

Wnioskowanie Statystyczne

Maksym Dunajewski, Idzi Łopatniuk

2025-01-24

#Załadowanie potrzebnych bibliotek

```
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(grid)
```

```
library(ggplot2)
```

Zbiór danych

Zbiór danych `players_stats.csv` zawiera statystyki indywidualne zawodników NBA z sezonu 2014/2015. Dane obejmują szeroki zakres metryk, takich jak liczba zdobytych punktów, zbiórek, asyst, skuteczność rzutów oraz inne szczegóły związane z grą zawodników.

Zmienne

Zmienna	Opis
Name	Imię i nazwisko gracza
X3PM	Średnia liczba celnych rzutów za trzy punkty na mecz
FTP	Skuteczność rzutów osobistych
OREB	Średnia liczba zbiórek ofensywnych na mecz
BLK	Średnia liczba bloków ofensywnych na mecz
Height	Wzrost zawodnika (cm)
Pos	Pozycja zawodnika (PG, SG, SF, PF, C)

Hipoteza

Czy wyżsi zawodnicy mają większą liczbę zbiórek ofensywnych (OREB) oraz bloków (BLK), ale jednocześnie osiągają gorsze wyniki w skuteczności, zdobywając mniej punktów za trzy (X3PM) i wykazując niższą skuteczność rzutów osobistych (FTP) w porównaniu do niższych zawodników?

Wczytanie i obróbka danych

```
player_stats <- read.csv(file = "players_stats.csv")
colnames(player_stats)[colnames(player_stats) == "FT."] <- "FTP"
```

Stworzenie nowej ramki danych z interesującymi nas zmiennymi

```
new_stats <- data.frame(
  Name = player_stats$Name,
  X3PM = player_stats$X3PM / player_stats$Games.Played,
  FTP = player_stats$FTP/100,
  OREB = player_stats$OREB / player_stats$Games.Played,
  BLK = player_stats$BLK / player_stats$Games.Played,
  Height = player_stats$Height,
  Pos = player_stats$Pos
)
```

#usunięcie wierszy zawierających wartości NA

```
player_stats_clean <- na.omit(new_stats)
```

#Podzielenie zawodników na dwie grupy wzrostowe

Obliczenie mediany wzrostu

```
median_height <- median(player_stats_clean$Height, na.rm = TRUE)
```

Tworzenie grup wzrostu

```
player_stats_clean$Height_Group <- ifelse(
  player_stats_clean$Height < median_height, "Niski", "Wysoki"
)
```

Obliczanie średnich wartości dla każdej grupy wzrostu

```
player_stats_clean %>%
  group_by(Height_Group) %>%
  summarise(
    mean_oreb = mean(OREB),
    mean_blk = mean(BLK),
    mean_3pm = mean(X3PM),
    mean ftp = mean(FTP)
  )
```

```
## # A tibble: 2 x 5
##   Height_Group mean_oreb mean_blk mean_3pm mean ftp
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 Niski          0.422     0.169     0.915     0.742
## 2 Wysoki         1.23      0.553     0.498     0.682
```

Wstępne wykresy dla naszych statystyk

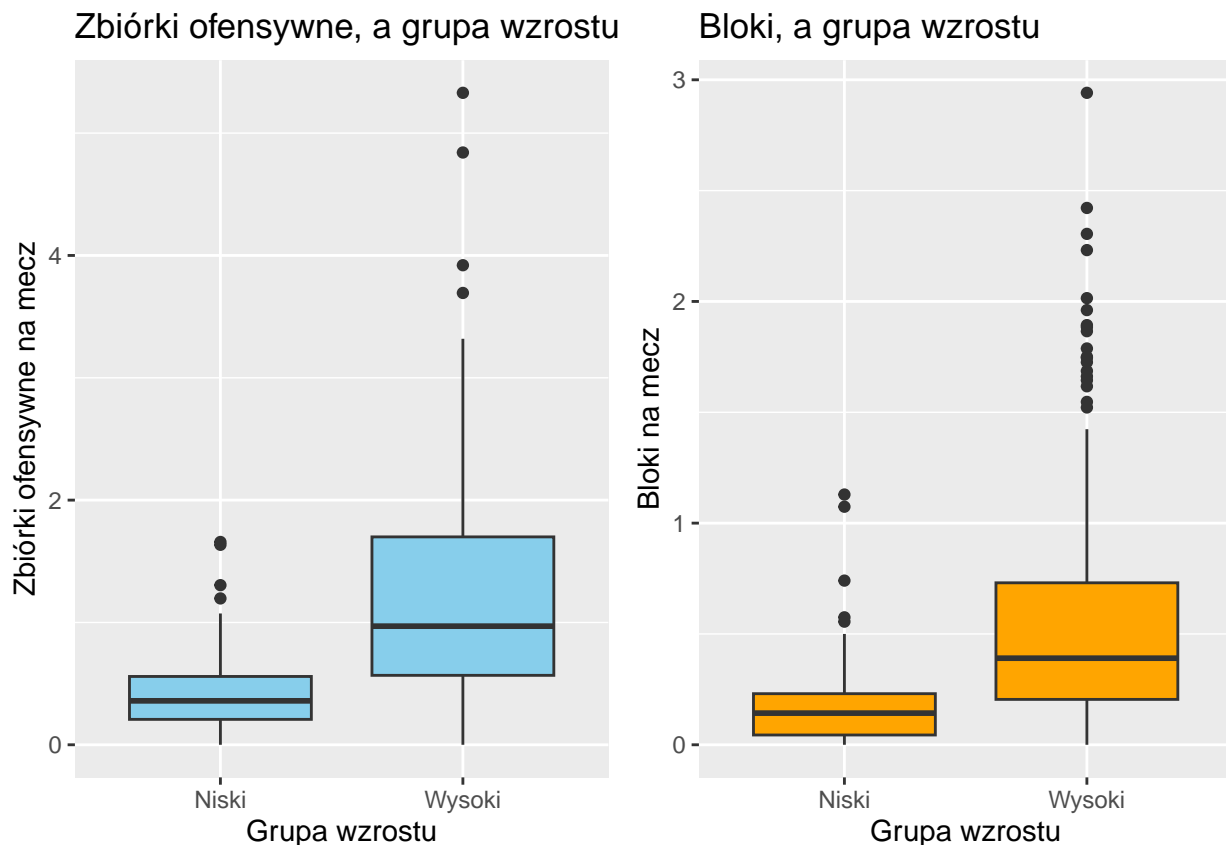
Boxploty

```
# Boxplot dla zbiorów ofensywnych
plot_oreb <- ggplot(player_stats_clean, aes(x = Height_Group, y = OREB)) +
  geom_boxplot(fill = "skyblue") +
  labs(title = "Zbiórki ofensywne, a grupa wzrostu", x = "Grupa wzrostu", y = "Zbiórki ofensywne na mecz")

# Boxplot dla bloków
plot_blk <- ggplot(player_stats_clean, aes(x = Height_Group, y = BLK)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Blokady, a grupa wzrostu", x = "Grupa wzrostu", y = "Blokady na mecz")

# Wyświetlenie dwóch wykresów obok siebie
grid.newpage()
pushViewport(viewport(layout = grid.layout(1, 2)))

print(plot_oreb, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(plot_blk, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
```

