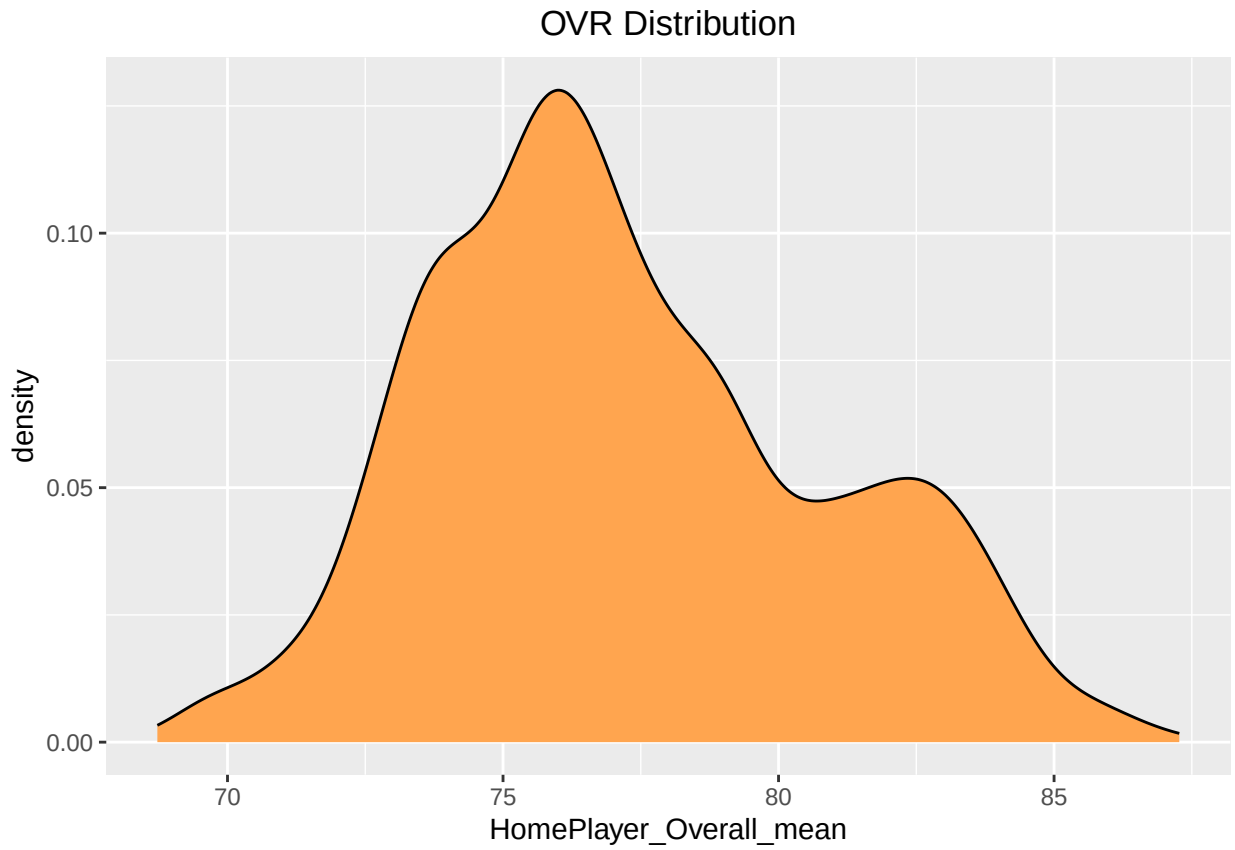
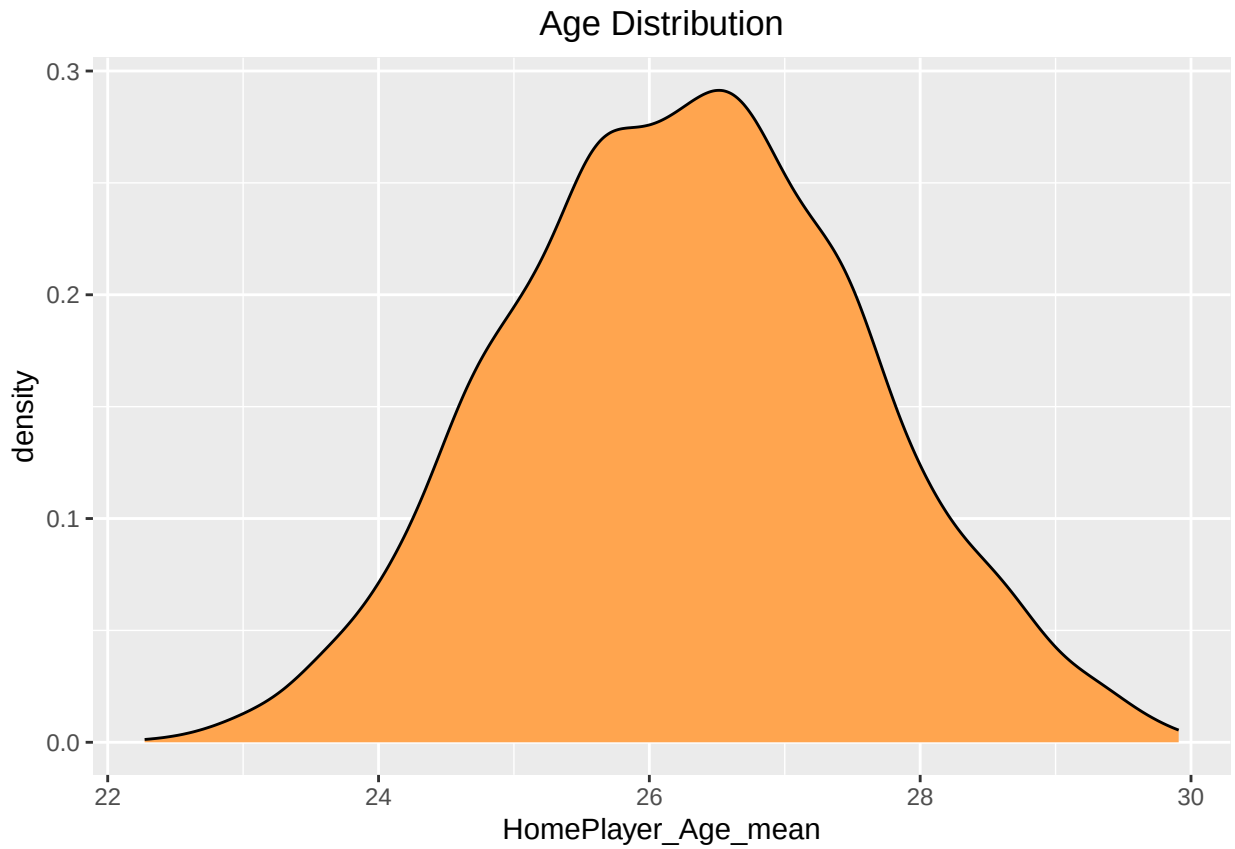
Exploring FIFA Ratings and motivation

