Analiza utakmica Bundeslige od sezone početka 2014-15 do 01.02.2017.

Alen Murtic, Mario Slatinac

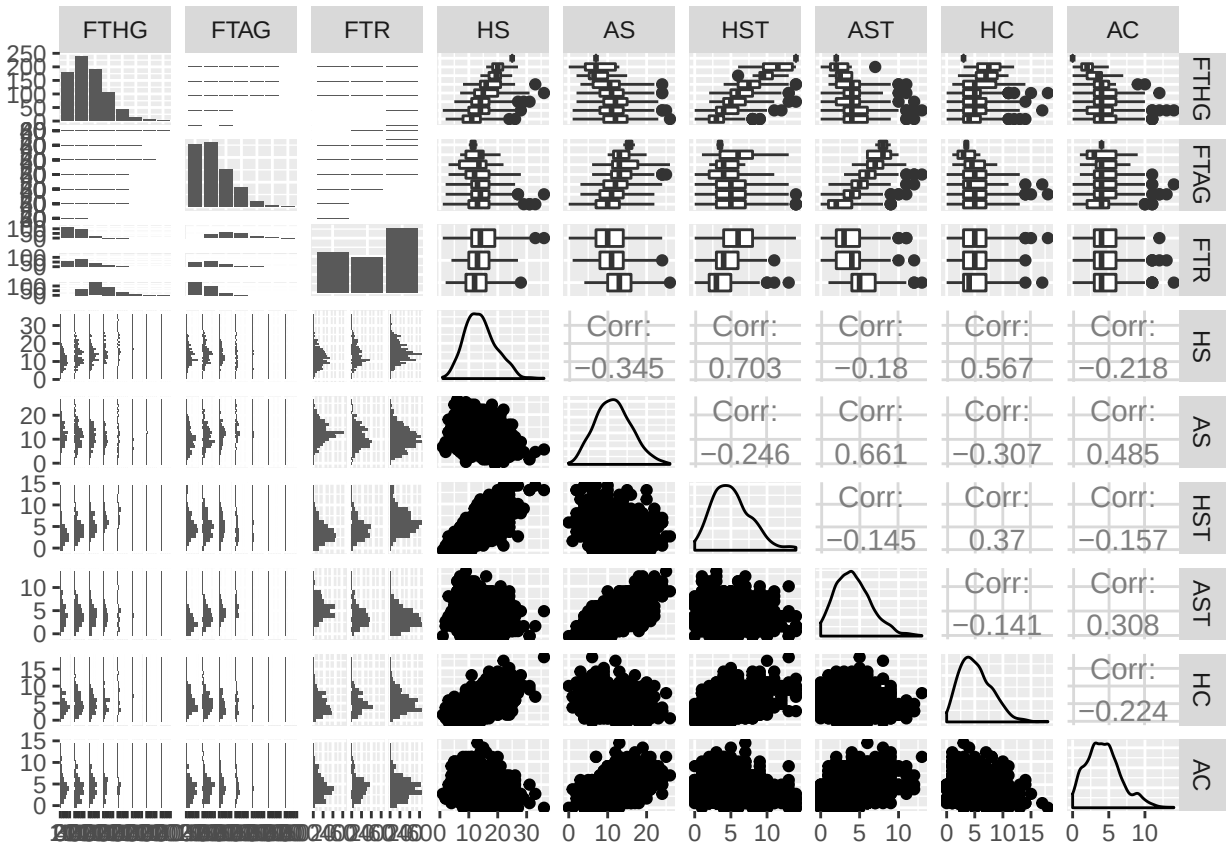
5 veljače 2017

Upoznavanje s podatkovnim skupom

Podatkovni skup preuzet je sa stranice http://www.football-data.co.uk/ te je detaljnije opisan u službenim uputama http://www.football-data.co.uk/notes.txt.

Najprije pozivamo naredbe summary, glimpse i ggpairs kako bi vidjeti osnovne podatke o varijablama.

```
##       Date                   HomeTeam            AwayTeam
##  Min.   :2014-08-22   Leverkusen   : 44   Dortmund     : 44
##  1st Qu.:2015-02-21   Schalke 04   : 44   Ein Frankfurt: 44
##  Median :2015-10-20   Werder Bremen: 44   FC Koln      : 44
##  Mean   :2015-10-16   Wolfsburg    : 44   Hamburg      : 44
##  3rd Qu.:2016-04-23   Augsburg     : 43   Hertha       : 44
##  Max.   :2017-01-29   Bayern Munich: 43   Augsburg     : 43
##                       (Other)      :512   (Other)      :511
##       FTHG     FTAG    FTR           HS              AS
##  1      :239   0:251   A:225   Min.   : 1.00   Min.   : 0.00
##  2      :189   1:262   D:193   1st Qu.:10.00   1st Qu.: 8.00
##  0      :178   2:153   H:356   Median :14.00   Median :11.00
##  3      :105   3: 77           Mean   :14.05   Mean   :11.52
##  4      : 42   4: 22           3rd Qu.:17.00   3rd Qu.:15.00
##  5      : 14   5:  7           Max.   :36.00   Max.   :26.00
##  (Other):  7   6:  2
##       HST              AST               HC               AC
##  Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 3.00   1st Qu.: 2.000   1st Qu.: 3.000   1st Qu.: 3.000
##  Median : 5.00   Median : 4.000   Median : 5.000   Median : 4.000
##  Mean   : 5.12   Mean   : 4.173   Mean   : 5.195   Mean   : 4.297
##  3rd Qu.: 7.00   3rd Qu.: 6.000   3rd Qu.: 7.000   3rd Qu.: 6.000
##  Max.   :14.00   Max.   :13.000   Max.   :18.000   Max.   :14.000
##
```



Zatim ćemo provjeriti koliko je različitih klubova igralo u Bundesligi u vremenskom periodu navedene tri sezone. Maksimalan mogući broj bio bi $18 + 3 + 3 = 24$, jer svake godine iz lige mogu ispasti tri kluba i biti zamijenjeni novim trima klubovima.

```
## U ligi je u navedene tri sezone nastupilo  21  različitih klubova. Oni su:
```

```
##  [1] "Bayern Munich" "Dortmund"      "Ein Frankfurt" "FC Koln"
##  [5] "Hannover"      "Hertha"        "Hoffenheim"    "M'gladbach"
##  [9] "Paderborn"     "Augsburg"      "Hamburg"       "Leverkusen"
## [13] "Schalke 04"    "Stuttgart"     "Werder Bremen" "Wolfsburg"
## [17] "Freiburg"      "Mainz"         "Darmstadt"     "Ingolstadt"
## [21] "RB Leipzig"
```

