

Εργασία 5

2024-04-05

Επιλογή Dataset

Το dataset που επέλεξα περιέχει μέσους όρους των στατιστικών όλων των παικτών που αγωνίστηκαν στην Euroleague την σεζόν 2021/21. Το επέλεξα γιατί εκτός από τα στατιστικά στοιχεία, περιέχει και αναλυτικά στοιχεία για τον κάθε παίκτη. Για κάθε παίκτη έχουμε ονοματεπώνυμο, ομάδα, ύψος, αριθμό φανέλας, ημερομηνία γέννησης, θέση, εθνικότητα, συνολικά παιχνίδια*, συνολικοί πόντοι*, μέσος όρος πόντων*, επιτυχημένα(στον στόχο)/εκτελεσμένα δίποντα και ποσοστό στα δίποντα, επιτυχημένα(στον στόχο)/εκτελεσμένα τρίποντα και ποσοστό στα τρίποντα, επιτυχημένες(στον στόχο)/εκτελεσμένες βολές και ποσοστό στις βολές, μέσο όρο στα ριμπάουντ, στα κλεψίματα, στις ασίστ και στα κοψίματα(μπλοκ). Επίσης άλλαξα τα column names καθώς υπήρχε θέμα στο κώδικα (2FG -> TWOFG, 2FG_% -> TWOFG_PER, 3FG -> THREEFG, 3FG_% -> THREEFG_PER, FT_% -> FT_PER).

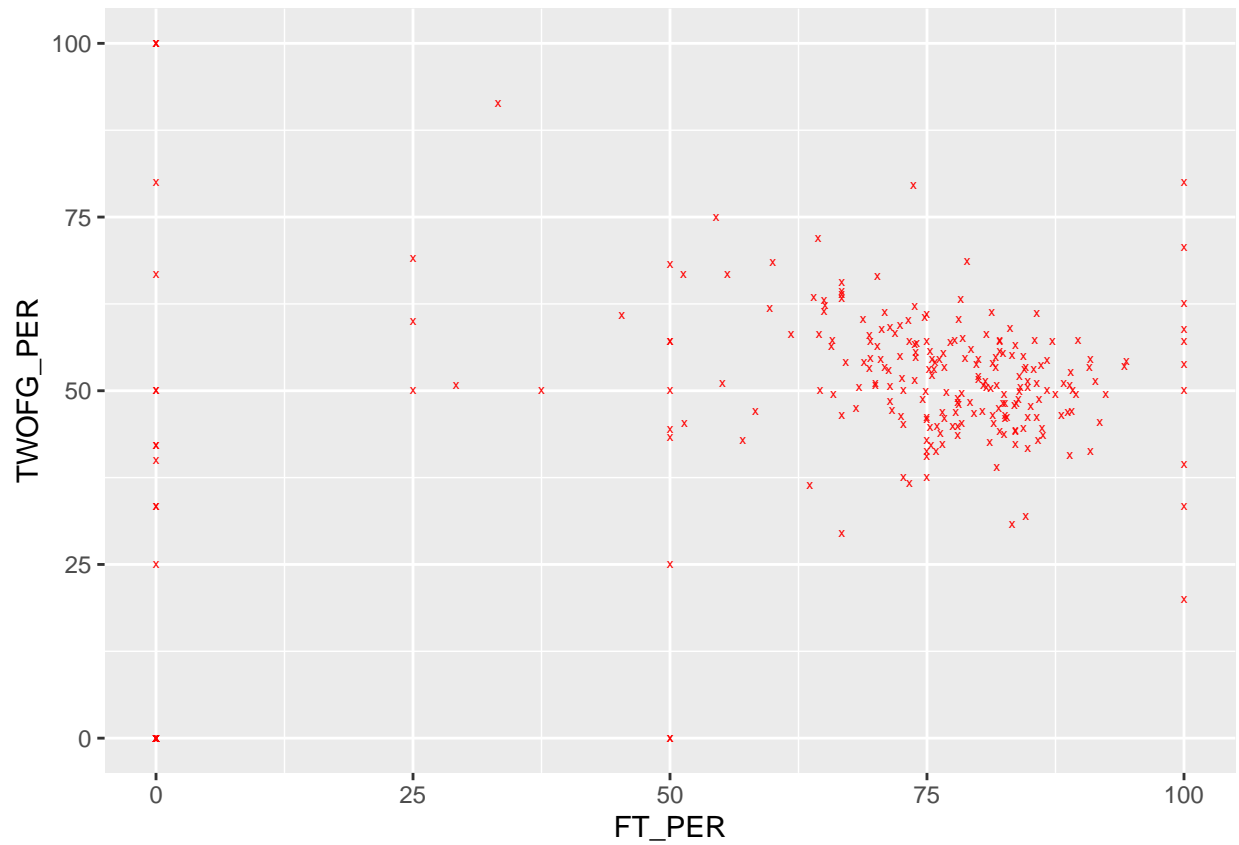
*(σε όλη τη καριέρα τους στη διοργάνωση)

