

Valor de Mercado dos Jogadores da UEFA Euro 2024

Andres Kindel Barbosa

Luiz Adriano Augusto dos Santos

Gustavo Beretta Gonçalves

Josef Konrad Holanda Brüseke

→ Introdução e objetivos:

A base de dados escolhida contém 10 informações diferentes sobre todos os 623 atletas participantes do campeonato europeu de futebol de seleções de 2024 (UEFA Euro 2024). Todas as análises foram realizadas com auxílio dos softwares estatísticos R e RStudio.

O objetivo deste trabalho é analisar diferentes dados sobre os jogadores e relacioná-los entre si para que seja possível tirar *insights* sobre os assuntos analisados, especialmente sobre a influência desses dados sobre o valor de mercado dos jogadores.

→ Materiais e métodos:

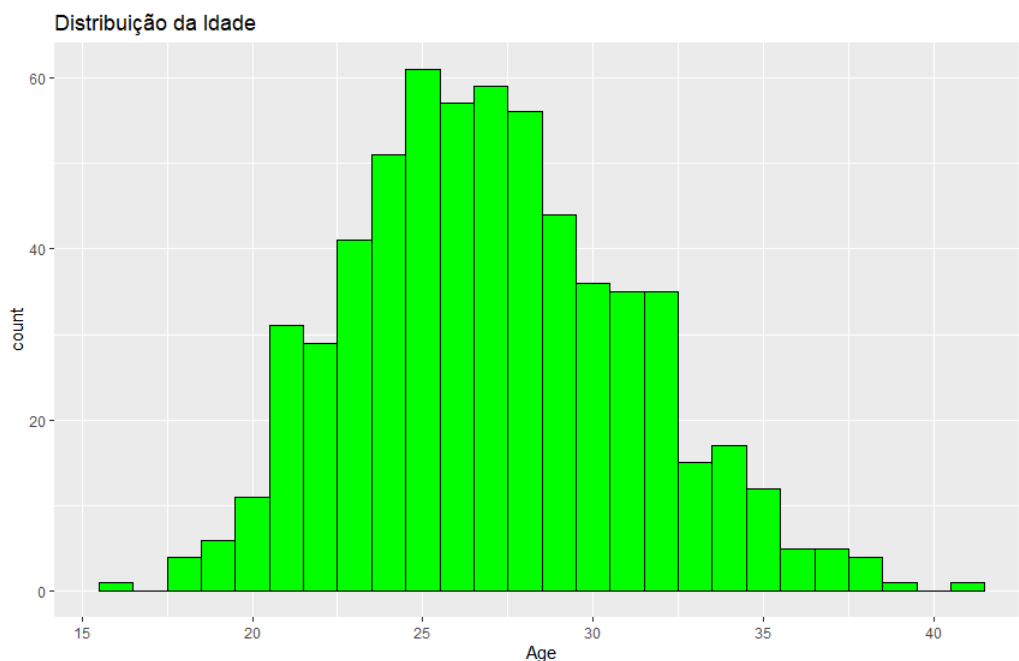
Para realizar uma análise aprofundada da influência sobre a variável MarketValue, foram selecionadas as seguintes variáveis: Age, Caps, Goals, Foot, Position e Country. Cada uma delas desempenha um papel crucial na determinação do valor de mercado de um jogador de futebol. Ao analisar essas variáveis, podemos obter uma visão mais clara de como cada uma contribui para a avaliação do valor de mercado dos jogadores, permitindo uma análise mais precisa e fundamentada.

Nome	Descrição	Tipo
Age	Idade do jogador	Quantitativa discreta
Caps	Partidas jogadas pelo jogador representando sua seleção	Quantitativa discreta
Goals	Gols marcados pelo jogador representando sua seleção	Quantitativa discreta

MarketValue	Valor de mercado do jogador (em euros)	Quantitativa discreta
Foot	Pé dominante do jogador	Qualitativa nominal
Position	Posição que o jogador atua	Qualitativa nominal
Country	Seleção que o jogador representa	Qualitativa nominal

→ Resultados e discussões:

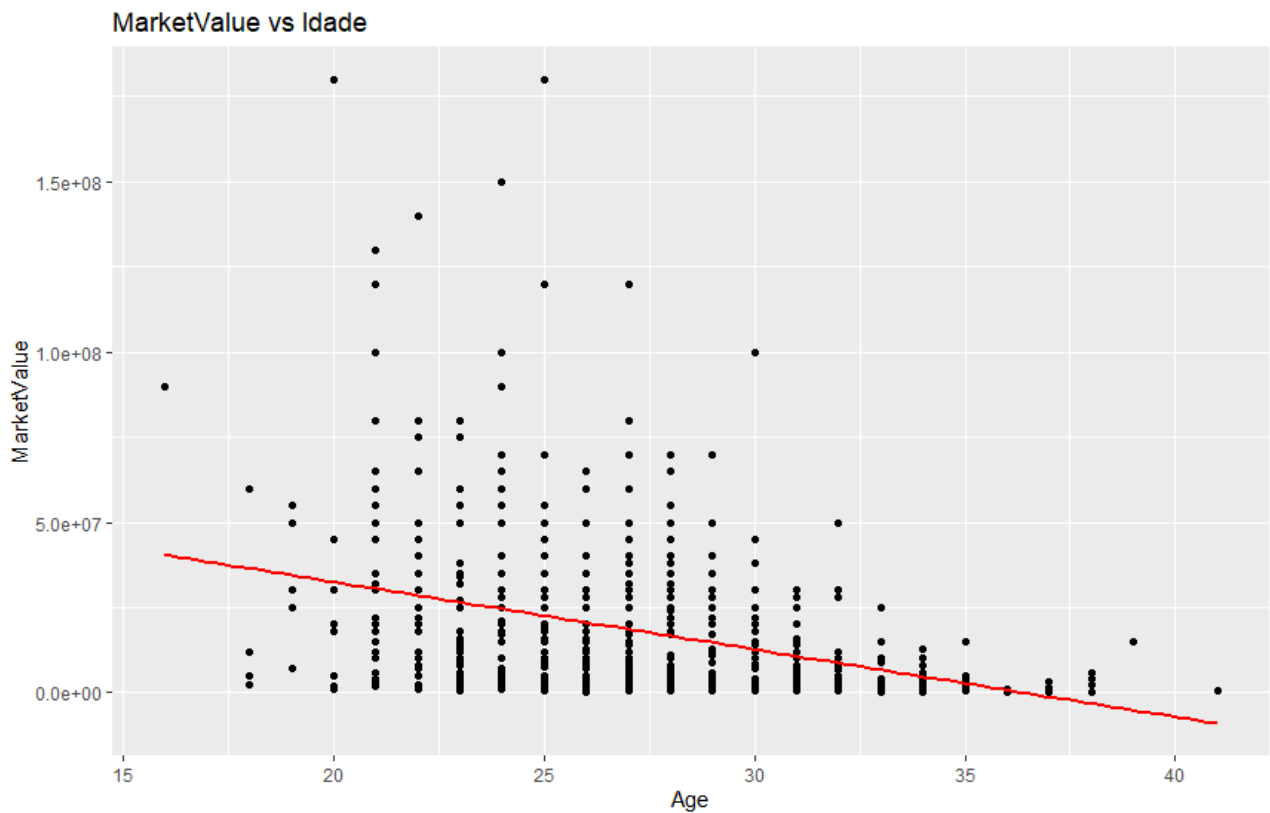
- **Distribuição dos jogadores pela sua idade**



Este gráfico mostra a distribuição da idade dos jogadores no dataset.

- A maior concentração de jogadores está na faixa etária de 20 a 30 anos, com um pico em torno dos 25 anos.
- Há uma diminuição gradual no número de jogadores conforme a idade aumenta após os 30 anos.
- Poucos jogadores têm menos de 20 ou mais de 35 anos, indicando que a maioria dos jogadores está em sua juventude ou em início de maturidade atlética.