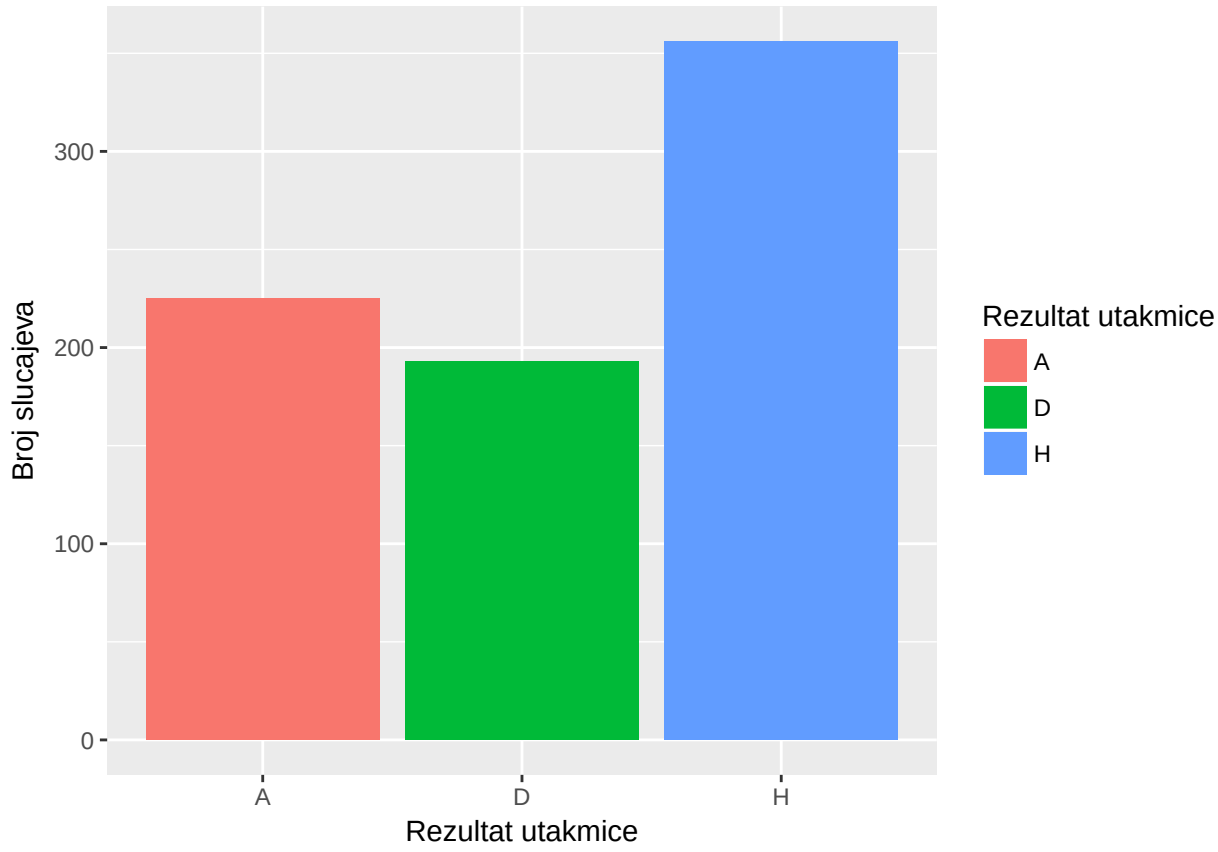
Analiza osnovnih podataka

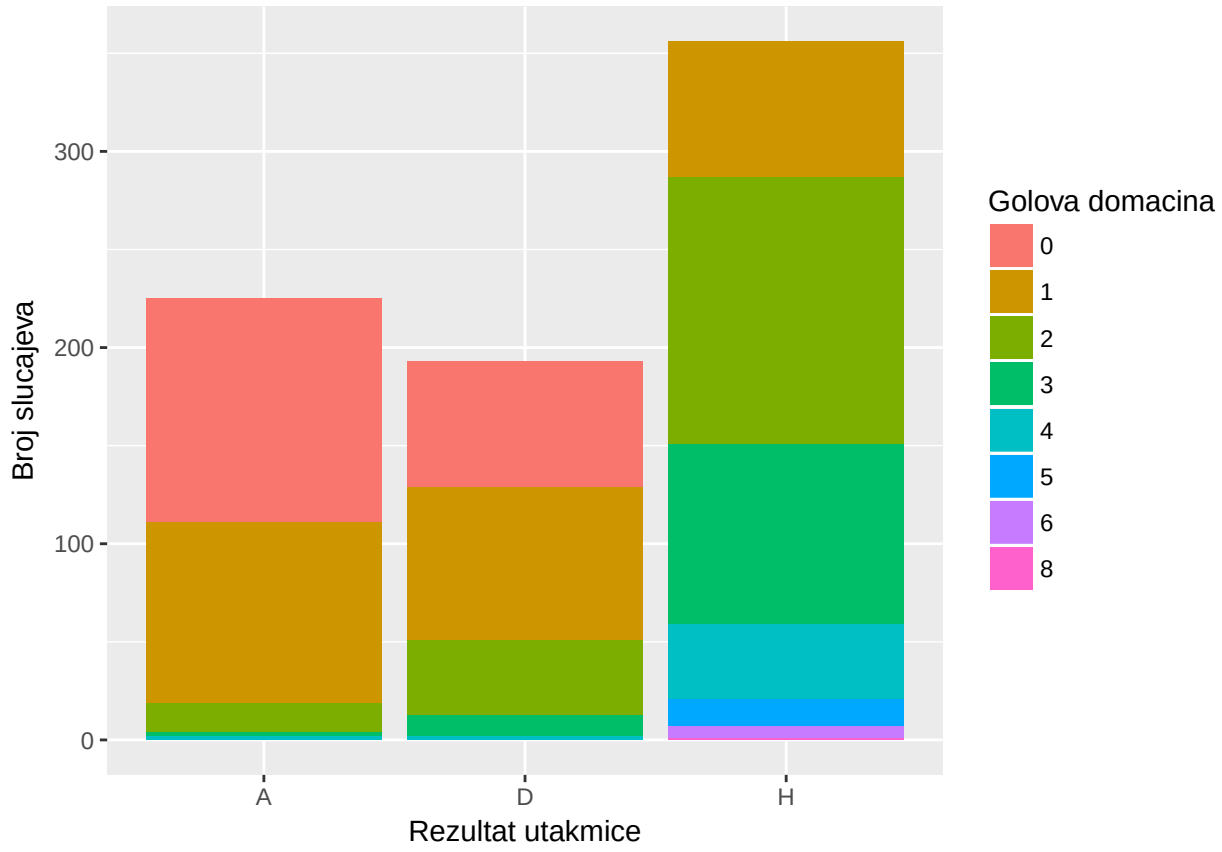
Analiza broja golova i rezultata

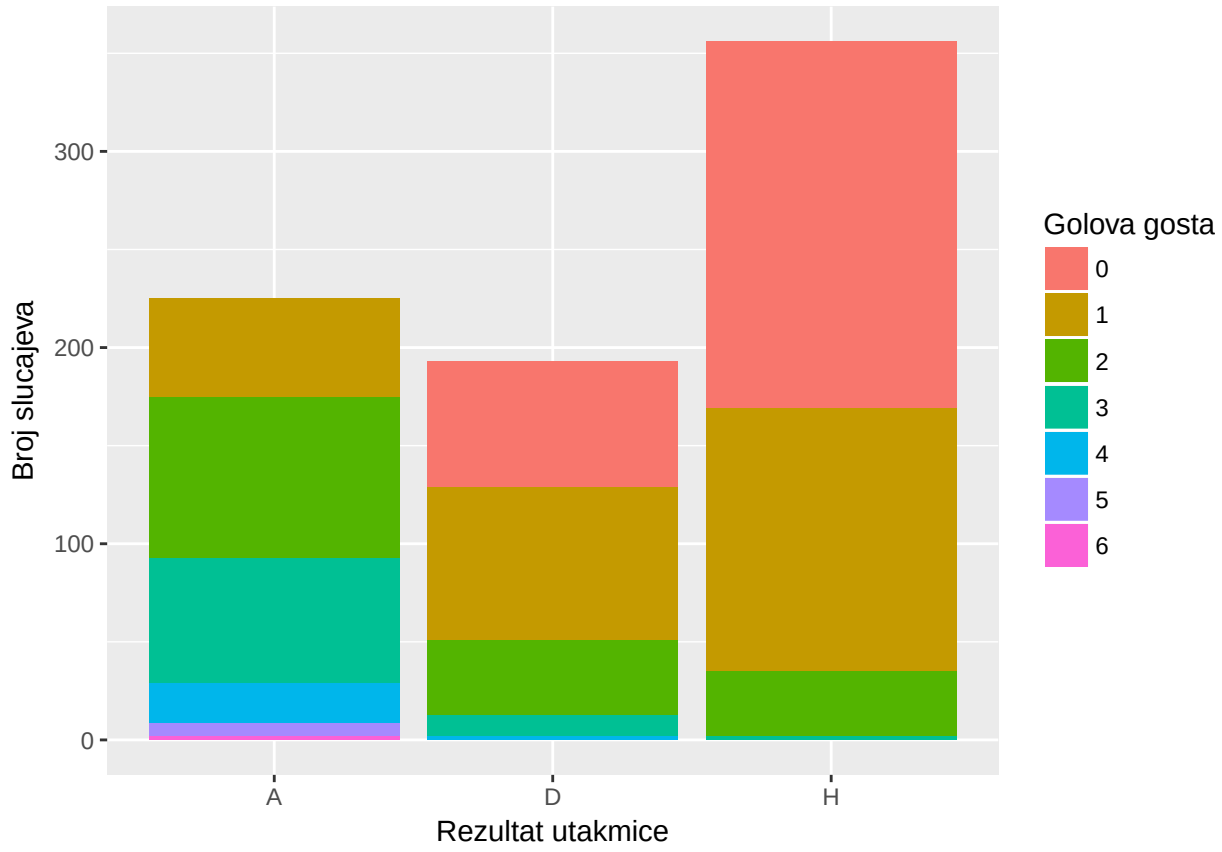
U sljedećih nekoliko grafova bit će prikazan broj golova domaćina/gosta u odnosu na ishod