Ενδεικτικά διαγράμματα

```
library(ggplot2)
library(readr)
euroleaguePlayers_average <- read_csv("C:/Users/UserA/Downloads/euroleaguePlayers_average.csv")

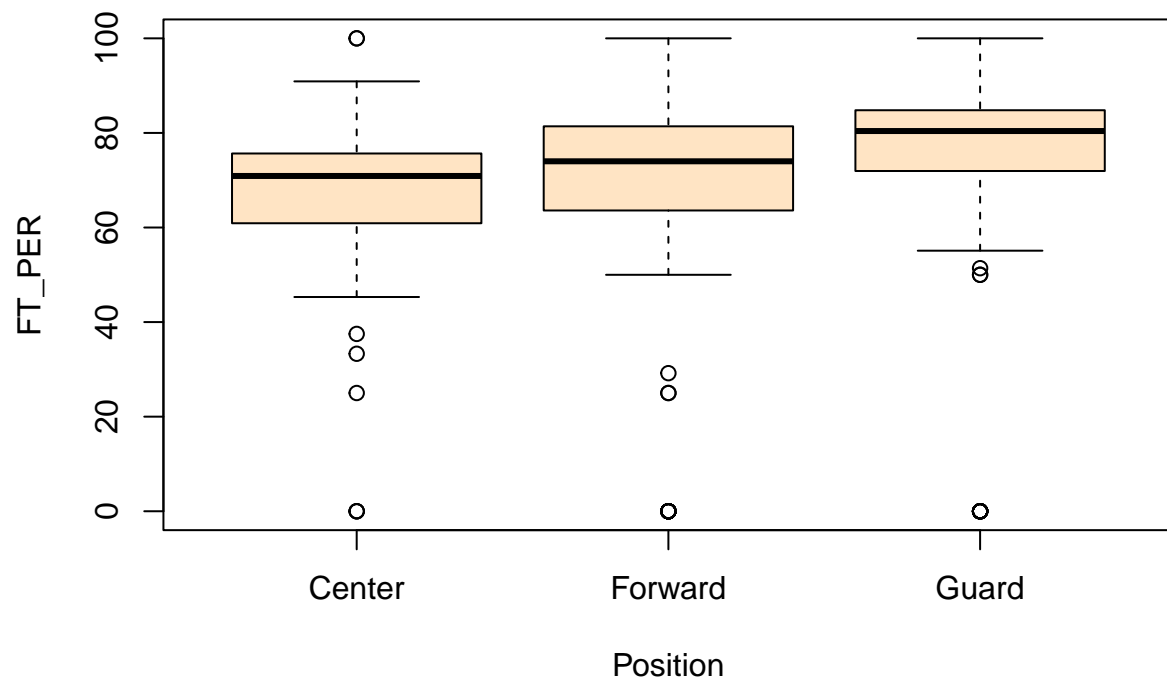
## Rows: 281 Columns: 22
## -- Column specification -----
## Delimiter: ","
## chr (10): full_name, surname, name, Club, Position, Born, Nationality, 2FG, ...
## dbl (12): Dorsal, Height, G, Pts, Avg, 2FG_%, 3FG_%, FT_%, Reb, St, As, Bl
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

colnames(euroleaguePlayers_average) <- c('full_name', 'surname', 'name', 'Club', 'Dorsal', 'Position',
ggplot(euroleaguePlayers_average, aes(FT_PER, TWOFG_PER))+geom_point(shape="x",color="red", na.rm = TRUE)
```