2024-01-25

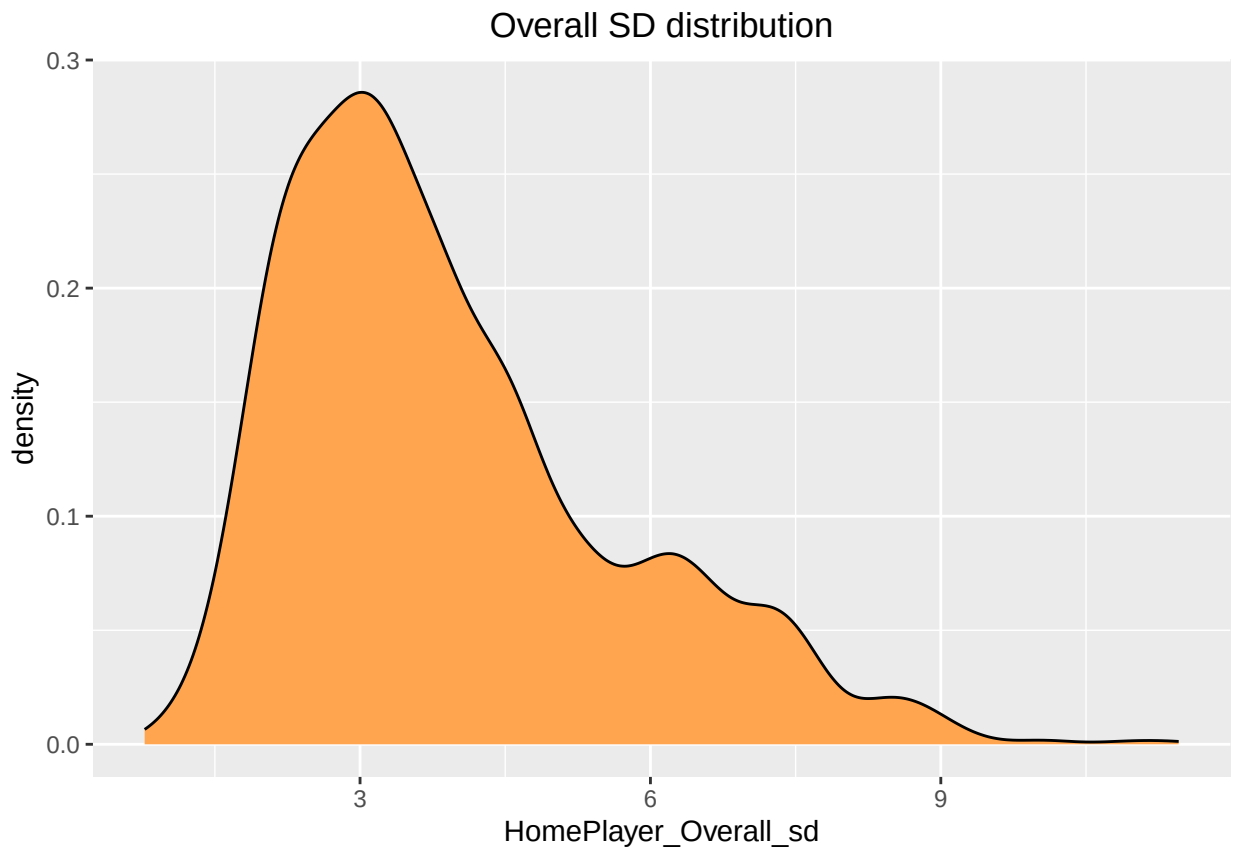
```
ggplot(data=data, aes(x=HomePlayer_Overall_mean)) +
  geom_density(fill="tan1") + labs(title="OVR Distribution") + theme(plot.title = element_text(hjust = 
```



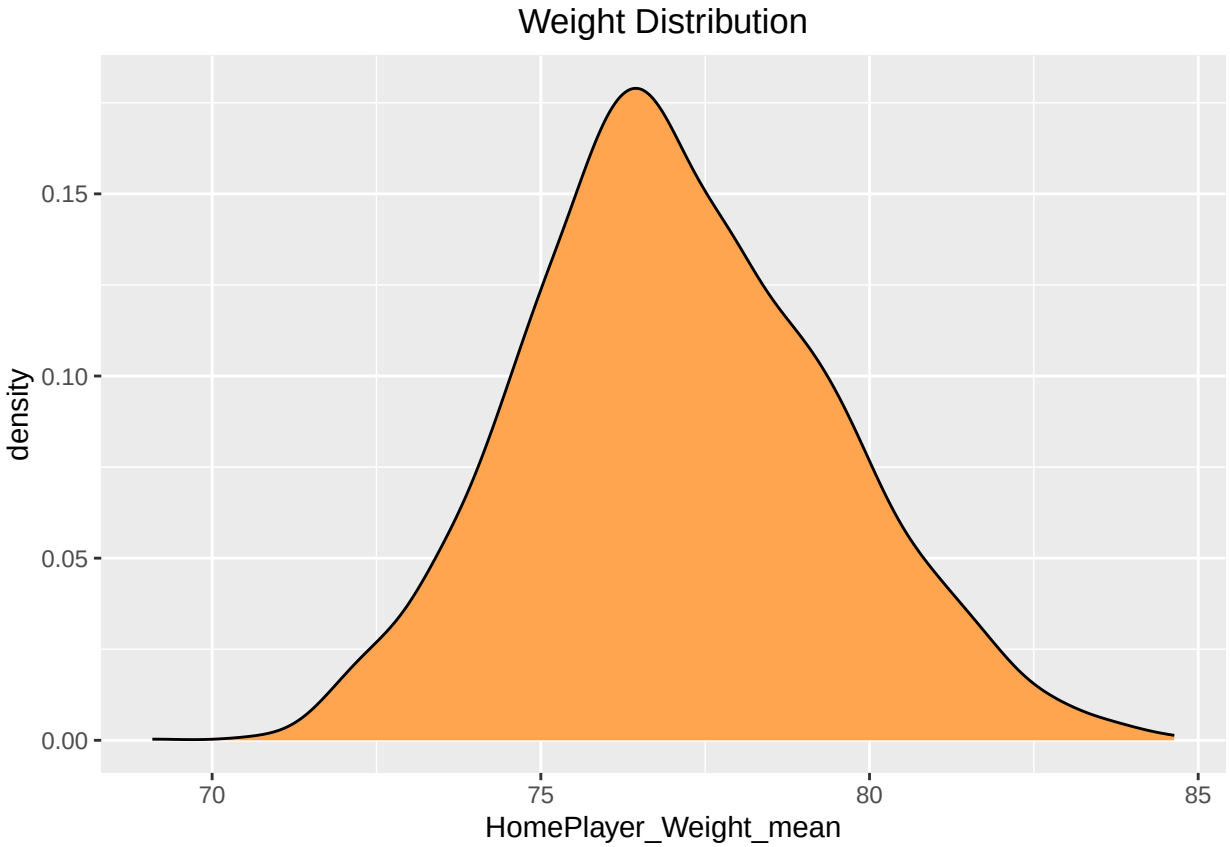

```
ggplot(data=data, aes(x=HomePlayer_Age_mean)) +
  geom_density(fill="tan1") + labs(title="Age Distribution") + theme(plot.title = element_text(hjust = 
```



