

Análise descritiva e estatística da base dos jogadores ativos da NBA

Marcos Perazo Viana

13 abril, 2025

Introdução

Este projeto utiliza como base de dados informações sobre jogadores ativos da NBA, obtidas no site oficial da NBA, na aba *Players*, no dia 22 de março de 2025. Os dados coletados incluem: nomes dos jogadores, equipes em que atuam, posições, altura, peso, data de nascimento, idade, média de pontos, rebotes e assistências por jogo da atual temporada, além de informações sobre o *Draft* e experiência profissional. Adicionalmente, foi criada uma variável indicando o mês de nascimento dos jogadores. Com esses dados, é possível investigar diversas correlações estatísticas relevantes que ajudam a compreender o desempenho e as características dos atletas. Por exemplo, pode-se analisar a relação entre altura dos jogadores e a média de rebotes por jogo. Partindo da hipótese de que jogadores mais altos possuem vantagem nesse quesito, essa análise busca confirmar ou refutar tal tendência. Outras correlações potencialmente interessantes incluem:

- Peso e média de pontos por jogo: Investiga se jogadores com maior massa corporal têm vantagens em marcar pontos, especialmente em posições como pivô.
- Assistências por jogo e pontos por jogo: Examina a relação entre jogadores que criam oportunidades para seus colegas de equipe e aqueles que executam as finalizações.
- Altura e assistências por jogo: Avalia se jogadores mais baixos, como armadores, têm maior tendência a contribuir com assistências.
- Média de pontos por jogo e rebotes por jogo: Analisa a relação entre a capacidade de pontuar e pegar rebotes, considerando as diferentes posições em quadra.

Essas análises têm como objetivo explorar padrões e características que podem enriquecer a compreensão sobre o desempenho dos jogadores da NBA.

Carregamento da base.

```
dados_tratados <- read_csv("Dados_auxiliares/dados_apos_coluna_mes.csv", quote = "\"",  
                           locale = locale(encoding = "UTF-8"), show_col_types = FALSE)
```

Observações iniciais:

1 - Fazendo uma análise preliminar, verificamos que a linha 507 não possui dados em nenhuma das colunas, e foi determinada a eliminação dessa linha.

```
dados <- dados %>% slice(-507)
```

2 - Após verificar as estatísticas individuais dos atletas que receberam NA nas colunas PPG, APG e RPG, verificamos que esses campos deveriam receber valor zero. O que foi realizado.

<div> <div> </div> <div> Games Schedule Watch News Stats Standings Playoffs Teams Players NBA Play Fantasy NBA Bet </div> </div> <div> League Pass Store Tickets Sign In </div>									
<div> <div>PLAYERS</div> <div> Players Home 2024-25 Audio Pronunciation Guide Player Stats Starting Lineups Free Agent Tracker Transactions </div> </div>									
<div> <div>LEAGUE ROSTER</div> <div> <div>Search Players</div> </div> </div>									
<div> <div> <div>All Players</div> <div>All Teams</div> <div>All Positions</div> <div>All Colleges</div> <div>All Countries</div> <div>Show Historic</div> </div> <div>573 Rows • Page <div>All</div> of 12</div> </div>									
PLAYER	TEAM	NUMBER	POSITION	HEIGHT	WEIGHT	LAST ATTENDED	COUNTRY		
Precious Achiuwa	NYK	5	F	6-8	243 lbs	Memphis	Nigeria		
Steven Adams	HOU	12	C	6-11	265 lbs	Pittsburgh	New Zealand		
Bam Adebayo	MIA	13	C-F	6-9	255 lbs	Kentucky	USA		
Ochai Agbaji	TOR	30	G	6-5	215 lbs	Kansas	USA		
Santi Aldama	MEM	7	F-C	7-0	215 lbs	Loyola-Maryland	Spain		
Trey Alexander	DEN	23	G	6-4	185 lbs	Creighton	USA		
Nickel Alexander-Walker	MIN	9	G	6-5	205 lbs	Virginia Tech	Canada		
Grayson Allen	PHX	8	G	6-4	198 lbs	Duke	USA		
Jarrett Allen	CLE	31	C	6-9	243 lbs	Texas	USA		
Jose Alvarado	NOP	15	G	6-0	179 lbs	Georgia Tech	USA		
Kyle Anderson	MIA	20	F-G	6-9	230 lbs	UCLA	USA		
Giannis Antetokounmpo	MIL	34	F	6-11	243 lbs	Filathlitikos	Greece		
Cole Anthony	ORL	50	G	6-2	185 lbs	North Carolina	USA		

Figure 1: Site da NBA

```
dados <- dados %>%
  mutate(
    PPG = replace_na(PPG, 0),
    RPG = replace_na(RPG, 0),
    APG = replace_na(APG, 0)
  )
```

Analisar a centralidade dos dados, dispersão, assimetria, bem como suas estatísticas de ordem, a fim de checar se há presença de outliers.

```
dados %>% dplyr::select(Altura, Peso, PPG, RPG, APG) %>%
  summarytools::descr() %>% kable()
```