utakmice. Dobiveni grafovi imaju smisla budući da je postoji prednost domaćeg terena, a najmanje je vjerojatno da obje momčadi zabiju jednake broj golova.

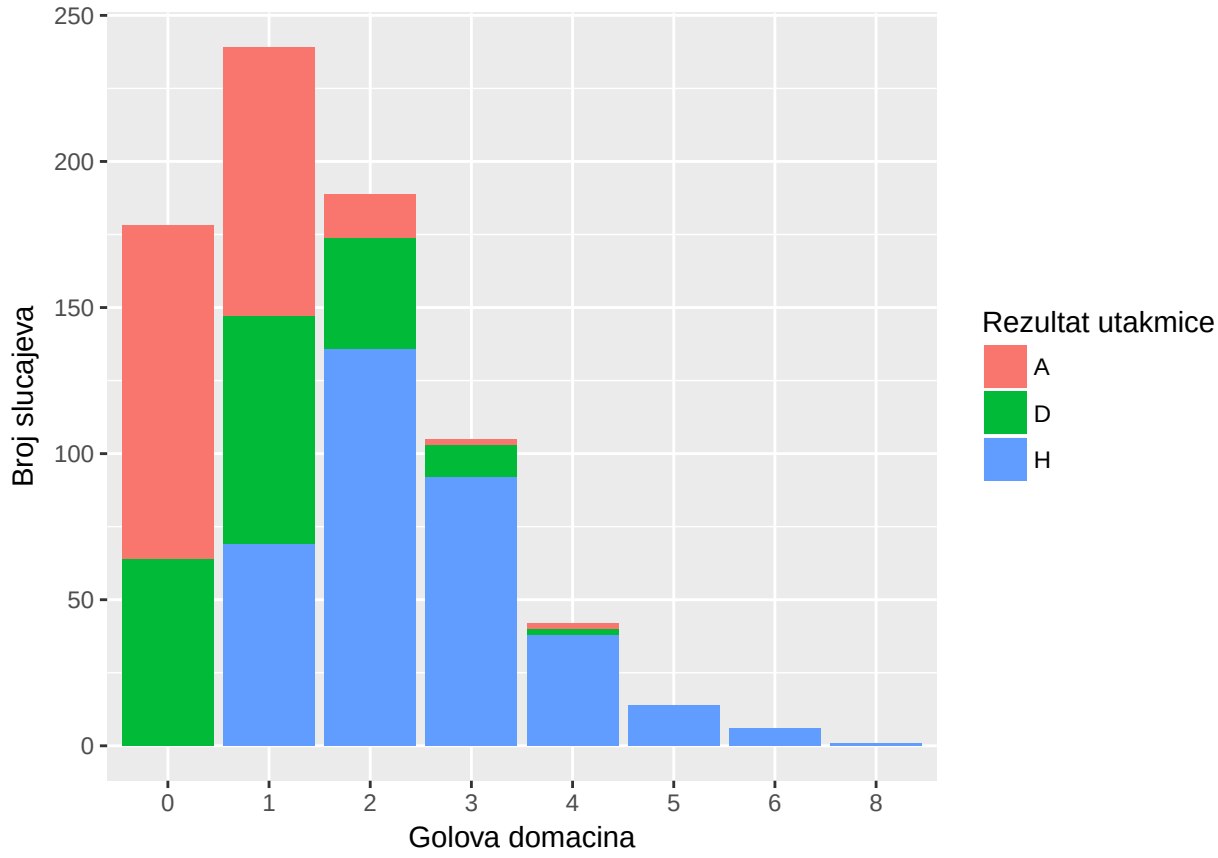
A predstavlja pobjedu gosta (away), D neodlučen rezultat (draw), a H pobjedu domaćina (home).