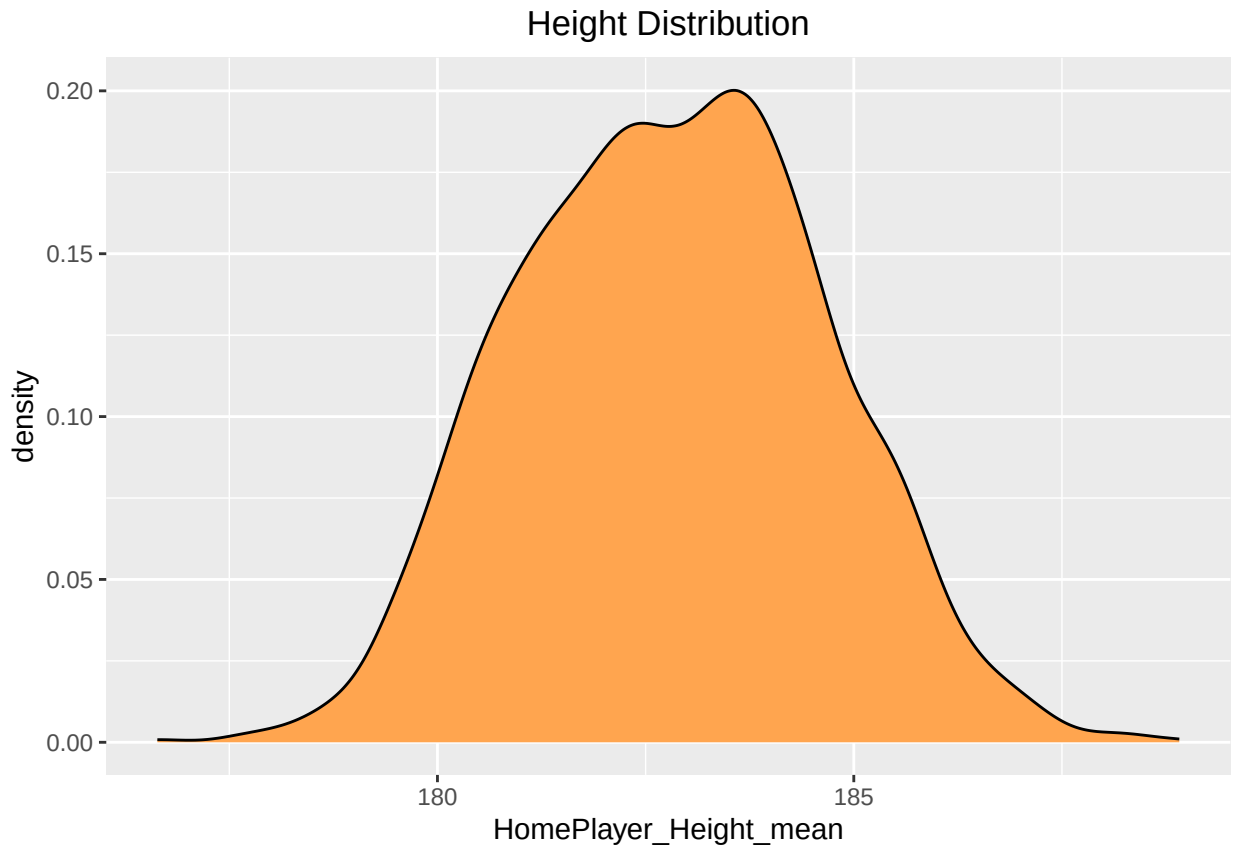

```
ggplot(data=data, aes(x=HomePlayer_Overall_sd)) +
  geom_density(fill="tan1") + labs(title="Overall SD distribution") + theme(plot.title = element_text(h
```




```
ggplot(data=data, aes(x=HomePlayer_Weight_mean)) +
  geom_density(fill="tan1") + labs(title="Weight Distribution") + theme(plot.title = element_text(hjust
```

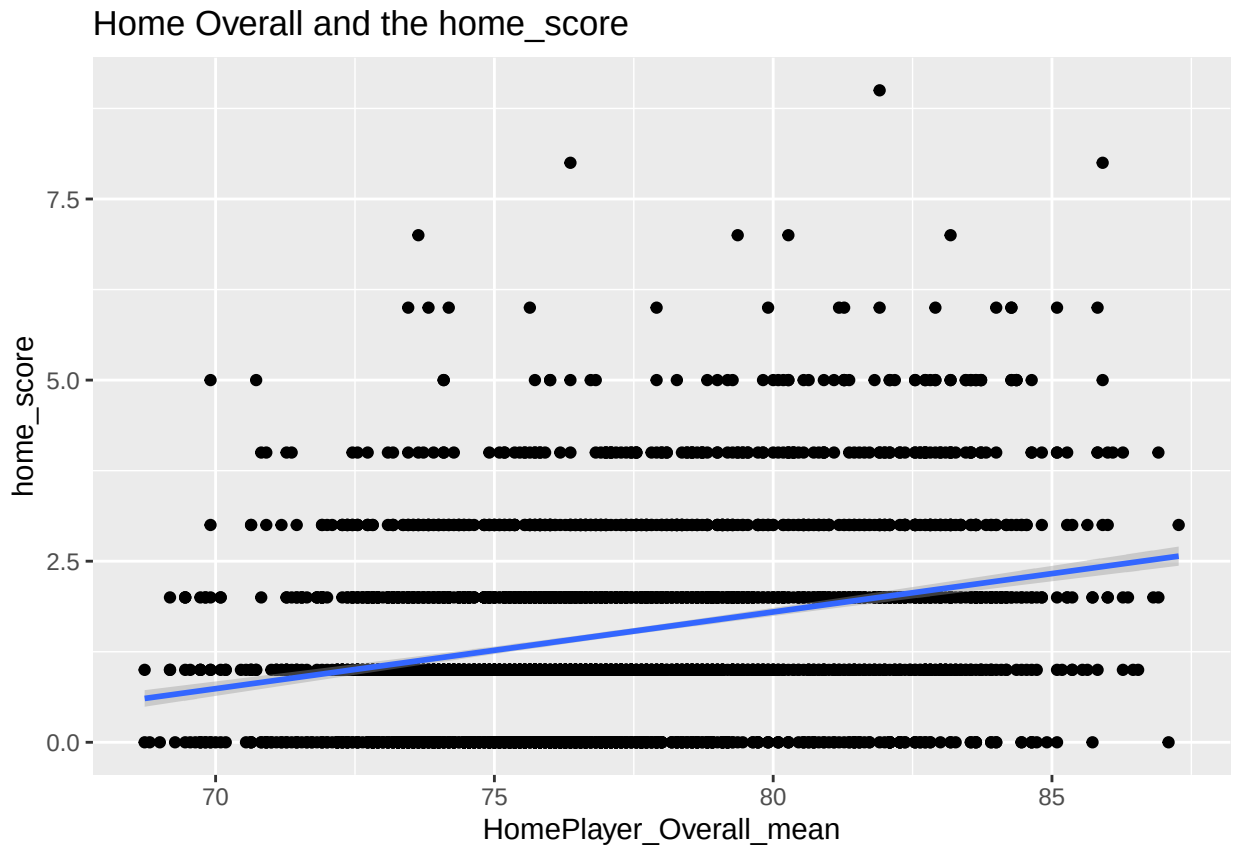



```
ggplot(data=data, aes(x=HomePlayer_Height_mean)) +
  geom_density(fill="tan1") + labs(title="Height Distribution") + theme(plot.title = element_text(hjust
```