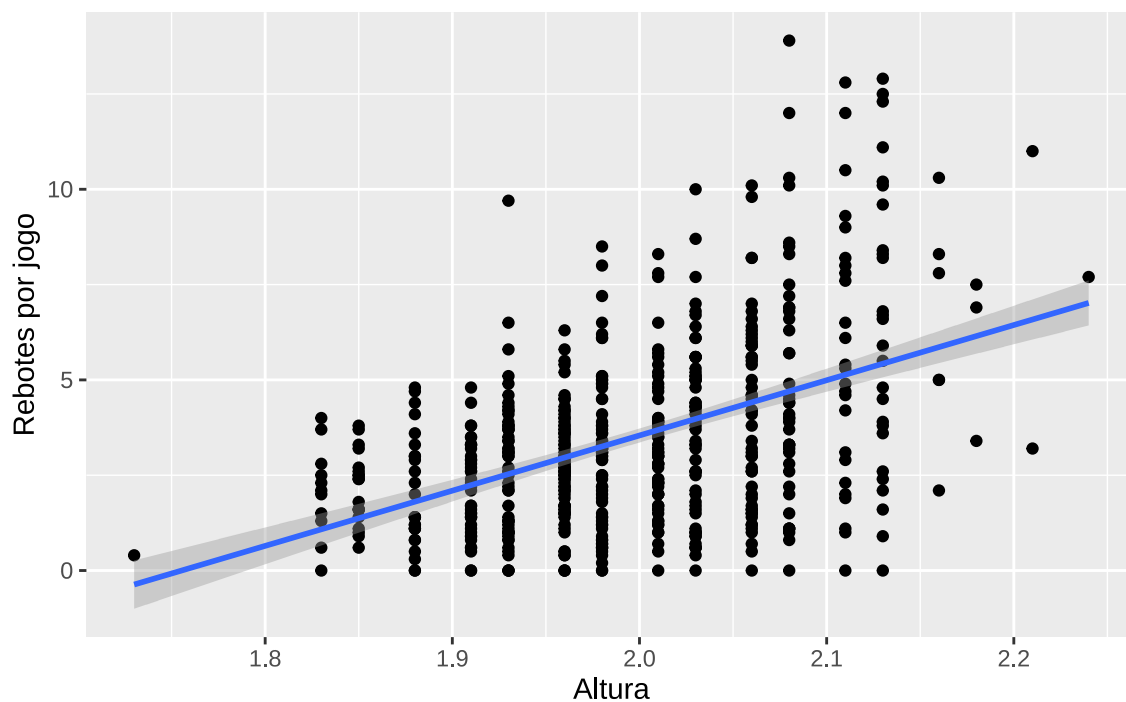
```
## Error in table(names(candidates))["tested"]: índice fora dos limites
```

	Altura	APG	Peso	PPG	RPG
Mean	1.9949296	2.0112676	97.5265018	8.5901408	3.4693662
Std.Dev	0.0794136	1.8405907	10.5926258	6.7465799	2.4857786
Min	1.7300000	0.0000000	72.0000000	0.0000000	0.0000000
Q1	1.9300000	0.7500000	90.0000000	3.5000000	1.6000000
Median	1.9800000	1.4000000	97.0000000	7.0000000	3.1000000
Q3	2.0600000	2.6500000	104.0000000	11.9000000	4.6000000
Max	2.2400000	11.4000000	138.0000000	32.9000000	13.9000000
MAD	0.0741300	1.1860800	10.3782000	5.7821400	2.2239000

	Altura	APG	Peso	PPG	RPG
IQR	0.1300000	1.8500000	14.0000000	8.4000000	3.0000000
CV	0.0398077	0.9151396	0.1086128	0.7853864	0.7164936
Skewness	0.0603177	1.6163724	0.4680604	1.0141186	1.2213483
SE.Skewness	0.1025083	0.1025083	0.1026883	0.1025083	0.1025083
Kurtosis	-0.2866479	2.9294536	0.1500945	0.4402431	1.7838539
N.Valid	568.0000000	568.0000000	566.0000000	568.0000000	568.0000000
N	568.0000000	568.0000000	568.0000000	568.0000000	568.0000000
Pct.Valid	100.0000000	100.0000000	99.6478873	100.0000000	100.0000000

Dispersão Altura x Rebotes

Gráfico com a dispersão entre Altura e Rebotes por jogo



Correlação entre as variáveis Altura e Rebotes por jogo

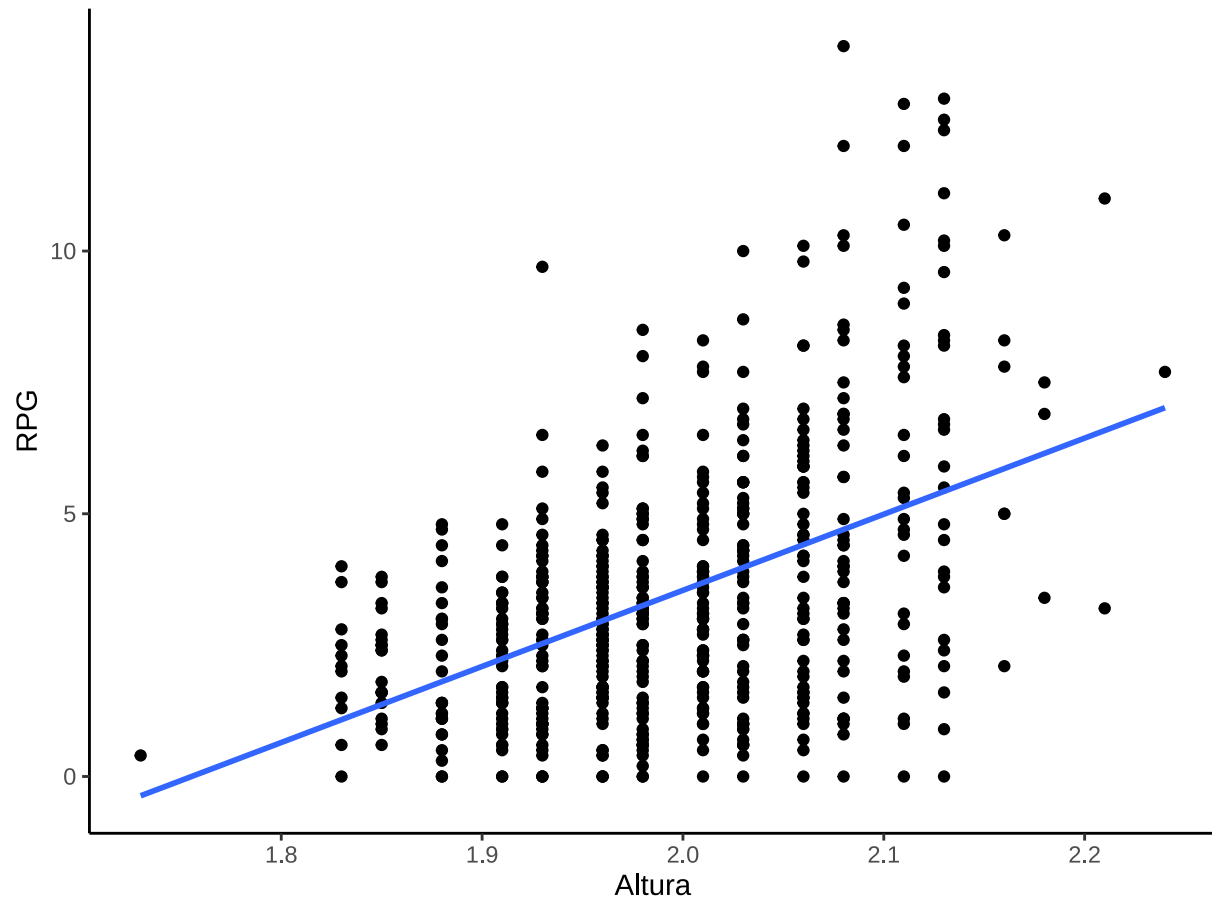
```
kable(cor(dados %>% dplyr::select(Altura, RPG)))
```

