

EPL_24

Deambrogio

4/25/2025

```
pass_EPL_24<- read.csv("pass_EPL_24.csv")
mod_EPL_24 <- pass_EPL_24 %>%
  group_by(Game_URL) %>%
  summarise(
    # compare the two xG columns (they're constant within each group)
    Home_better = as.integer(first(Home_xG) > first(Away_xG)),
    # extract the Cmp_Total for Home and for Away and take the ratio
    HomePass_to_AwayPass = Cmp_Total[Home_Away == "Home"] /
                          Cmp_Total[Home_Away == "Away"],
    HomePass_to_AwayPass_short = Att_Short[Home_Away == "Home"] /
                          Att_Short[Home_Away == "Away"]
  ) %>%
  ungroup() %>%
  select(Game_URL, Home_better, HomePass_to_AwayPass, HomePass_to_AwayPass_short)

# print to check
head(mod_EPL_24)
```

```
## # A tibble: 6 x 4
##   Game_URL                Home_better HomePass_to_AwayPass HomePass_to_AwayPass~1
##   <chr>                  <int>          <dbl>          <dbl>
## 1 https://fbref.com/en/~      1            1.71            1.66
## 2 https://fbref.com/en/~      0            0.462           0.394
## 3 https://fbref.com/en/~      1            1.27            1.70
## 4 https://fbref.com/en/~      1            3.03            3.57
## 5 https://fbref.com/en/~      0            0.452           0.448
## 6 https://fbref.com/en/~      0            1.82            1.70
## # i abbreviated name: 1: HomePass_to_AwayPass_short
```

```
# Fit logistic model
log_mod <- glm(
  Home_better ~ HomePass_to_AwayPass,
  data = mod_EPL_24,
  family = 'binomial'
)

# View the coefficient estimates, p-values, etc.
summary(log_mod)
```

```
##
## Call:
```

```
## glm(formula = Home_better ~ HomePass_to_AwayPass, family = "binomial",
##      data = mod_EPL_24)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -2.3298  -1.1468   0.7184   1.0343   1.3165
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.4440     0.1972  -2.251   0.0244 *
## HomePass_to_AwayPass  0.6644     0.1376   4.830 1.37e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 510.67  on 379  degrees of freedom
## Residual deviance: 480.96  on 378  degrees of freedom
## AIC: 484.96
##
## Number of Fisher Scoring iterations: 4
```

```
exp(coef(log_mod))
```

```
##           (Intercept) HomePass_to_AwayPass
##           0.6414939      1.9434154
```

```
1/(1+exp(0.4440-0.6644))
```

```
## [1] 0.554878
```

```
1/(1+exp(0.4440-0.6644*1.5))
```

```
## [1] 0.6347386
```

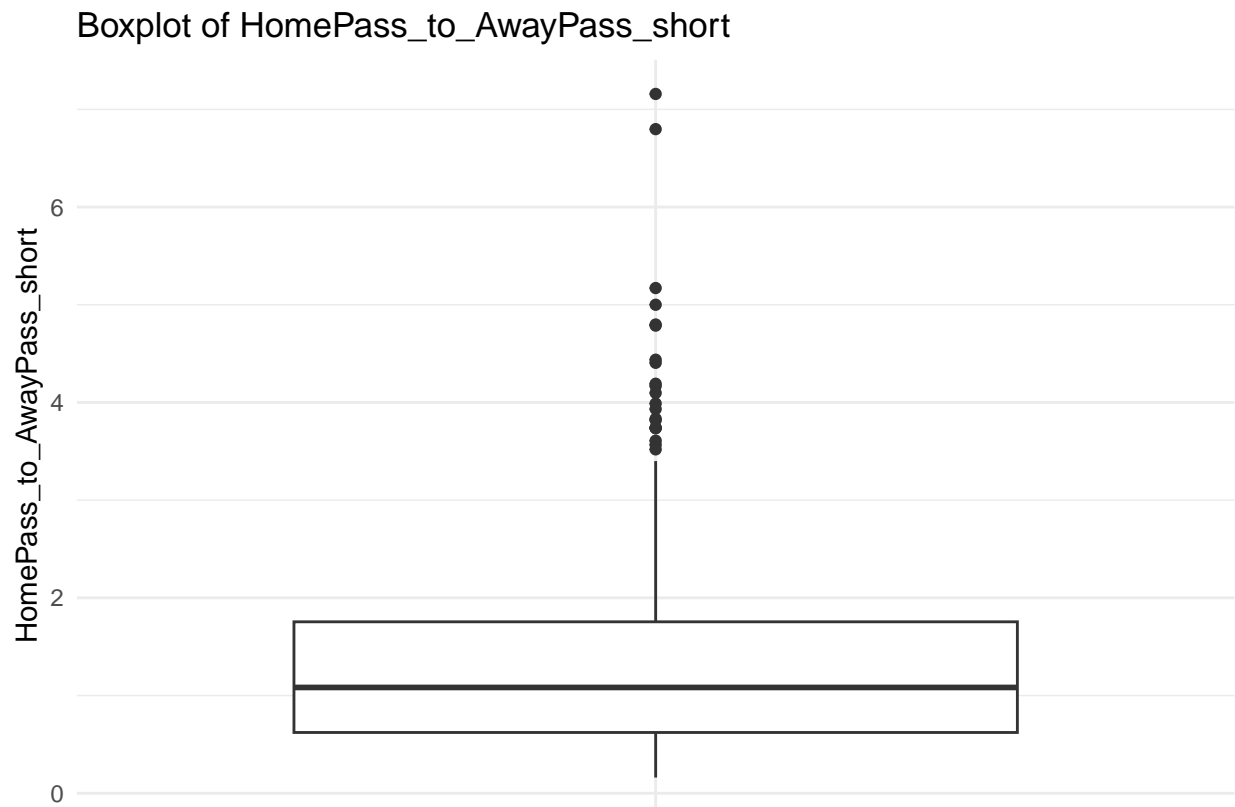
```
# Fit logistic model
log_mod2 <- glm(
  Home_better ~ HomePass_to_AwayPass_short,
  data      = mod_EPL_24,
  family    = 'binomial'
)

# View the coefficient estimates, p-values, etc.
summary(log_mod2)
```