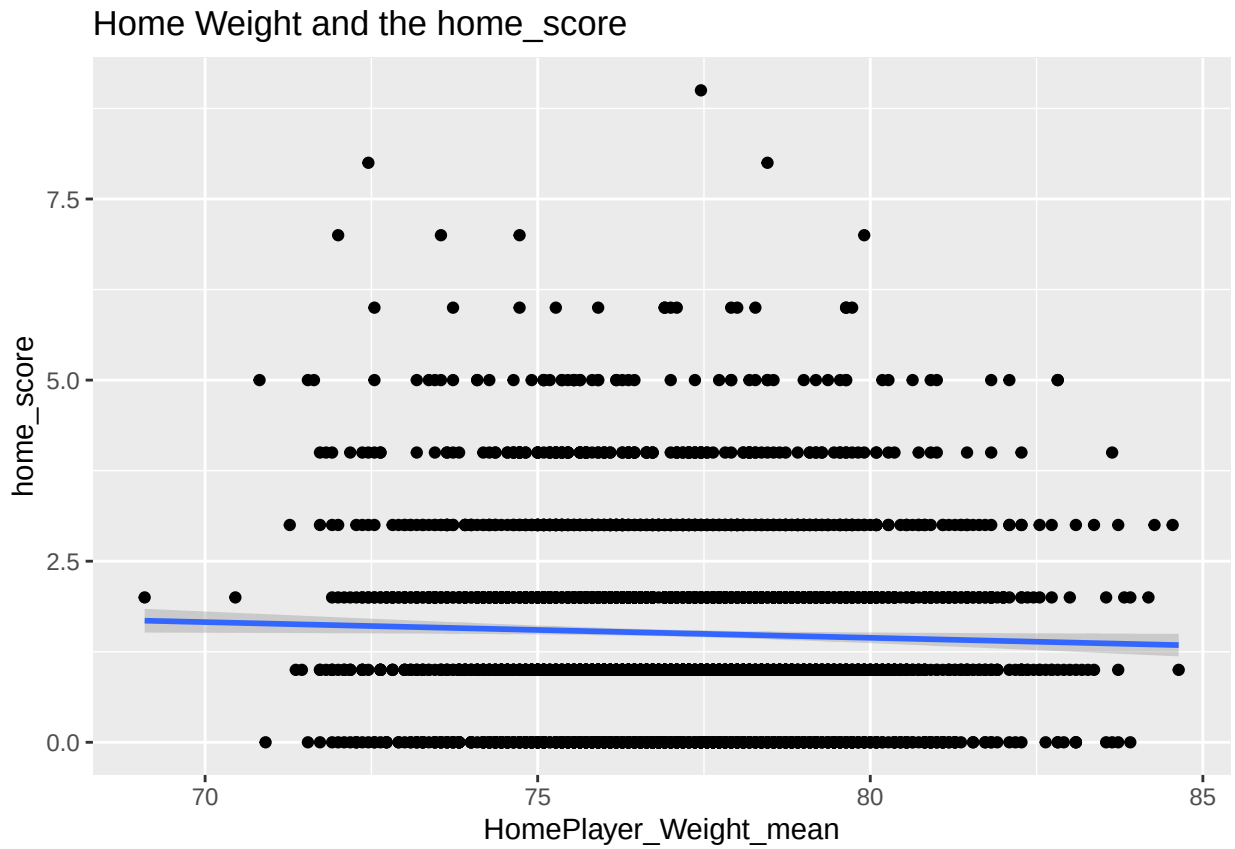

```
ggplot(data = data, aes(x = HomePlayer_Overall_mean,
                        y = home_score)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Overall and the home_score")
```

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



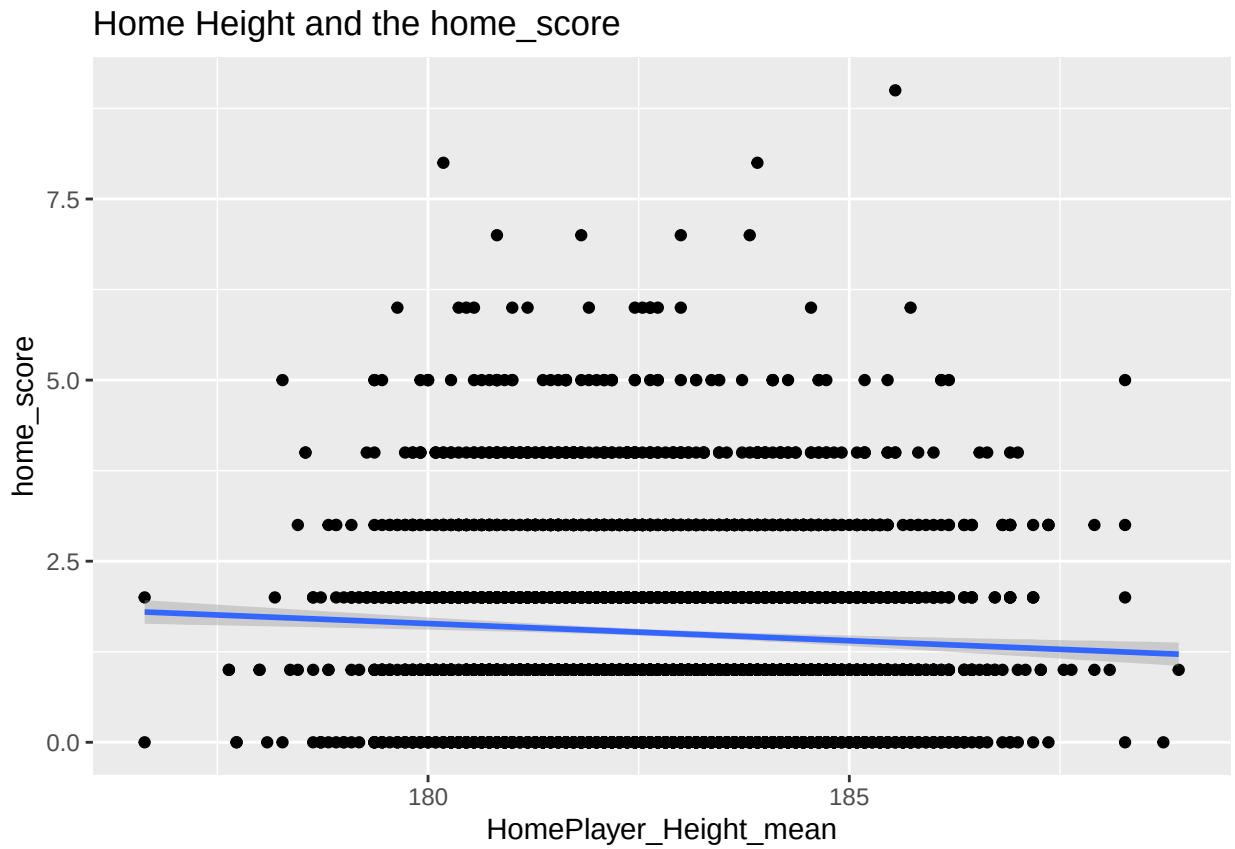

```
ggplot(data = data, aes(x = HomePlayer_Weight_mean,
                        y = home_score)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Weight and the home_score")
```

