Basketball Statistical Analysis

Web Scraping and Dataset Creation

- Scraped data from basketball reference site
- The dataset contains per game statistics
- Table creation

```
## # A tibble: 842 x 30
##    Rk    Player Pos   Age   Tm    G     GS    MP    FG    FGA   `FG%` `3P`
##    <chr> <chr>  <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
##  1 1     Preci~ C     22    TOR   73    28    23.6  3.6   8.3   .439  0.8
##  2 2     Steve~ C     28    MEM   76    75    26.3  2.8   5.1   .547  0.0
##  3 3     Bam A~ C     24    MIA   56    56    32.6  7.3   13.0  .557  0.0
##  4 4     Santi~ PF    21    MEM   32    0     11.3  1.7   4.1   .402  0.2
##  5 5     LaMar~ C     36    BRK   47    12    22.3  5.4   9.7   .550  0.3
##  6 6     Nicke~ SG    23    TOT   65    21    22.6  3.9   10.5  .372  1.6
##  7 6     Nicke~ SG    23    NOP   50    19    26.3  4.7   12.6  .375  1.9
##  8 6     Nicke~ SG    23    UTA   15    2     9.9   1.1   3.2   .333  0.7
##  9 7     Grays~ SG    26    MIL   66    61    27.3  3.9   8.6   .448  2.4
## 10 8     Jarre~ C     23    CLE   56    56    32.3  6.6   9.7   .677  0.0
## # ... with 832 more rows, and 18 more variables: `3PA` <chr>, `3P%` <chr>,
## #   `2P` <chr>, `2PA` <chr>, `2P%` <chr>, `eFG%` <chr>, FT <chr>, FTA <chr>,
## #   `FT%` <chr>, ORB <chr>, DRB <chr>, TRB <chr>, AST <chr>, STL <chr>,
## #   BLK <chr>, TOV <chr>, PF <chr>, PTS <chr>
```

