweight	-0.027917	0.131144	-0.213	0.831425
WeightClassFlyweight	-0.003566	0.159466	-0.022	0.982161
WeightClassHeavyweight	-0.045397	0.208875	-0.217	0.827942
WeightClassLight Heavyweight	-0.176252	0.192860	-0.914	0.360776

```

WeightClassLightweight      -0.136914  0.126242 -1.085  0.278128
WeightClassMiddleweight    -0.276099  0.164307 -1.680  0.092883 .
WeightClassWelterweight   -0.281531  0.145431 -1.936  0.052888 .
WeightClassWomen's Bantamweight -0.177841  0.191071 -0.931  0.351979
WeightClassWomen's Featherweight  0.230400  0.512452  0.450  0.652996
WeightClassWomen's Flyweight    -0.122585  0.184749 -0.664  0.506998
WeightClassWomen's Strawweight   0.011955  0.187268  0.064  0.949100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

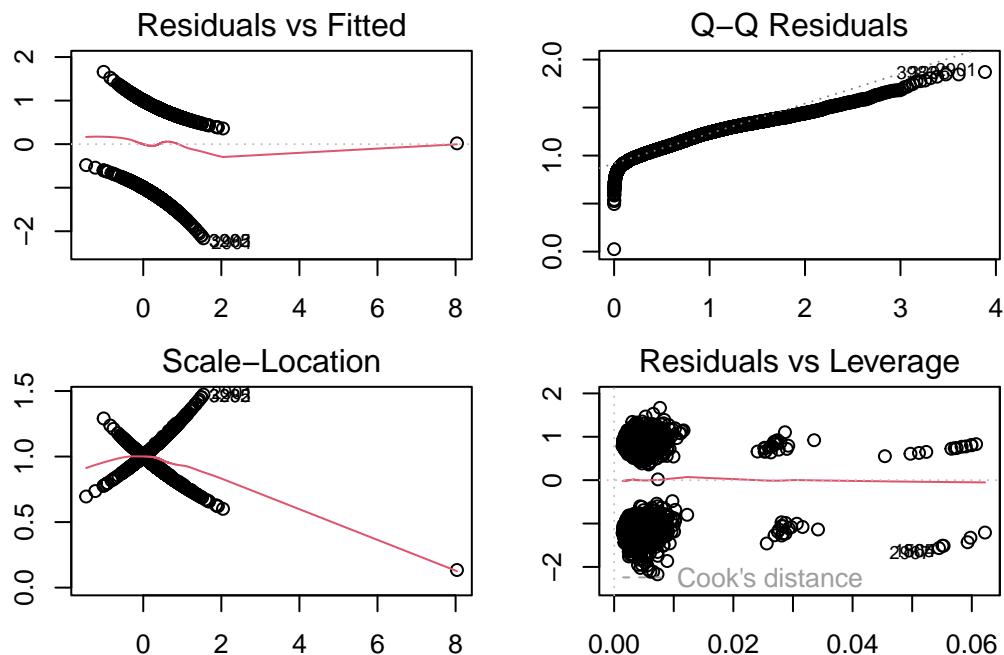
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6674.5 on 4894 degrees of freedom
Residual deviance: 6566.3 on 4875 degrees of freedom
AIC: 6606.3

Number of Fisher Scoring iterations: 4

```
par(mfrow=c(2,2), mar = c(2,2,2,2))  
plot(sim_logistic_model)
```



Logistic Regression with Interaction Terms

```
ext_logistic_model <- glm(  
  Outcome ~  
    LogRedSubAttempts +  
    LogBlueSubAttempts +  
    LogBlueReach +  
    LogRedReach +  
    LogBlueSigStr +  
    LogRedSigStr +  
    LogFightTime +  
    WeightClass +  
    LogRedSubAttempts*LogBlueSubAttempts +  
    LogRedSubAttempts*LogBlueReach +  
    LogBlueSubAttempts*LogRedReach +  
    LogBlueSigStr*LogRedSigStr +  
    LogRedSubAttempts*LogFightTime +  
    LogBlueSubAttempts*LogFightTime +  
    WeightClass*LogRedSubAttempts +  
    WeightClass*LogBlueSubAttempts,  
  data = ufc_data,  
  family = binomial  
)  
  