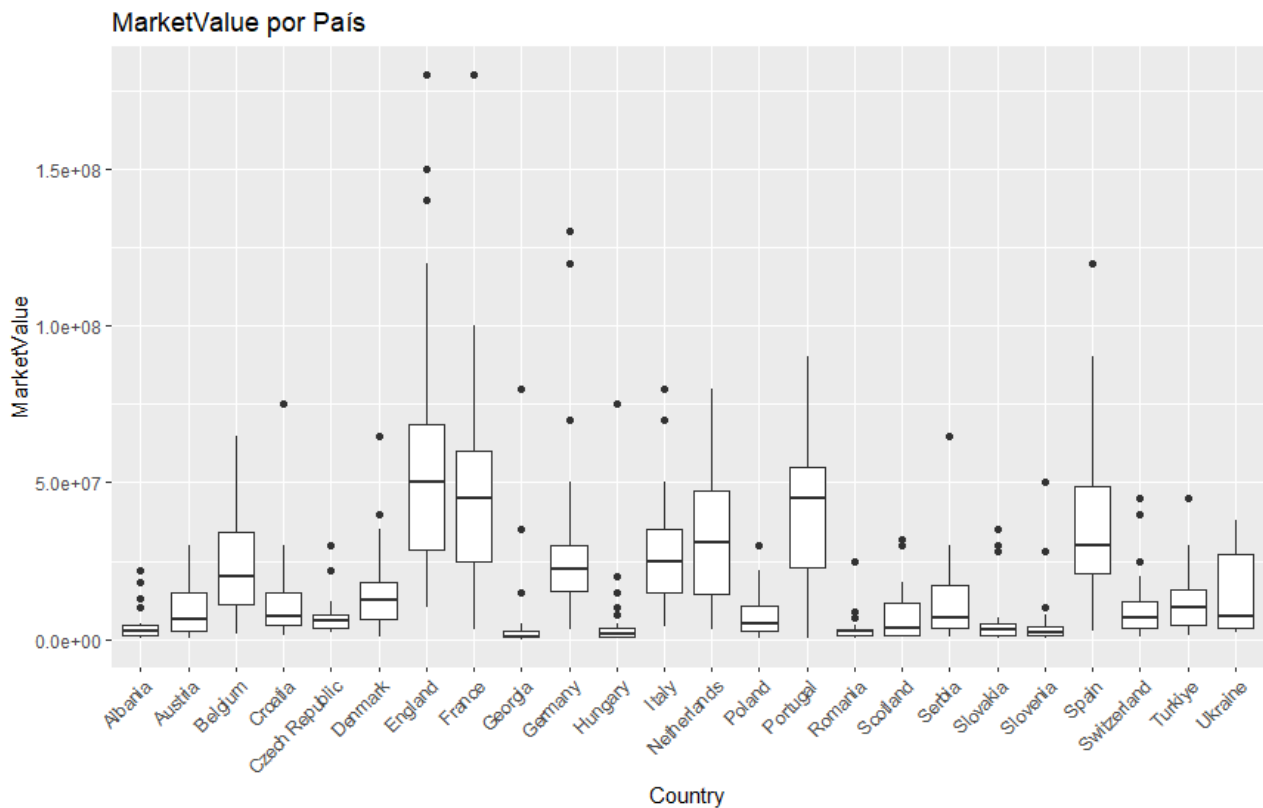
- **Relação entre a idade dos jogadores e seu valor de mercado**



Este gráfico apresenta a relação entre a idade dos jogadores e o valor de mercado.

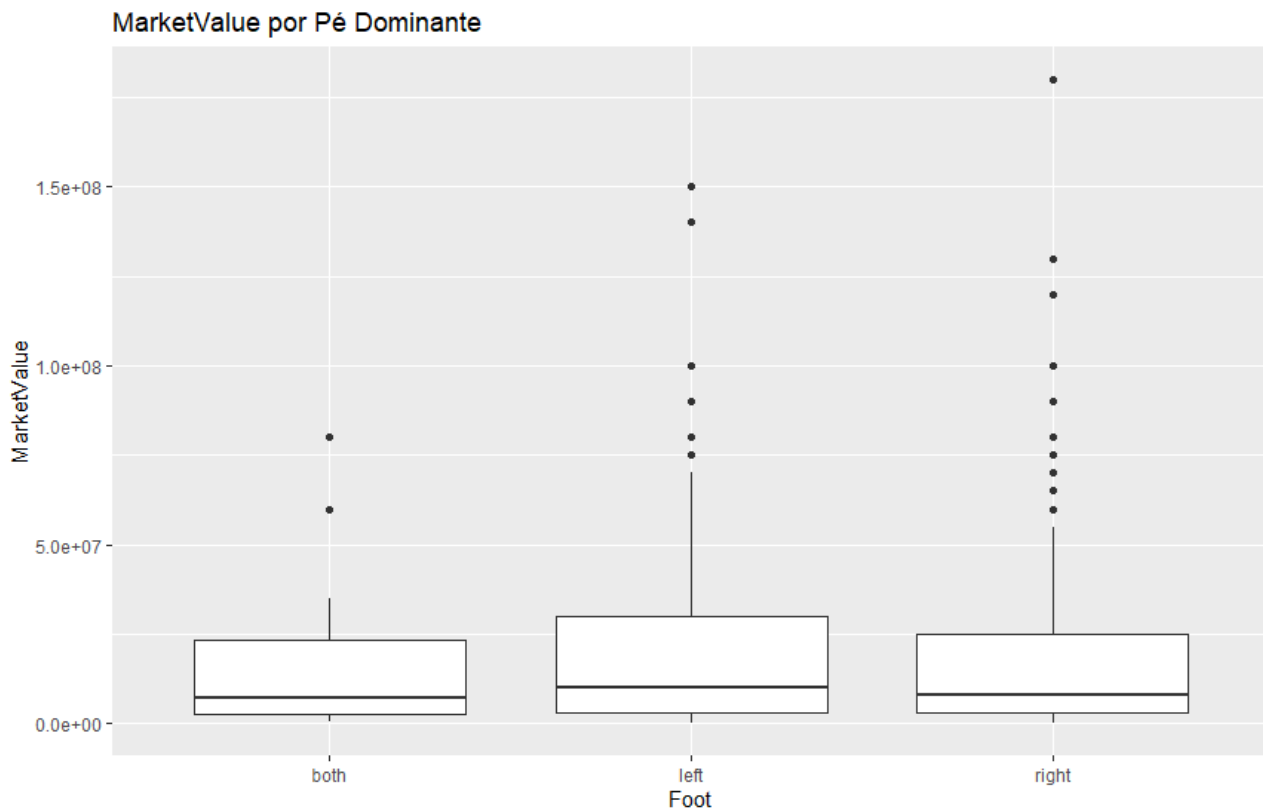
- Há uma tendência de declínio no valor de mercado conforme a idade dos jogadores aumenta.
- Jogadores mais jovens (entre 20 e 25 anos) têm uma maior variação de valor de mercado, com alguns jogadores de alto valor.
- A partir dos 30 anos, o valor de mercado dos jogadores tende a diminuir significativamente, indicando que a juventude é um fator importante para o valor de mercado.

- Valor de mercado de jogadores por país



Os valores de mercado médio dos jogadores variam entre os países. A **Inglaterra** é o país com o maior valor médio, seguido por **Portugal**, **França** e **Holanda**. Os países com os menores valores de mercado médio são a **Geórgia**, a **Eslovênia** e a **Hungria**.

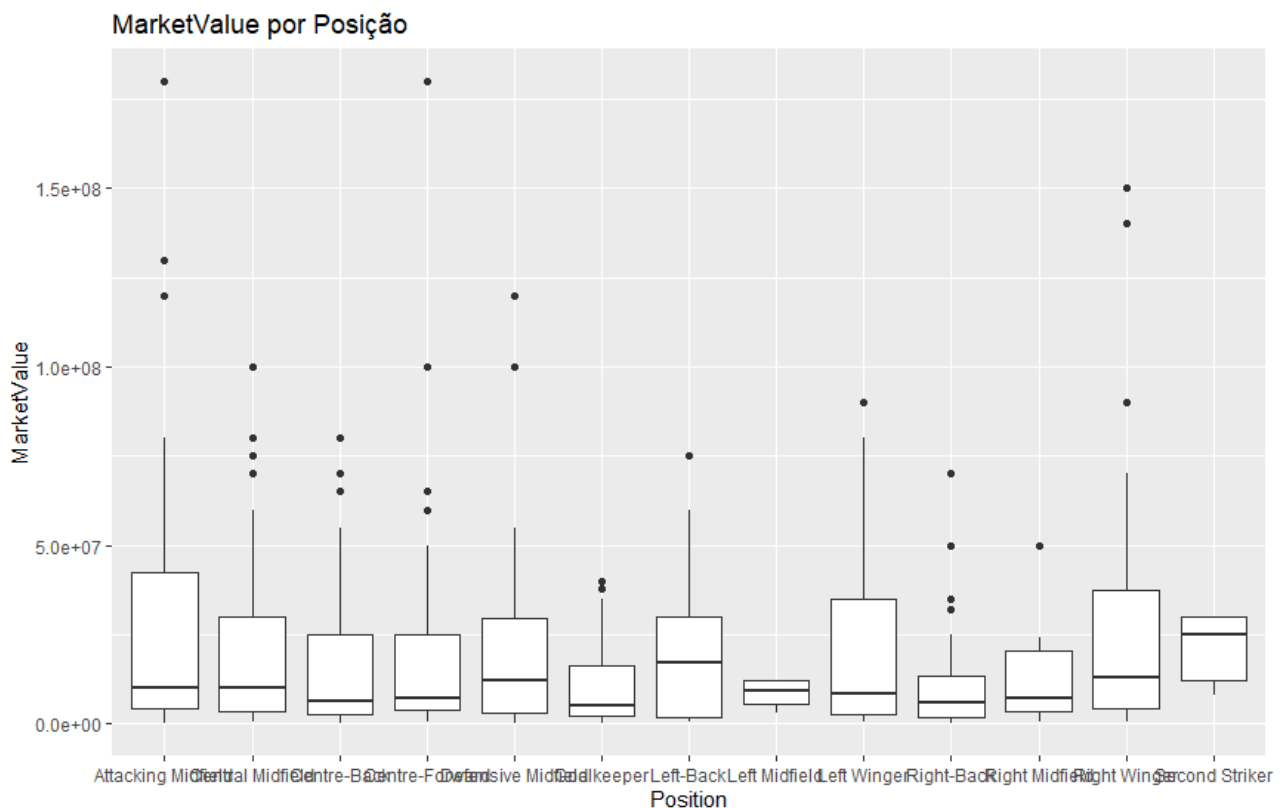
Valor de mercado de jogadores por pé dominante