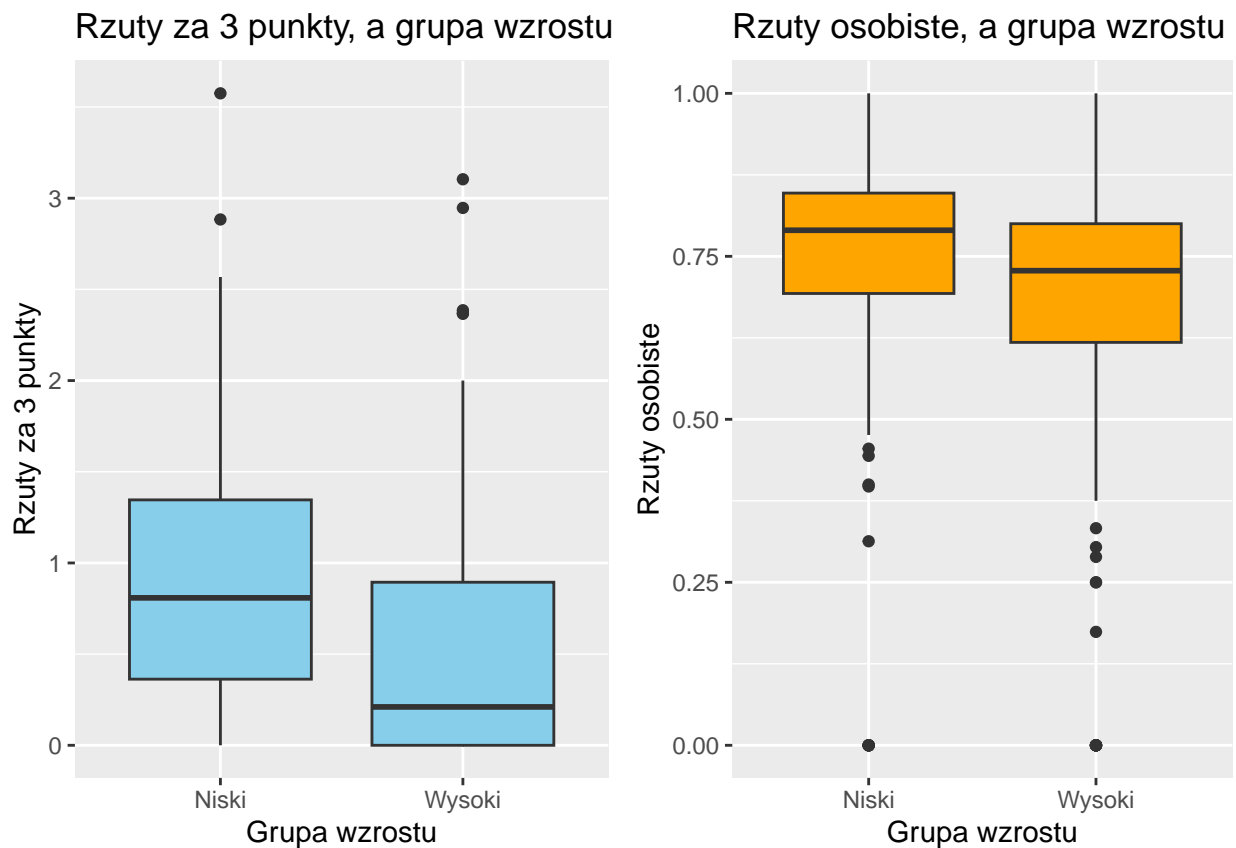


```
# Boxplot dla rzutów za 3 punkty
plot_oreb <- ggplot(player_stats_clean, aes(x = Height_Group, y = X3PM)) +
  geom_boxplot(fill = "skyblue") +
  labs(title = "Rzuty za 3 punkty, a grupa wzrostu", x = "Grupa wzrostu", y = "Rzuty za 3 punkty")
```

```
# Boxplot dla rzutów osobistych
plot_blk <- ggplot(player_stats_clean, aes(x = Height_Group, y = FTP)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Rzuty osobiste, a grupa wzrostu", x = "Grupa wzrostu", y = "Rzuty osobiste")

# Wyświetlenie dwóch wykresów obok siebie
grid.newpage()
pushViewport(viewport(layout = grid.layout(1, 2)))

print(plot_oreb, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(plot_blk, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
```



Przeprowadzenie testów ANOVA, aby sprawdzić czy różnice między grupami wzrostu są statystycznie istotne.

```
# ANOVA dla zbiorów ofensywnych
anova_oreb <- aov(OREB ~ Height_Group, data = player_stats_clean)
summary(anova_oreb)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Height_Group  1  66.49   66.49   127.4 <2e-16 ***
## Residuals    420  219.20    0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ANOVA dla bloków
anova_blk <- aov(BLK ~ Height_Group, data = player_stats_clean)
summary(anova_blk)

##               Df Sum Sq Mean Sq F value Pr(>F)
## Height_Group   1  15.03  15.034    90.12 <2e-16 ***
## Residuals    420   70.06   0.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ANOVA dla celnych rzutów za trzy
anova_3pm <- aov(X3PM ~ Height_Group, data = player_stats_clean)
summary(anova_3pm)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Height_Group   1   17.8  17.797    41.75 2.88e-10 ***
## Residuals    420  179.0   0.426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ANOVA dla rzutów wolnych
anova_ftp <- aov(FTP ~ Height_Group, data = player_stats_clean)
summary(anova_ftp)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Height_Group   1   0.37  0.3705    10.38 0.00138 **
## Residuals    420  14.99   0.0357
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wyniki testów ANOVA

Wyniki testów ANOVA wykazały istotne statystycznie różnice między grupami zawodników (**Niski** i **Wysoki**) we wszystkich analizowanych zmiennych:

1. **Zbiórki ofensywne (OREB):** Wyżsi zawodnicy osiągają znacząco więcej zbiórek ofensywnych (średnia = 1.23) niż niżsi zawodnicy (średnia = 0.42, $p < 0.001$).
2. **Bloki (BLK):** Wyżsi zawodnicy wykonują więcej bloków (średnia = 0.55) w porównaniu do niższych zawodników (średnia = 0.17, $p < 0.001$).
3. **Punkty za trzy (X3PM):** Niżsi zawodnicy zdobywają więcej punktów za trzy (średnia = 0.92) niż wyżsi zawodnicy (średnia = 0.50, $p < 0.001$).
4. **Skuteczność rzutów osobistych (FTP):** Niżsi zawodnicy mają wyższą skuteczność rzutów osobistych (średnia = 0.74) w porównaniu do wyższych zawodników (średnia = 0.68, $p < 0.01$).

Zależności między wzrostem a statystykami

Aby dokładniej zbadać wpływ wzrostu na poszczególne statystyki, przeprowadzono dalszą analizę regresji, która pozwala lepiej zrozumieć te zależności:

- Zbiórki ofensywne (**OREB**): Wyższy wzrost jest istotnie związany z większą liczbą zbiórek ofensywnych. Każdy dodatkowy cal wzrostu zwiększa średnią liczbę zbiórek o 0.051 ($p < 0.001$).
- Bloki (**BLK**): Wyższy wzrost istotnie zwiększa liczbę bloków. Każdy dodatkowy cal wzrostu zwiększa liczbę bloków o 0.026 ($p < 0.001$).
- Punkty za trzy (**X3PM**): Wyższy wzrost istotnie zmniejsza liczbę punktów zdobywanych za trzy. Każdy dodatkowy cal wzrostu zmniejsza średnią liczbę punktów o 0.028 ($p < 0.001$).
- Skuteczność rzutów osobistych (**FTP**): Wyższy wzrost wiąże się z nieznacznie niższą skutecznością rzutów osobistych. Każdy dodatkowy cal wzrostu obniża skuteczność o 0.358% ($p < 0.01$).

Te wyniki potwierdzają, że wzrost zawodnika ma znaczący wpływ na jego statystyki, ale szczegółowa analiza zależności między wzrostem a statystykami dla różnych pozycji jest konieczna, aby uwzględnić rolę i zadania zawodników w drużynie.

```
# Regresja dla zbiórek ofensywnych
model_oreb <- lm(OREB ~ Height, data = player_stats_clean)
summary(model_oreb)

##
## Call:
## lm(formula = OREB ~ Height, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1710 -0.3999 -0.1045  0.3053  3.9228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -9.170542   0.764516  -11.99  <2e-16 ***
## Height       0.050973   0.003868   13.18  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6937 on 420 degrees of freedom
## Multiple R-squared:  0.2925, Adjusted R-squared:  0.2908
## F-statistic: 173.6 on 1 and 420 DF, p-value: < 2.2e-16

# Regresja dla bloków
model_blk <- lm(BLK ~ Height, data = player_stats_clean)
summary(model_blk)

##
## Call:
## lm(formula = BLK ~ Height, data = player_stats_clean)
##
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -0.72912 -0.24024 -0.06734  0.13066  2.34773
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.823151   0.425606  -11.33  <2e-16 ***
## Height      0.026422   0.002154   12.27  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3862 on 420 degrees of freedom
## Multiple R-squared:  0.2639, Adjusted R-squared:  0.2621
## F-statistic: 150.5 on 1 and 420 DF,  p-value: < 2.2e-16
```

```
# Regresja dla celnych rzutów za trzy
model_3pm <- lm(X3PM ~ Height, data = player_stats_clean)
summary(model_3pm)
```

```
##
## Call:
## lm(formula = X3PM ~ Height, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1683 -0.4560 -0.1988  0.3665  2.6204
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.297328   0.702622   8.963  < 2e-16 ***
## Height      -0.028495   0.003555  -8.015  1.1e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6376 on 420 degrees of freedom
## Multiple R-squared:  0.1327, Adjusted R-squared:  0.1306
## F-statistic: 64.24 on 1 and 420 DF,  p-value: 1.101e-14
```