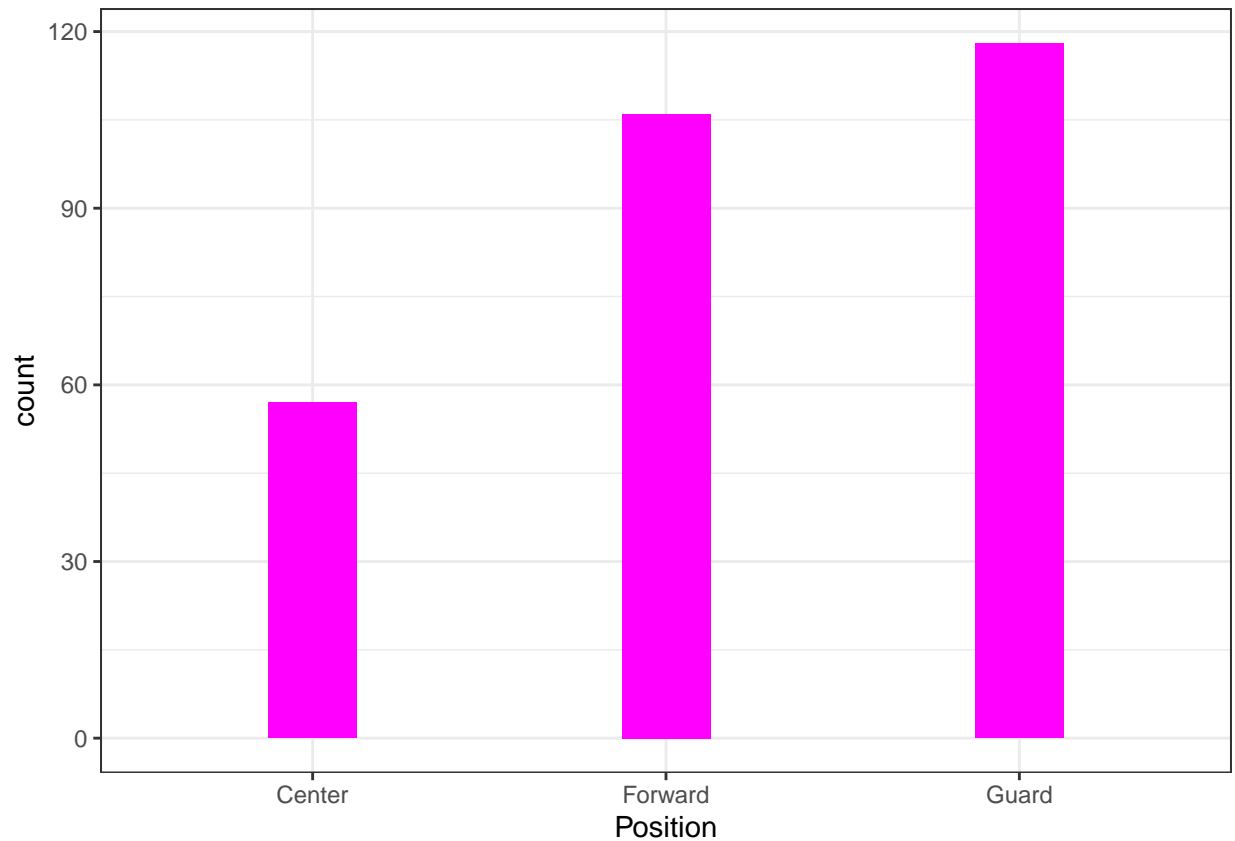
Σε αυτό το διάγραμμα διασπόρας παρατηρούμε ότι όσοι παίκτες έχουν γύρω στο 75% στις βολές έχουν και σχετικά καλό ποσοστό στα δίποντα (γύρω στο 50%).

```
with(euroleaguePlayers_average,boxplot(FT_PER ~ Position, col="bisque"))
```



Σε αυτό το διάγραμμα παρατηρούμε το ποσοστό των παικτών στις βόλες σε σχέση με τη θέση τους.