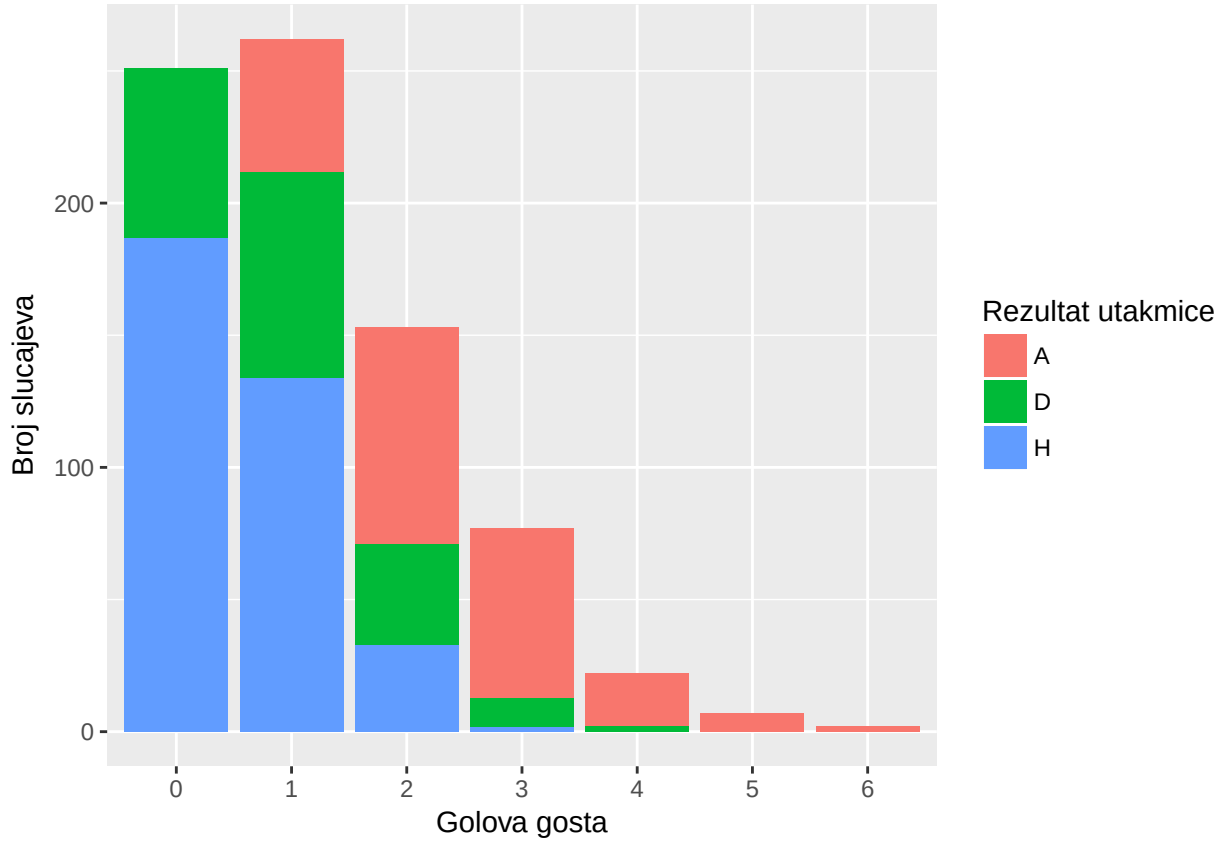
Prilično očita informacija na grafovima je da bilo koja, domaća ili gostujuća, momčad mora zabiti više od 1 gola da joj najvjerojatniji ishod ne bi bio poraz. I to semantički ima smisla jer je Bundesliga jedna od najatraktivnijih europskih liga po broj golova.


300
200
100
0
Broj slucajeva
A
D
H
Rezultat utakmice
Rezultat utakmice
A
D
H


Broj slucajeva
300
200
100
0
A
D
H
Rezultat utakmice
Golova domacina
0
1
2
3
4
5
6
8


300
200
100
0
Broj slucajeva
A
D
H
Rezultat utakmice
Golova gosta
0
1
2
3
4
5
6


250
200
150
100
50
0
Broj slucajeva
0
1
2
3
4
5
6
8
Golova domacina
Rezultat utakmice
A
D
H