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



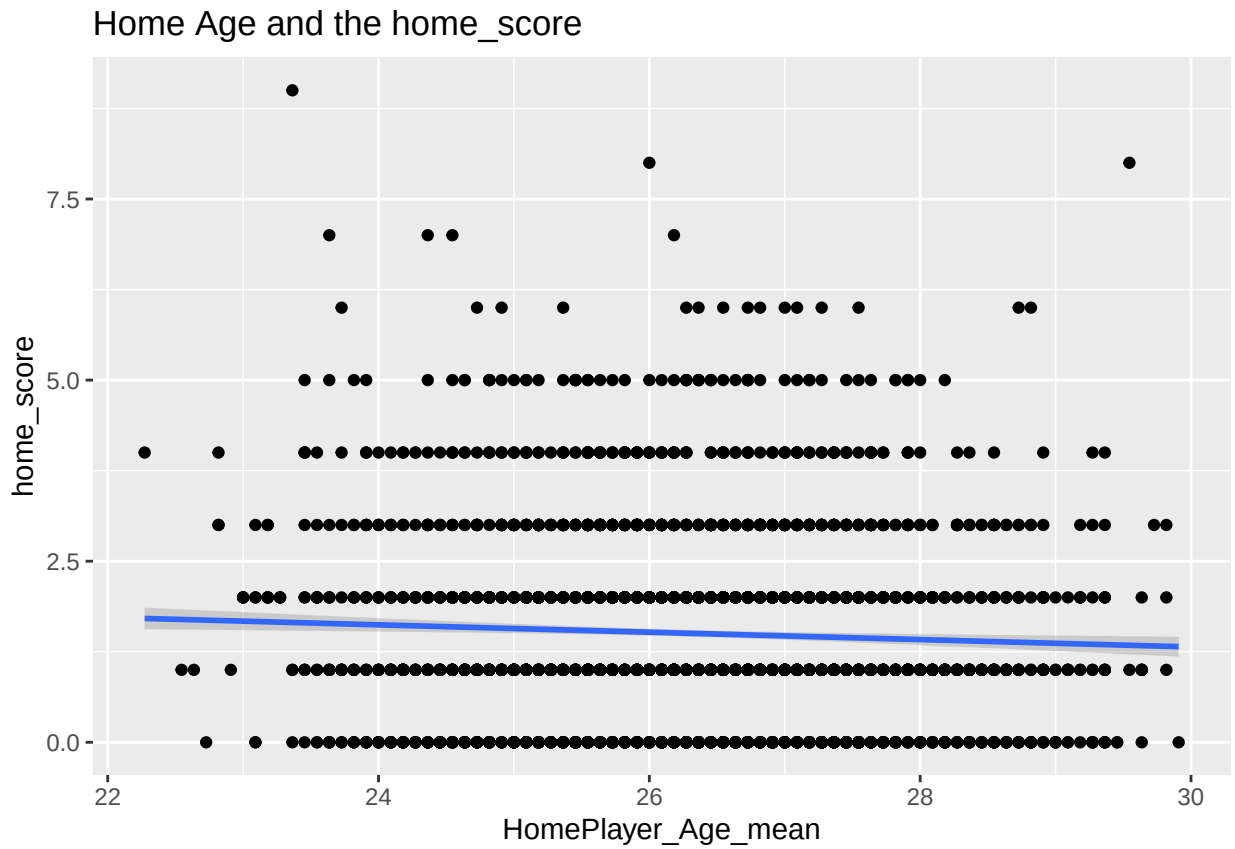

```
ggplot(data = data, aes(x = HomePlayer_Height_mean,
                        y = home_score)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Height and the home_score")
```

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




```
ggplot(data = data, aes(x = HomePlayer_Age_mean,
                        y = home_score)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Age and the home_score")
```