A análise do boxplot do valor de mercado de jogadores por pé dominante sugere que o pé dominante pode influenciar o valor de mercado dos jogadores, com jogadores canhotos tendo, em média, maior valor de mercado do que jogadores destros. Isso faz sentido, muito por conta da disponibilidade de jogadores canhotos no mercado, que é menor do que de destros. Os dados da própria competição mostram isso: 150 canhotos e 436 destros.

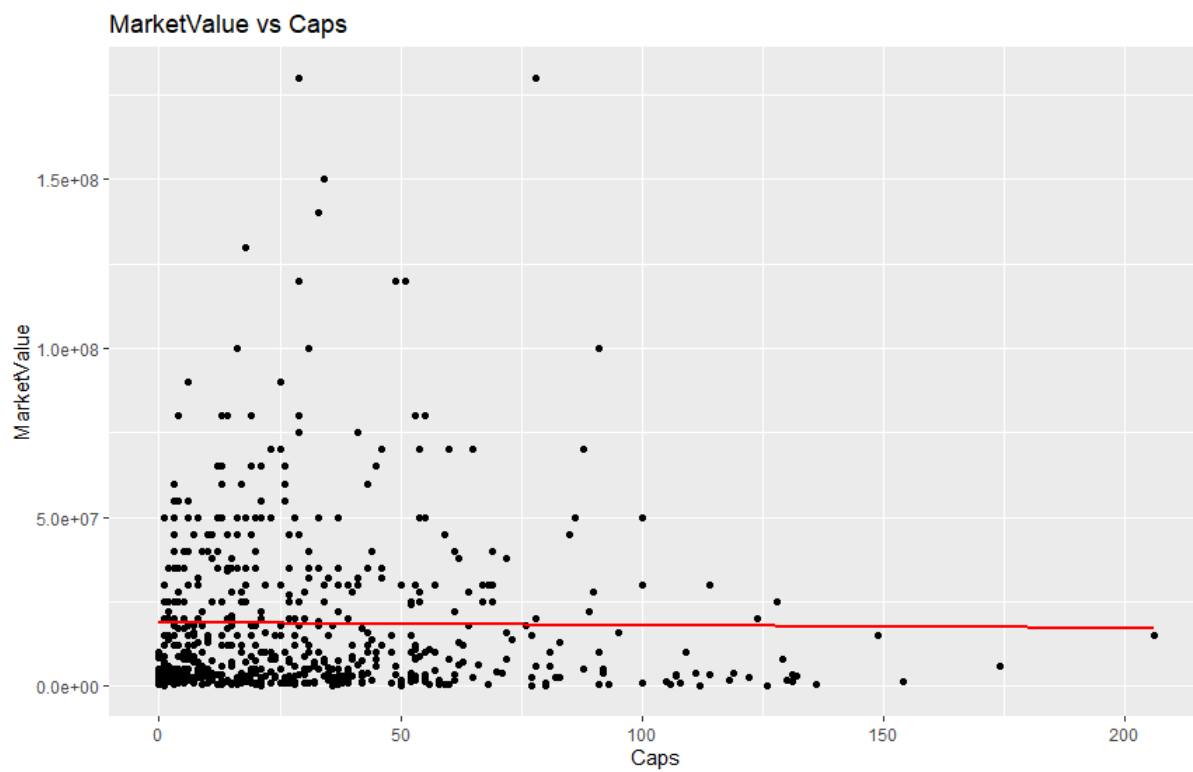
No entanto, é contra intuitivo o dado de que jogadores ambidestros têm valor médio menor do que canhotos e destros. É importante ressaltar que o pé dominante é apenas um dos fatores que influenciam o valor de mercado de um jogador, e que outros fatores também devem ser considerados na avaliação.

- **Valor de mercado de jogadores por posição**



- ❖ **Medianas Próximas dos Quartis:** Observa-se que a maioria das medianas não está centralizada nas caixas, mas sim próximas de um dos quartis (principalmente Q1). Isso indica que a distribuição dos valores de mercado não é simétrica. Com a mediana da maioria das posições mais próximas de Q1, logo a maior parte dos jogadores tem valores de mercado mais próximos ao valor mínimo.
- ❖ **Valores Mínimos de Mercado:** O valor mínimo dos jogadores em diferentes posições é bastante semelhante, exceto para a posição de "Second Striker", que tem um valor mínimo um pouco mais alto comparado às outras posições.
- ❖ **Posições com Maiores Medianas:** As posições de "Left Back" e "Second Striker" têm medianas mais altas comparadas às outras posições. Isso indica que, em média, os jogadores nessas posições possuem valores de mercado mais altos.

- Valor de mercado por partidas jogadas

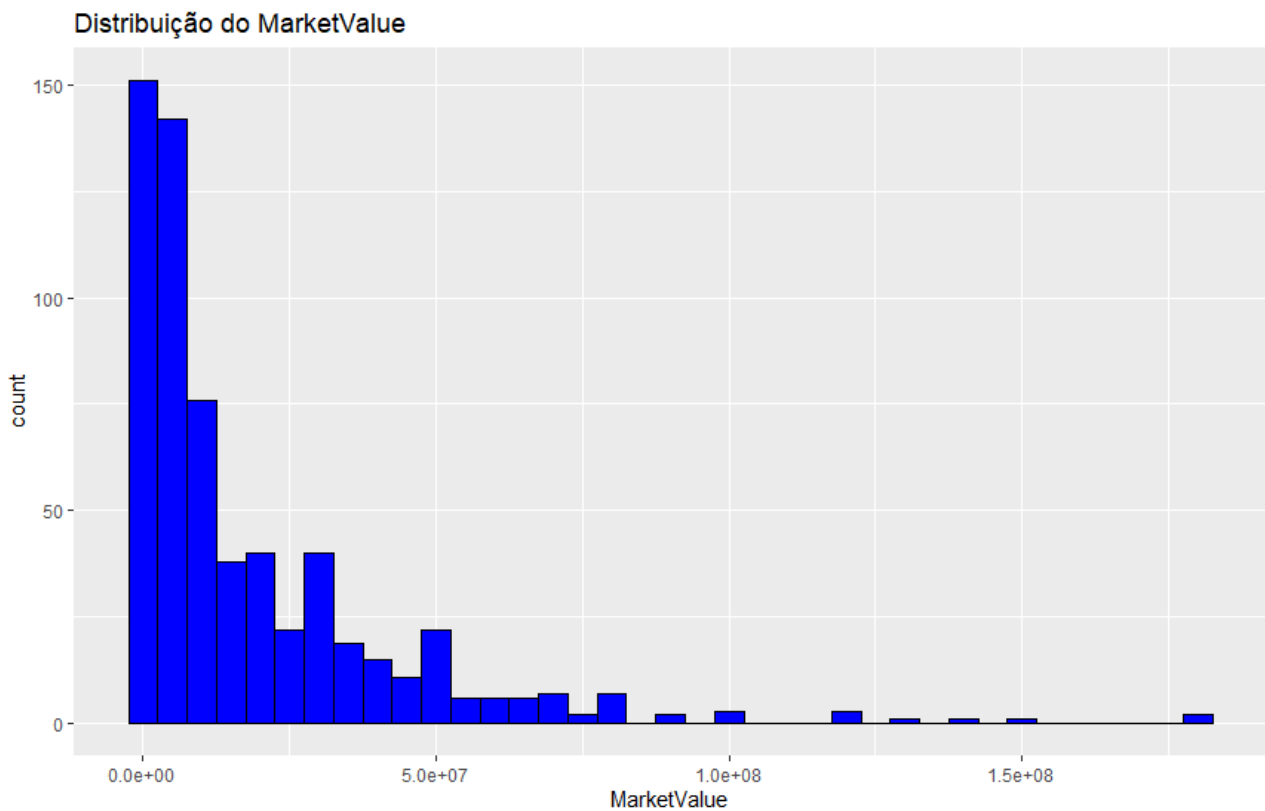


Com base na análise do gráfico, observa-se que não existe relação clara entre o valor de mercado de um jogador e o número de partidas jogadas.