```
# Regresja dla rzutów wolnych
model_ftpm <- lm(FTP ~ Height, data = player_stats_clean)
summary(model_ftpm)
```

```
##
## Call:
## lm(formula = FTP ~ Height, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76048 -0.05680  0.04748  0.11110  0.34702
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.414410   0.207926   6.802 3.55e-11 ***
## Height      -0.003583   0.001052  -3.406 0.000723 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1887 on 420 degrees of freedom
## Multiple R-squared:  0.02688,    Adjusted R-squared:  0.02456
## F-statistic: 11.6 on 1 and 420 DF,  p-value: 0.0007232
```

Wpływ wzrostu na parametry gry

1. **Zbiórki ofensywne (OREB):** Wyższy wzrost jest istotnie związany z większą liczbą zbiórek ofensywnych ($p < 0.001$). Każdy dodatkowy cal wzrostu zwiększa średnią liczbę zbiórek o 0.051. Model wyjaśnia 29.25% zmienności ($R^2 = 0.2925$).
2. **Bloki (BLK):** Wyższy wzrost istotnie zwiększa liczbę bloków ($p < 0.001$). Każdy dodatkowy cal wzrostu zwiększa liczbę bloków o 0.026. Model wyjaśnia 26.39% zmienności ($R^2 = 0.2639$).
3. **Punkty za trzy (X3PM):** Wyższy wzrost istotnie zmniejsza liczbę punktów zdobywanych za trzy ($p < 0.001$). Każdy dodatkowy cal zmniejsza średnią liczbę punktów o 0.028. Model wyjaśnia 13.27% zmienności ($R^2 = 0.1327$).
4. **Skuteczność rzutów osobistych (FTP):** Wyższy wzrost wiąże się z nieznacznie niższą skutecznością rzutów osobistych ($p < 0.01$). Każdy dodatkowy cal obniża skuteczność o 0.358%. Model wyjaśnia 2.69% zmienności ($R^2 = 0.02688$).

Wyniki naszej analizy wyraźnie wskazują, że wzrost zawodnika istotnie wpływa na kluczowe parametry gry, takie jak liczba zbiórek ofensywnych (OREB), liczba bloków (BLK), skuteczność rzutów osobistych (FTP) oraz liczba punktów zdobywanych rzutami za trzy punkty (3PM). Jednak, aby uzyskać pełny obraz, konieczne jest bliższe przyjrzenie się, jak te zależności rozkładają się w kontekście konkretnych pozycji zawodników.

Czyli wszystko jasne?

Zawodnicy w koszykówce są przydzielani do określonych pozycji, takich jak **PG (Point Guard)**, **SG (Shooting Guard)**, **SF (Small Forward)**, **PF (Power Forward)** i **C (Center)**. Każda z tych pozycji wiąże się z odmiennymi zadaniami i rolami w drużynie:

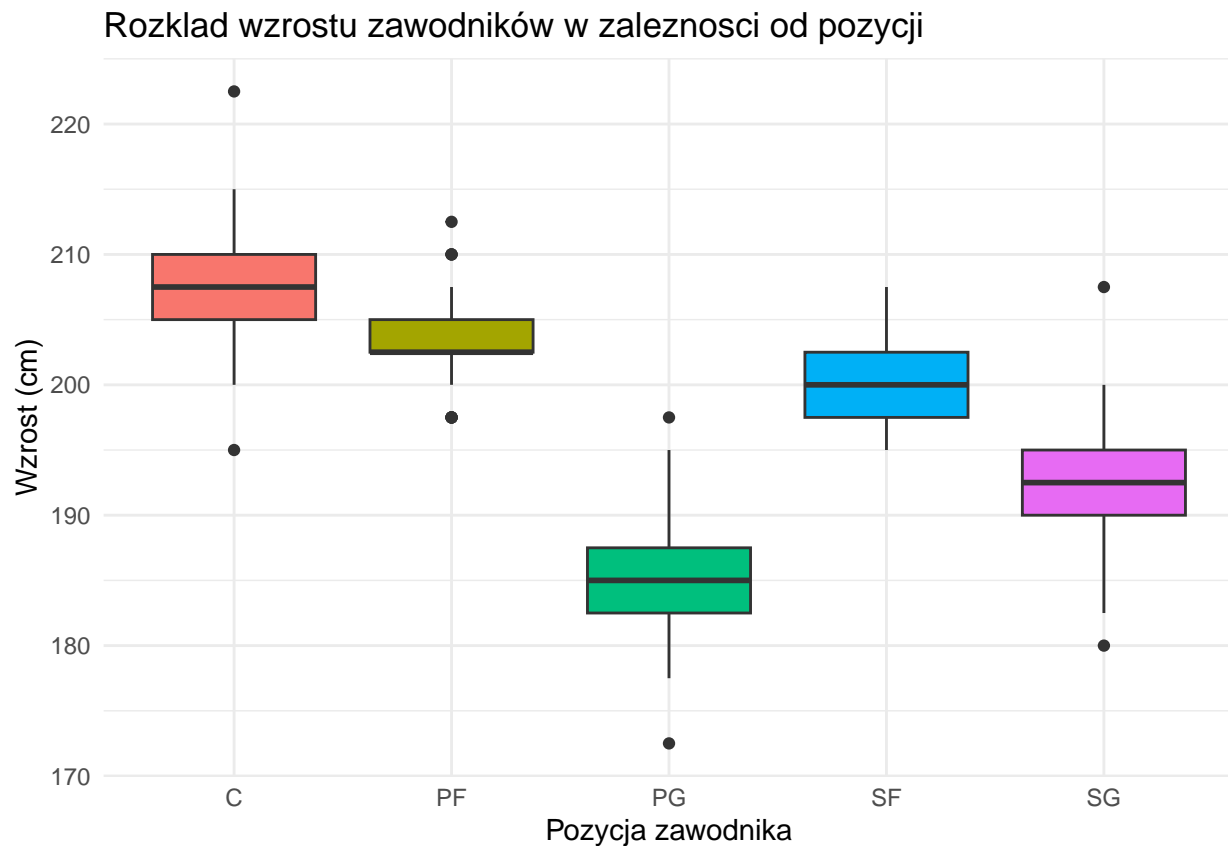
- **Rozgrywający (PG):** Skupiają się na prowadzeniu gry i tworzeniu okazji strzeleckich dla innych zawodników, co przekłada się na niższe wartości dla zbiórek i bloków, ale wyższą skuteczność rzutów za trzy punkty i osobistych.
- **Rzucający obrońcy (SG):** Koncentrują się na zdobywaniu punktów, często z dystansu, dlatego ich parametry skuteczności (FTP, 3PM) mogą być wyższe niż u zawodników grających pod koszem.
- **Niscy skrzydłowi (SF):** Pełnią wszechstronne role, co może powodować większe zróżnicowanie w ich statystykach, zarówno w zdobywaniu punktów, jak i w zbiórkach.
- **Silni skrzydłowi (PF) oraz centrzy (C):** Z racji swojej fizyczności i wzrostu, dominują w strefie podkoszowej, co skutkuje wyższymi średnimi w zbiórkach i blokach, ale niższą skutecznością rzutów osobistych i za trzy punkty.

Zadania przypisane do każdej pozycji w znacznym stopniu determinują priorytety zawodników i wpływają na ich statystyki. Dlatego, aby w pełni zrozumieć wpływ wzrostu na parametry gry, konieczne jest uwzględnienie pozycji zawodnika w dalszych testach. Analiza rozkładu tych zmiennych dla każdej pozycji pozwoli odpowiedzieć na pytanie, czy zaobserwowane różnice wynikają głównie z wzrostu, czy też z roli, jaką zawodnik pełni w drużynie.