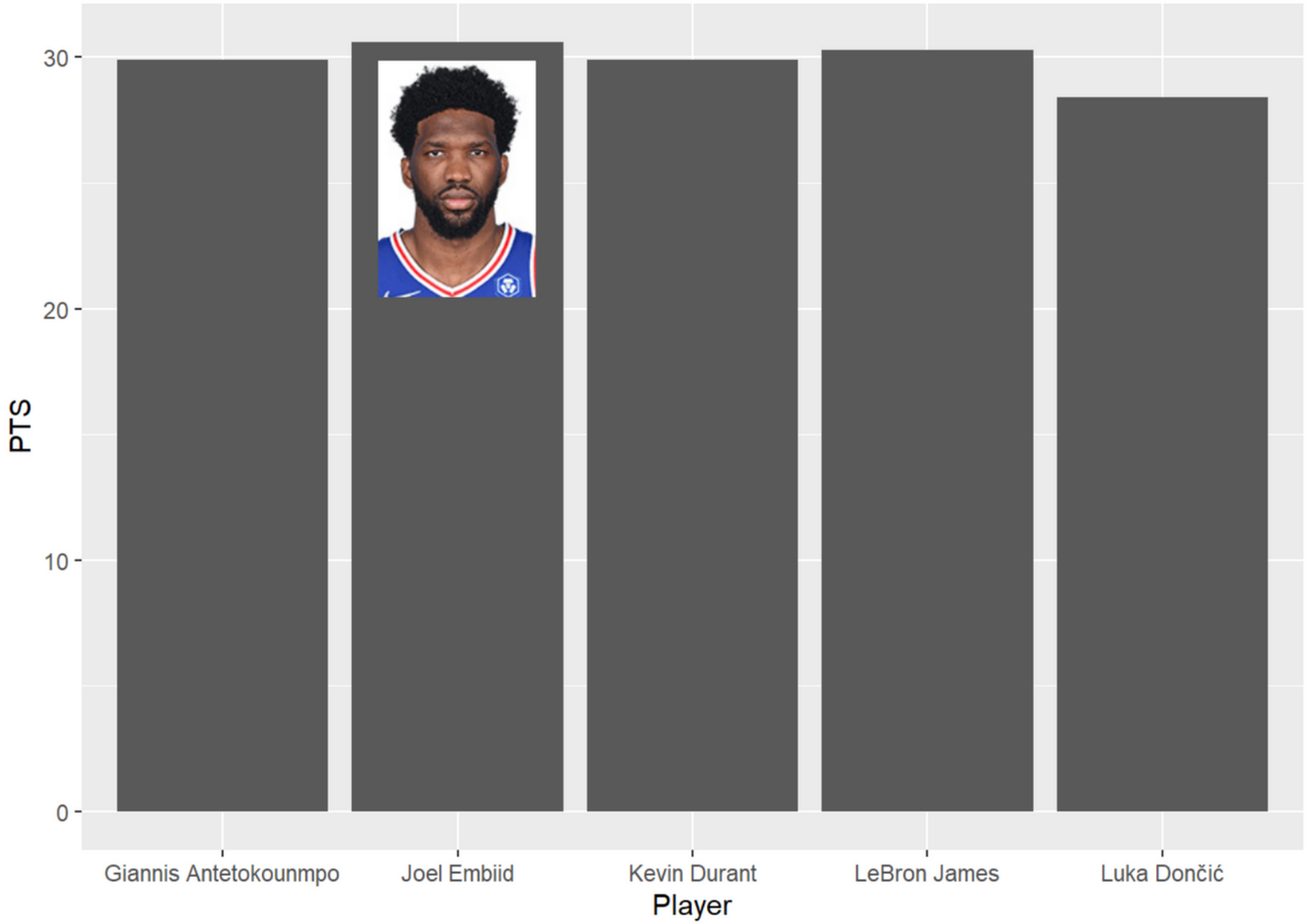
Data Cleaning

- **Using the glimpse functions to check if any change in the variables names was needed**
- **Converting char data to integer data**
- **Eliminating label lines from the tibble**

Data Analysis

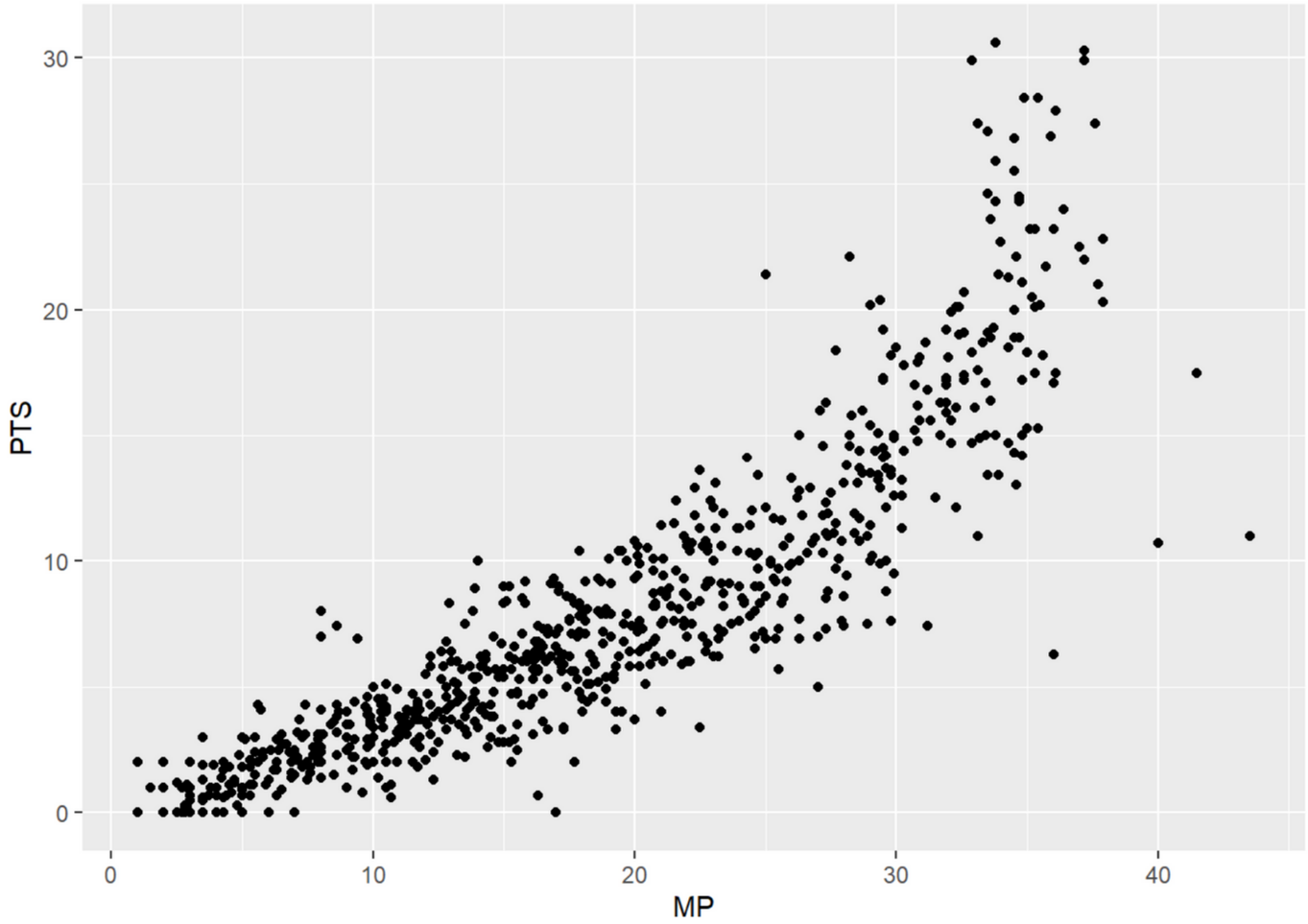
Top 5 scorer

First, we analyzed the top 5 scorers in the dataset. Then we created a barplot with the names of the top scorers and their scores. Moreover, we scraped the image of the player with the highest score and inserted it on the player bar.