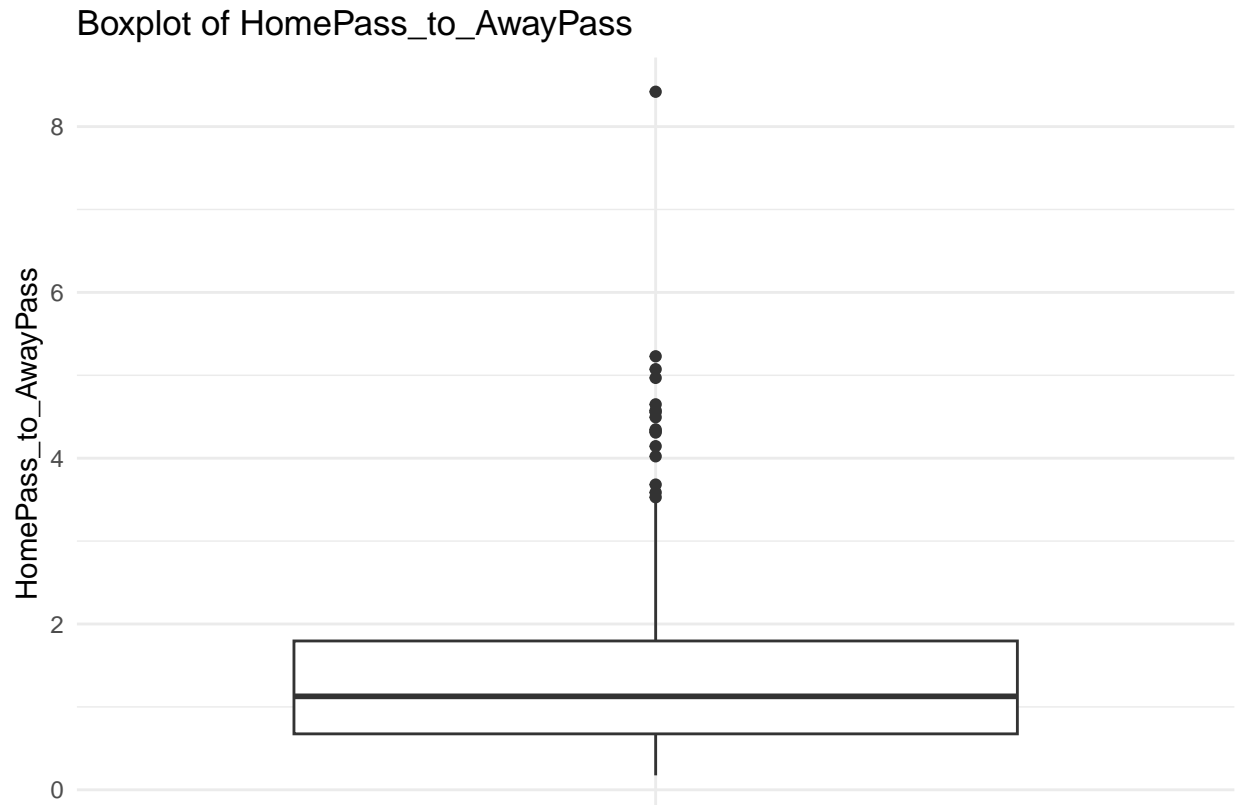
```
##
## Call:
## glm(formula = Home_better ~ HomePass_to_AwayPass_short, family = "binomial",
##      data = mod_EPL_24)
##
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.3658 -1.1703  0.7365   1.0396   1.2894
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.3556    0.1904  -1.868   0.0618 .
## HomePass_to_AwayPass_short  0.5978    0.1312   4.556 5.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 510.67  on 379  degrees of freedom
## Residual deviance: 484.55  on 378  degrees of freedom
## AIC: 488.55
##
## Number of Fisher Scoring iterations: 4
```