# Model summary  
summary(ext_logistic_model)
```

Call:

```
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +  
  LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr +  
  LogFightTime + WeightClass + LogRedSubAttempts * LogBlueSubAttempts +  
  LogRedSubAttempts * LogBlueReach + LogBlueSubAttempts * LogRedReach +  
  LogBlueSigStr * LogRedSigStr + LogRedSubAttempts * LogFightTime +  
  LogBlueSubAttempts * LogFightTime + WeightClass * LogRedSubAttempts +  
  WeightClass * LogBlueSubAttempts, family = binomial, data = ufc_data)
```

Coefficients:

	Estimate	Std. Error
(Intercept)	-10.373594	9.186084
LogRedSubAttempts	-2.989190	13.310459
LogBlueSubAttempts	19.055198	13.570897

LogBlueReach	-1.820556	1.290773
LogRedReach	3.868457	1.254986
LogBlueSigStr	-0.382357	0.121462
LogRedSigStr	0.362629	0.118344
LogFightTime	0.035372	0.056990
WeightClassCatch Weight	-0.751441	0.643138
WeightClassFeatherweight	-0.068785	0.235111
WeightClassFlyweight	0.052292	0.288037
WeightClassHeavyweight	-0.320825	0.312685
WeightClassLight Heavyweight	-0.527903	0.301680
WeightClassLightweight	-0.314287	0.218430
WeightClassMiddleweight	-0.499386	0.266803
WeightClassWelterweight	-0.532639	0.243589
WeightClassWomen's Bantamweight	0.013955	0.324225
WeightClassWomen's Featherweight	-0.442344	0.780278
WeightClassWomen's Flyweight	-0.308384	0.339526
WeightClassWomen's Strawweight	0.338865	0.309612
LogRedSubAttempts:LogBlueSubAttempts	0.003080	0.241082
LogRedSubAttempts:LogBlueReach	0.822186	2.574780
LogBlueSubAttempts:LogRedReach	-3.977559	2.626342
LogBlueSigStr:LogRedSigStr	0.006222	0.042187
LogRedSubAttempts:LogFightTime	-0.149367	0.095073
LogBlueSubAttempts:LogFightTime	0.160785	0.091539
LogRedSubAttempts:WeightClassCatch Weight	1.322386	0.887436
LogRedSubAttempts:WeightClassFeatherweight	-0.223685	0.380819
LogRedSubAttempts:WeightClassFlyweight	0.247776	0.450956
LogRedSubAttempts:WeightClassHeavyweight	-0.123642	0.562719
LogRedSubAttempts:WeightClassLight Heavyweight	-0.056204	0.529709
LogRedSubAttempts:WeightClassLightweight	0.057712	0.363045
LogRedSubAttempts:WeightClassMiddleweight	0.182205	0.430051
LogRedSubAttempts:WeightClassWelterweight	0.116370	0.406849
LogRedSubAttempts:WeightClassWomen's Bantamweight	-0.365599	0.598914
LogRedSubAttempts:WeightClassWomen's Featherweight	1.643231	2.304218
LogRedSubAttempts:WeightClassWomen's Flyweight	-0.182690	0.550541
LogRedSubAttempts:WeightClassWomen's Strawweight	-0.127528	0.508712
LogBlueSubAttempts:WeightClassCatch Weight	0.279012	0.852528