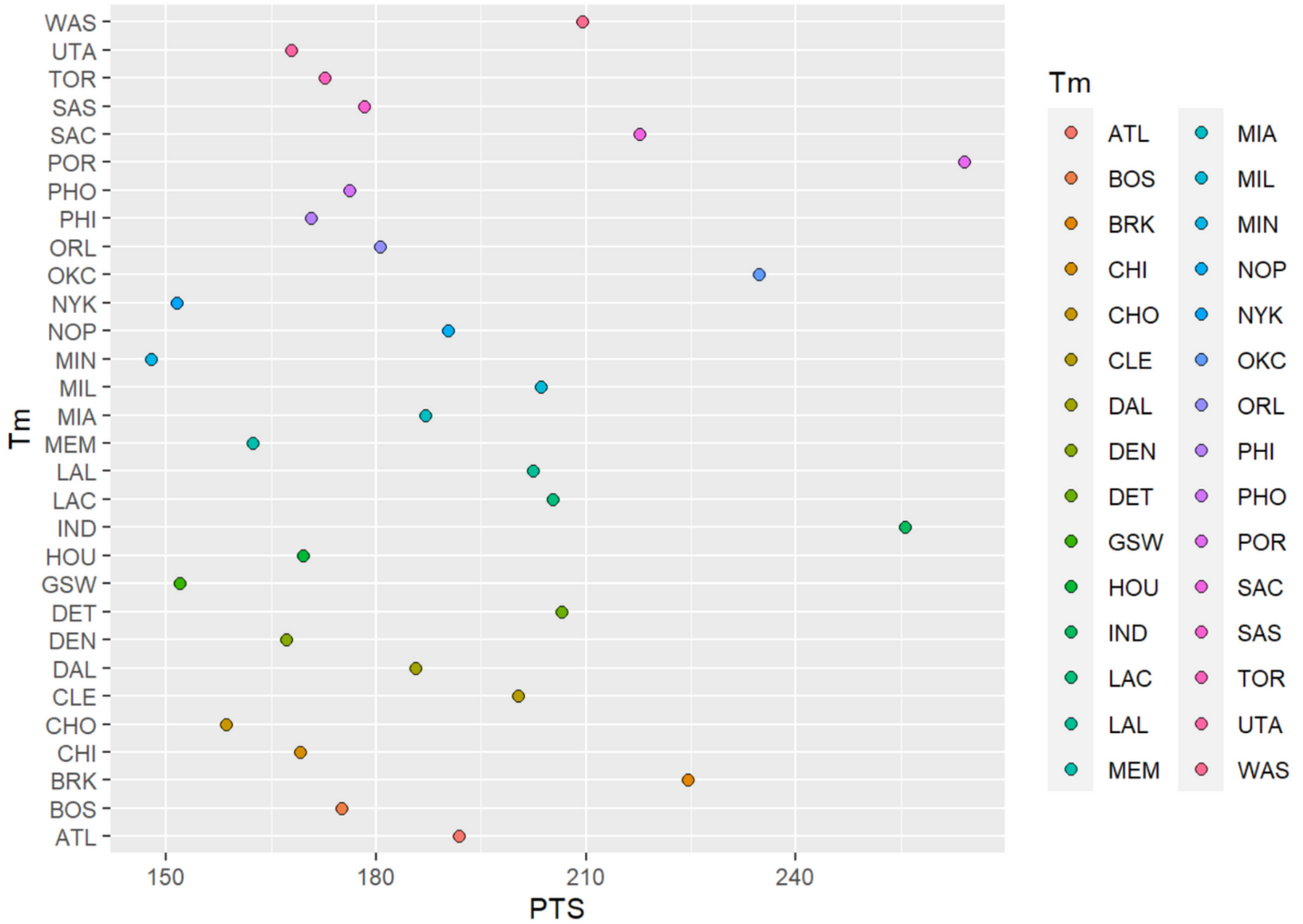
Relationship between PTS and MP

- Positive correlation



Top teams by points

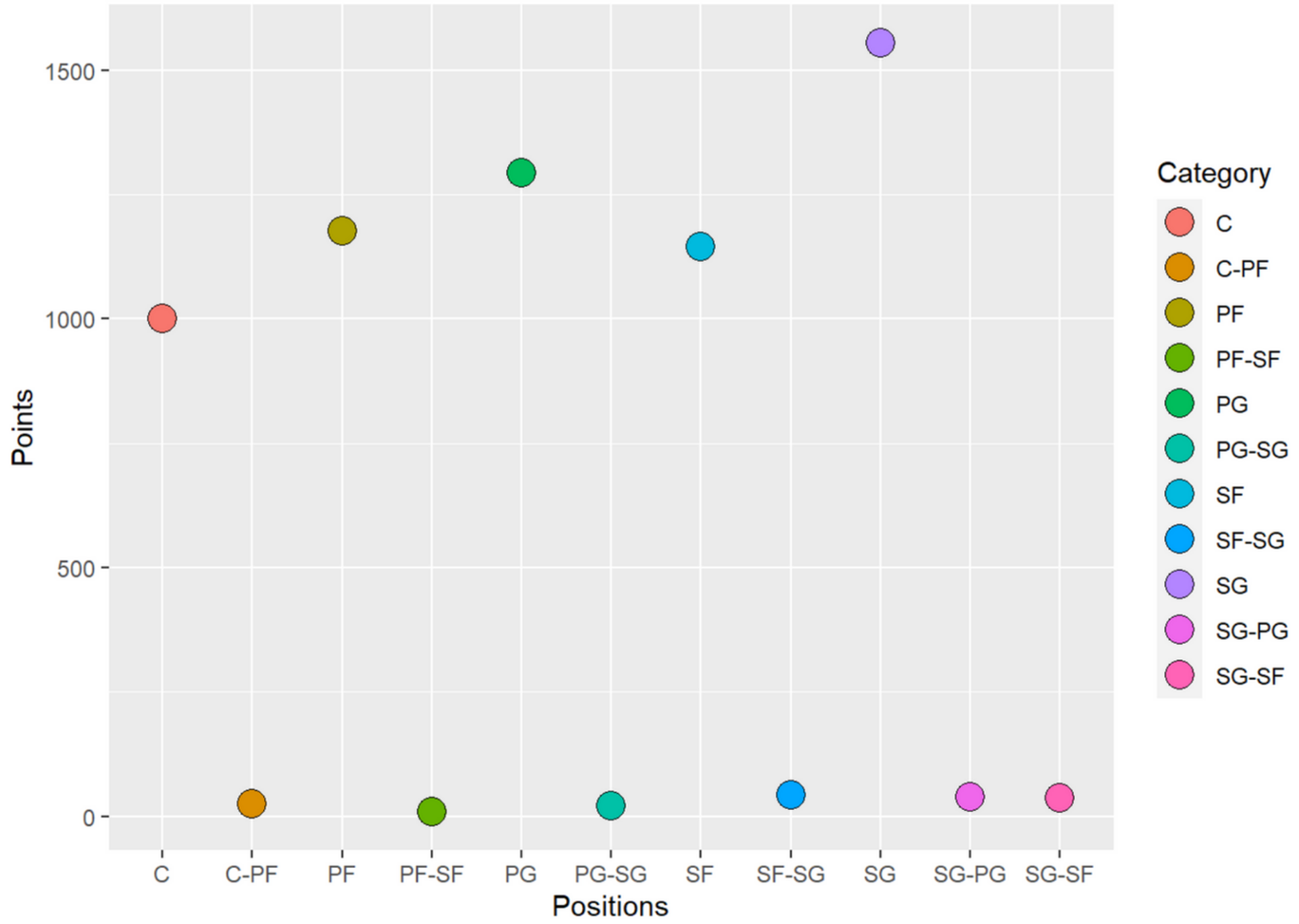
```
##    Tm   PTS
## 1 POR 264.1
## 2 IND 255.6
## 3 OKC 234.7
## 4 BRK 224.6
## 5 SAC 217.7
## 6 WAS 209.5
```