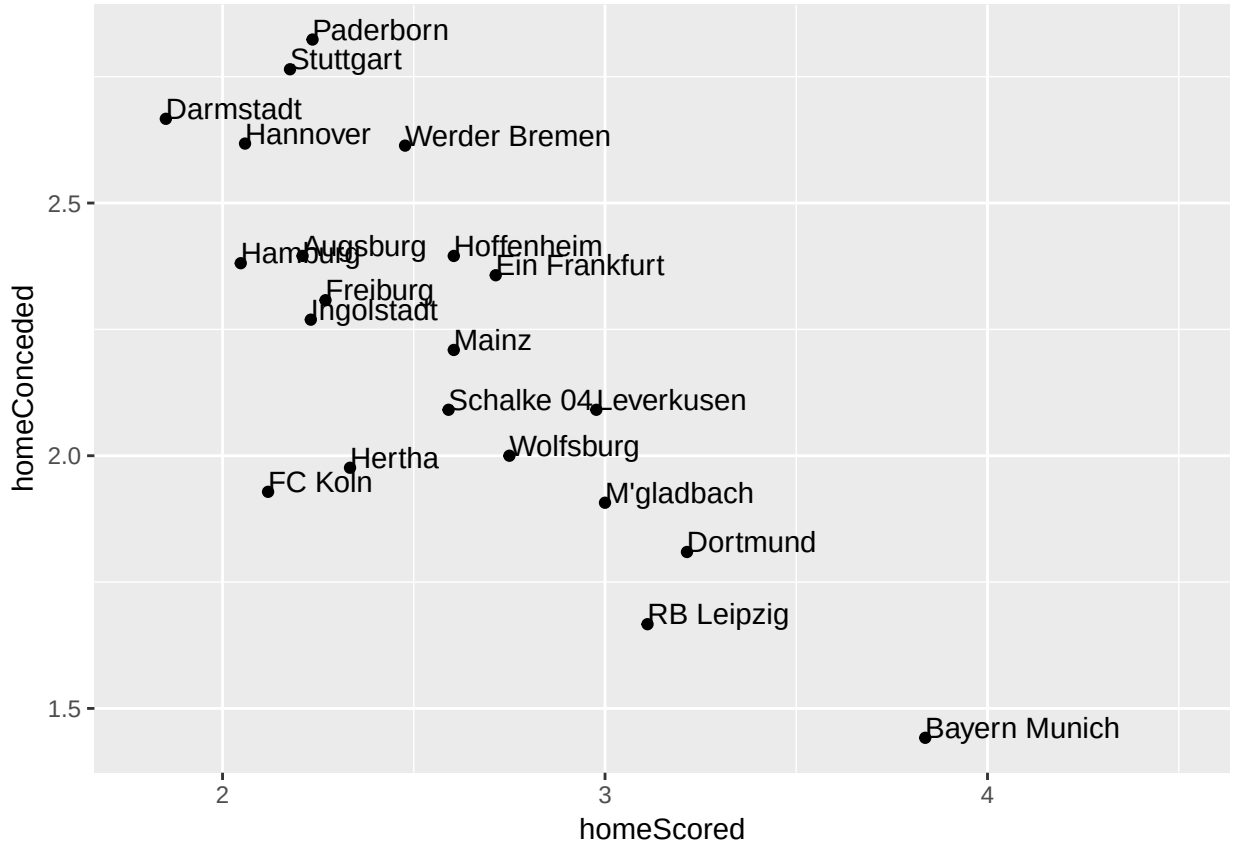


Nakon čega dobivamo konkretne brojke pojavljivanja pojedinog ishoda utakmice:

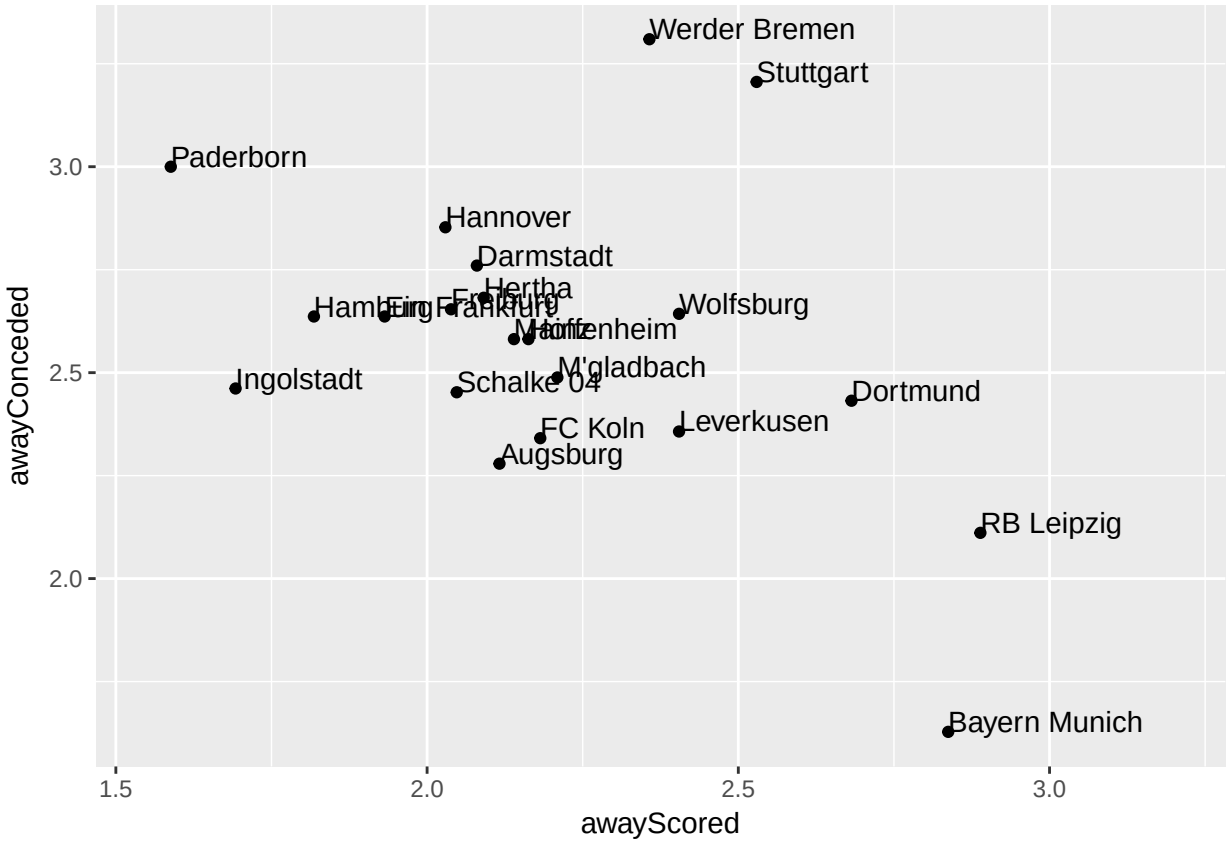
```
select(bl, FTR) %>% group_by(FTR) %>% summarize(Count = n()) -> results
results
```

```
## # A tibble: 3 × 2
##      FTR Count
##   <fctr> <int>
## 1      A   225
## 2      D   193
## 3      H   356
```