MODELO DE REGRESSÃO LINEAR MÚLTIPLA COM INTERAÇÃO:

$$\begin{aligned} \text{Log}(\text{MarketValue}) \sim & \text{Age} + \text{Log}(\text{Caps}+1) + \text{Log}(\text{Goals}+1) + \text{Country} + \\ & \text{Age}:\text{Log}(\text{Caps}+1) + \text{Log}(\text{Caps}+1):\text{Log}(\text{Goals}+1) \end{aligned}$$

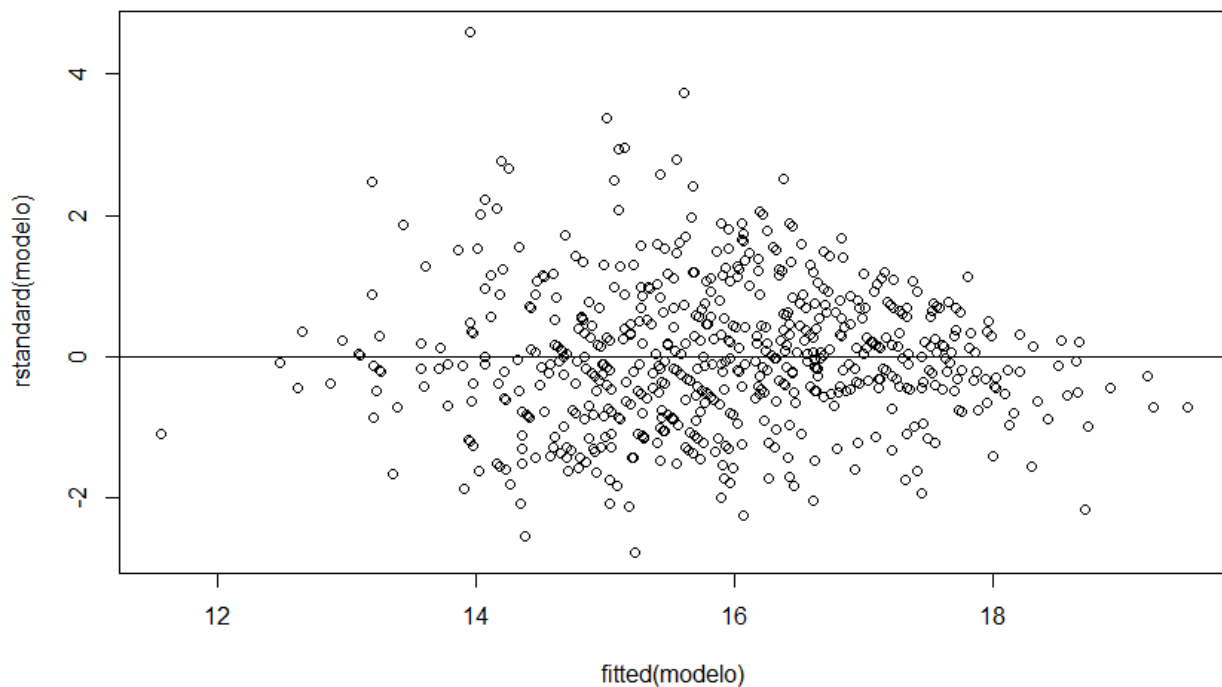
Primeiramente, foi realizada uma transformação logarítmica em algumas das variáveis quantitativas. Como tínhamos alguns dados muito discrepantes, essa transformação estabiliza a variância e reduz a assimetria do modelo. Essa discrepância pode ser observada no gráfico abaixo:



Já o motivo das variáveis Caps (número de jogos) e Goals (gols) estarem sendo somadas a 1 antes de sofrerem a transformação é porque diversos jogadores têm 0 partidas jogadas ou gols marcados, o que impossibilitaria o cálculo do log (log(0) não existe).

Resíduos

Os resíduos são a diferença entre os valores observados e os valores ajustados pelo modelo.



A distribuição dos resíduos parece estar centrada em torno de zero, com uma leve assimetria positiva, indicando que há alguns jogadores cujo valor de mercado é subestimado pelo modelo.

Coefficientes

Os coeficientes da regressão indicam a mudança esperada na variável dependente logarítmica ($\log(\text{MarketValue})$) para uma determinada mudança na variável independente, mantendo todas as outras constantes.

- **Age:** Cada ano adicional de idade diminui o MarketValue em 7,9%, sugerindo que jogadores mais velhos tendem a ter menor valor de mercado.
- **$\log(\text{Caps} + 1)$:** Cada aumento de 1% em Caps aumenta o MarketValue em 1,7%, sugerindo que jogadores com mais partidas jogadas pelas suas seleções tendem a ter maior valor de mercado.

- **log(Goals + 1):** Cada aumento de 1% em Goals diminui o MarketValue em 0,33%, sugerindo que jogadores com mais gols marcados pelas suas seleções tendem a ter menor valor de mercado, o que é contraintuitivo e pode sugerir multicolinearidade ou uma relação complexa. Vale destacar que, apesar de significativa, essa variável possui menor significância que as demais (0.027099).

Países

Os coeficientes dos países indicam como o valor de mercado logarítmico de jogadores de cada país difere do país de referência (Albânia). O país de referência é esse por uma simples questão de ordem alfabética.

- **CountryEngland:** Jogadores da Inglaterra têm o maior coeficiente (**2.763075x**).
- Outros países também mostram coeficientes significativos, como **França (2.493242x)**, **Alemanha (2.435525x)**, **Espanha (2.486148x)** e **Portugal(2.268831x)** (justamente as favoritas para serem campeãs do torneio, segundo as casas de aposta.)
- **Hungria, Romênia, Eslovênia e Eslováquia** não têm coeficientes significativos.
- **Geórgia (-0.546464x)** é o único país com coeficiente negativo significativo (é também a única estreante do torneio).

Interações:

- **Age:log(Caps + 1) (-0.047930):** tem um efeito negativo, indicando que o impacto das partidas jogadas diminui com a idade.
- **log(Caps + 1):(Goals + 1) (0.127253):** tem um efeito positivo, sugerindo que o efeito combinado das partidas jogadas e dos gols marcados aumenta o valor de mercado.

Estatísticas do Modelo

- **Multiple R-squared (0.7686):** Aproximadamente 76.86% da variabilidade no log(MarketValue) é explicada pelo modelo.
- **Adjusted R-squared (0.7577):** Ajustado para o número de variáveis no modelo.
- **F-statistic (70.45), p-value (< 2.2e-16):** O modelo é altamente significativo, rejeitando a hipótese nula de que todos os coeficientes são zero.

→ Considerações finais:

Os resultados destacaram os principais aspectos que influenciam no valor de mercado de um jogador de futebol. Claramente, a idade é o fator principal, o que faz total sentido, já que um jogador novo tem muito tempo para se dedicar ao futebol no futuro, enquanto um jogador mais velho pode se aposentar em poucos anos, trazendo menor retorno financeiro para o clube que o compre.

Foi interessante analisar a variável Goals, que indicou que quanto mais gols um jogador marca menor seu valor de mercado, o que, a princípio, não faz sentido algum. A multicolinearidade pode explicar esse acontecimento e conseguimos entender melhor ao analisar a interação entre Caps e Goals:

Essa interação indicou que o efeito combinado das partidas jogadas e dos gols marcados aumenta o valor de mercado, o que faz sentido, pois um jogador marcar muito gols em poucas partidas pode ser “sorte”, mas, para manter um alto número de gols jogando muitas partidas, ele precisa ser realmente bom, o que reflete no seu valor de mercado mais elevado.

Outra interação interessante de analisar foi a de Age com Caps, que indicou que o impacto das partidas jogadas diminui com a idade, o que também faz sentido, já que, como dito anteriormente, o que realmente importa para o valor de mercado de um jogador é o potencial de entrega futura, que é muito maior para jogadores mais novos.