Top scorer by position

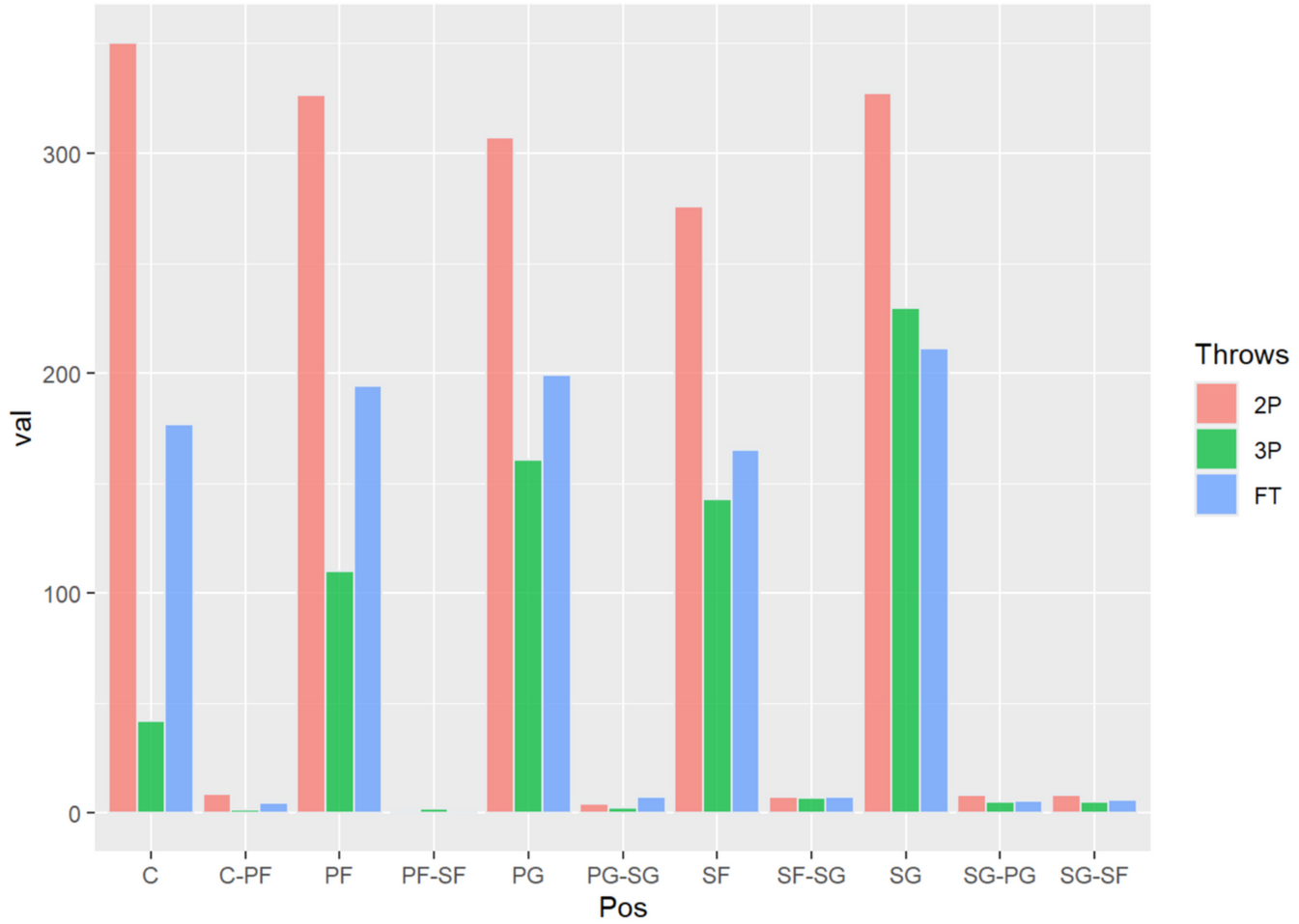
Top 3 categories:

- Shooting guards
- Point Guards
- Power Forward



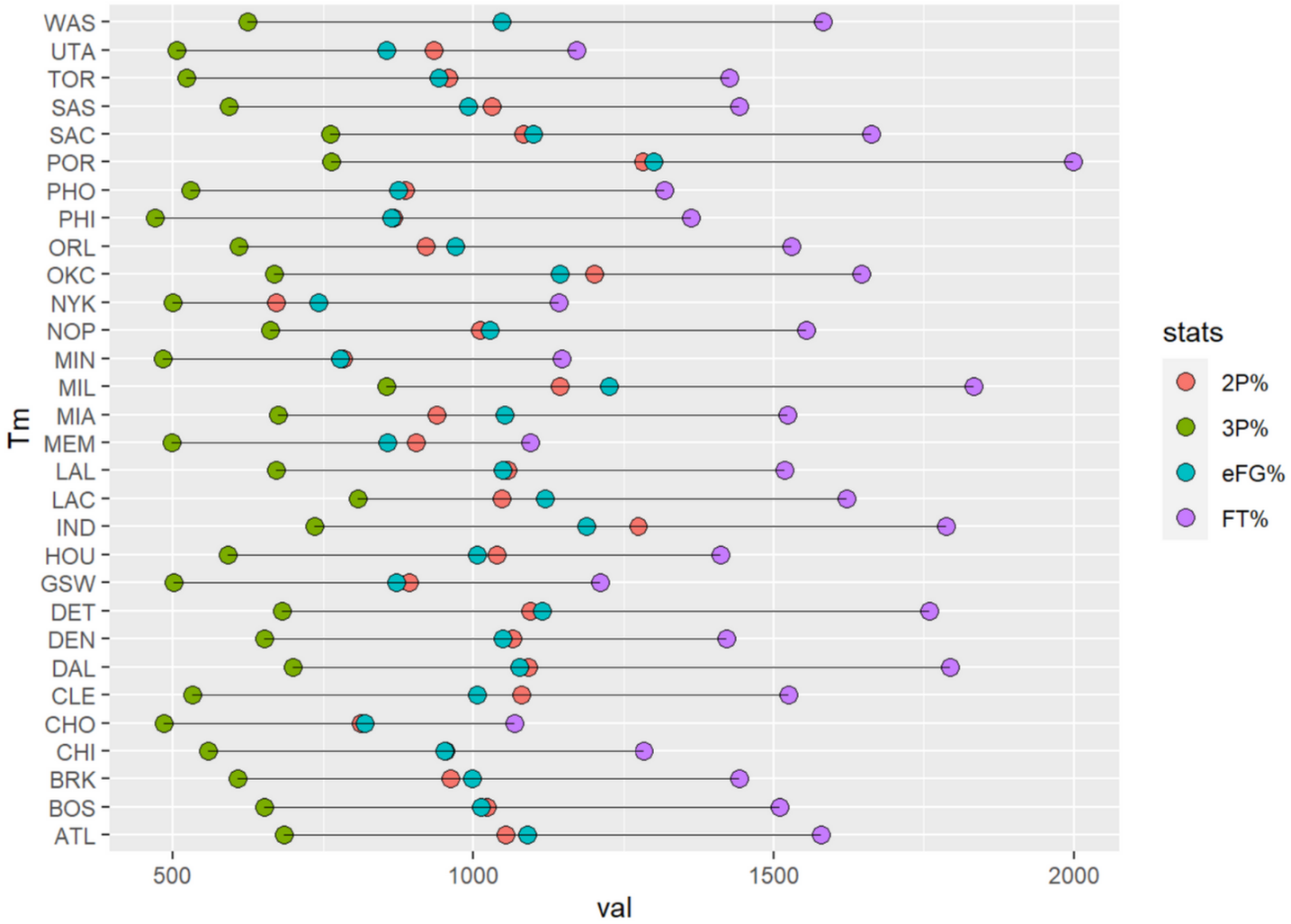
Number of FT, 3P, 2P by position

While 2 pointers and free throws are scored by each type of player, 3 pointers are mainly scored by shooting guards.



Goal percentages by team

- The highest percentage shots are free throws
- The next highest percentage statistics are field goal percentage and 2 point percentage
- 3 point percentage is the lowest