LogBlueSubAttempts:WeightClassFeatherweight	0.390901	0.350969
LogBlueSubAttempts:WeightClassFlyweight	-0.330356	0.434667
LogBlueSubAttempts:WeightClassHeavyweight	1.086664	0.546974
LogBlueSubAttempts:WeightClassLight Heavyweight	1.266863	0.531801
LogBlueSubAttempts:WeightClassLightweight	0.468200	0.349187
LogBlueSubAttempts:WeightClassMiddleweight	0.469678	0.420813
LogBlueSubAttempts:WeightClassWelterweight	0.613189	0.372908

	z	value	Pr(> z)
LogBlueSubAttempts:WeightClassWomen's Bantamweight	-0.248851	0.546607	
LogBlueSubAttempts:WeightClassWomen's Featherweight	2.747369	3.519170	
LogBlueSubAttempts:WeightClassWomen's Flyweight	0.689807	0.527995	
LogBlueSubAttempts:WeightClassWomen's Strawweight	-0.867951	0.510445	
(Intercept)	-1.129	0.25878	
LogRedSubAttempts	-0.225	0.82231	
LogBlueSubAttempts	1.404	0.16028	
LogBlueReach	-1.410	0.15841	
LogRedReach	3.082	0.00205 **	
LogBlueSigStr	-3.148	0.00164 **	
LogRedSigStr	3.064	0.00218 **	
LogFightTime	0.621	0.53482	
WeightClassCatch Weight	-1.168	0.24265	
WeightClassFeatherweight	-0.293	0.76985	
WeightClassFlyweight	0.182	0.85594	
WeightClassHeavyweight	-1.026	0.30488	
WeightClassLight Heavyweight	-1.750	0.08014 .	
WeightClassLightweight	-1.439	0.15020	
WeightClassMiddleweight	-1.872	0.06124 .	
WeightClassWelterweight	-2.187	0.02877 *	
WeightClassWomen's Bantamweight	0.043	0.96567	
WeightClassWomen's Featherweight	-0.567	0.57078	
WeightClassWomen's Flyweight	-0.908	0.36373	
WeightClassWomen's Strawweight	1.094	0.27374	
LogRedSubAttempts:LogBlueSubAttempts	0.013	0.98981	
LogRedSubAttempts:LogBlueReach	0.319	0.74948	
LogBlueSubAttempts:LogRedReach	-1.514	0.12990	
LogBlueSigStr:LogRedSigStr	0.147	0.88275	
LogRedSubAttempts:LogFightTime	-1.571	0.11617	
LogBlueSubAttempts:LogFightTime	1.756	0.07901 .	
LogRedSubAttempts:WeightClassCatch Weight	1.490	0.13619	
LogRedSubAttempts:WeightClassFeatherweight	-0.587	0.55695	
LogRedSubAttempts:WeightClassFlyweight	0.549	0.58270	
LogRedSubAttempts:WeightClassHeavyweight	-0.220	0.82609	
LogRedSubAttempts:WeightClassLight Heavyweight	-0.106	0.91550	
LogRedSubAttempts:WeightClassLightweight	0.159	0.87369	
LogRedSubAttempts:WeightClassMiddleweight	0.424	0.67180	
LogRedSubAttempts:WeightClassWelterweight	0.286	0.77486	
LogRedSubAttempts:WeightClassWomen's Bantamweight	-0.610	0.54157	
LogRedSubAttempts:WeightClassWomen's Featherweight	0.713	0.47576	
LogRedSubAttempts:WeightClassWomen's Flyweight	-0.332	0.74001	
LogRedSubAttempts:WeightClassWomen's Strawweight	-0.251	0.80205	