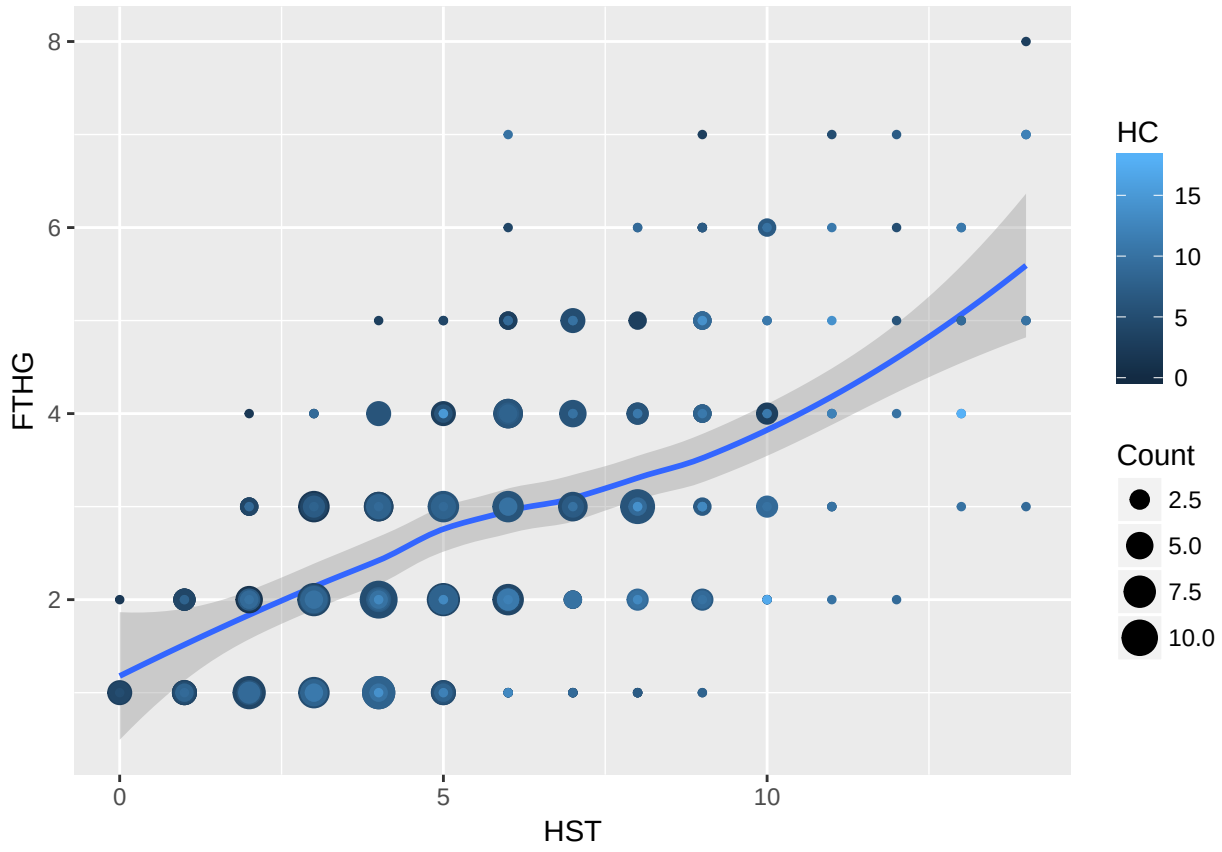
```
bl$FTHG <- as.numeric(bl$FTHG)
bl$FTAG <- as.numeric(bl$FTAG)

group_by(bl, HomeTeam) %>% summarize(homeScored = mean(FTHG), homeConceded = mean(FTAG), hgames = n(), 

ggplot(homeTeamGoals, aes(x = homeScored, y = homeConceded)) + geom_text(aes(label = HomeTeam),hjust=0,
```



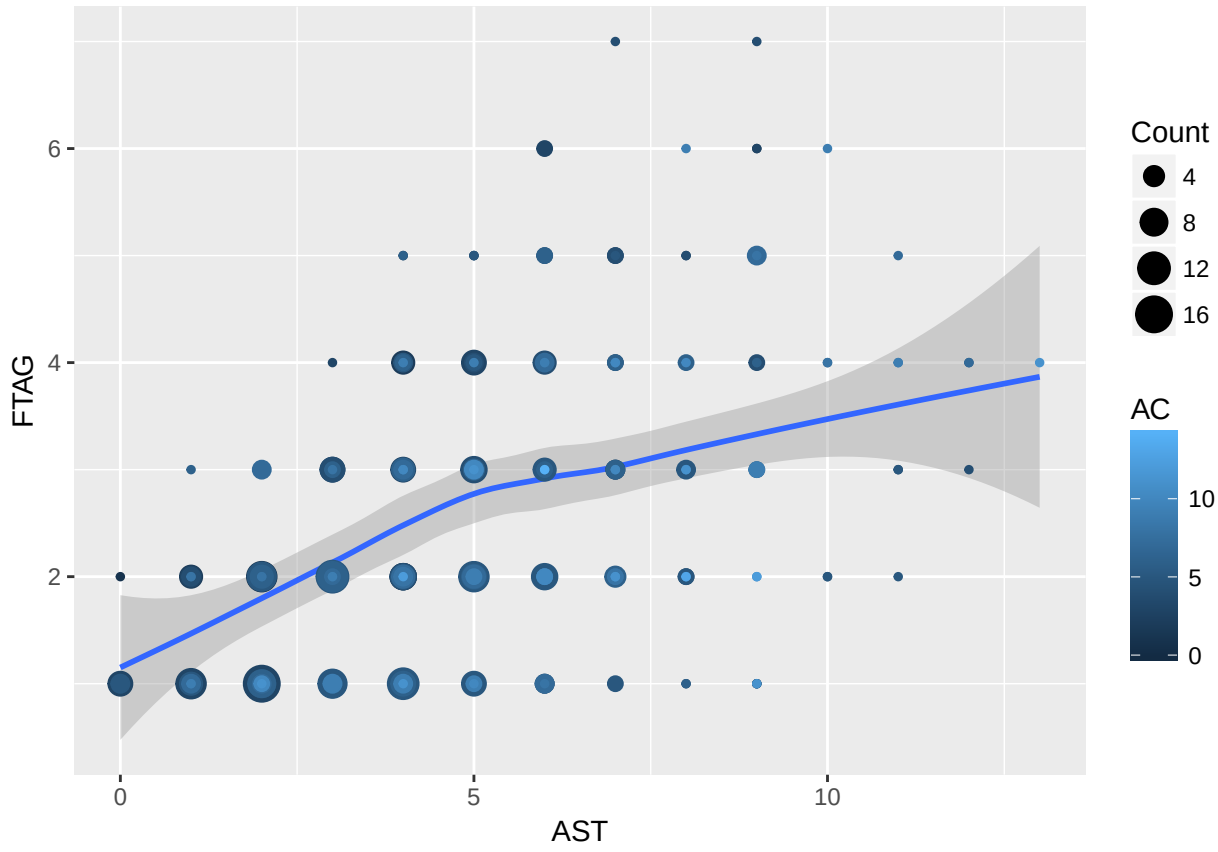

```
group_by(bl, AwayTeam) %>% summarize(awayScored = mean(FTAG), awayConceded = mean(FTHG), agames = n(), a

ggplot(awayTeamGoals, aes(x = awayScored, y = awayConceded)) + geom_text(aes(label = AwayTeam),hjust=0,
```