Os coeficientes dos países indicaram que algumas seleções contribuem para o valor de mercado de seus jogadores serem mais elevados, entretanto, essa variável deve ser interpretada entendendo o contexto do futebol mundial. Países com tradição no futebol e grandes populações tem mais chance de relevarem bons jogadores e com alto valor de mercado. Mas não é porque um jogador nasceu na Inglaterra que isso vai fazer ele ter um valor mais alto.

→ Referências bibliográficas:

DIZDAREVIC, DAMIR. UEFA EURO 2024 - Players. 2024. Disponível em: <<https://www.kaggle.com/datasets/damirdizdarevic/uefa-euro-2024-players>>. Acesso em: 22, Junho, 2024.

STATPLACE. Análise de Regressão Linear Múltipla - Definição. Disponível em: <<https://www.youtube.com/watch?v=1MbQpCNafWc&t=133s>>. Acesso em: 22, Junho, 2024.

PORTUGUÊS, DOCSITY. Econometria - Regressão Linear Simples - Modelo Log Nível. Disponível em: <<https://www.youtube.com/watch?v=XsrBKAYN17o>>. Acesso em: 25, Junho, 2024.

PORTUGUÊS, DOCSITY. Econometria - Regressão Linear Simples - Modelo Log Log. Disponível em: <<https://www.youtube.com/watch?v=3p1EiXD6Yts>>. Acesso em: 25, Junho, 2024.

trabalho.R

```
# Carregar as bibliotecas necessárias
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(car)

## Carregando pacotes exigidos: carData

##
## Attaching package: 'car'

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

# Carregar a base de dados
data <- read.csv("euro2024_players.csv")

# Visualizar as primeiras linhas dos dados
head(data)

##           Name      Position Age      Club Height
Foot Caps
## 1 Marc-André ter Stegen  Goalkeeper  32      FC Barcelona   187
right  40
## 2      Manuel Neuer  Goalkeeper  38      Bayern Munich   193
right 119
## 3      Oliver Baumann  Goalkeeper  34 TSG 1899 Hoffenheim   187
right   0
## 4      Nico Schlotterbeck Centre-Back  24 Borussia Dortmund   191
left  12
## 5      Jonathan Tah Centre-Back  28 Bayer 04 Leverkusen   195
right  25
## 6      Antonio Rüdiger Centre-Back  31      Real Madrid   190
right  69
##   Goals MarketValue Country
```

```
## 1      0      28000000 Germany
## 2      0      4000000 Germany
## 3      0      3000000 Germany
## 4      0      40000000 Germany
## 5      0      30000000 Germany
## 6      3      25000000 Germany
```

Descrição dos dados

```
summary(data)
```

```
##      Name      Position      Age      Club
## Length:623    Length:623    Min.   :16.00 Length:623
## Class :character Class :character 1st Qu.:24.00 Class
## Mode  :character Mode  :character Median :27.00 Mode
##                               Mean   :27.04
##                               3rd Qu.:30.00
##                               Max.   :41.00
##      Height      Foot      Caps      Goals
## Min.   :167.0    Length:623    Min.   :  0.00 Min.   :  0.000
## 1st Qu.:180.0    Class :character 1st Qu.:  7.00 1st Qu.:  0.000
## Median :185.0    Mode  :character Median : 21.00 Median :  1.000
## Mean   :184.2                      Mean   : 30.34 Mean   :  4.152
## 3rd Qu.:189.0                      3rd Qu.: 42.00 3rd Qu.:  4.000
## Max.   :202.0                      Max.   :206.00 Max.   :128.000
##      MarketValue      Country
## Min.   :    50000    Length:623
## 1st Qu.:   290000    Class :character
## Median :   900000    Mode  :character
## Mean   : 18409029
## 3rd Qu.: 25000000
## Max.   :180000000
```

```
table(data$Foot)
```

```
##
##      - both left right
##      3      3      31      150      436
```

Converter a variável de mercado (MarketValue) em numérica (se não for)
data\$MarketValue <- as.numeric(gsub("[^0-9]", "", data\$MarketValue))

Transformar variáveis qualitativas em fatores
data\$Position <- as.factor(data\$Position)

Transformar campos vazios, "-", ou "N/A" na coluna Foot em NA
data\$Foot <- ifelse(data\$Foot == "" | data\$Foot == "-" | data\$Foot == "N/A", NA, data\$Foot)
data\$Foot <- as.factor(data\$Foot)

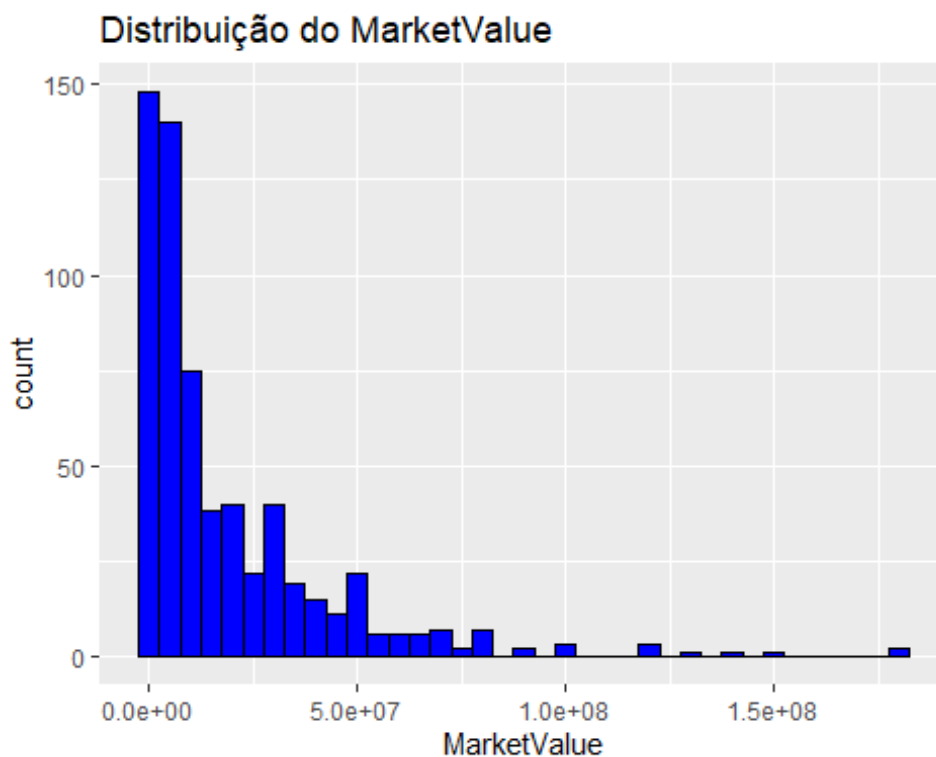
```

# Transformar a coluna Country em fator
data$Country <- as.factor(data$Country)

# Remover linhas com valores NA na coluna Foot para as análises
data <- data[!is.na(data$Foot), ]

# Visualização inicial dos dados
# Histogramas
ggplot(data, aes(x = MarketValue)) +
  geom_histogram(binwidth = 5000000, fill = "blue", color = "black") +
  ggtitle("Distribuição do MarketValue")

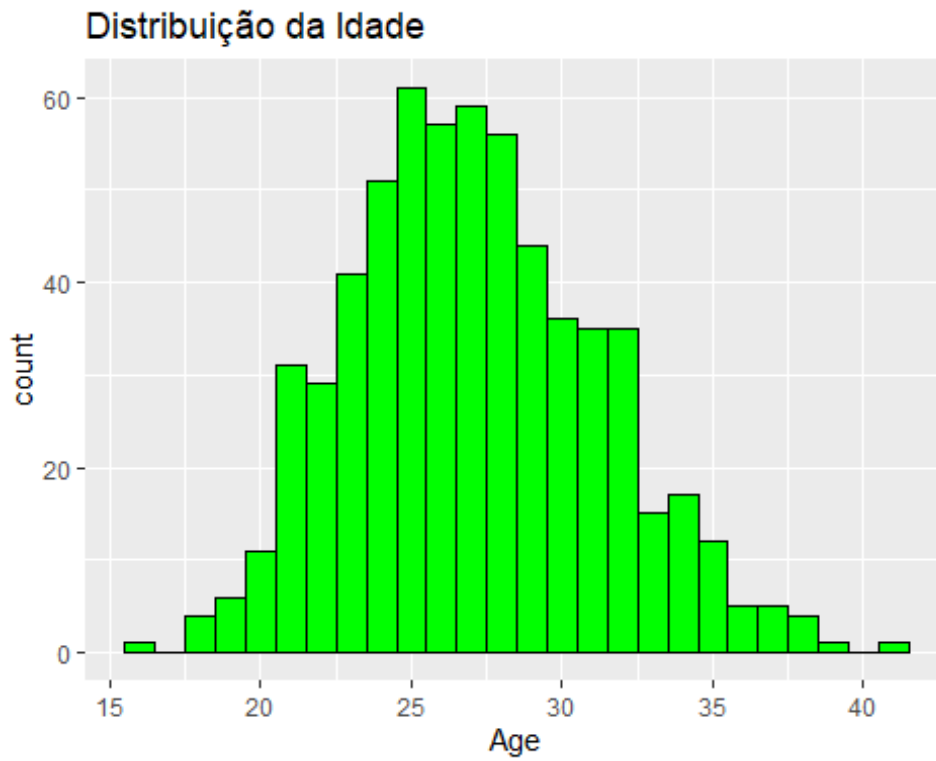
```