	Altura	RPG
Altura	1.0000000	0.4626776
RPG	0.4626776	1.0000000

Analisando scatter entre as variáveis Altura e Rebotes por jogo

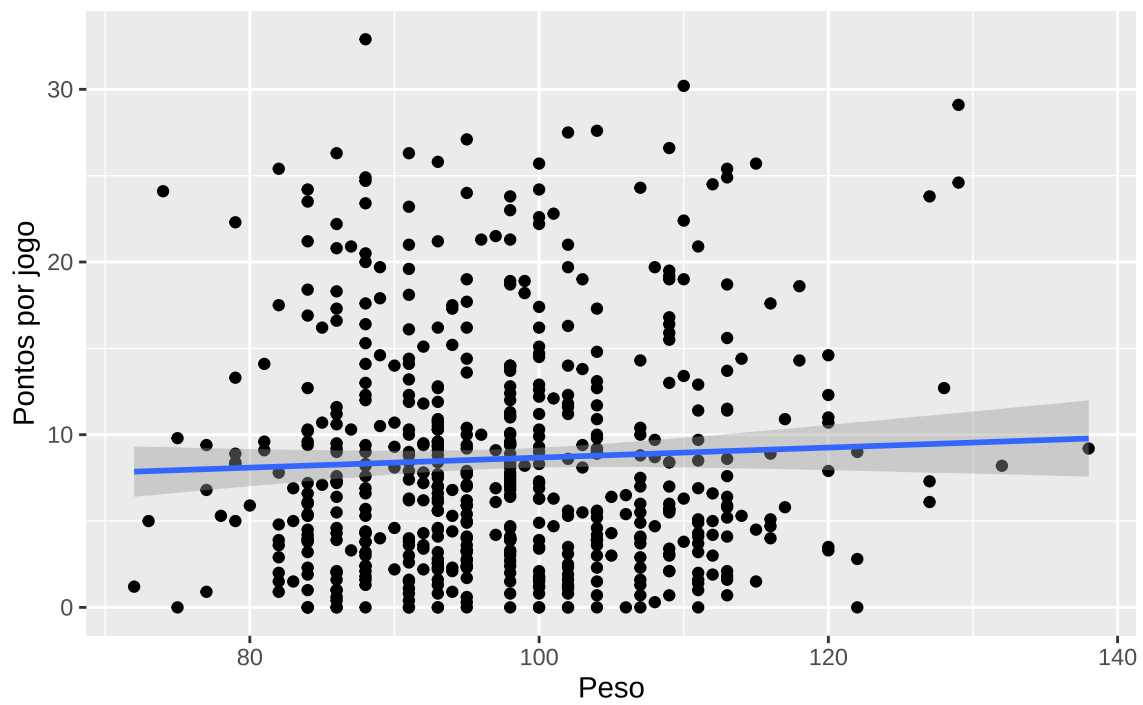
```
dados %>% dplyr::select(Altura, RPG) %>% ggplot(aes(x=Altura, y =RPG)) + geom_point() + geom_smooth(method="lm")
```