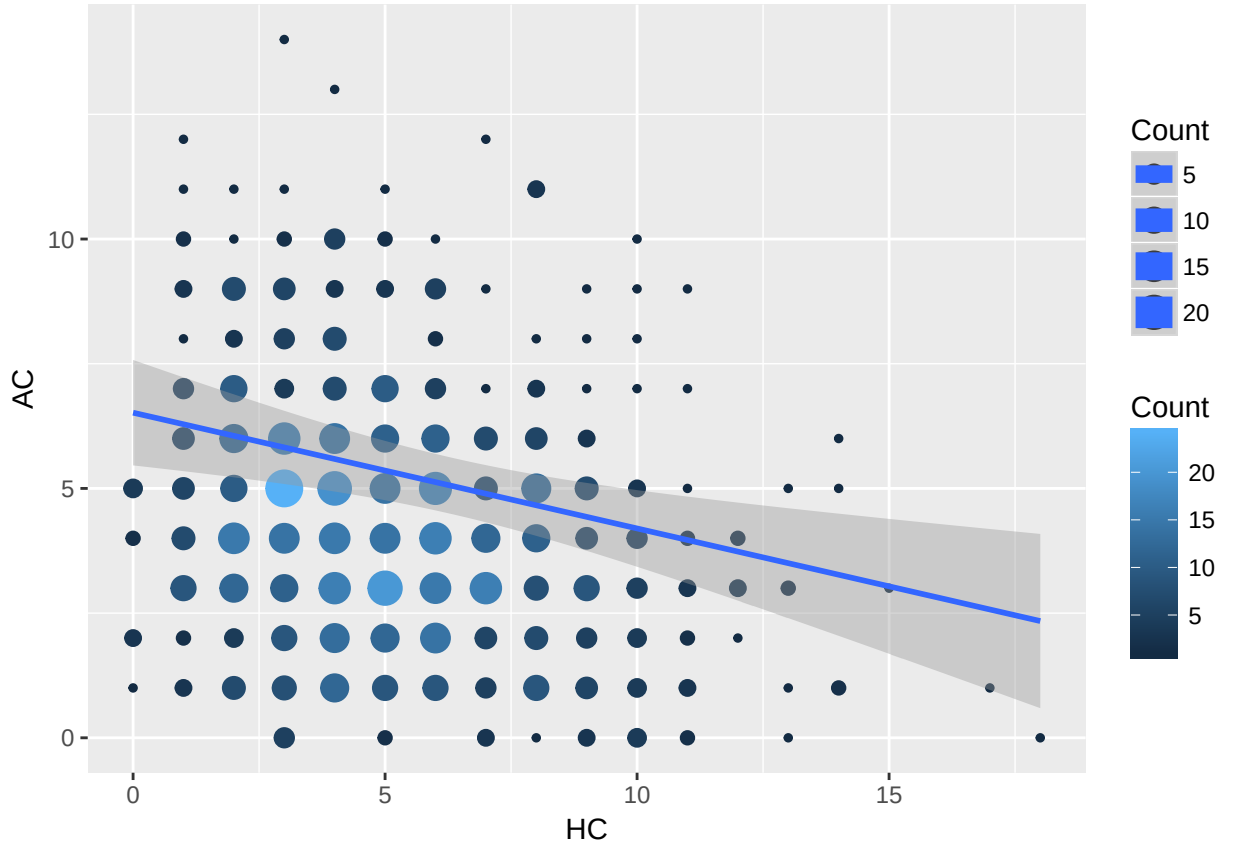

```
group_by(bl, HST, FTHG, HC) %>% summarise(Count = n()) -> temp
ggplot(temp, aes(x = HST, y = FTHG)) + geom_smooth(method = loess) + geom_point(aes(size = Count, color
```



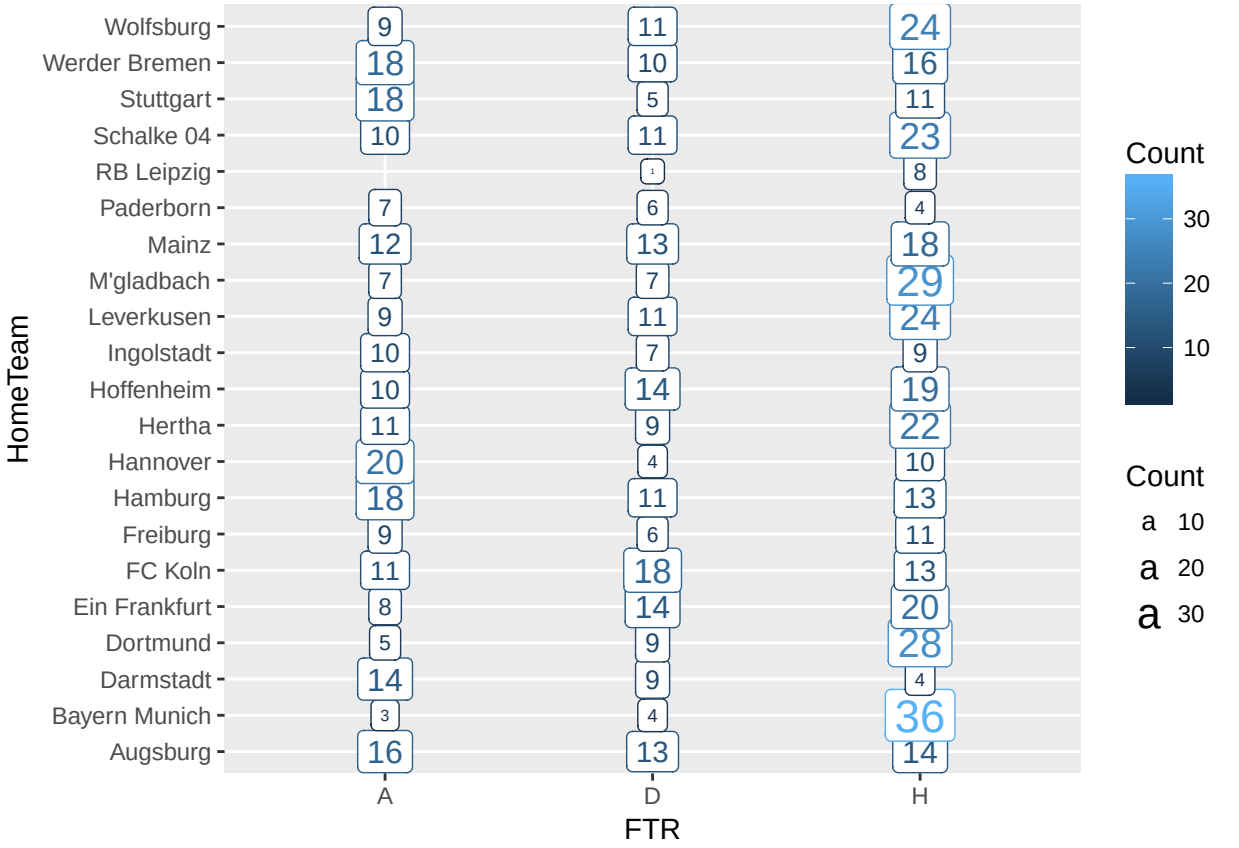

```
group_by(bl, AST, FTAG, AC) %>% summarise(Count = n()) -> temp
ggplot(temp, aes(x = AST, y = FTAG)) + geom_smooth(method = loess) + geom_point(aes(size = Count, color
```