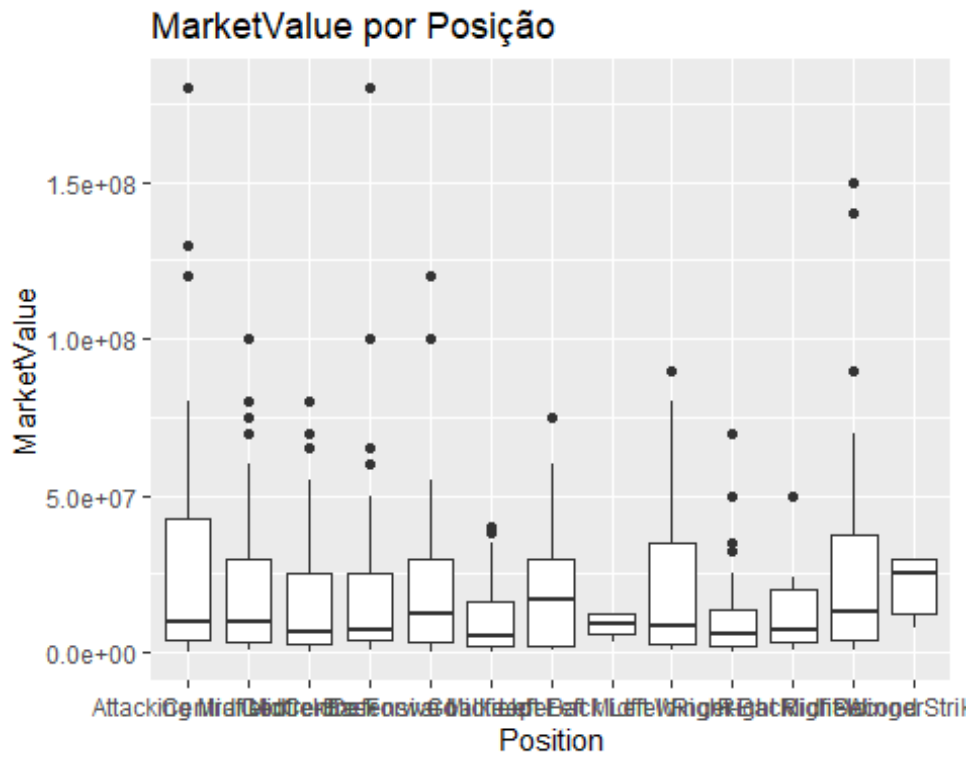
```

ggplot(data, aes(x = Age)) +
  geom_histogram(binwidth = 1, fill = "green", color = "black") +
  ggtitle("Distribuição da Idade")

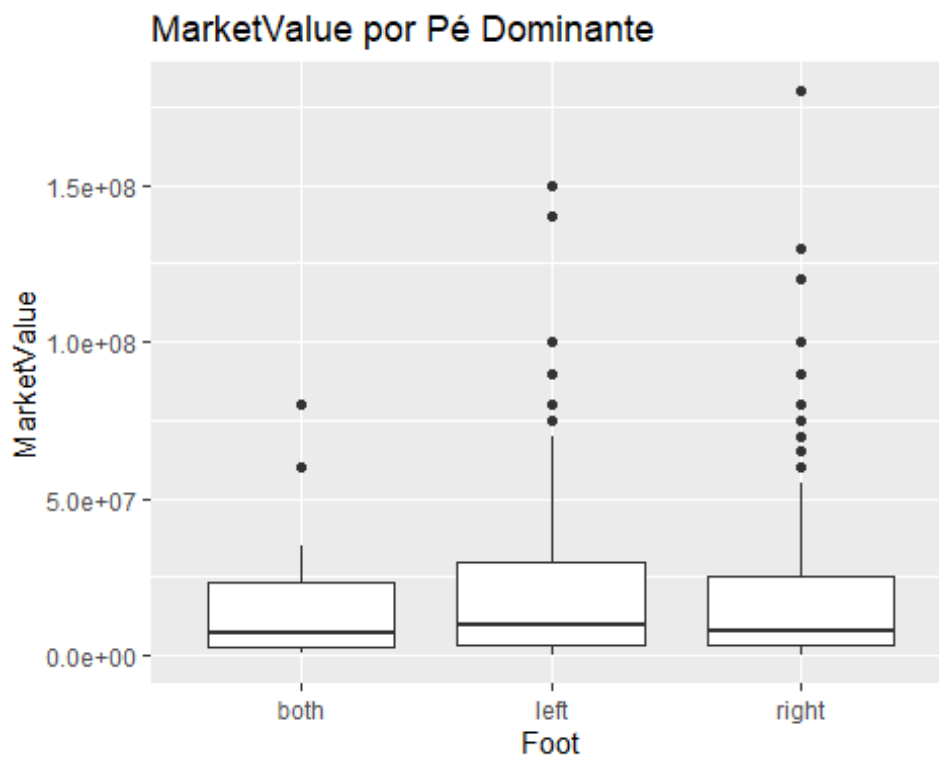
```



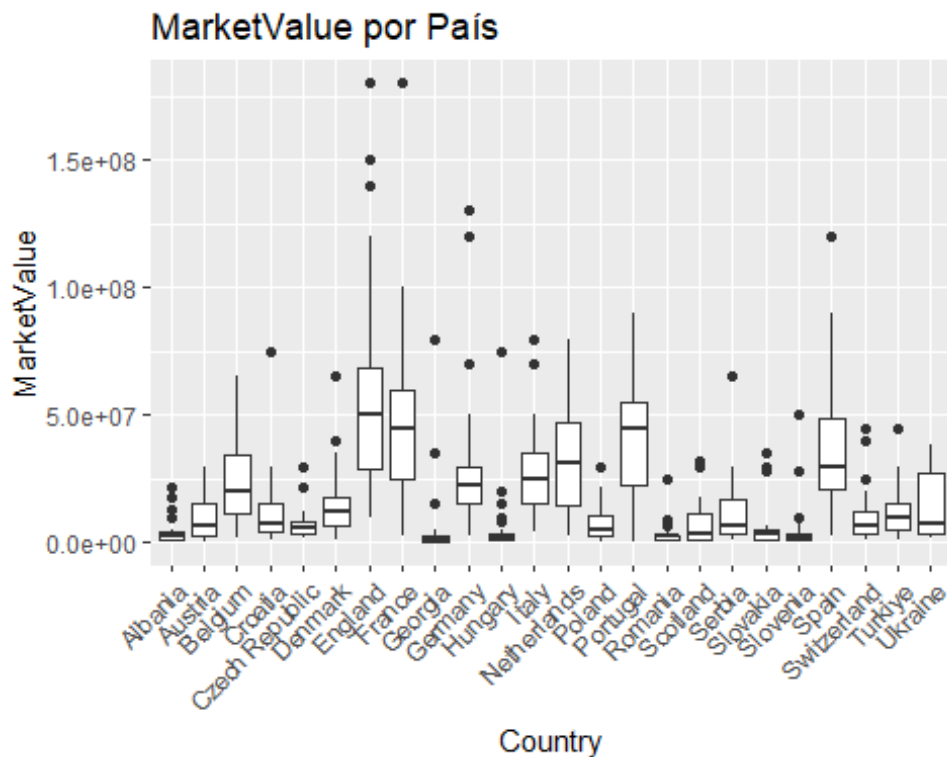
```
# Boxplots
ggplot(data, aes(x = Position, y = MarketValue)) +
  geom_boxplot() +
  ggtitle("MarketValue por Posição")
```

```
ggplot(data, aes(x = Foot, y = MarketValue)) +
  geom_boxplot() +
  ggtitle("MarketValue por Pé Dominante")
```