```

LogBlueSubAttempts:WeightClassCatch Weight          0.327  0.74346
LogBlueSubAttempts:WeightClassFeatherweight       1.114  0.26537
LogBlueSubAttempts:WeightClassFlyweight         -0.760  0.44724
LogBlueSubAttempts:WeightClassHeavyweight        1.987  0.04696 *
LogBlueSubAttempts:WeightClassLight Heavyweight   2.382  0.01721 *
LogBlueSubAttempts:WeightClassLightweight        1.341  0.17998
LogBlueSubAttempts:WeightClassMiddleweight       1.116  0.26437
LogBlueSubAttempts:WeightClassWelterweight      1.644  0.10010
LogBlueSubAttempts:WeightClassWomen's Bantamweight -0.455  0.64892
LogBlueSubAttempts:WeightClassWomen's Featherweight 0.781  0.43499
LogBlueSubAttempts:WeightClassWomen's Flyweight    1.306  0.19139
LogBlueSubAttempts:WeightClassWomen's Strawweight   -1.700  0.08906 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 6674.5  on 4894  degrees of freedom
Residual deviance: 6537.9  on 4845  degrees of freedom
AIC: 6637.9

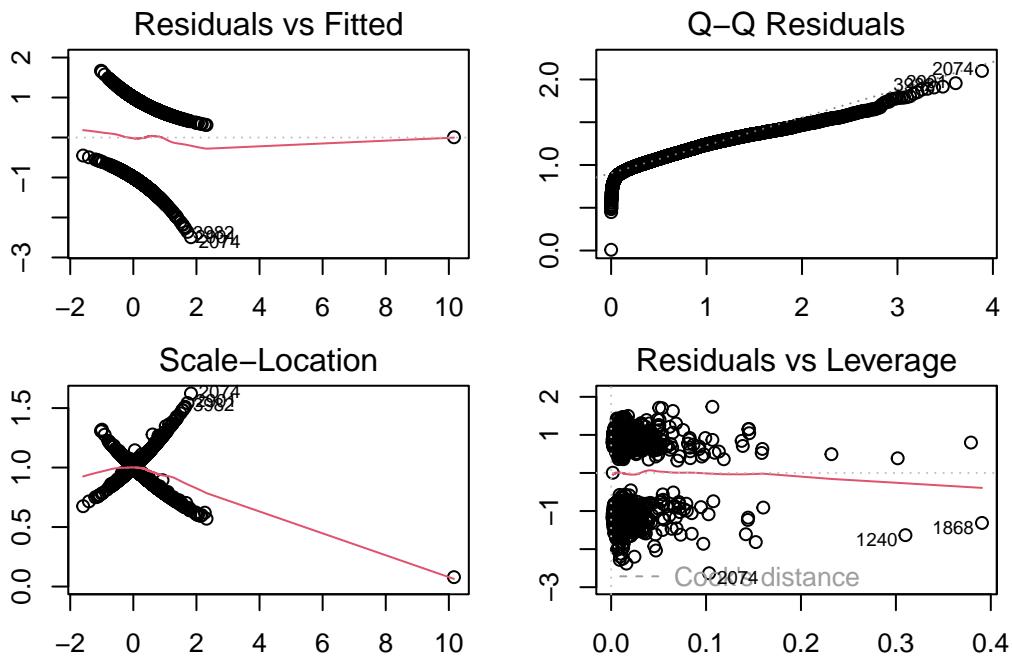
```

Number of Fisher Scoring iterations: 5

```

par(mfrow=c(2,2), mar = c(2,2,2,2))
plot(ext_logistic_model)

```



Step 4: Model Refinement

Stepwise Selection: Backward and Forward