```
# ANOVA dla wzrostu w zależności od pozycji
anova_height <- aov(Height ~ Pos, data = player_stats_clean)
summary(anova_height)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Pos           4  26322    6580   469.9 <2e-16 ***
## Residuals    417   5839      14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Boxplot wzrostu dla każdej pozycji
ggplot(player_stats_clean, aes(x = Pos, y = Height, fill = Pos)) +
  geom_boxplot() +
  labs(
    title = "Rozkład wzrostu zawodników w zależności od pozycji",
    x = "Pozycja zawodnika",
    y = "Wzrost (cm)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```



Wpływ wzrostu na rolę zawodnika w drużynie

Obserwując zarówno rozkład wzrostu zawodników w zależności od pozycji na boisku, jak i wyniki analizy ANOVA, można zauważyć wyraźny wpływ wzrostu na przypisanie gracza do określonej roli w drużynie. **Centrzy (C)** wyróżniają się jako najwyżsi zawodnicy, podczas gdy **rozgrywający (PG)** mają najniższy wzrost, co odzwierciedla różnorodność zadań przypisanych do tych pozycji. Zawodnicy na pozostałych pozycjach, takich jak **SF (niski skrzydłowy)**, **SG (rzucający obrońca)** i **PF (silny skrzydłowy)**, zajmują miejsca pośrednie pod względem wzrostu. Ta zależność między wzrostem a pozycją znacząco wpływa na analizowane parametry gry, takie jak liczba rzutów za trzy punkty, zbiórki ofensywne czy skuteczność rzutów osobistych.

Na przykład, **centrzy**, z racji swojego wzrostu i roli w strefie podkoszowej, oddają znacznie mniej rzutów za trzy punkty, co może zakłócać analizę wpływu samego wzrostu na tę statystykę. Podobnie, **niżsi zawodnicy**, grający na pozycjach **PG** czy **SG**, mają większą swobodę w wykonywaniu rzutów z dystansu, co może sugerować, że różnice w statystykach wynikają bardziej z pozycji niż ze wzrostu.

```
# ANOVA dla każdej statystyki w zależności od pozycji
anova_oreb <- aov(OREB ~ Pos, data = player_stats_clean)
summary(anova_oreb)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Pos              4   118.8    29.7   74.21 <2e-16 ***
## Residuals      417   166.9     0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova_blk <- aov(BLK ~ Pos, data = player_stats_clean)
summary(anova_blk)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Pos              4    29.5    7.375   55.31 <2e-16 ***
## Residuals      417    55.6    0.133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

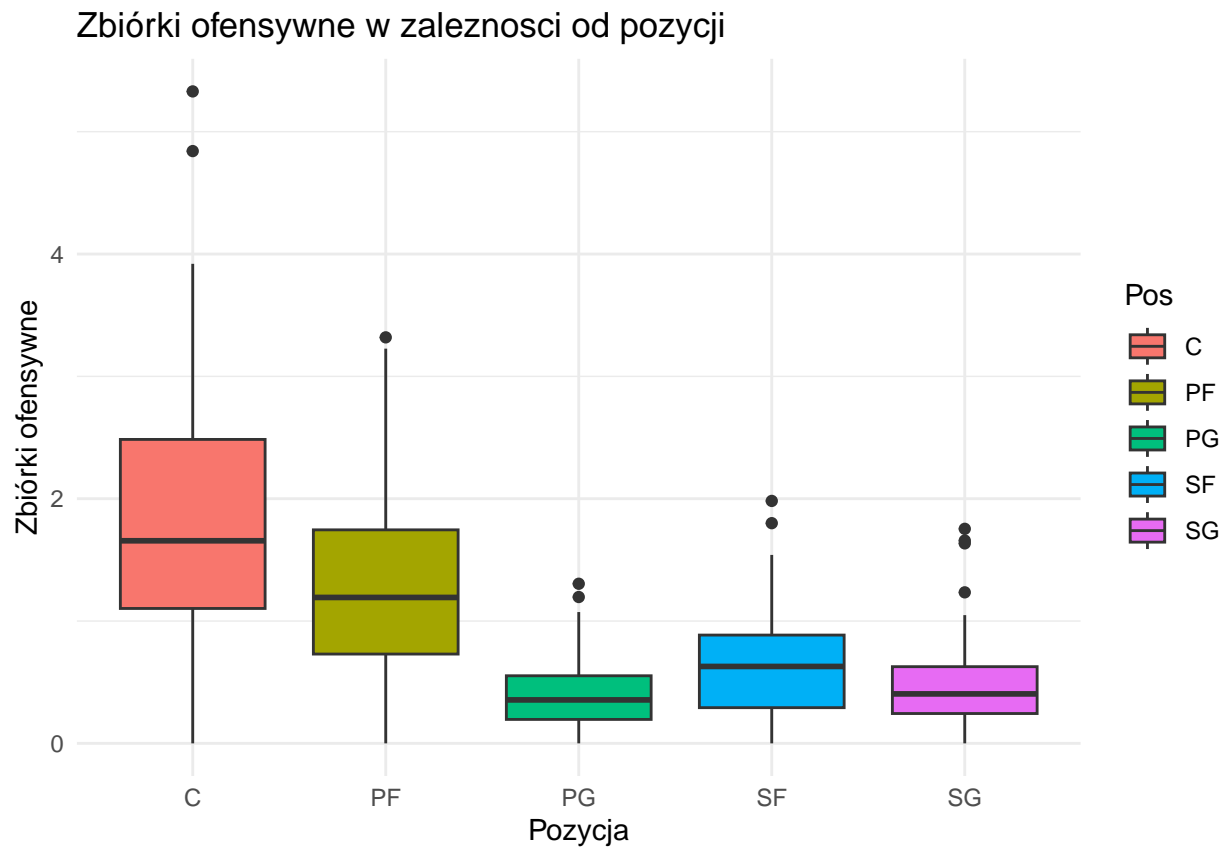
```
anova_3pm <- aov(X3PM ~ Pos, data = player_stats_clean)
summary(anova_3pm)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Pos              4    53.5   13.374   38.91 <2e-16 ***
## Residuals      417   143.3    0.344
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

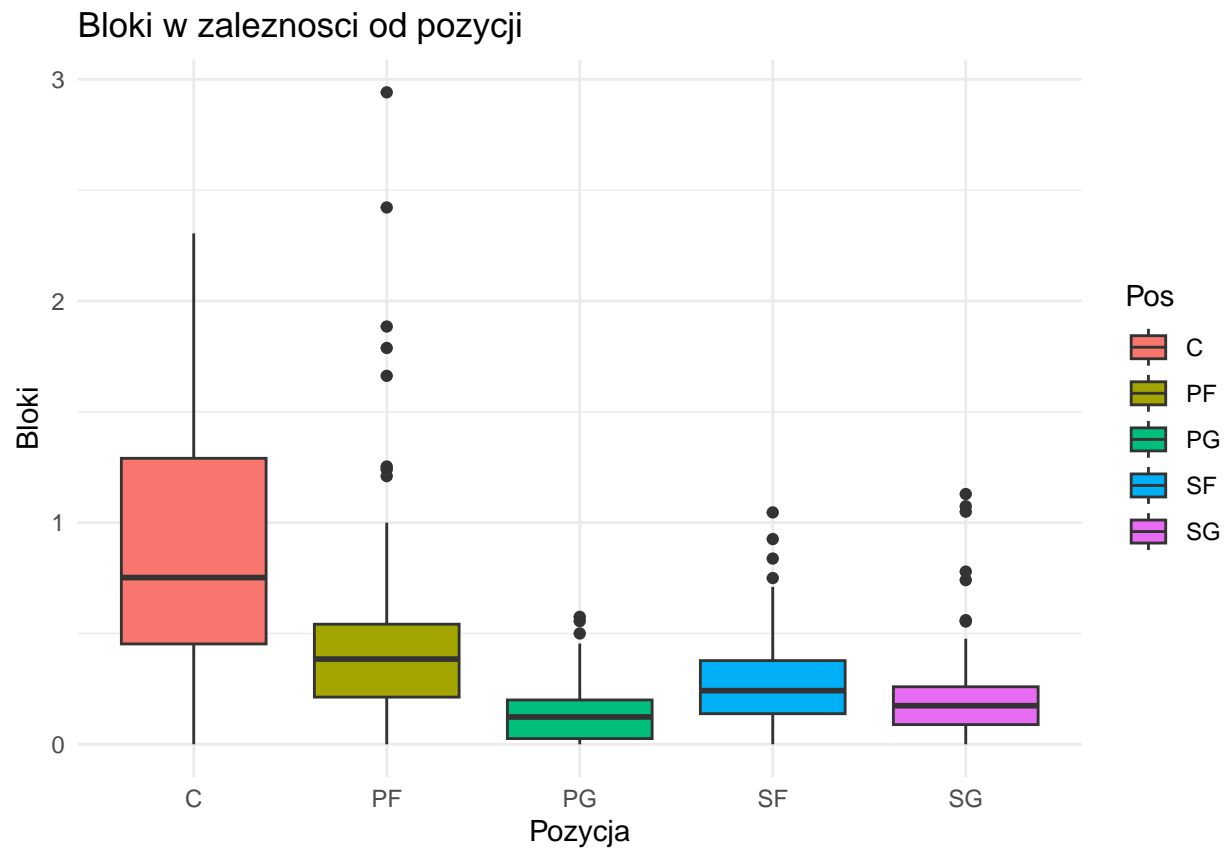
```
anova_ftp <- aov(FTP ~ Pos, data = player_stats_clean)
summary(anova_ftp)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Pos              4  0.668 0.16692   4.736 0.000954 ***
## Residuals      417 14.696 0.03524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Boxploty dla każdej statystyki w zależności od pozycji
ggplot(player_stats_clean, aes(x = Pos, y = OREB, fill = Pos)) +
  geom_boxplot() +
  labs(title = "Zbiórki ofensywne w zależności od pozycji", x = "Pozycja", y = "Zbiórki ofensywne") +
  theme_minimal()
```