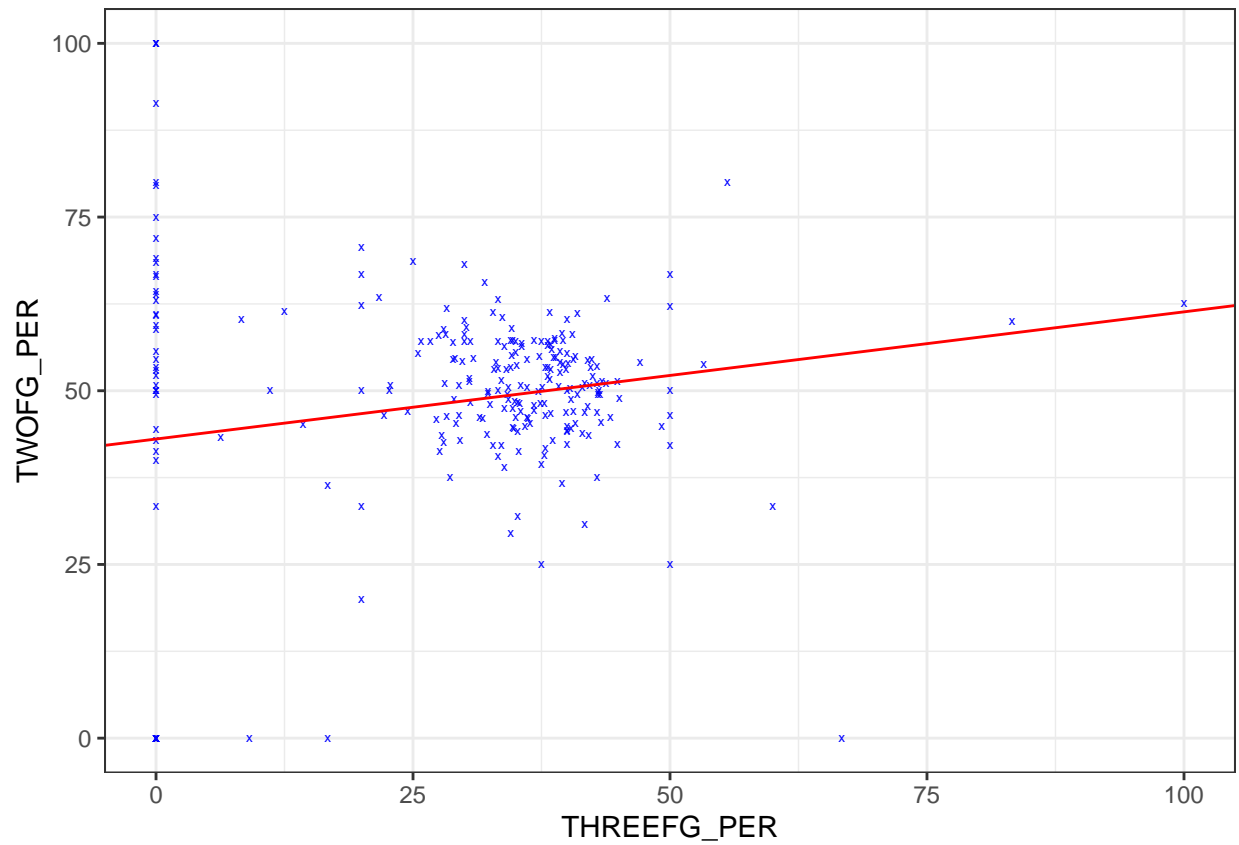
```
ggplot(euroleaguePlayers_average, aes(Position)) + geom_bar(fill="magenta", position="dodge", width=0.2)
```



Σε αυτό το διάγραμμα παρατηρούμε τον αριθμό των παικτών ανάλογα με τη θέση στην οποία παίζουν.