```
# Stepwise backward selection
step_model <- step(ext_logistic_model, direction = "both")
```

Start: AIC=6637.86

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 WeightClass + LogRedSubAttempts * LogBlueSubAttempts + LogRedSubAttempts *
 LogBlueReach + LogBlueSubAttempts * LogRedReach + LogBlueSigStr *
 LogRedSigStr + LogRedSubAttempts * LogFightTime + LogBlueSubAttempts *
 LogFightTime + WeightClass * LogRedSubAttempts + WeightClass *
 LogBlueSubAttempts

	Df	Deviance	AIC
- LogRedSubAttempts:WeightClass	12	6544.1	6620.1
- LogBlueSubAttempts:WeightClass	12	6553.3	6629.3
- LogRedSubAttempts:LogBlueSubAttempts	1	6537.9	6635.9
- LogBlueSigStr:LogRedSigStr	1	6537.9	6635.9
- LogRedSubAttempts:LogBlueReach	1	6538.0	6636.0

<none>		6537.9	6637.9
- LogBlueSubAttempts:LogRedReach	1	6540.2	6638.2
- LogRedSubAttempts:LogFightTime	1	6540.4	6638.4
- LogBlueSubAttempts:LogFightTime	1	6541.0	6639.0

Step: AIC=6620.12

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 WeightClass + LogRedSubAttempts:LogBlueSubAttempts + LogRedSubAttempts:LogBlueReach +
 LogBlueSubAttempts:LogRedReach + LogBlueSigStr:LogRedSigStr +
 LogRedSubAttempts:LogFightTime + LogBlueSubAttempts:LogFightTime +
 LogBlueSubAttempts:WeightClass

	Df	Deviance	AIC
- LogBlueSubAttempts:WeightClass	12	6559.4	6611.4
- LogRedSubAttempts:LogBlueSubAttempts	1	6544.1	6618.1
- LogBlueSigStr:LogRedSigStr	1	6544.2	6618.2
- LogRedSubAttempts:LogBlueReach	1	6545.2	6619.2
<none>		6544.1	6620.1
- LogBlueSubAttempts:LogRedReach	1	6546.4	6620.4
- LogRedSubAttempts:LogFightTime	1	6546.4	6620.4
- LogBlueSubAttempts:LogFightTime	1	6547.1	6621.1
+ LogRedSubAttempts:WeightClass	12	6537.9	6637.9

Step: AIC=6611.36

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 WeightClass + LogRedSubAttempts:LogBlueSubAttempts + LogRedSubAttempts:LogBlueReach +
 LogBlueSubAttempts:LogRedReach + LogBlueSigStr:LogRedSigStr +
 LogRedSubAttempts:LogFightTime + LogBlueSubAttempts:LogFightTime

	Df	Deviance	AIC
- WeightClass	12	6571.4	6599.4
- LogBlueSigStr:LogRedSigStr	1	6559.4	6609.4
- LogRedSubAttempts:LogBlueSubAttempts	1	6559.4	6609.4
- LogRedSubAttempts:LogBlueReach	1	6560.6	6610.6
- LogBlueSubAttempts:LogRedReach	1	6560.8	6610.8
<none>		6559.4	6611.4
- LogBlueSubAttempts:LogFightTime	1	6561.4	6611.4
- LogRedSubAttempts:LogFightTime	1	6561.5	6611.5
+ LogBlueSubAttempts:WeightClass	12	6544.1	6620.1
+ LogRedSubAttempts:WeightClass	12	6553.3	6629.3

Step: AIC=6599.42

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
LogRedSubAttempts:LogBlueSubAttempts + LogRedSubAttempts:LogBlueReach +
LogBlueSubAttempts:LogRedReach + LogBlueSigStr:LogRedSigStr +
LogRedSubAttempts:LogFightTime + LogBlueSubAttempts:LogFightTime

	Df	Deviance	AIC
- LogBlueSigStr:LogRedSigStr	1	6571.5	6597.5
- LogRedSubAttempts:LogBlueSubAttempts	1	6571.5	6597.5
- LogRedSubAttempts:LogBlueReach	1	6572.1	6598.1
- LogBlueSubAttempts:LogRedReach	1	6572.4	6598.4
- LogRedSubAttempts:LogFightTime	1	6573.2	6599.2
- LogBlueSubAttempts:LogFightTime	1	6573.4	6599.4
<none>		6571.4	6599.4
+ WeightClass	12	6559.4	6611.4

Step: AIC=6597.48

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
LogRedSubAttempts:LogBlueSubAttempts + LogRedSubAttempts:LogBlueReach +
LogBlueSubAttempts:LogRedReach + LogRedSubAttempts:LogFightTime +
LogBlueSubAttempts:LogFightTime

	Df	Deviance	AIC
- LogRedSubAttempts:LogBlueSubAttempts	1	6571.5	6595.5
- LogRedSubAttempts:LogBlueReach	1	6572.2	6596.2
- LogBlueSubAttempts:LogRedReach	1	6572.4	6596.4