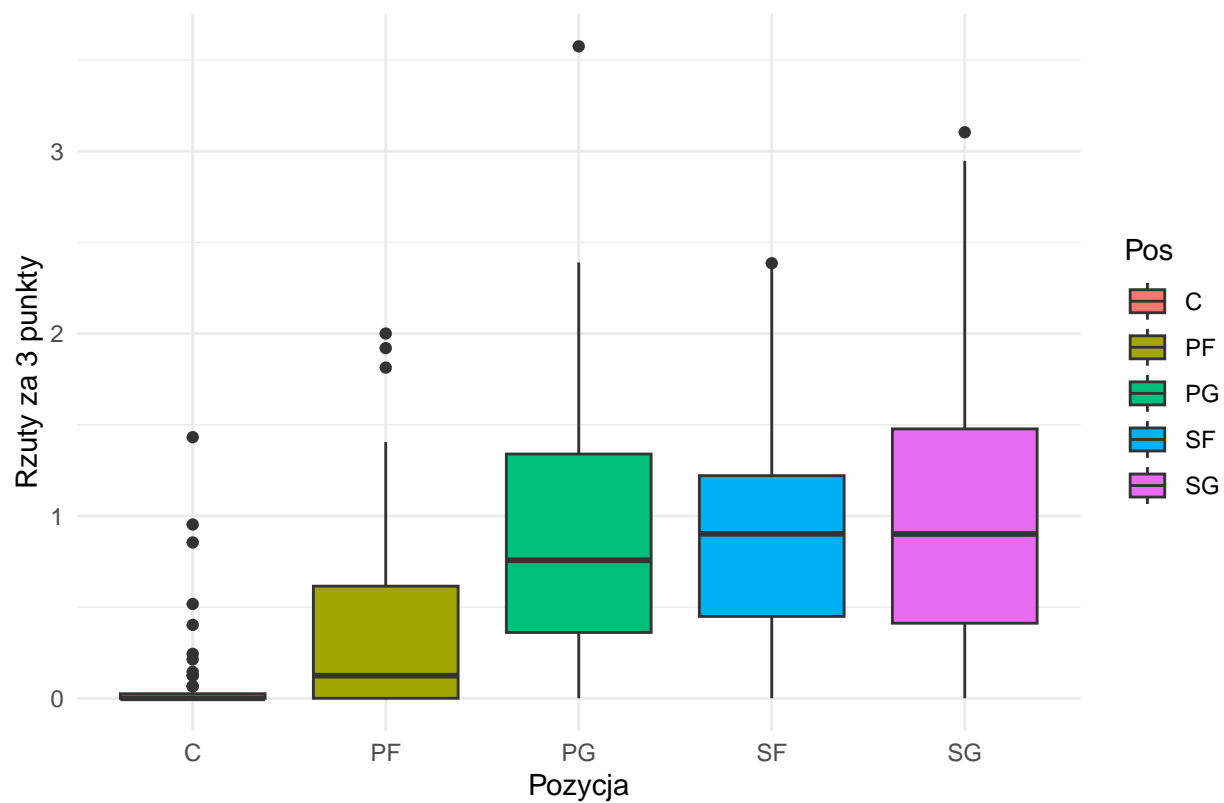


```
ggplot(player_stats_clean, aes(x = Pos, y = BLK, fill = Pos)) +
  geom_boxplot() +
  labs(title = "Blok w zależności od pozycji", x = "Pozycja", y = "Blok") +
  theme_minimal()
```

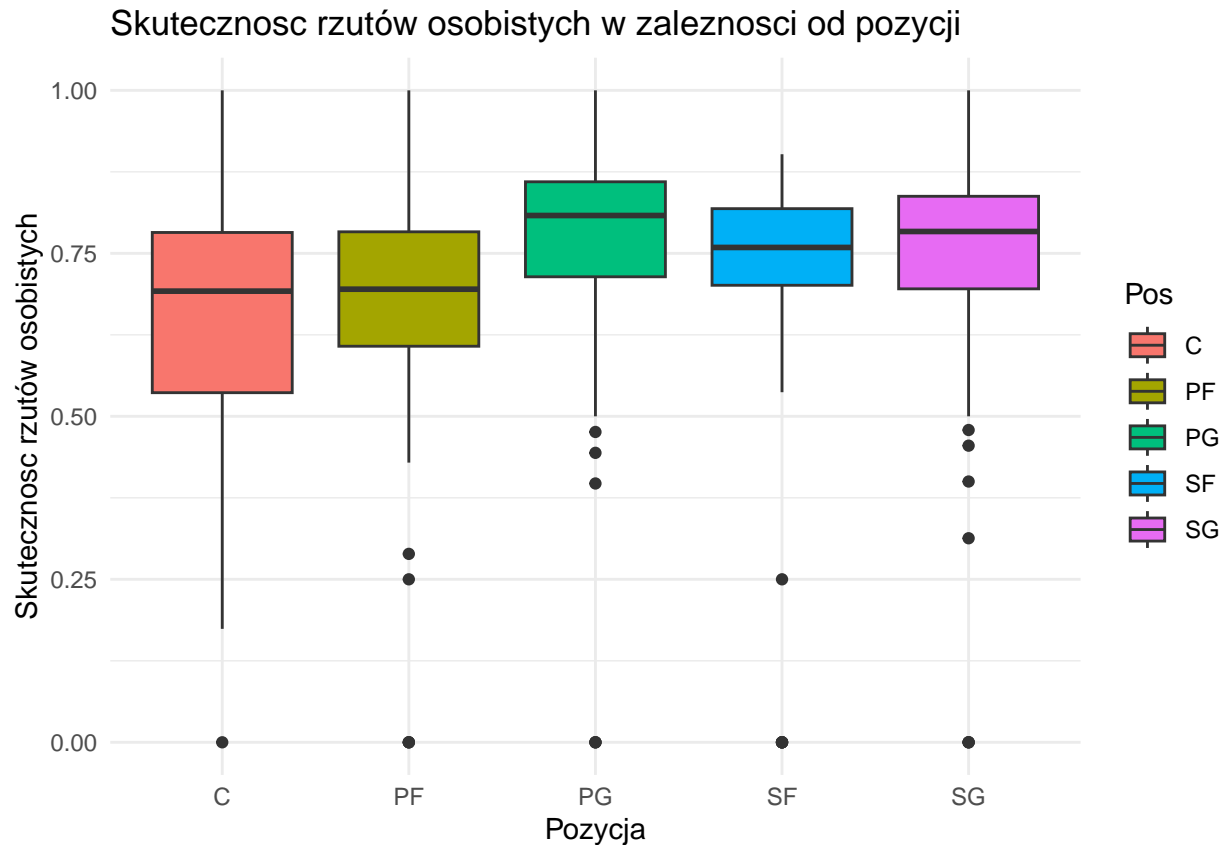


```
ggplot(player_stats_clean, aes(x = Pos, y = X3PM, fill = Pos)) +
  geom_boxplot() +
  labs(title = "Rzuty za 3 punkty w zależności od pozycji", x = "Pozycja", y = "Rzuty za 3 punkty") +
  theme_minimal()
```

Rzuty za 3 punkty w zależności od pozycji



```
ggplot(player_stats_clean, aes(x = Pos, y = FTP, fill = Pos)) +
  geom_boxplot() +
  labs(title = "Skuteczność rzutów osobistych w zależności od pozycji", x = "Pozycja", y = "Skuteczność")
  theme_minimal()
```



Jak wynika zarówno z wykresów, jak i przeprowadzonych testów statystycznych, różnice w statystykach między pozycjami są zauważalne i statystycznie istotne.