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



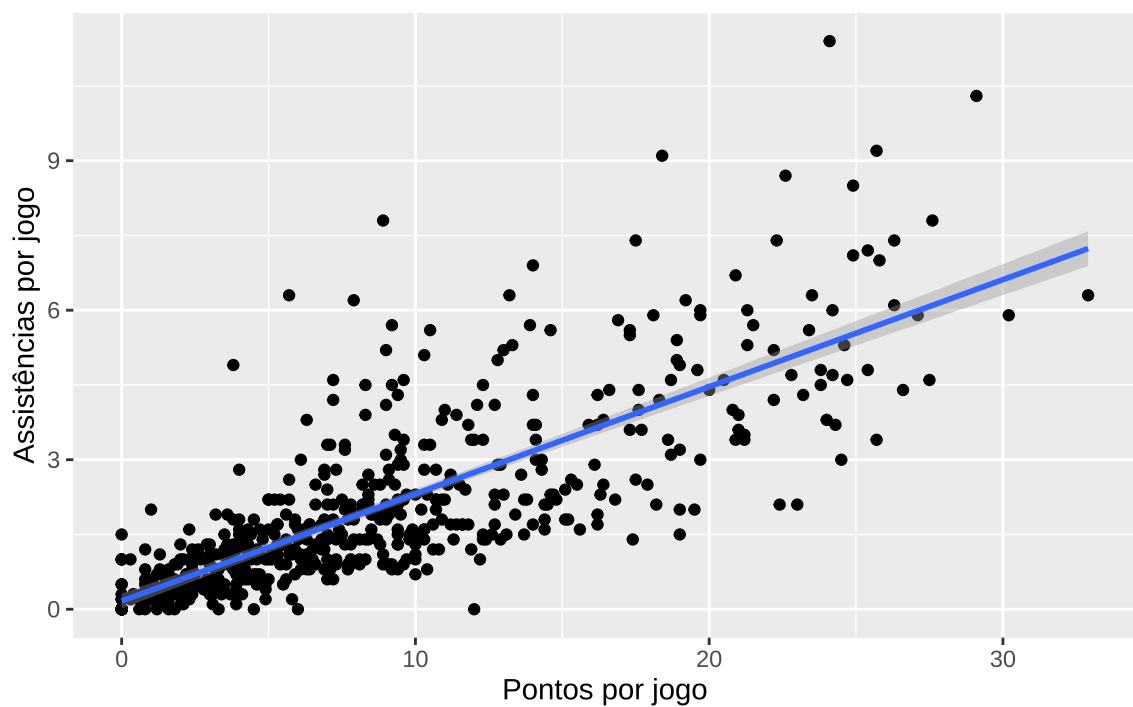
Dispersão Pontos x Peso

Gráfico com a dispersão entre Peso e Pontos por jogo



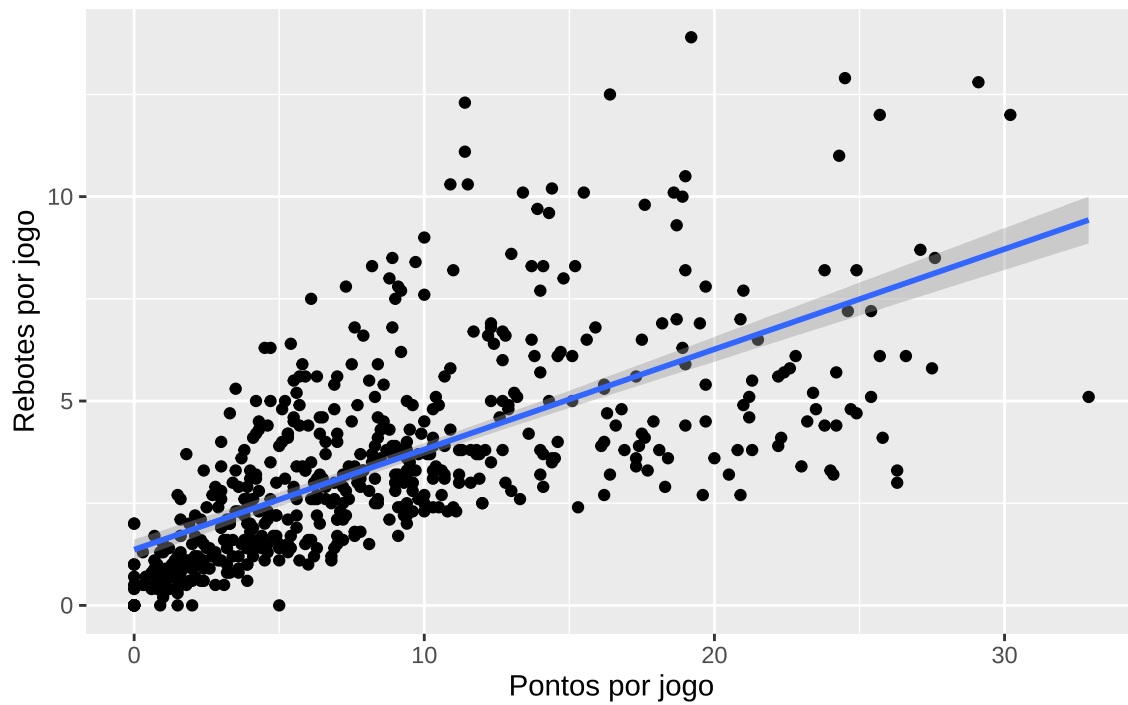
Dispersão Pontos x Assistências

Gráfico com a dispersão entre Pontos por jogo e Assistências por jog



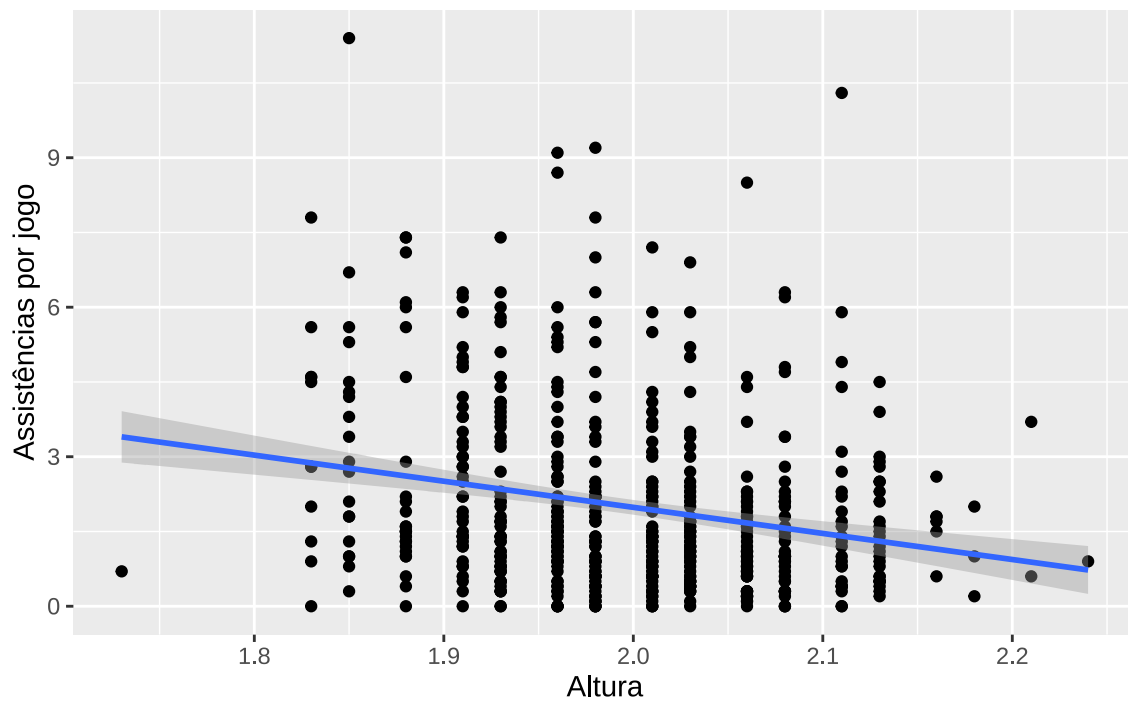
Dispersão Pontos x Rebotes

Gráfico com a dispersão entre Pontos por jogo e Rebotes por jogo