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

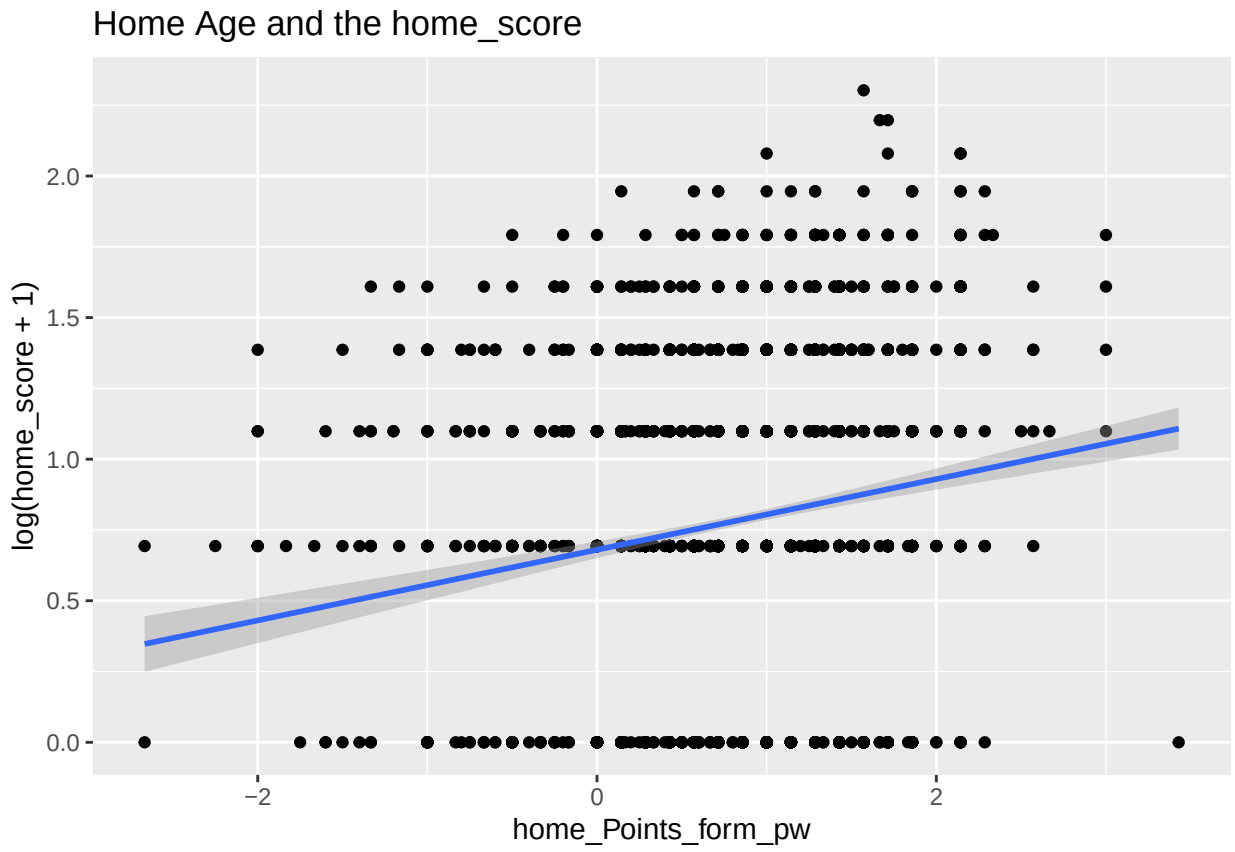



```
ggplot(data = data, aes(x = home_GD_prior- away_GD_prior,
                        y = home_score)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Age and the home_score")
```

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

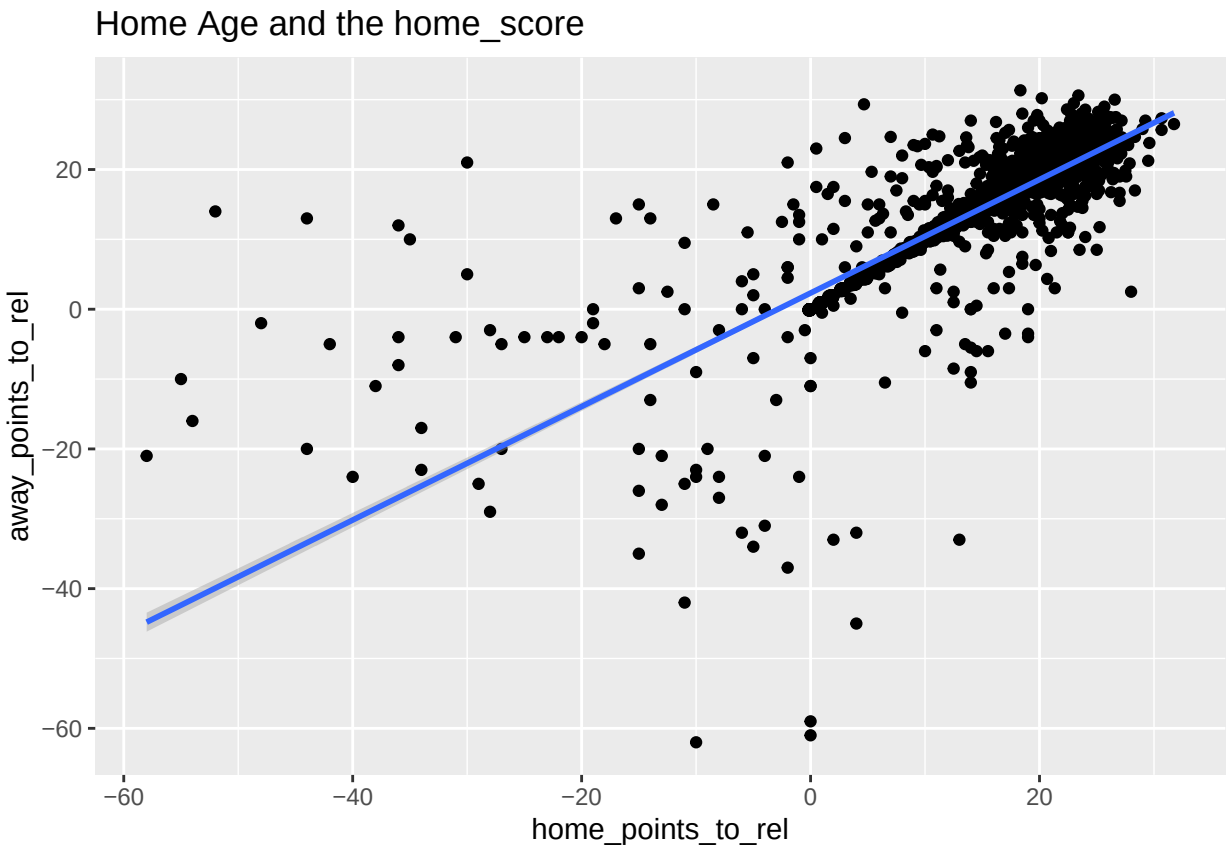

```
ggplot(data = data, aes(x = home_Points_form_pw,
                        y = log(home_score+1))) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Age and the home_score")
```

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




```
ggplot(data = data, aes(x = home_points_to_rel,
                        y = away_points_to_rel)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home Age and the home_score")
```

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



Now we will see different models to check the following:

1. Does the standard deviation have a positive effect on the teams performance?
2. How is the motivation to not get relegated affects the team different than the motivation to reach top 4 / win the league
3. Which is better in estimating form? Goal differences before the game or points gained