- LogRedSubAttempts:LogFightTime	1	6573.3	6597.3
- LogBlueSubAttempts:LogFightTime	1	6573.4	6597.4
<none>		6571.5	6597.5
+ LogBlueSigStr:LogRedSigStr	1	6571.4	6599.4
+ WeightClass	12	6559.4	6609.4
- LogRedSigStr	1	6628.5	6652.5
- LogBlueSigStr	1	6631.5	6655.5

Step: AIC=6595.54

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
LogRedSubAttempts:LogBlueReach + LogBlueSubAttempts:LogRedReach +
LogRedSubAttempts:LogFightTime + LogBlueSubAttempts:LogFightTime

Df	Deviance	AIC
----	----------	-----

- LogRedSubAttempts:LogBlueReach	1	6572.2	6594.2
- LogBlueSubAttempts:LogRedReach	1	6572.5	6594.5
- LogRedSubAttempts:LogFightTime	1	6573.4	6595.4
- LogBlueSubAttempts:LogFightTime	1	6573.5	6595.5
<none>		6571.5	6595.5
+ LogRedSubAttempts:LogBlueSubAttempts	1	6571.5	6597.5
+ LogBlueSigStr:LogRedSigStr	1	6571.5	6597.5
+ WeightClass	12	6559.4	6607.4
- LogRedSigStr	1	6628.6	6650.6
- LogBlueSigStr	1	6631.6	6653.6

Step: AIC=6594.21

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 LogBlueSubAttempts:LogRedReach + LogRedSubAttempts:LogFightTime +
 LogBlueSubAttempts:LogFightTime

	Df	Deviance	AIC
- LogBlueSubAttempts:LogRedReach	1	6573.2	6593.2
<none>		6572.2	6594.2
- LogBlueSubAttempts:LogFightTime	1	6574.2	6594.2
- LogRedSubAttempts:LogFightTime	1	6574.3	6594.3
+ LogRedSubAttempts:LogBlueReach	1	6571.5	6595.5
+ LogRedSubAttempts:LogBlueSubAttempts	1	6572.2	6596.2
+ LogBlueSigStr:LogRedSigStr	1	6572.2	6596.2
- LogBlueReach	1	6580.1	6600.1
+ WeightClass	12	6560.7	6606.7
- LogRedSigStr	1	6629.5	6649.5
- LogBlueSigStr	1	6632.6	6652.6

Step: AIC=6593.24

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 LogRedSubAttempts:LogFightTime + LogBlueSubAttempts:LogFightTime

	Df	Deviance	AIC
- LogBlueSubAttempts:LogFightTime	1	6575.0	6593.0
<none>		6573.2	6593.2
- LogRedSubAttempts:LogFightTime	1	6575.3	6593.3
+ LogBlueSubAttempts:LogRedReach	1	6572.2	6594.2
+ LogRedSubAttempts:LogBlueReach	1	6572.5	6594.5
+ LogBlueSigStr:LogRedSigStr	1	6573.2	6595.2
+ LogRedSubAttempts:LogBlueSubAttempts	1	6573.2	6595.2

- LogRedReach	1	6580.8	6598.8
- LogBlueReach	1	6581.1	6599.1
+ WeightClass	12	6562.2	6606.2
- LogRedSigStr	1	6630.6	6648.6
- LogBlueSigStr	1	6633.7	6651.7

Step: AIC=6593.01

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
 LogRedSubAttempts:LogFightTime

	Df	Deviance	AIC
- LogRedSubAttempts:LogFightTime	1	6576.9	6592.9
<none>		6575.0	6593.0
+ LogBlueSubAttempts:LogFightTime	1	6573.2	6593.2
+ LogRedSubAttempts:LogBlueReach	1	6574.2	6594.2
+ LogBlueSubAttempts:LogRedReach	1	6574.2	6594.2
+ LogRedSubAttempts:LogBlueSubAttempts	1	6575.0	6595.0
+ LogBlueSigStr:LogRedSigStr	1	6575.0	6595.0
- LogRedReach	1	6582.7	6598.7
- LogBlueReach	1	6582.8	6598.8
- LogBlueSubAttempts	1	6586.8	6602.8
+ WeightClass	12	6564.1	6606.1
- LogRedSigStr	1	6632.4	6648.4
- LogBlueSigStr	1	6635.5	6651.5

Step: AIC=6592.94

Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
 LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime

	Df	Deviance	AIC
- LogFightTime	1	6577.8	6591.8
<none>		6576.9	6592.9
+ LogRedSubAttempts:LogFightTime	1	6575.0	6593.0
+ LogBlueSubAttempts:LogFightTime	1	6575.3	6593.3
+ LogRedSubAttempts:LogBlueReach	1	6575.9	6593.9
+ LogBlueSubAttempts:LogRedReach	1	6576.2	6594.2
+ LogRedSubAttempts:LogBlueSubAttempts	1	6576.9	6594.9
+ LogBlueSigStr:LogRedSigStr	1	6576.9	6594.9
- LogRedReach	1	6584.6	6598.6
- LogBlueReach	1	6584.8	6598.8
- LogBlueSubAttempts	1	6589.3	6603.3
+ WeightClass	12	6566.3	6606.3