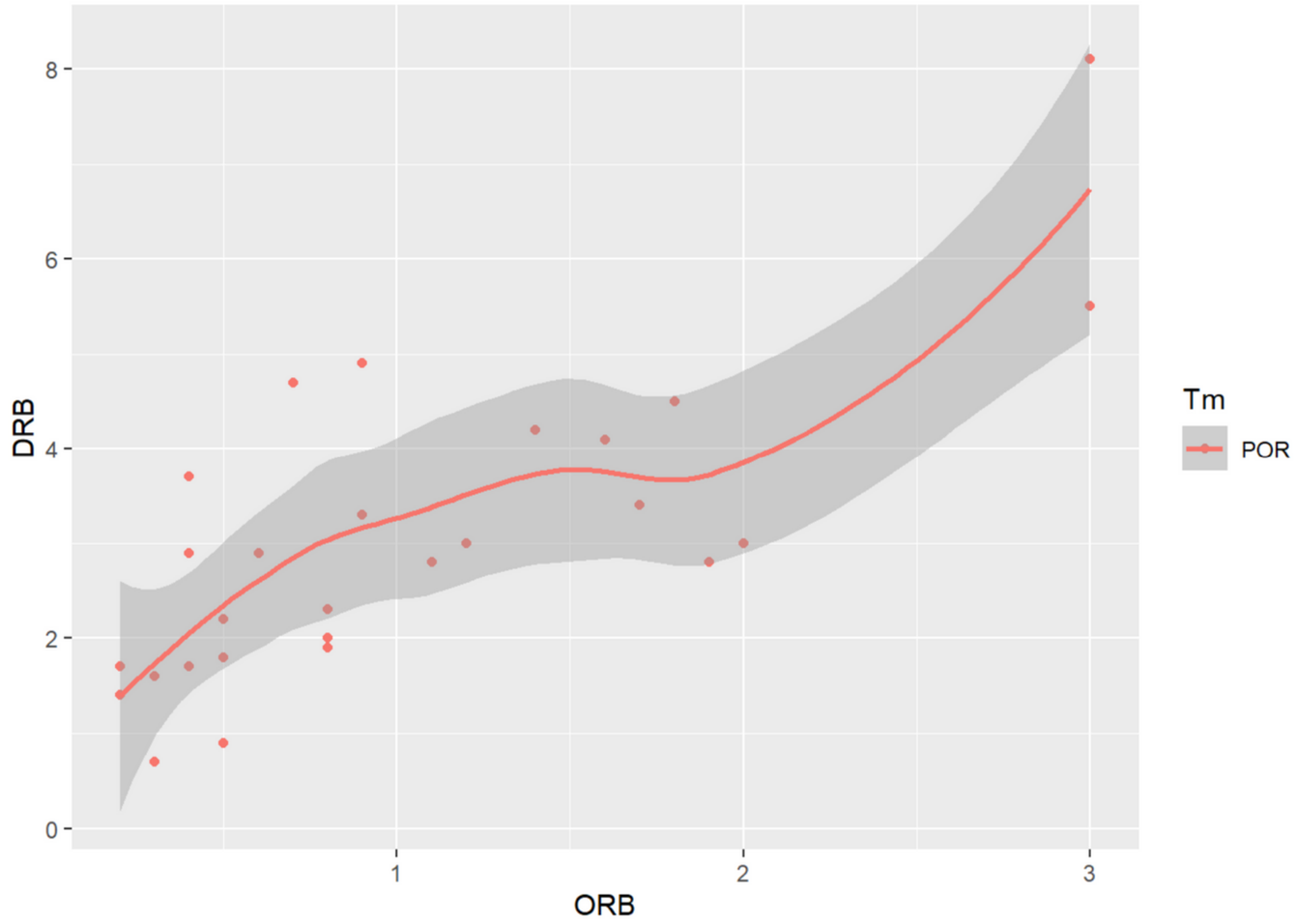
Regression

Finally,we performed a linear regression to analyze the relationship between the total number of points and the different rebounding variables and the assists variable.

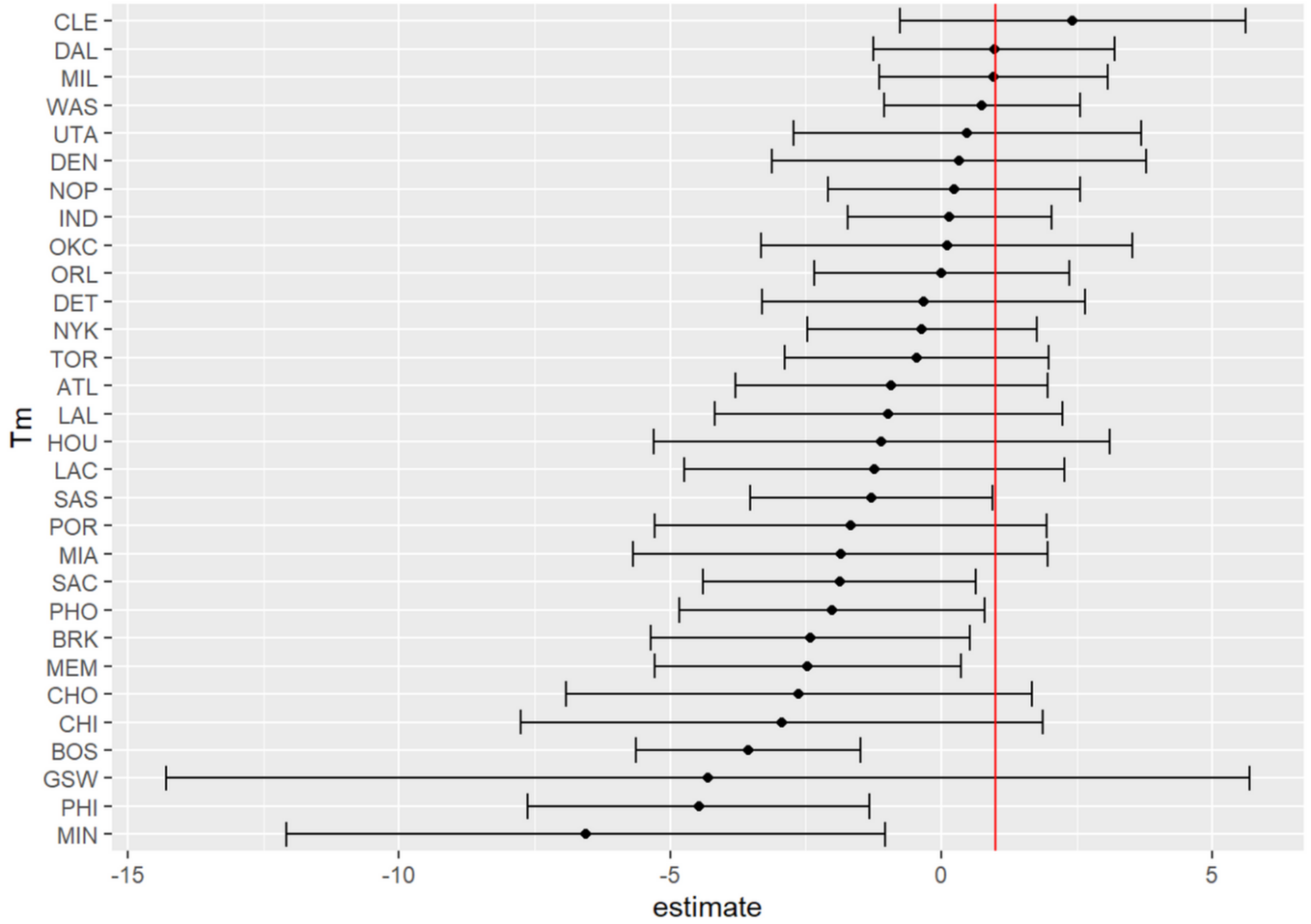
```
## Call:
## lm(formula = PTS ~ ORB + DRB + AST, data = NBA_table)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.1811  -1.8833  -0.3817   1.6670  12.6543
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.83541    0.20590   4.057 5.44e-05 ***
## ORB         -0.87666    0.22768  -3.850 0.000127 ***
## DRB          2.03947    0.11069  18.426  < 2e-16 ***
## AST          1.41052    0.08078  17.461  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.282 on 808 degrees of freedom
## Multiple R-squared:  0.7151, Adjusted R-squared:  0.7141
## F-statistic: 676.1 on 3 and 808 DF,  p-value: < 2.2e-16
```