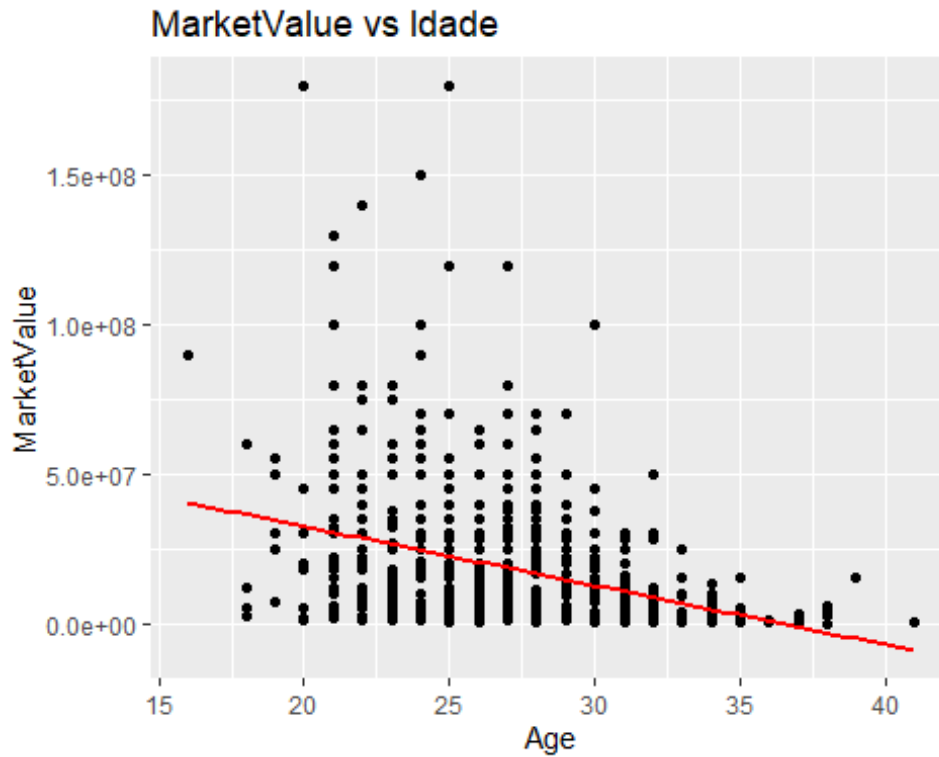


```
# Análise adicional por país
ggplot(data, aes(x = Country, y = MarketValue)) +
  geom_boxplot() +
  ggtitle("MarketValue por País") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



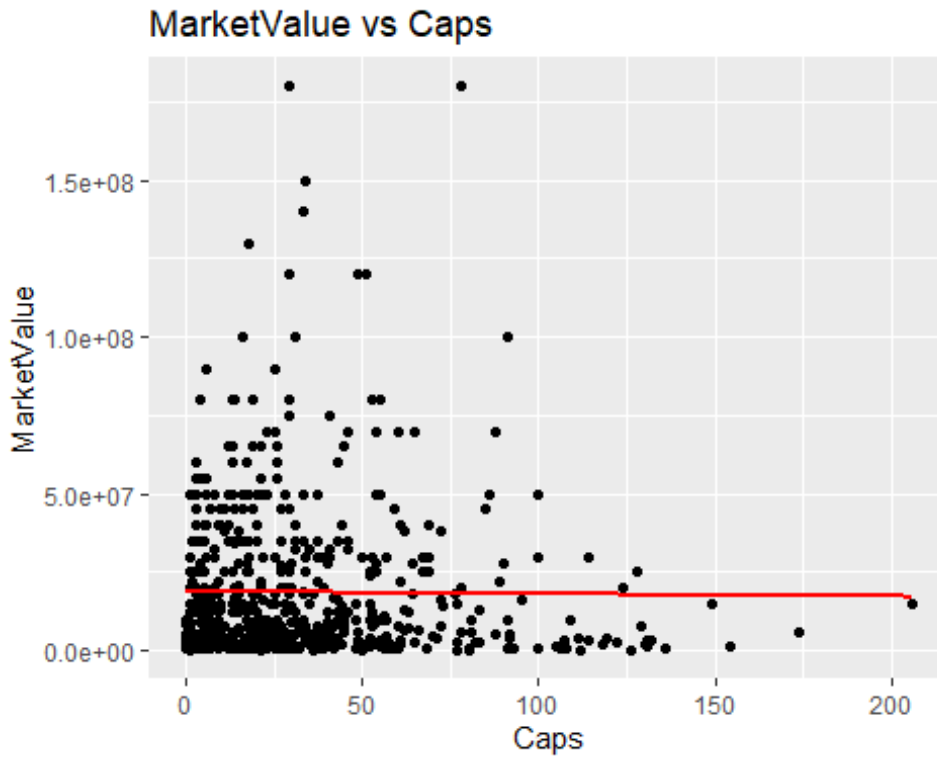
```
# Análise por idade
ggplot(data, aes(x = Age, y = MarketValue)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("MarketValue vs Idade")

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



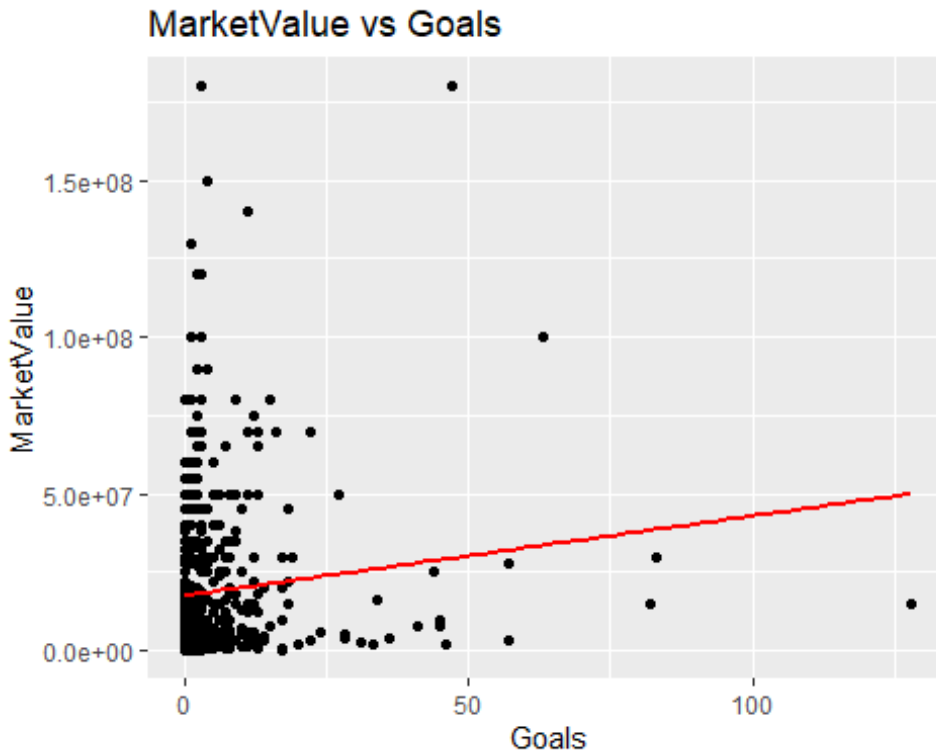
```
# Análise por número de partidas jogadas (Caps)
ggplot(data, aes(x = Caps, y = MarketValue)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("MarketValue vs Caps")
```

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



```
# Análise por número de gols marcados (Goals)
ggplot(data, aes(x = Goals, y = MarketValue)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  ggtitle("MarketValue vs Goals")
```