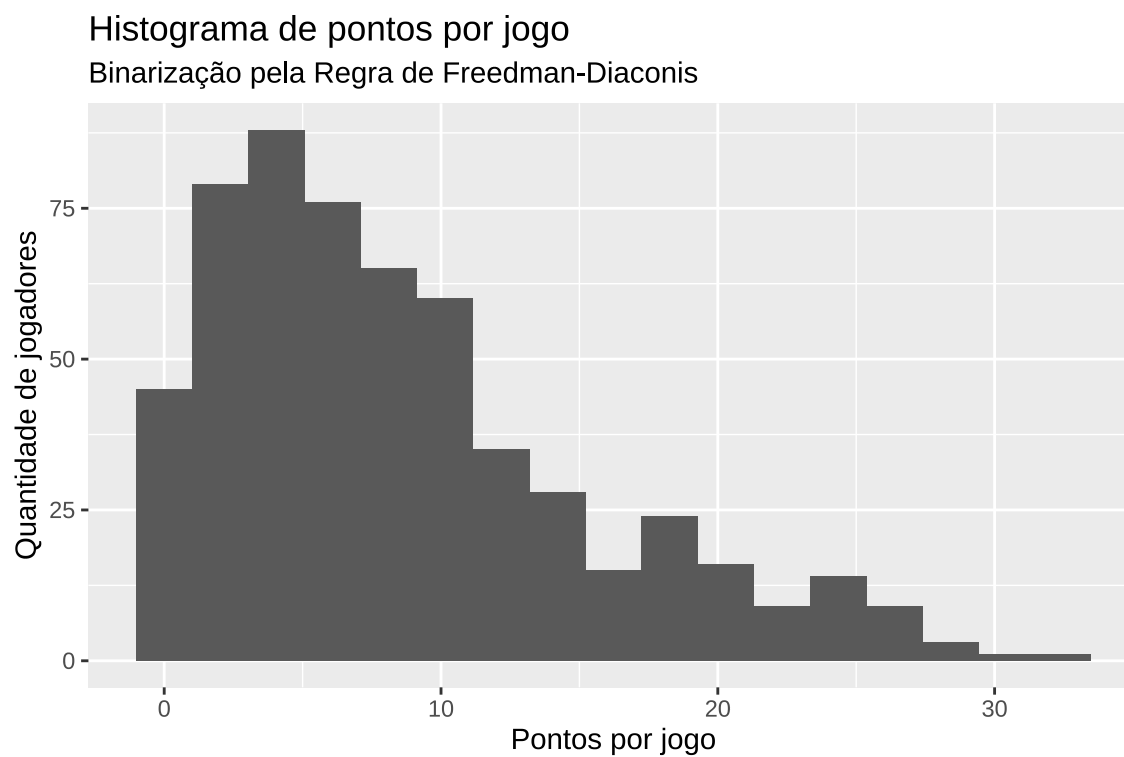
Dispersão Altura x Assistências

Gráfico com a dispersão entre Altura e Assistências por jogo



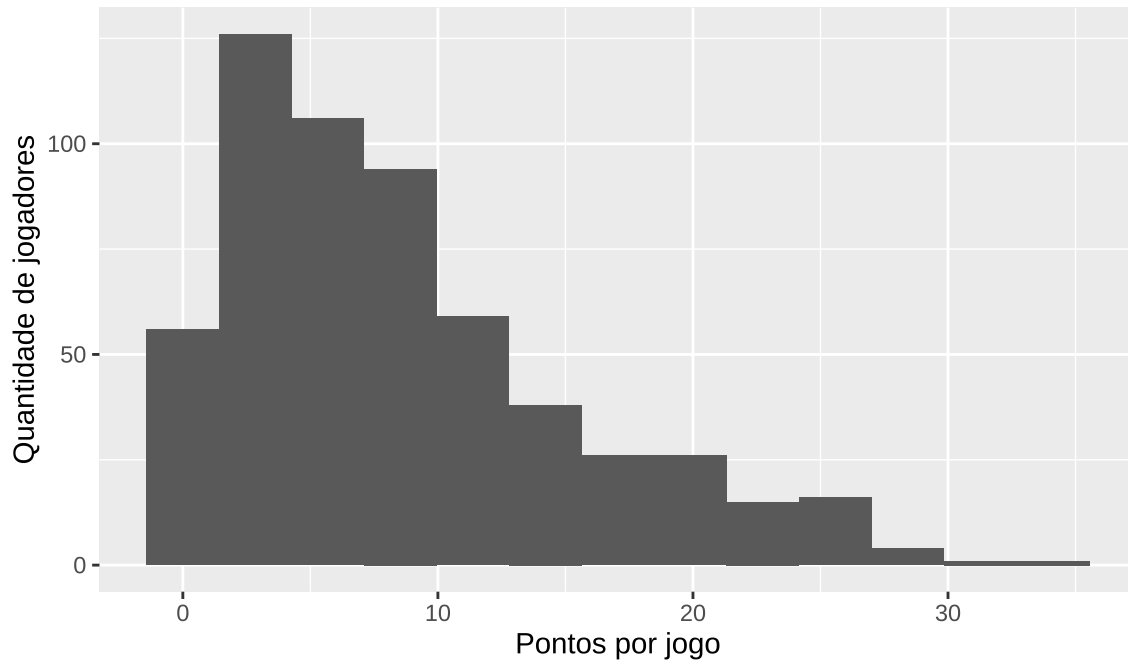
Definindo as funcoes geradoras de binwidths com as Regras de Freedman-Diaconis e Sturge

```
fd <- function(x) {  
  n <- length(x)  
  return((2*IQR(x))/n^(1/3))  
}  
  
sr <- function(x) {  
  n <- length(x)  
  return((3.49*sd(x))/n^(1/3))  
}
```