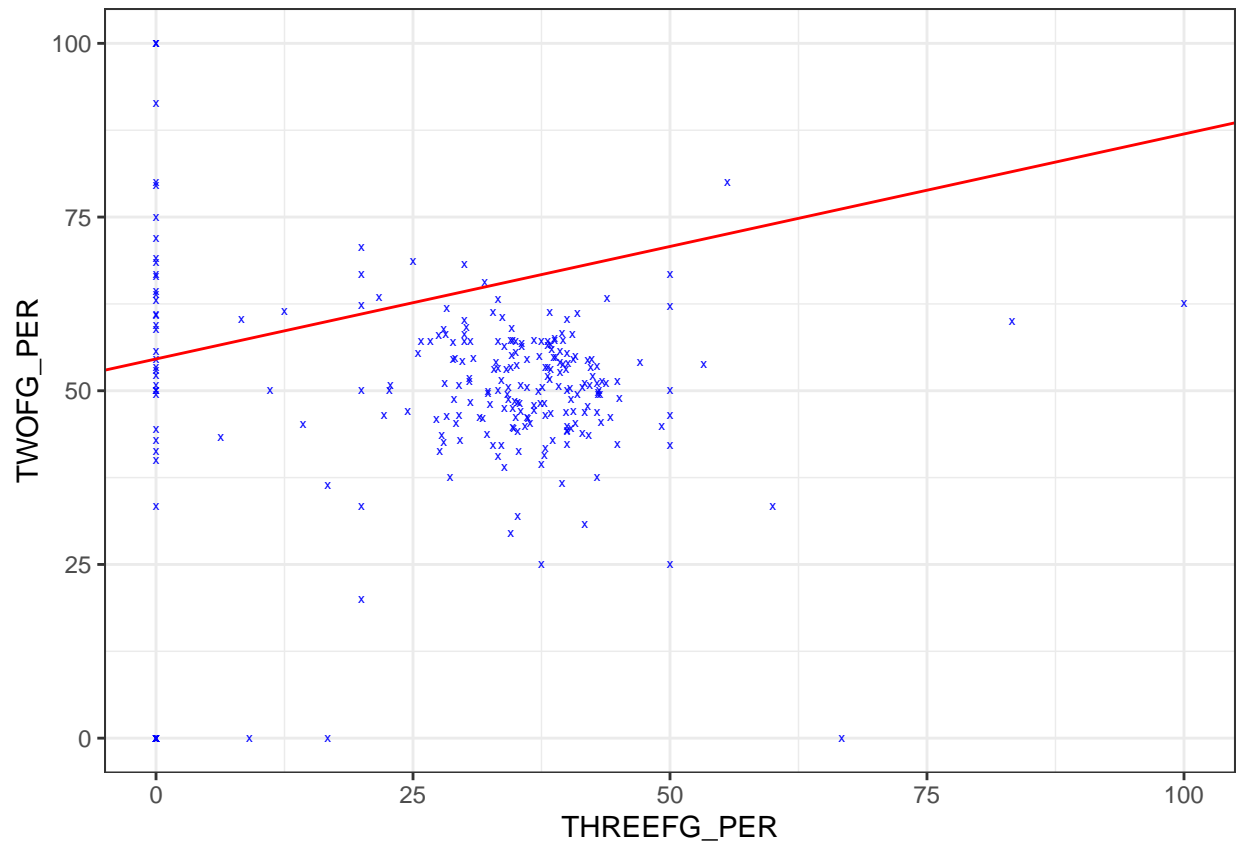
Γραμμική παλινδρόμηση με μία μεταβλητή

```
model1 <- lm(TWOFG_PER ~ THREEFG_PER, euroleaguePlayers_average)
ggplot(euroleaguePlayers_average, aes(THREEFG_PER, TWOFG_PER)) + geom_point(shape="x", color="blue", na.rm=TRUE)
```



Γραμμική παλινδρόμηση με δύο μεταβλητές

```
model2 <- lm(TWOFG_PER ~ THREEFG_PER + Position, euroleaguePlayers_average)
ggplot(euroleaguePlayers_average, aes(THREEFG_PER, TWOFG_PER)) + geom_point(shape="x", color="blue", na.rm=TRUE)
```



Συσχετίσεις, SSE και R-Squared

```
cor(euroleaguePlayers_average$TWOFG_PER,euroleaguePlayers_average$THREEFG_PER, use = "complete.obs")
```

```
## [1] 0.1784833
```

```
cor(euroleaguePlayers_average$TWOFG_PER,euroleaguePlayers_average$FT_PER, use = "complete.obs")
```

```
## [1] 0.4182876
```

```
cor(euroleaguePlayers_average$FT_PER,euroleaguePlayers_average$THREEFG_PER, use = "complete.obs")
```

```
## [1] 0.3890867
```

```
SSE1 <- sum(model1$residuals^2)
SSE1
```

```
## [1] 75376.06
```

```
SSE2 <- sum(model2$residuals^2)
SSE2
```

```
## [1] 61088.59
```

```
summary(model1)$r.squared
```

```
## [1] 0.03185629
```

```
summary(model2)$r.squared
```

```
## [1] 0.2153671
```

Σχολιασμός

Στην γραμμική παλινδρόμηση με μία μεταβλητή παρατηρούμε ότι όσο αυξάνεται η τιμή της μεταβλητής THREEFG_PER τόσο αυξάνεται και της μεταβλητής TWOFG_PER. Στην γραμμική παλινδρόμηση με δύο μεταβλητές παρατηρούμε ότι η ευθεία μετατοπίστηκε προς τα πάνω και αυτό δικαιολογείται από την μείωση της τιμής του SSE αλλά και από την αύξηση της τιμής του R-Squared. Επομένως, με την προσθήκη της μεταβλητής Position το μοντέλο έγινε καλύτερο.