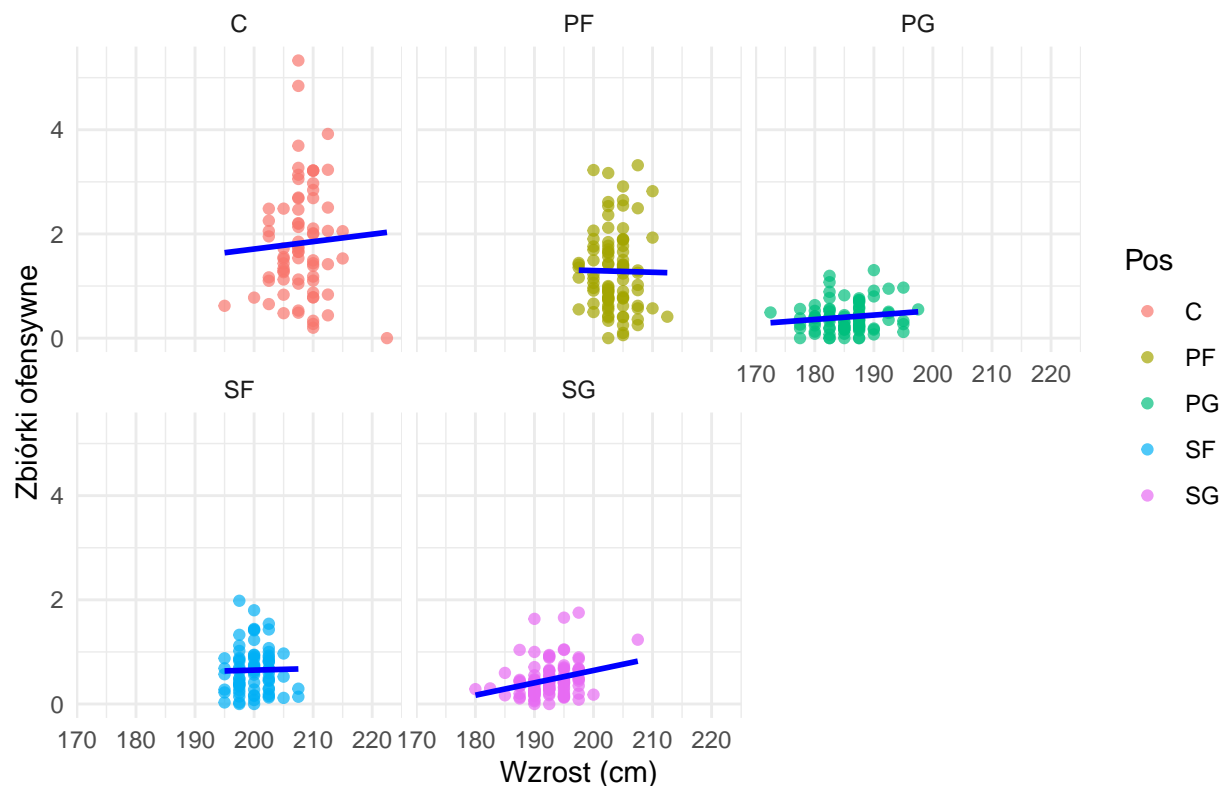
Aby wyeliminować te potencjalne zakłócenia i dokładniej przeanalizować wpływ wzrostu na statystyki gry, konieczne jest porównanie zawodników o różnym wzroście, ale grających na tej samej pozycji. Taka analiza pozwoli oddzielić wpływ wzrostu od zadań związanych z konkretną rolą na boisku i da bardziej precyzyjny obraz zależności między wzrostem a kluczowymi parametrami gry.

```
# Zależność wzrostu i zbiórek ofensywnych
plot_oreb <- ggplot(player_stats_clean, aes(x = Height, y = OREB, color = Pos)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  facet_wrap(~Pos) +
  labs(
    title = "Zależność wzrostu i zbiórek ofensywnych dla każdej pozycji",
    x = "Wzrost (cm)",
    y = "Zbiórki ofensywne"
  ) +
  theme_minimal()

print(plot_oreb)
```

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

Zależność wzrostu i zbiórek ofensywnych dla każdej pozycji



Zależność wzrostu i zbiórek ofensywnych (OREB)

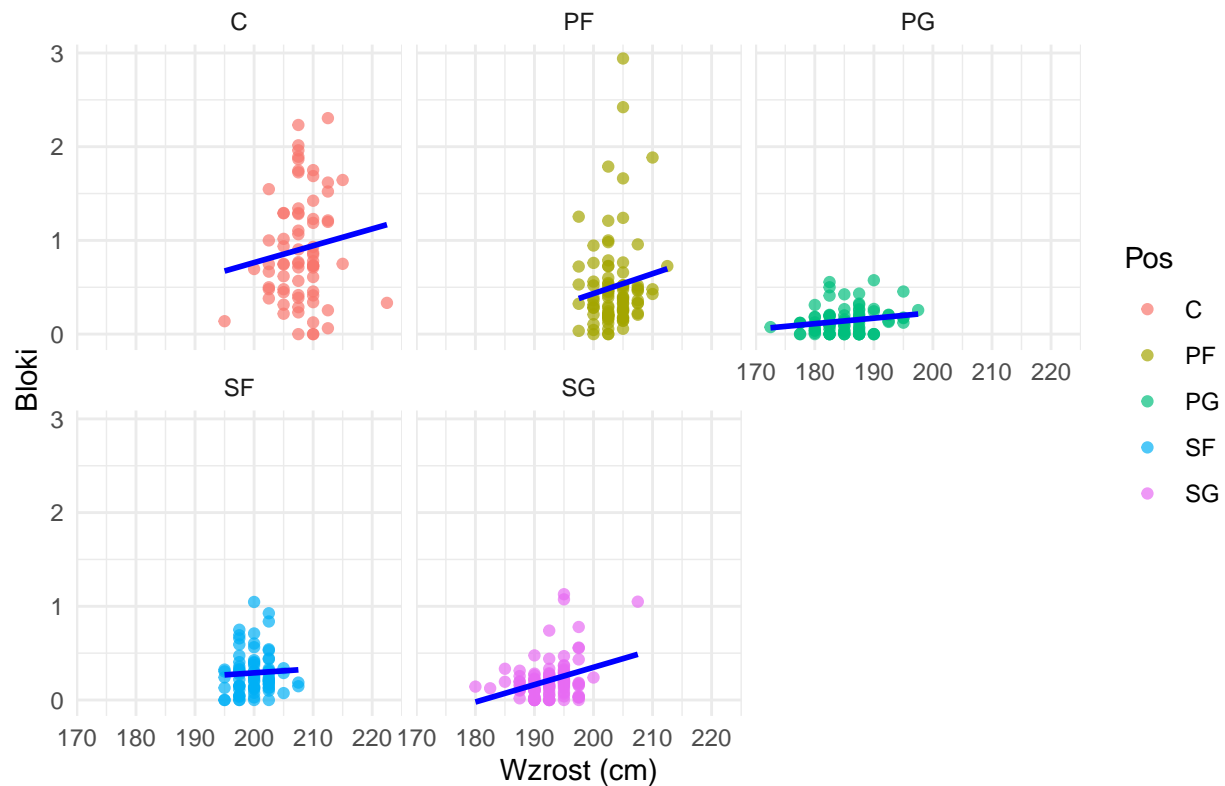
- **Centrzy (C) i rzucający obrońcy (SG):** Wyrażna pozytywna zależność – wyżsi zawodnicy na tych pozycjach zbierają więcej piłek ofensywnych dzięki swojej przewadze fizycznej i grze bliżej kosza.
- **Niscy skrzydłowi (SF), silni skrzydłowi (PF) i rozgrywający (PG):** Zależność niezauważalna – zbiórki ofensywne na tych pozycjach zależą bardziej od dynamiki i techniki niż od wzrostu.

```
# Zależność wzrostu i bloków
plot_blk <- ggplot(player_stats_clean, aes(x = Height, y = BLK, color = Pos)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  facet_wrap(~Pos) +
  labs(
    title = "Zależność wzrostu i bloków dla każdej pozycji",
    x = "Wzrost (cm)",
    y = "Blokki"
  ) +
  theme_minimal()

print(plot_blk)
```

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

Zależność wzrostu i bloków dla każdej pozycji



Zależność wzrostu i bloków (BLK)