```

- LogRedSubAttempts           1   6597.5 6611.5
- LogRedSigStr                1   6634.1 6648.1
- LogBlueSigStr               1   6637.7 6651.7

Step: AIC=6591.77
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
         LogRedReach + LogBlueSigStr + LogRedSigStr

```

	Df	Deviance	AIC
<none>		6577.8	6591.8
+ LogRedSubAttempts:LogBlueReach	1	6576.7	6592.7
+ LogFightTime	1	6576.9	6592.9
+ LogBlueSubAttempts:LogRedReach	1	6577.0	6593.0
+ LogBlueSigStr:LogRedSigStr	1	6577.7	6593.7
+ LogRedSubAttempts:LogBlueSubAttempts	1	6577.7	6593.7
- LogRedReach	1	6585.2	6597.2
- LogBlueReach	1	6586.0	6598.0
- LogBlueSubAttempts	1	6590.4	6602.4
+ WeightClass	12	6567.5	6605.5
- LogRedSubAttempts	1	6597.9	6609.9
- LogRedSigStr	1	6635.0	6647.0
- LogBlueSigStr	1	6638.4	6650.4

```
summary(step_model)
```

Call:

```
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_data)
```

Coefficients:

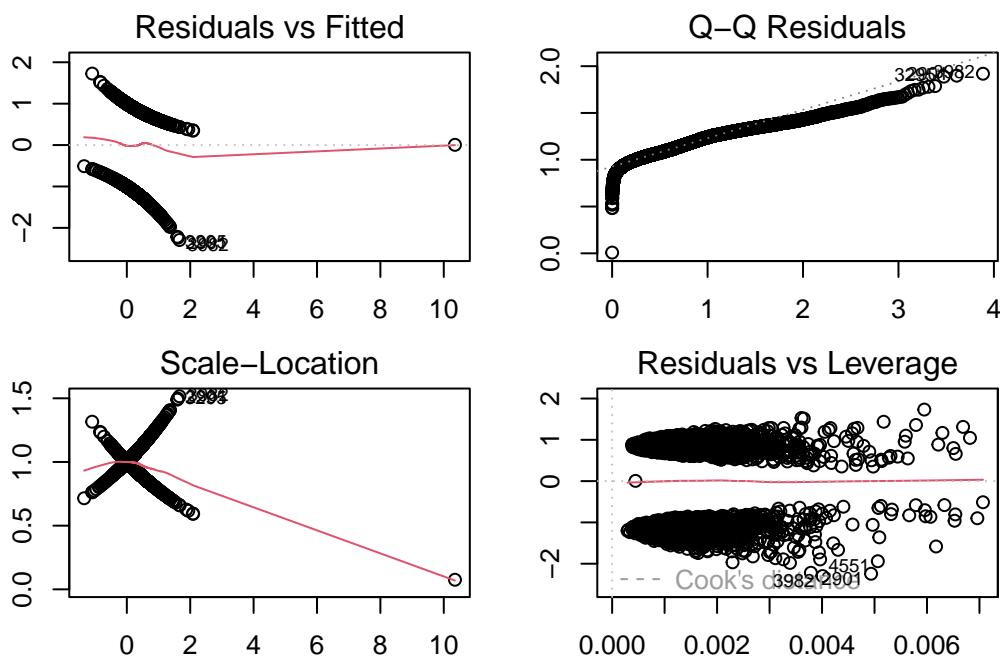
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.36477	2.73096	0.134	0.893744
LogRedSubAttempts	0.39057	0.08768	4.454	8.41e-06 ***
LogBlueSubAttempts	-0.29815	0.08385	-3.556	0.000377 ***
LogBlueReach	-1.90853	0.71404	-2.673	0.007521 **
LogRedReach	1.88597	0.69892	2.698	0.006967 **
LogBlueSigStr	-0.36729	0.04821	-7.618	2.57e-14 ***
LogRedSigStr	0.36873	0.04979	7.405	1.31e-13 ***
<hr/>				
Signif. codes:	0	'***'	0.001	'**'
	0.01	'*'	0.05	'. '
	0.1	' '	1	

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6674.5 on 4894 degrees of freedom
Residual deviance: 6577.8 on 4888 degrees of freedom
AIC: 6591.8

Number of Fisher Scoring iterations: 4

```
par(mfrow=c(2,2), mar = c(2,2,2,2))
plot(step_model)
```



Removing Influential Points

```
# Calculate Cook's distance and leverage
cooks_distance <- cooks.distance(step_model)
hat_values <- hatvalues(step_model)
residuals <- residuals(step_model, type = "deviance")

# Thresholds
n <- nrow(ufc_data)
```

```

p <- length(coef(step_model)) - 1
cooks_threshold <- 4 / n
leverage_threshold <- 2 * (p + 1) / n

# Identify influential points
influential_points <- which(cooks_distance > cooks_threshold |
                             hat_values > leverage_threshold |
                             abs(residuals) > 2)

# Remove influential points
ufc_data_filtered <- ufc_data[-influential_points, ]

# Refit the model
final_model <- glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
                     LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
                     family = binomial, data = ufc_data_filtered)

# Model summary
summary(final_model)

```

Call:

```
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_data_filtered)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.68061	2.87556	0.237	0.812897
LogRedSubAttempts	0.43736	0.09732	4.494	6.98e-06 ***
LogBlueSubAttempts	-0.34346	0.09384	-3.660	0.000252 ***
LogBlueReach	-2.10785	0.76588	-2.752	0.005920 **
LogRedReach	2.03037	0.74314	2.732	0.006292 **
LogBlueSigStr	-0.46782	0.05704	-8.201	2.37e-16 ***
LogRedSigStr	0.46156	0.05861	7.876	3.39e-15 ***

Signif. codes:	0 ***	0.001 **	0.01 *	0.05 .
	''	'	'	'