First we will see which is better at estimating form: points form or GD form over the past 5-6 (per week)

```
data <- data%>%
  mutate(GD = home_score-away_score)
GD_model <- felm(data=data, formula = GD ~
                   HomePlayer_Overall_mean+
                 HomePlayer_Overall_sd +
                   AwayPlayer_Overall_mean+
                   AwayPlayer_Overall_sd+
                   home_GD_form_pw +
                   away_GD_form_pw
                  | home_team_name + away_team_name | 0 | home_team_name)

Points_model <- felm(data=data, formula = GD ~
                   HomePlayer_Overall_mean+
                 HomePlayer_Overall_sd +
                   AwayPlayer_Overall_mean+
                   AwayPlayer_Overall_sd+
                   home_Points_form_pw +
```

```
                away_Points_form_pw
               | home_team_name + away_team_name | 0 | home_team_name)

summary(GD_model)
```

```
##
## Call:
##    felm(formula = GD ~ HomePlayer_Overall_mean + HomePlayer_Overall_sd +      AwayPlayer_Overall_mean
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7618 -1.0386 -0.0212  1.0493  7.9060
##
## Coefficients:
##                          Estimate Cluster s.e. t value Pr(>|t|)
## HomePlayer_Overall_mean  0.04022      0.01610   2.499  0.01798 *
## HomePlayer_Overall_sd    0.07088      0.02649   2.676  0.01180 *
## AwayPlayer_Overall_mean -0.04364      0.02202  -1.982  0.05637 .
## AwayPlayer_Overall_sd   -0.02199      0.02078  -1.058  0.29807
## home_GD_form_pw          0.08652      0.04851   1.784  0.08429 .
## away_GD_form_pw         -0.21837      0.05348  -4.083  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.646 on 2971 degrees of freedom
## Multiple R-squared(full model): 0.2448   Adjusted R-squared: 0.2275
## Multiple R-squared(proj model): 0.01331   Adjusted R-squared: -0.009269
## F-statistic(full model, *iid*):14.16 on 68 and 2971 DF, p-value: < 2.2e-16
## F-statistic(proj model):  6.23 on 6 and 31 DF, p-value: 0.0002253
```

```
summary(Points_model)
```

```
##
## Call:
##    felm(formula = GD ~ HomePlayer_Overall_mean + HomePlayer_Overall_sd +      AwayPlayer_Overall_mean
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8721 -1.0598 -0.0232  1.0390  7.8976
##
## Coefficients:
##                          Estimate Cluster s.e. t value Pr(>|t|)
## HomePlayer_Overall_mean  0.04221      0.01635   2.581   0.0148 *
## HomePlayer_Overall_sd    0.06922      0.02723   2.542   0.0162 *
## AwayPlayer_Overall_mean -0.04593      0.02202  -2.085   0.0454 *
## AwayPlayer_Overall_sd   -0.02688      0.02009  -1.338   0.1906
## home_Points_form_pw      0.10381      0.04979   2.085   0.0454 *
## away_Points_form_pw     -0.13316      0.08593  -1.550   0.1314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.65 on 2971 degrees of freedom
## Multiple R-squared(full model): 0.2411   Adjusted R-squared: 0.2237
```

```
## Multiple R-squared(proj model): 0.008456   Adjusted R-squared: -0.01424
## F-statistic(full model, *iid*):13.88 on 68 and 2971 DF, p-value: < 2.2e-16
## F-statistic(proj model): 4.341 on 6 and 31 DF, p-value: 0.002726
```

Now we see looking at the adjusted R squared for the full model, it's better off to use to GD which makes sense because there will be instances where 2 teams are in the same form point wise but one team scores more goals.

Now we will add motivation to the mix:

```
staying_league_model <- felm(data=data, formula = GD ~
                HomePlayer_Overall_mean + AwayPlayer_Overall_mean+
              HomePlayer_Overall_sd+
                AwayPlayer_Overall_sd+
                +home_GD_form_pw + away_GD_form_pw+
                home_points_to_rel + away_points_to_rel
               | home_team_name + away_team_name | 0 |home_team_name)


tr_model <- felm(data=data, formula = GD ~
                HomePlayer_Overall_mean + AwayPlayer_Overall_mean+
              HomePlayer_Overall_sd+
                AwayPlayer_Overall_sd+
                 +home_GD_form_pw + away_GD_form_pw+
                home_points_to_championship + away_points_to_championship
               | home_team_name + away_team_name | 0 | home_team_name)

gi_model <- felm(data=data, formula = GD ~
                HomePlayer_Overall_mean + AwayPlayer_Overall_mean+
              HomePlayer_Overall_sd+
                AwayPlayer_Overall_sd+
                 +home_GD_form_pw + away_GD_form_pw+
                home_match_importance + away_match_importance
               | home_team_name + away_team_name | 0 | home_team_name)

summary(gi_model)
```

```
##
## Call:
##    felm(formula = GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean +      HomePlayer_Overall_s
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7565 -1.0336 -0.0194  1.0516  7.9023
##
## Coefficients:
##                          Estimate Cluster s.e. t value Pr(>|t|)
## HomePlayer_Overall_mean  0.03957      0.01602   2.470 0.019240 *
## AwayPlayer_Overall_mean -0.04359      0.02227  -1.958 0.059342 .
## HomePlayer_Overall_sd    0.07062      0.02670   2.645 0.012719 *
## AwayPlayer_Overall_sd   -0.02167      0.02065  -1.049 0.302074
## home_GD_form_pw          0.08607      0.04878   1.765 0.087491 .
## away_GD_form_pw         -0.21617      0.05322  -4.062 0.000307 ***
## home_match_importance   -0.10055      0.08780  -1.145 0.260895
```

```
## away_match_importance    0.11685      0.09993   1.169 0.251177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.646 on 2969 degrees of freedom
## Multiple R-squared(full model): 0.2452   Adjusted R-squared: 0.2274
## Multiple R-squared(proj model): 0.01375   Adjusted R-squared: -0.009507
## F-statistic(full model, *iid*):13.78 on 70 and 2969 DF, p-value: < 2.2e-16
## F-statistic(proj model): 4.938 on 8 and 31 DF, p-value: 0.0005327
```

```
summary(tr_model)
```

```
##
## Call:
##    felm(formula = GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean +      HomePlayer_Overall_s
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7419 -1.0283 -0.0165  1.0365  7.9212
##
## Coefficients:
##                              Estimate Cluster s.e. t value Pr(>|t|)
## HomePlayer_Overall_mean      0.041293     0.016030   2.576 0.014988 *
## AwayPlayer_Overall_mean     -0.042186     0.021998  -1.918 0.064394 .
## HomePlayer_Overall_sd        0.066671     0.026333   2.532 0.016630 *
## AwayPlayer_Overall_sd       -0.025572     0.021417  -1.194 0.241548
## home_GD_form_pw              0.082225     0.050088   1.642 0.110781
## away_GD_form_pw             -0.212584     0.052842  -4.023 0.000343 ***
## home_points_to_championship -0.007794     0.007826  -0.996 0.327019
## away_points_to_championship  0.009505     0.007893   1.204 0.237620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.646 on 2969 degrees of freedom
## Multiple R-squared(full model): 0.2457   Adjusted R-squared: 0.2279
## Multiple R-squared(proj model): 0.01449   Adjusted R-squared: -0.008745
## F-statistic(full model, *iid*):13.82 on 70 and 2969 DF, p-value: < 2.2e-16
## F-statistic(proj model): 5.593 on 8 and 31 DF, p-value: 0.0002019
```

```
summary(staying_league_model)
```

```
##
## Call:
##    felm(formula = GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean +      HomePlayer_Overall_s
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7670 -1.0324 -0.0101  1.0400  7.9111
##
## Coefficients:
##                           Estimate Cluster s.e. t value Pr(>|t|)
## HomePlayer_Overall_mean  0.040650     0.016234   2.504 0.017755 *
## AwayPlayer_Overall_mean -0.043090     0.021934  -1.965 0.058487 .
```

```
## HomePlayer_Overall_sd    0.070378     0.026463   2.660 0.012273 *
## AwayPlayer_Overall_sd   -0.021887     0.021783  -1.005 0.322778
## home_GD_form_pw          0.080000     0.049799   1.606 0.118313
## away_GD_form_pw         -0.211521     0.053062  -3.986 0.000379 ***
## home_points_to_rel      -0.008169     0.007610  -1.073 0.291340
## away_points_to_rel       0.009348     0.008864   1.055 0.299730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.646 on 2969 degrees of freedom
## Multiple R-squared(full model): 0.2454   Adjusted R-squared: 0.2276
## Multiple R-squared(proj model): 0.01403   Adjusted R-squared: -0.009216
## F-statistic(full model, *iid*):13.79 on 70 and 2969 DF, p-value: < 2.2e-16
## F-statistic(proj model): 5.669 on 8 and 31 DF, p-value: 0.0001811
```

```
linearHypothesis(gi_model, c("home_match_importance= 0", "away_match_importance=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## home_match_importance = 0
## away_match_importance = 0
##
## Model 1: restricted model
## Model 2: GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean + HomePlayer_Overall_sd +
##     AwayPlayer_Overall_sd + +home_GD_form_pw + away_GD_form_pw +
##     home_match_importance + away_match_importance | home_team_name +
##     away_team_name | 0 | home_team_name
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1     33
## 2     31  2 1.415     0.4929
```

```
linearHypothesis(tr_model, c("home_points_to_championship= 0", "away_points_to_championship=0"))
```