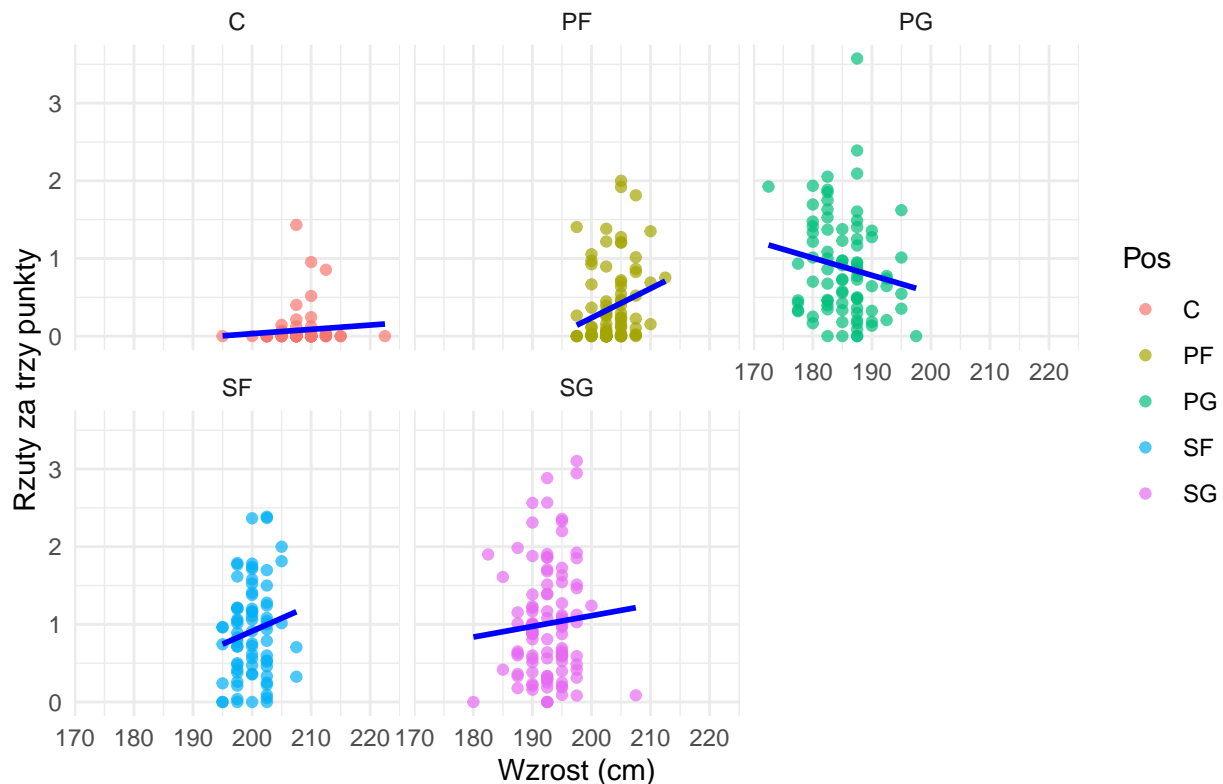
- **Centrzy (C), silni skrzydłowi (PF) i rzucający obrońcy (SG):** Wyraźna pozytywna zależność – wyżsi zawodnicy na tych pozycjach częściej blokują rzuty. Centrzy i silni skrzydłowi dominują w strefie podkoszowej, a rzucający obrońcy wykorzystują przewagę wzrostu w pojedynkach na obwodzie.
- **Niscy skrzydłowi (SF) i rozgrywający (PG):** Brak wyraźnej zależności – liczba bloków jest niska, niezależnie od wzrostu, co wynika z ich zadań skupionych na obronie obwodu i kreowaniu gry.

```
# Zależność wzrostu i rzutów za trzy punkty
plot_3pm <- ggplot(player_stats_clean, aes(x = Height, y = X3PM, color = Pos)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  facet_wrap(~Pos) +
  labs(
    title = "Zależność wzrostu i rzutów za trzy punkty dla każdej pozycji",
    x = "Wzrost (cm)",
    y = "Rzuty za trzy punkty"
  ) +
  theme_minimal()

print(plot_3pm)
```

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

Zależność wzrostu i rzutów za trzy punkty dla każdej pozycji



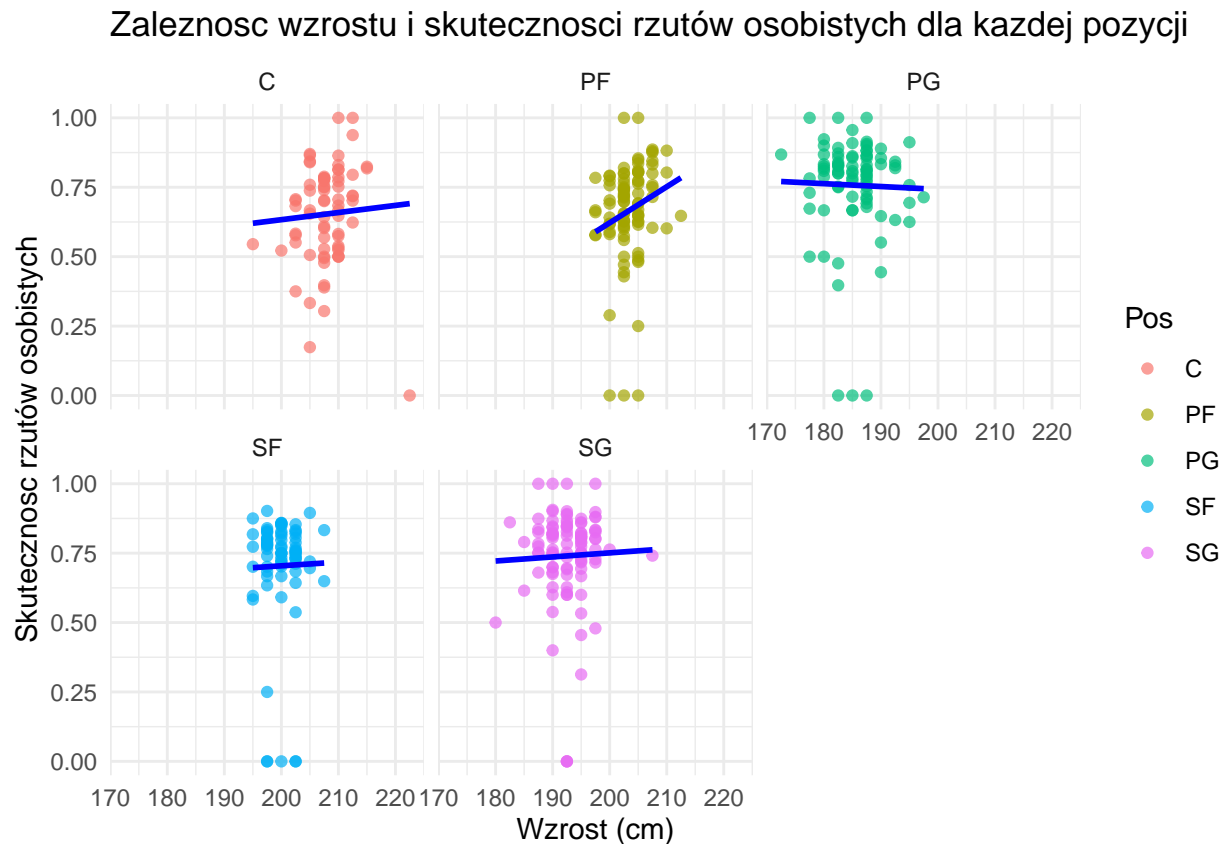
Zależność wzrostu i trafień za trzy punkty (X3PM)

- **Centrzy (C):** Zależność słabo zauważalna – centrzy rzadko trafiają za trzy punkty, a wzrost nie ma istotnego wpływu na ich skuteczność.
- **Niscy skrzydłowi (SF), rzucający obrońcy (SG) i silni skrzydłowi (PF):** Wyżsi zawodnicy częściej trafiają za trzy punkty. Może to wynikać z rozwijających się umiejętności rzutowych nawet wśród wyższych graczy na tych pozycjach.
- **Rozgrywający (PG):** Zależność odwrotna – niżsi zawodnicy trafiają więcej rzutów za trzy punkty, co jest zgodne z ich rolą na obwodzie i większą liczbą prób.

```
# Zależność wzrostu i skuteczności rzutów osobistych
plot_ftp <- ggplot(player_stats_clean, aes(x = Height, y = FTP, color = Pos)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  facet_wrap(~Pos) +
  labs(
    title = "Zależność wzrostu i skuteczności rzutów osobistych dla każdej pozycji",
    x = "Wzrost (cm)",
    y = "Skuteczność rzutów osobistych"
  ) +
  theme_minimal()

print(plot_ftp)
```