Subgroup regression analysis

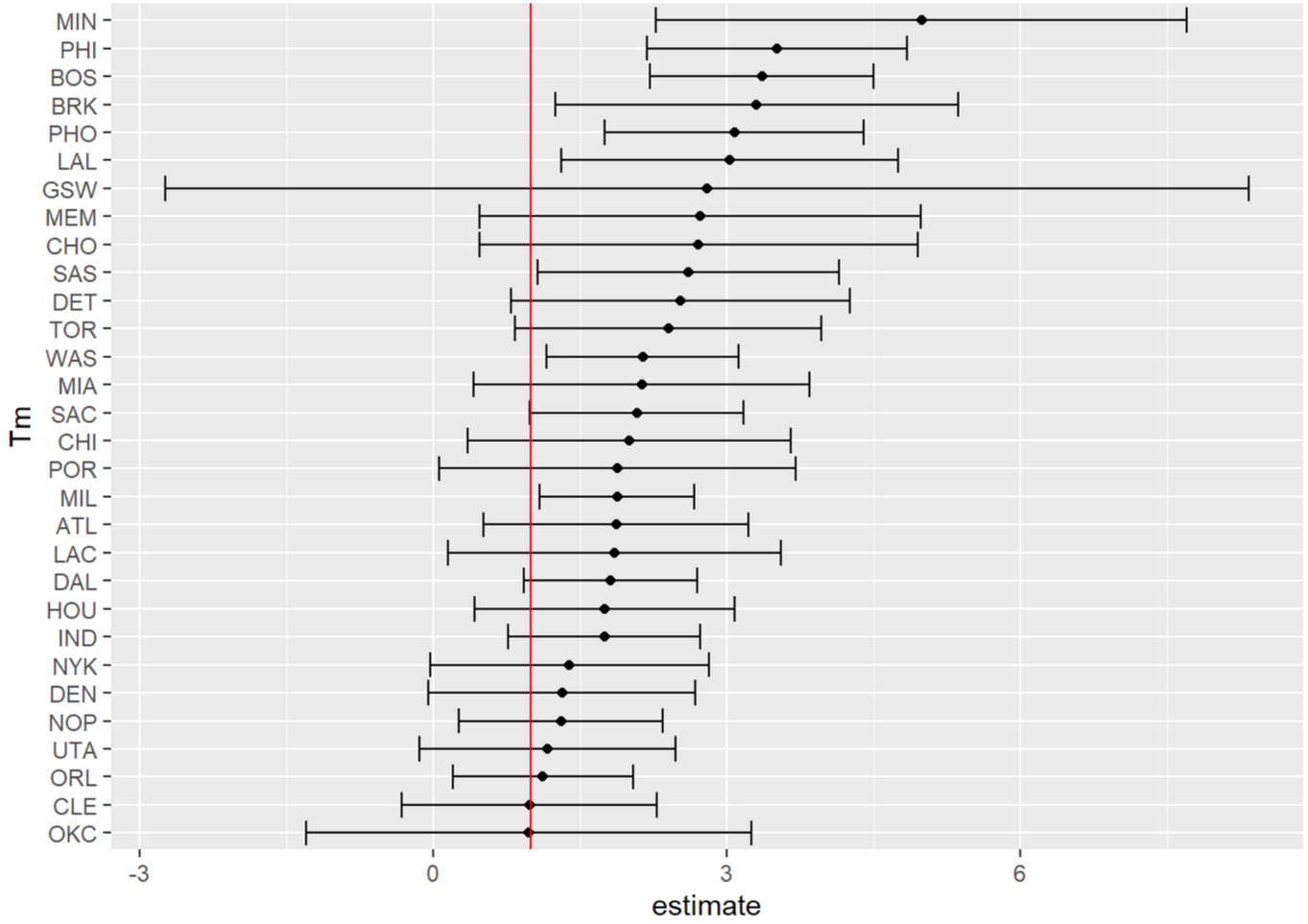
- We grouped by team and we picked portland because of its interesting output



Estimated coefficients for offensive rebounds



Estimated coefficients for defensive rebounds



Estimated coefficients for assists