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



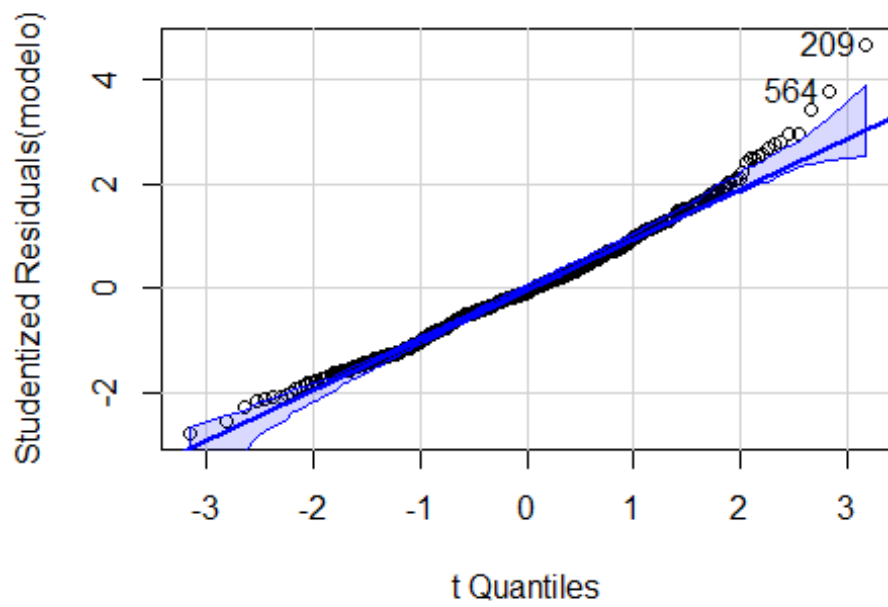
```
# Montando modelo de regressão múltipla
modelo <- lm(log(MarketValue) ~ Age + log(Caps + 1) + log(Goals + 1) +
Country + Age:log(Caps + 1) + log(Caps + 1):log(Goals + 1), data = data)

# Análises do modelo
summary(modelo)

##
## Call:
## lm(formula = log(MarketValue) ~ Age + log(Caps + 1) + log(Goals +
##      1) + Country + Age:log(Caps + 1) + log(Caps + 1):log(Goals +
##      1), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9254 -0.4383 -0.0481  0.4063  3.1857
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   15.767753    0.533857  29.536 < 2e-16 ***
## Age           -0.077626    0.020652  -3.759 0.000188 ***
## log(Caps + 1)  1.664231    0.171153   9.724 < 2e-16 ***
## log(Goals + 1) -0.327680    0.153799  -2.131 0.033539 *
## CountryAustria  0.891222    0.200238   4.451 1.02e-05 ***
## CountryBelgium  1.718910    0.203554   8.444 2.39e-16 ***
## CountryCroatia  1.153922    0.198954   5.800 1.08e-08 ***
## CountryCzech Republic 0.892824    0.205312   4.349 1.61e-05 ***
```

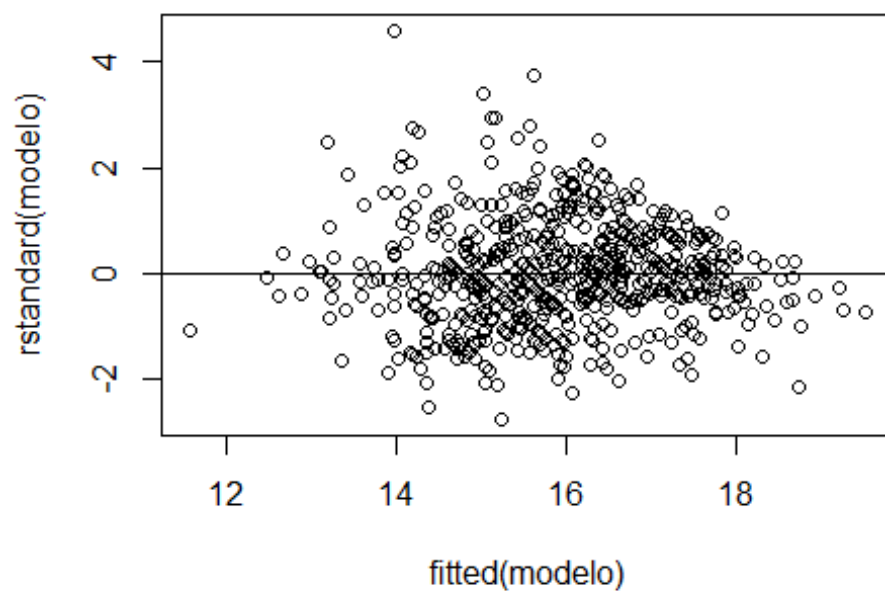
```
## CountryDenmark      1.363213    0.198181    6.879 1.55e-11 ***
## CountryEngland      2.762747    0.198158   13.942 < 2e-16 ***
## CountryFrance       2.493443    0.200731   12.422 < 2e-16 ***
## CountryGeorgia     -0.546689    0.197347   -2.770 0.005779 **
## CountryGermany      2.434042    0.198869   12.239 < 2e-16 ***
## CountryHungary     -0.050026    0.197423   -0.253 0.800051
## CountryItaly        2.224108    0.198452   11.207 < 2e-16 ***
## CountryNetherlands  2.090934    0.198502   10.534 < 2e-16 ***
## CountryPoland       0.816189    0.196094    4.162 3.62e-05 ***
## CountryPortugal     2.270430    0.198561   11.434 < 2e-16 ***
## CountryRomania     -0.143771    0.199621   -0.720 0.471677
## CountryScotland     0.758499    0.198286    3.825 0.000145 ***
## CountrySerbia       1.012122    0.198052    5.110 4.35e-07 ***
## CountrySlovakia     0.220617    0.197802    1.115 0.265160
## CountrySlovenia    -0.125522    0.197524   -0.635 0.525364
## CountrySpain        2.483333    0.198353   12.520 < 2e-16 ***
## CountrySwitzerland  1.068312    0.198953    5.370 1.14e-07 ***
## CountryTurkiye      1.086307    0.197927    5.488 6.04e-08 ***
## CountryUkraine      0.984059    0.198606    4.955 9.48e-07 ***
## Age:log(Caps + 1)   -0.048073    0.006386   -7.528 1.94e-13 ***
## log(Caps + 1):log(Goals + 1) 0.125689    0.037870    3.319 0.000959 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7113 on 588 degrees of freedom
## Multiple R-squared:  0.7678, Adjusted R-squared:  0.7567
## F-statistic: 69.43 on 28 and 588 DF, p-value: < 2.2e-16
```

```
qqPlot(modelo)
```

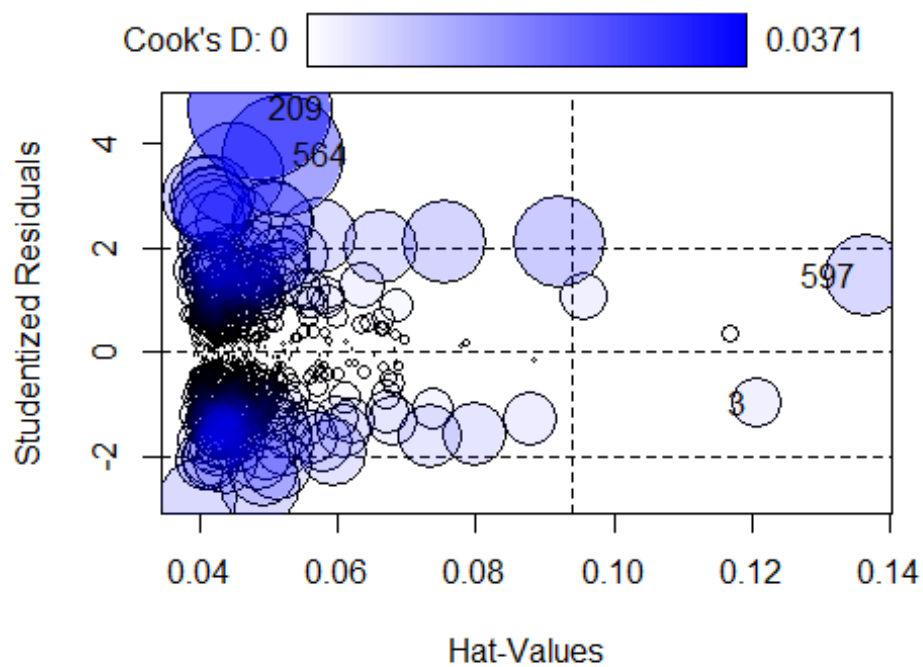


```
## 209 564
## 209 561

plot(fitted(modelo), rstandard(modelo))
abline(0, 0)
```



```
# Análise de influência  
influencePlot(modelo)
```




```
##      StudRes      Hat      CookD
## 3    -0.9720641 0.12064154 0.004470576
## 209  4.6725088 0.04856091 0.037109721
## 564  3.7749160 0.05186911 0.026289320
## 597  1.4708138 0.13652143 0.011770862
```

Diagnóstico de colinearidade

`vif(modelo)`

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

```
##              GVIF Df GVIF^(1/(2*Df))
## Age              8.838155 1      2.972904
## log(Caps + 1)    47.091724 1      6.862341
## log(Goals + 1)   30.421521 1      5.515571
## Country          1.212691 23      1.004201
## Age:log(Caps + 1) 79.373457 1      8.909178
## log(Caps + 1):log(Goals + 1) 33.992809 1      5.830335
```