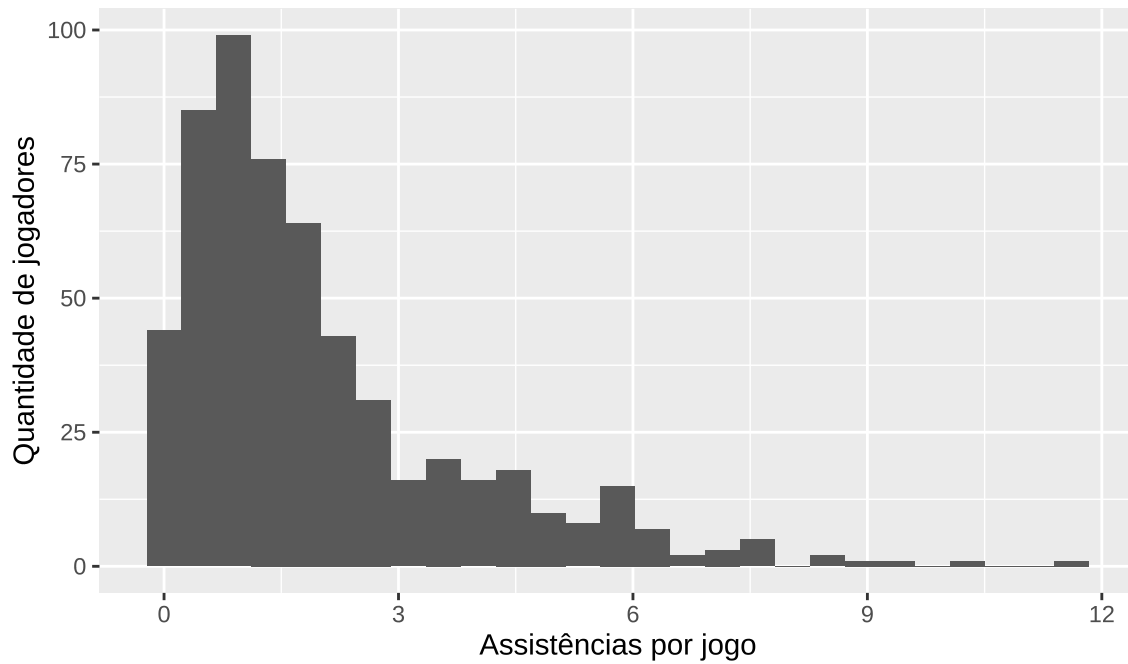
Histograma de pontos por jogo

Binarização pela Regra de Sturge



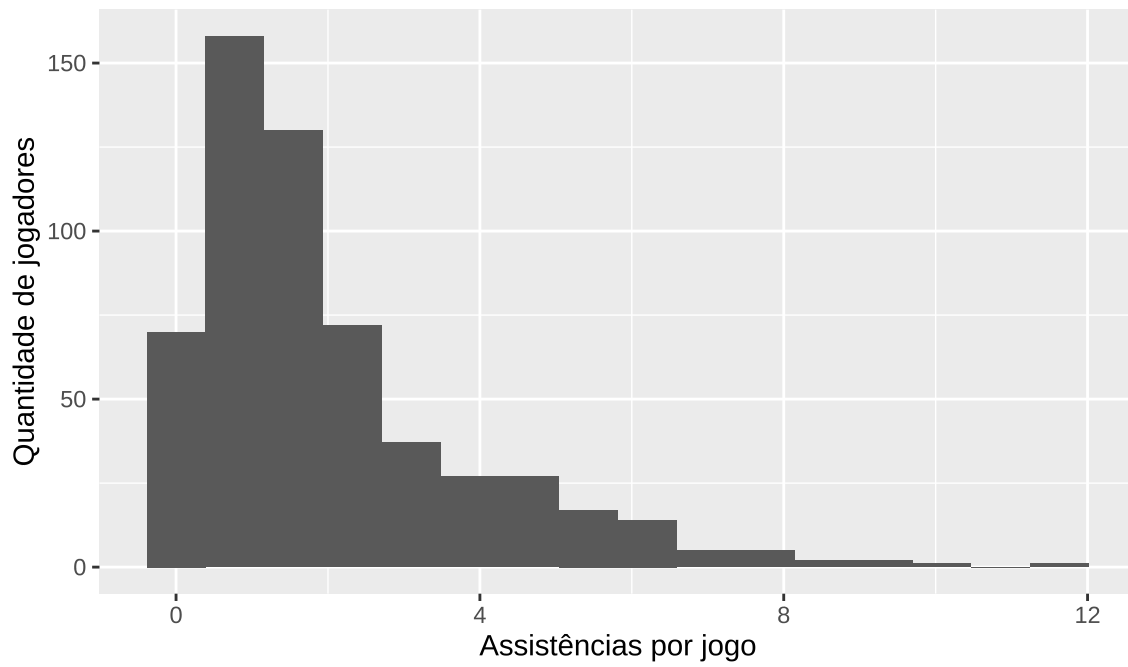
Histograma de assistências por jogo

Binarização pela Regra de Freedman-Diaconis



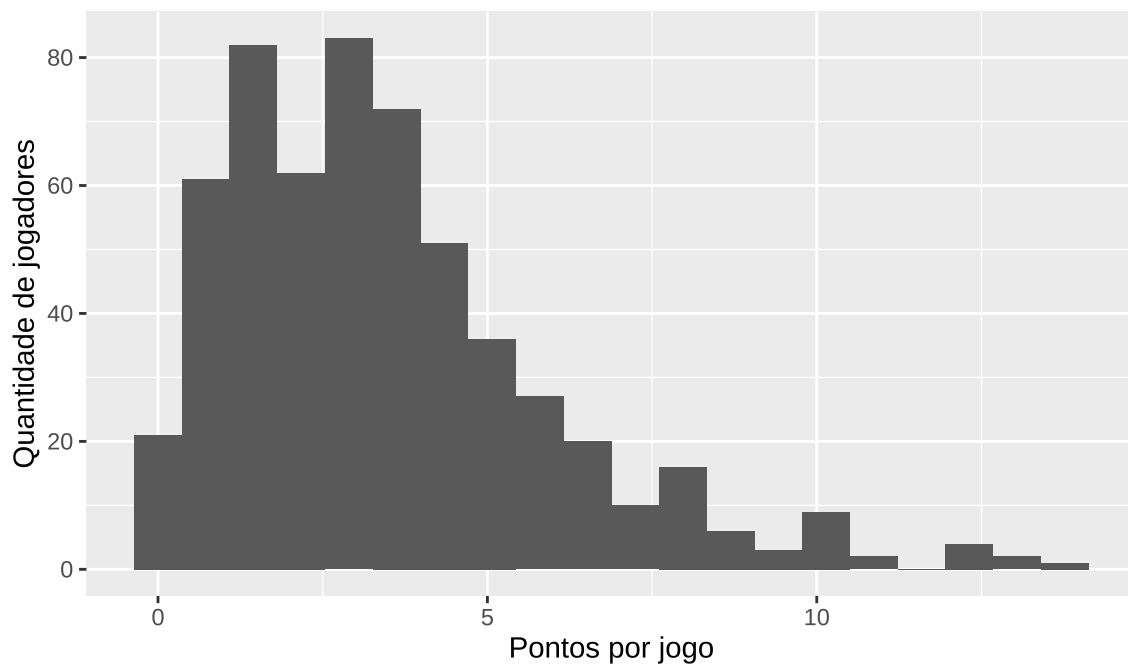
Histograma de assistências por jogo