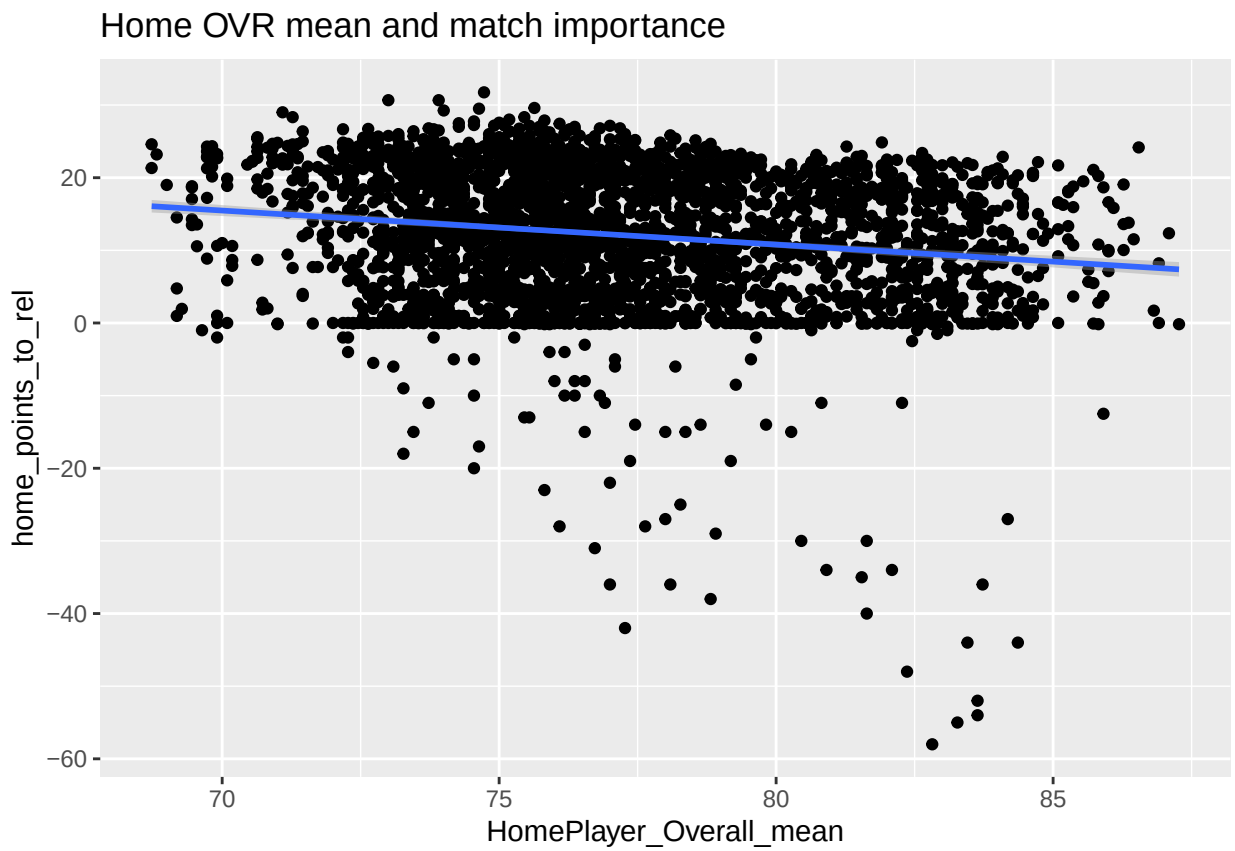
```
## Linear hypothesis test
##
## Hypothesis:
## home_points_to_championship = 0
## away_points_to_championship = 0
##
## Model 1: restricted model
## Model 2: GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean + HomePlayer_Overall_sd +
##     AwayPlayer_Overall_sd + +home_GD_form_pw + away_GD_form_pw +
##     home_points_to_championship + away_points_to_championship |
##     home_team_name + away_team_name | 0 | home_team_name
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1     33
## 2     31  2 2.8826     0.2366
```

```r
linearHypothesis(staying_league_model, c("home_points_to_rel= 0", "away_points_to_rel=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## home_points_to_rel = 0
## away_points_to_rel = 0
##
## Model 1: restricted model
## Model 2: GD ~ HomePlayer_Overall_mean + AwayPlayer_Overall_mean + HomePlayer_Overall_sd +
##     AwayPlayer_Overall_sd + +home_GD_form_pw + away_GD_form_pw +
##     home_points_to_rel + away_points_to_rel | home_team_name +
##     away_team_name | 0 | home_team_name
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1     33
## 2     31  2 1.207     0.5469
```

```r
ggplot(data = data, aes(x = HomePlayer_Overall_mean,
                        y = home_points_to_rel)) +
  geom_point() + stat_smooth(method = "lm") +
  ggtitle("Home OVR mean and match importance")
```

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



Naively we can say there is a problem when motivation, but that can be due to the high correlations between both the teams quality, and their position in the league, and we can gain information mainley from non linear models that can use such information to predict better.