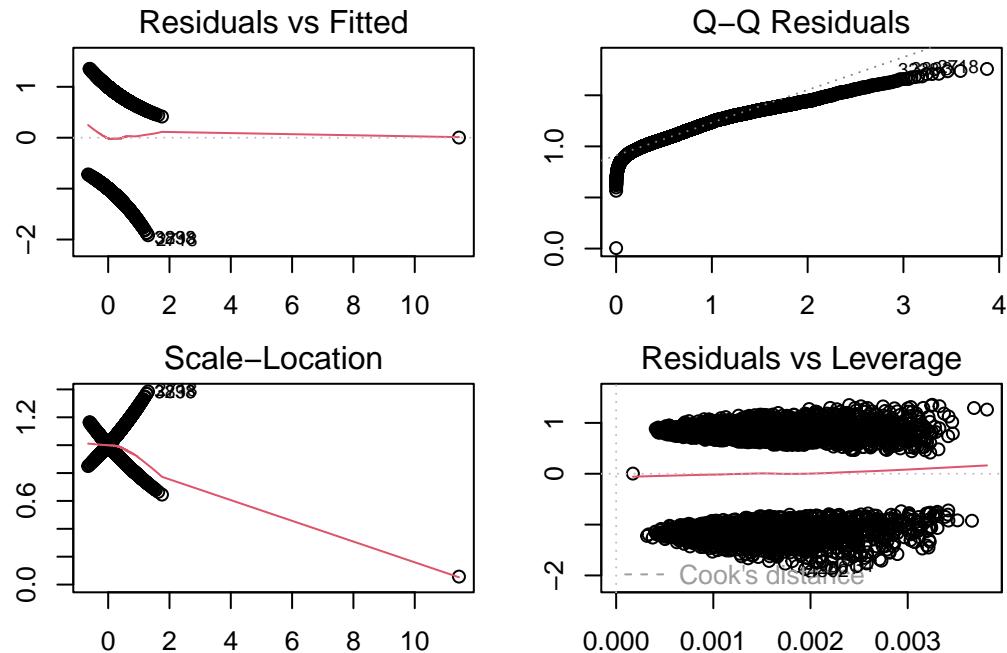
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 6324.0 on 4643 degrees of freedom
Residual deviance: 6221.7 on 4637 degrees of freedom

AIC: 6235.7

Number of Fisher Scoring iterations: 4

```
par(mfrow=c(2,2), mar = c(2,2,2,2))
plot(final_model)
```



Interpretation of the Logistic Regression Results

1. Intercept (0.68061, p = 0.8129):

The intercept represents the log-odds of a fight outcome (Red win) when all independent variables are at their baseline (log-transformed value of 0). However, the intercept is not statistically significant, indicating it does not add meaningful information about the outcome on its own.

2. LogRedSubAttempts (Estimate = 0.43736, p < 0.001):

A unit increase in the log-transformed submission attempts by the Red fighter increases the log-odds of Red winning by 0.437, holding all other variables constant. This is statistically significant and suggests that higher submission attempts by the Red fighter are associated with a higher probability of winning.

3. LogBlueSubAttempts (Estimate = -0.34346, p < 0.001):

A unit increase in the log-transformed submission attempts by the Blue fighter decreases the log-odds of Red winning by 0.343, holding other variables constant. This indicates that more submission attempts by Blue correlate negatively with Red's chances of winning.

4. LogBlueReach (-2.10785, p < 0.01) and LogRedReach (2.03037, p < 0.01):

- A unit increase in Blue's reach reduces the log-odds of Red winning by 2.11, while a similar increase for Red increases the log-odds of Red winning by 2.03.
- Both variables are statistically significant, indicating reach plays an important role in determining the outcome.

5. LogBlueSigStr (-0.46782, p < 0.001) and LogRedSigStr (0.46156, p < 0.001):

- A unit increase in Blue's log-transformed significant strikes landed reduces the log-odds of Red winning by 0.46782.
- Similarly, an increase in Red's significant strikes landed improves their chances of winning by 0.46156.
- These findings align with the intuition that significant strikes are strongly correlated with winning probability.

6. Model Performance (Residual Deviance and AIC):

- The residual deviance (6221.7) is lower than the null deviance (6324.0), suggesting the model fits the data well.
- The AIC (6235.7) provides a metric for model comparison, with lower values indicating better fit.

Relationship to the Research Question

The research question asked whether fight outcomes are associated with the number of submission attempts made by a fighter. The findings clearly show that submission attempts by both the Red and Blue fighters significantly affect the fight outcome.

- **Positive Impact of Red Submission Attempts:** The more submission attempts the Red fighter makes, the higher their chances of winning.
- **Negative Impact of Blue Submission Attempts:** Conversely, Blue's submission attempts lower the probability of a Red win, reinforcing the competitive nature of the fight.

These results support the hypothesis that submission attempts are indeed a critical factor influencing fight outcomes. However, other variables such as reach and significant strikes also play substantial roles, suggesting that submission attempts should be considered within the broader context of fight strategy and performance metrics.