Binarização pela Regra de Sturge



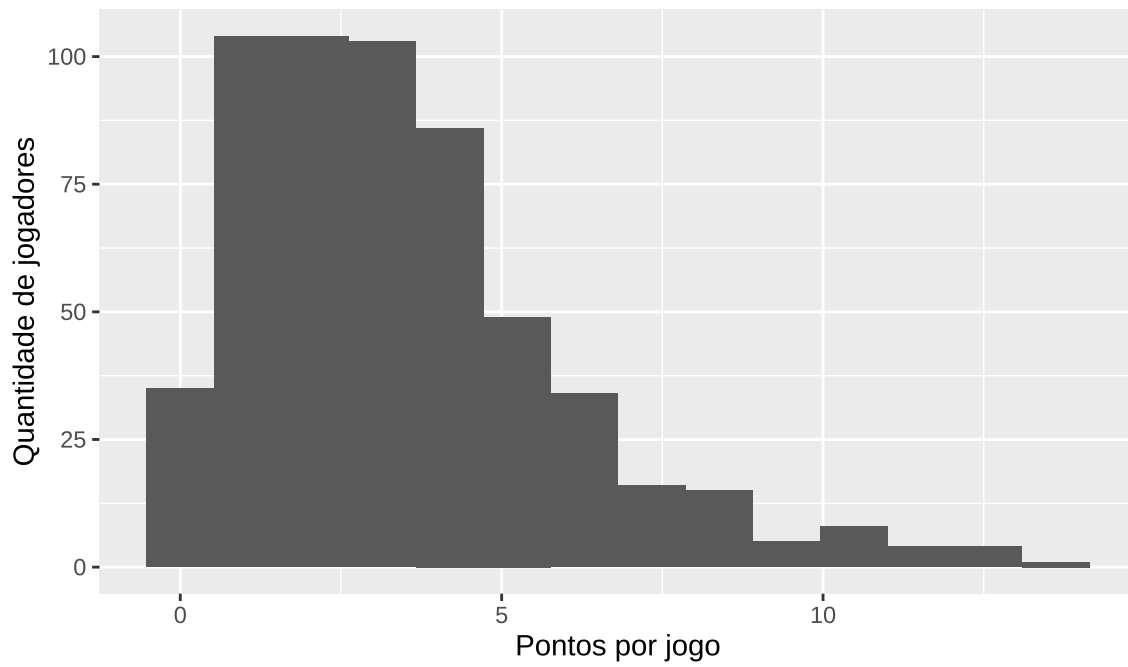
Histograma de rebotes por jogo

Binarização pela Regra de Freedman-Diaconis



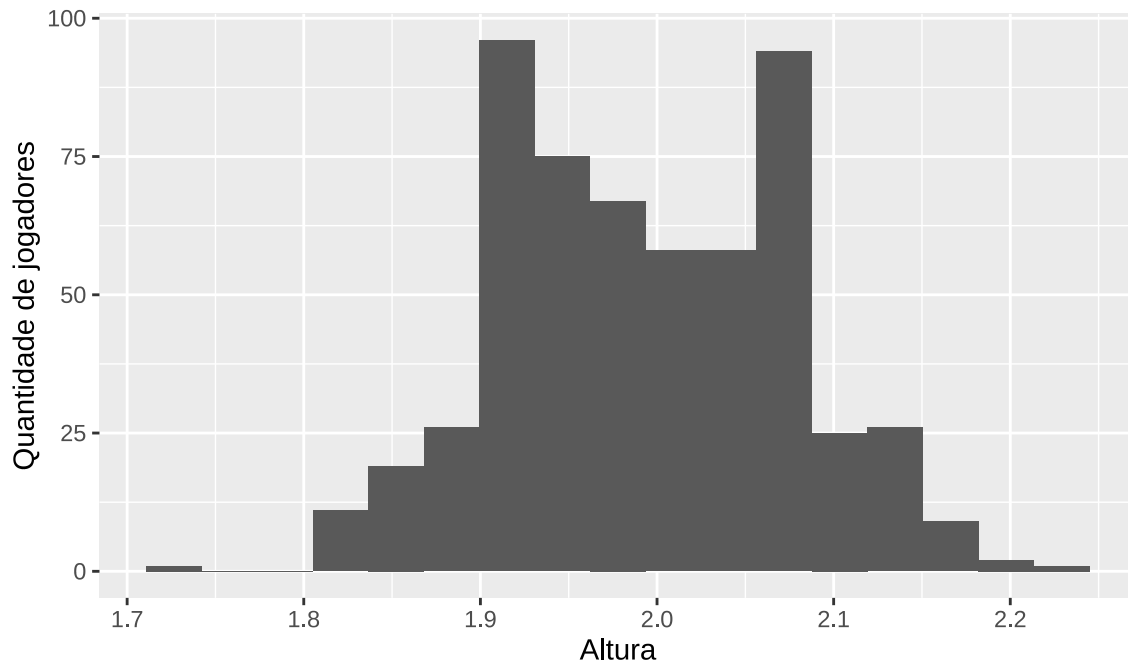
Histograma de rebotes por jogo

Binarização pela com Regra de Sturge



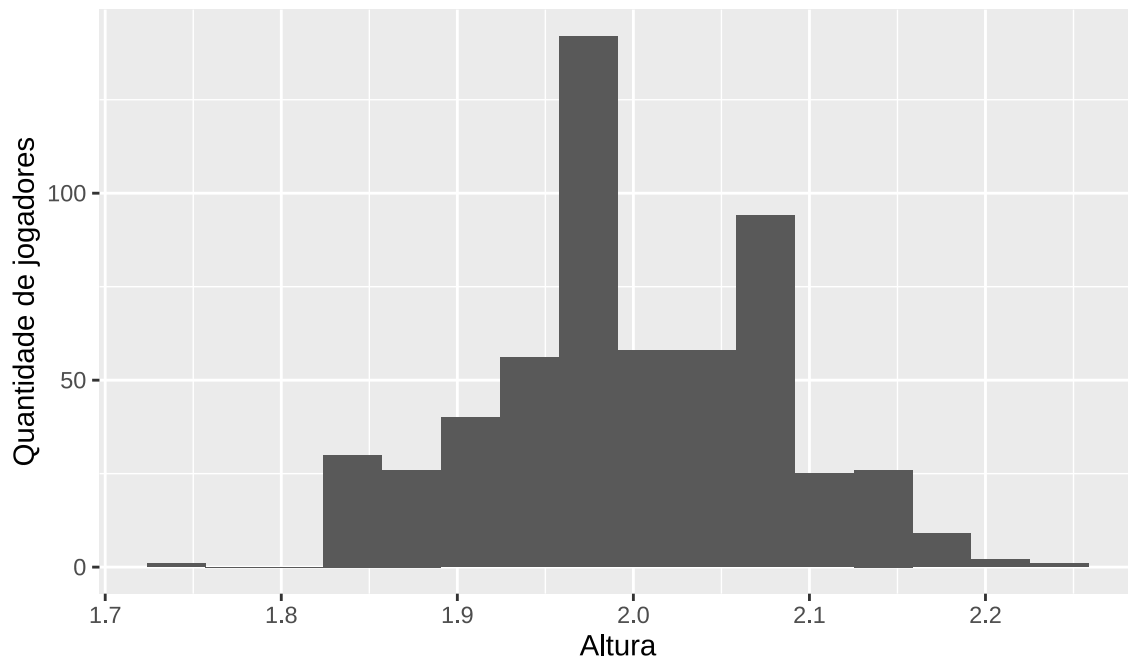