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



Zależność wzrostu i skuteczności rzutów osobistych (FTP)

- **Silni skrzydłowi (PF):** Wyraźna pozytywna zależność – wyżsi zawodnicy na tej pozycji mają wyższą skuteczność rzutów osobistych, co może wynikać z ich lepszej techniki lub większego doświadczenia w sytuacjach rzutowych.
- **Centrzy (C), rozgrywający (PG), rzucający obrońcy (SG) i niscy skrzydłowi (SF):** Zależność bardzo słaba lub niezauważalna – skuteczność rzutów osobistych jest w dużej mierze niezależna od wzrostu na tych pozycjach.

```
# Regresja wieloraka dla zbiorów ofensywnych
model_oreb_multi <- lm(OREB ~ Height + Pos, data = player_stats_clean)
summary(model_oreb_multi)
```

```
##
## Call:
## lm(formula = OREB ~ Height + Pos, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9879 -0.3307 -0.0620  0.2491  3.5108
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.52489    1.72147  -0.305    0.761
## Height      0.01129    0.00827   1.366    0.173
## PosPF       -0.48857    0.10641  -4.591 5.84e-06 ***
## PosPG       -1.16507    0.21244  -5.484 7.22e-08 ***
## PosSF       -1.08303    0.12314  -8.795 < 2e-16 ***
## PosSG       -1.18181    0.16019  -7.378 8.82e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.632 on 416 degrees of freedom
## Multiple R-squared:  0.4184, Adjusted R-squared:  0.4114
## F-statistic: 59.86 on 5 and 416 DF,  p-value: < 2.2e-16
```

```
# Regresja wieloraka dla bloków
```

```
model_blk_multi <- lm(BLK ~ Height + Pos, data = player_stats_clean)
summary(model_blk_multi)
```

```
##
## Call:
## lm(formula = BLK ~ Height + Pos, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93358 -0.17715 -0.05215  0.09773  2.41401
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.85723    0.98653  -1.883 0.060453 .
## Height      0.01329    0.00474   2.804 0.005284 **
## PosPF       -0.33997    0.06098  -5.575 4.46e-08 ***
## PosPG       -0.46342    0.12174  -3.807 0.000162 ***
## PosSF       -0.51070    0.07057  -7.237 2.23e-12 ***
## PosSG       -0.49051    0.09180  -5.343 1.51e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3622 on 416 degrees of freedom
## Multiple R-squared:  0.3588, Adjusted R-squared:  0.3511
## F-statistic: 46.55 on 5 and 416 DF,  p-value: < 2.2e-16
```

```
# Regresja wieloraka dla celnych rzutów za trzy
```

```
model_3pm_multi <- lm(X3PM ~ Height + Pos, data = player_stats_clean)
summary(model_3pm_multi)
```

```
##
## Call:
## lm(formula = X3PM ~ Height + Pos, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01442 -0.36630 -0.07306  0.26598  2.67586
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.192688   1.597737  -0.746   0.4558
## Height      0.006100   0.007676   0.795   0.4273
## PosPF       0.319733   0.098763   3.237   0.0013 **
## PosPG       0.948078   0.197166   4.809 2.13e-06 ***
## PosSF       0.883943   0.114287   7.734 7.92e-14 ***
## PosSG       1.027779   0.148676   6.913 1.79e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5866 on 416 degrees of freedom
## Multiple R-squared:  0.2729, Adjusted R-squared:  0.2641
## F-statistic: 31.23 on 5 and 416 DF,  p-value: < 2.2e-16

# Regresja wieloraka dla rzutów wolnych
model_ftp_multi <- lm(FTP ~ Height + Pos, data = player_stats_clean)
summary(model_ftp_multi)
```

```
##
## Call:
## lm(formula = FTP ~ Height + Pos, data = player_stats_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.76259 -0.05311  0.04490  0.11008  0.34130
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.146046   0.511374   0.286  0.77533
## Height      0.002441   0.002457   0.994  0.32097
## PosPF       0.024816   0.031610   0.785  0.43287
## PosPG       0.158818   0.063105   2.517  0.01222 *
## PosSF       0.070126   0.036579   1.917  0.05591 .
## PosSG       0.123951   0.047586   2.605  0.00952 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1877 on 416 degrees of freedom
## Multiple R-squared:  0.04572, Adjusted R-squared:  0.03425
## F-statistic: 3.986 on 5 and 416 DF,  p-value: 0.001525
```

Analiza wpływu wzrostu i pozycji na statystyki gry

- **Zbiórki ofensywne (OREB):** Wzrost nie jest istotnym predyktorem ($p = 0.173$), natomiast pozycja zawodnika odgrywa kluczową rolę – zawodnicy na pozycjach PG, SF, SG i PF zbierają mniej piłek niż centry ($p < 0.001$).
- **Bloki (BLK):** Zarówno wzrost ($p = 0.005$), jak i pozycja ($p < 0.001$) mają istotny wpływ. Wyżsi zawodnicy i gracze na pozycji C wykonują więcej bloków.

- **Punkty za trzy (X3PM):** Wzrost nie ma znaczenia ($p = 0.427$), ale zawodnicy na pozycjach PG, SF, SG i PF zdobywają więcej punktów za trzy niż centry ($p < 0.01$).
- **Skuteczność rzutów osobistych (FTP):** Wzrost nie jest istotny ($p = 0.321$), a wyższą skuteczność rzutów osobistych mają gracze na pozycjach PG i SG ($p < 0.05$).

Końcowe przemyślenia

Analiza wykazała, że wzrost zawodnika odgrywa istotną rolę w koszykówce, zarówno bezpośrednio, jak i pośrednio poprzez wpływ na przypisanie zawodnika do określonej roli na boisku. Wyżsi zawodnicy mają przewagę w statystykach związanych z grą pod koszem, takich jak bloki (BLK) i zbiórki ofensywne (OREB). Wynika to z ich fizycznych predyspozycji, które pozwalają im skuteczniej rywalizować w strefie podkoszowej, oraz z przypisanych im zadań defensywnych i zbiórkowych.

Jednak zależności w innych statystykach, takich jak punkty za trzy (X3PM) i skuteczność rzutów osobistych (FTP), są bardziej złożone. Na pierwszy rzut oka wzrost ma na nie niekorzystny wpływ. Wyżsi zawodnicy, zwłaszcza centry i silni skrzydłowi, rzadziej trafiają za trzy i mają niższą skuteczność rzutów osobistych niż ich niżsi koledzy z drużyny. Kluczowym czynnikiem jest jednak pozycja zawodnika i wynikające z niej zadania:

- Wyżsi zawodnicy są naturalnie przypisywani do ról podkoszowych (C, PF), co ogranicza ich udział w rzutach z dystansu.
- Niżsi zawodnicy, jak rozgrywający (PG) i rzucający obrońcy (SG), mają role skoncentrowane na kreowaniu i realizacji rzutów z dystansu, co tłumaczy ich lepsze wyniki w X3PM i FTP.

Ostatecznie wzrost sam w sobie nie zawsze jest decydujący, ale wpływa na dobór roli, która z kolei determinuje kluczowe statystyki gry. Bez uwzględnienia tego kontekstu, zależności między wzrostem a wynikami mogą być mylące. Podsumowując:

- Wzrost korzystnie wpływa na statystyki związane z grą pod koszem (bloki i zbiórki).
- W przypadku rzutów za trzy i skuteczności osobistych, większe znaczenie ma pozycja zawodnika i jego rola na boisku, na które wzrost ma wpływ pośredni.

Nasza analiza pokazuje, że wpływ wzrostu zawodnika na statystyki gry w koszykówce jest zauważalny, szczególnie w kontekście zbiórek ofensywnych i bloków. Jednak zależności te nie są proste i wymagają uwzględnienia dodatkowych czynników, takich jak pozycja zawodnika oraz wynikające z niej zadania na boisku. W szczególności statystyki związane z rzutami za trzy punkty i skutecznością rzutów osobistych są bardziej zależne od roli zawodnika niż od jego wzrostu. Warto podkreślić, że nasza analiza skupiała się na podstawowych zależnościach i nie obejmowała bardziej szczegółowych czynników, które mogłyby pogłębić te wnioski.