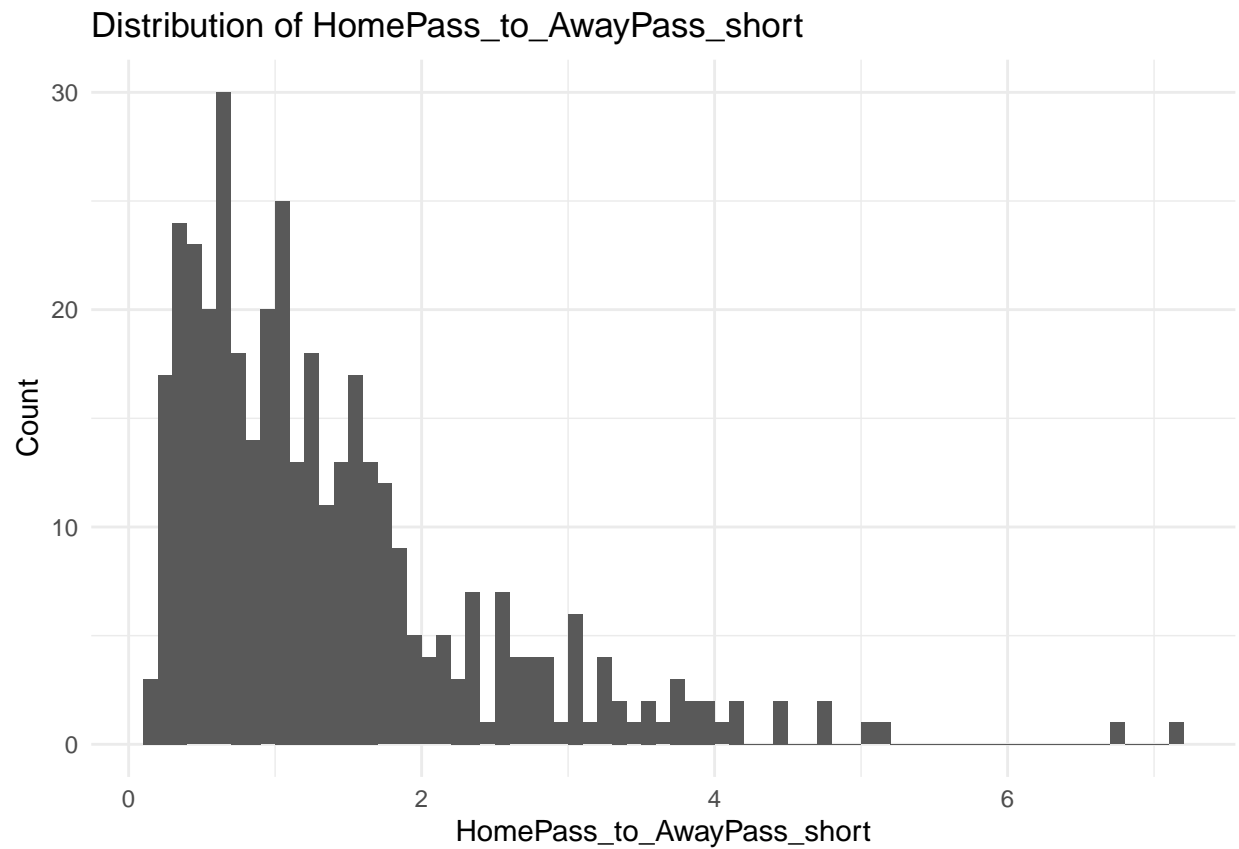
```
ggplot(mod_EPL_24, aes(x = "", y = HomePass_to_AwayPass_short)) +
  geom_boxplot() +
  labs(
    x      = NULL,
    y      = "HomePass_to_AwayPass_short",
    title  = "Boxplot of HomePass_to_AwayPass_short"
  ) +
  theme_minimal()
```



```
# Boxplot for HomePass_to_AwayPass
ggplot(mod_EPL_24, aes(x = "", y = HomePass_to_AwayPass)) +
  geom_boxplot() +
  labs(
    x = NULL,
    y = "HomePass_to_AwayPass",
    title = "Boxplot of HomePass_to_AwayPass"
  ) +
  theme_minimal()
```



```
ggplot(mod_EPL_24, aes(x = HomePass_to_AwayPass_short)) +
  geom_histogram(binwidth = 0.1, boundary = 0, closed = "left") +
  labs(
    x = "HomePass_to_AwayPass_short",
    y = "Count",
    title = "Distribution of HomePass_to_AwayPass_short"
  ) +
  theme_minimal()
```



```
# Histogram for HomePass_to_AwayPass
ggplot(mod_EPL_24, aes(x = HomePass_to_AwayPass)) +
  geom_histogram(binwidth = 0.1, boundary = 0, closed = "left") +
  labs(
    x = "HomePass_to_AwayPass",
    y = "Count",
    title = "Distribution of HomePass_to_AwayPass"
  ) +
  theme_minimal()
```