Histograma de Altura

Binarização pela com Regra de Freedman-Diaconis



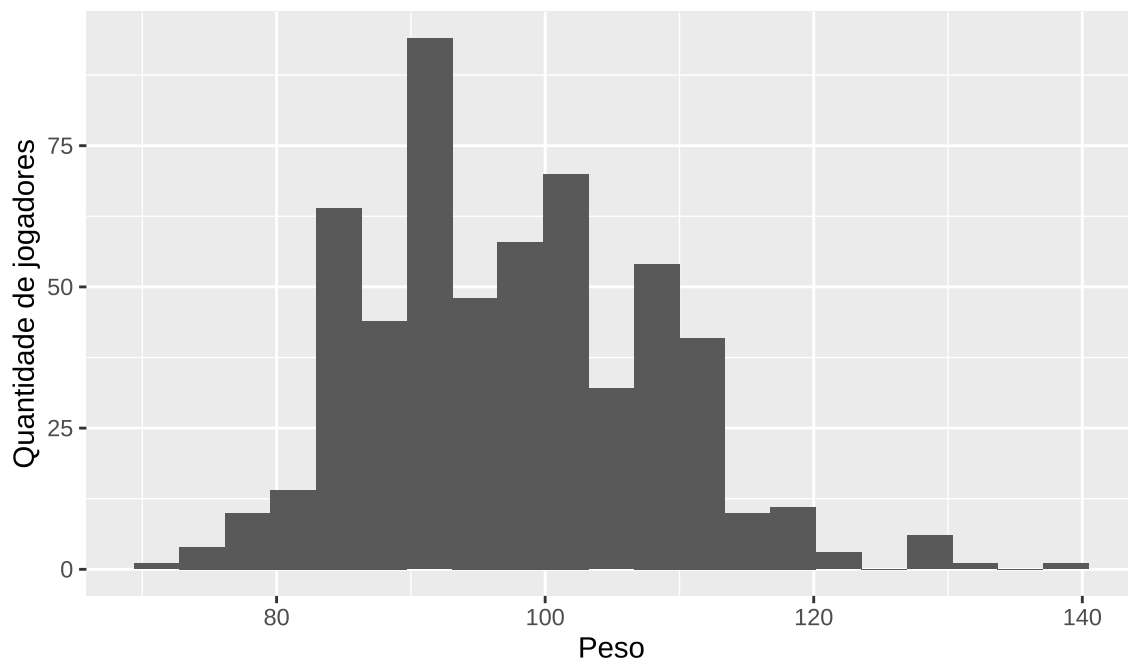
Histograma de Altura

Binarização pela com Regra de Sturge



Histograma de Peso

Binarização pela com Regra de Freedman-Diaconis



Histograma de Peso

Binarização pela com Regra de Sturge