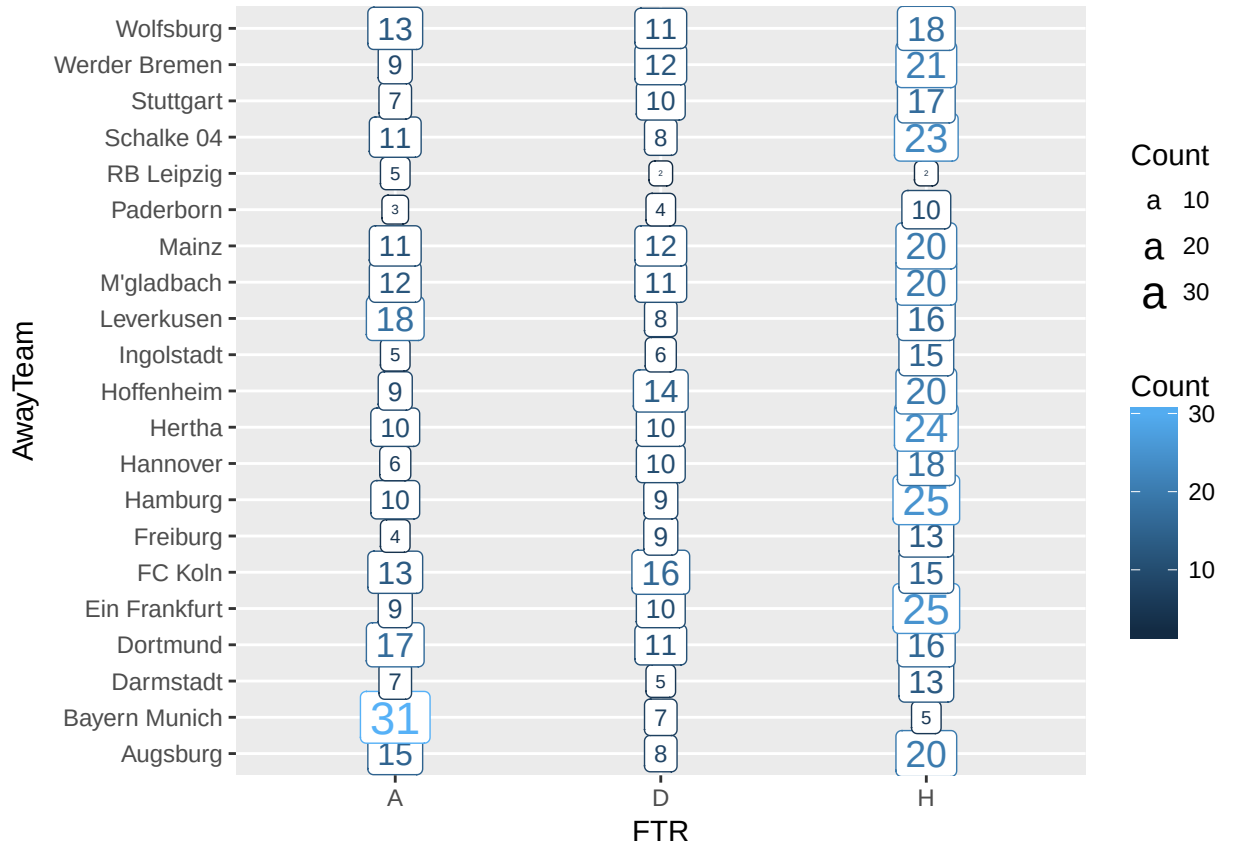

```
group_by(bl, HC, AC) %>% summarise(Count = n()) -> temp
ggplot(temp, aes(x = HC, y = AC, color = Count, size = Count)) + geom_point() + geom_smooth(method = lm
```



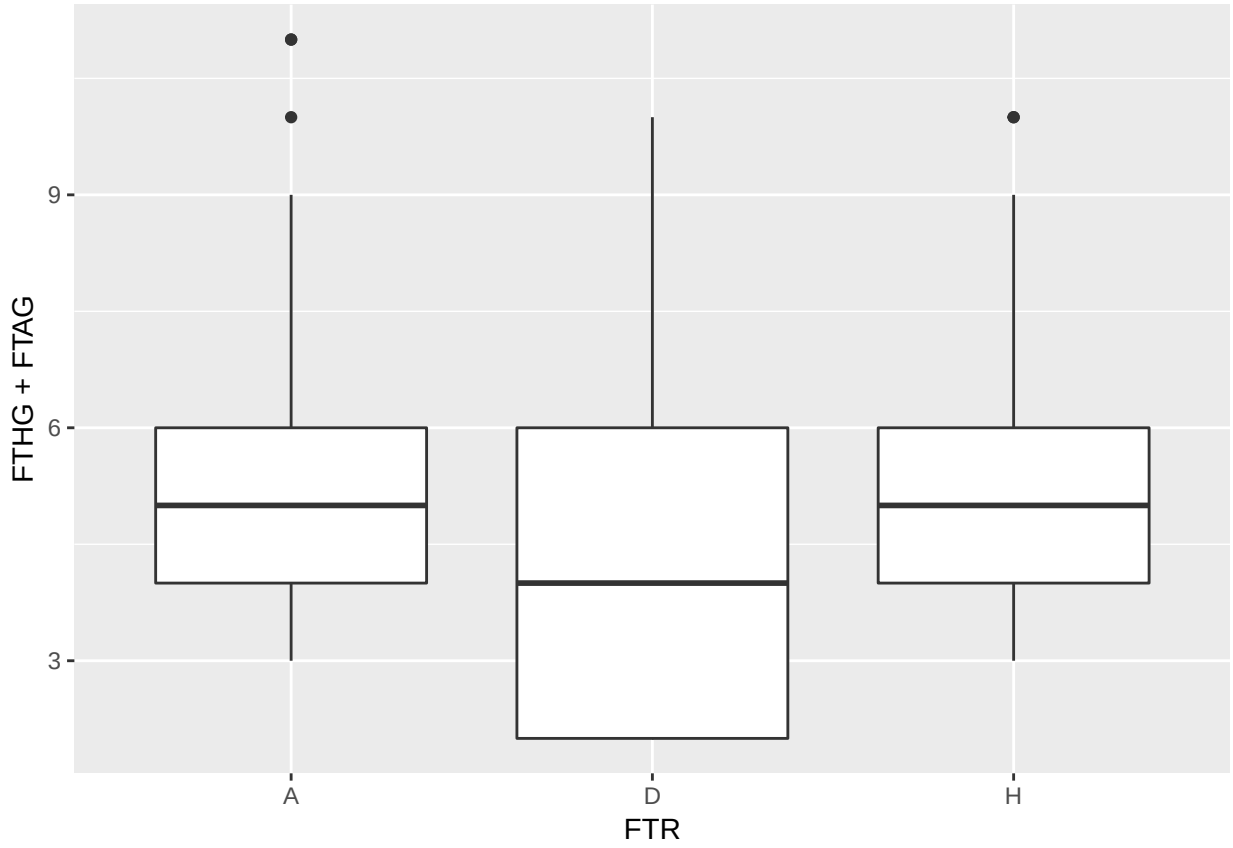

```
group_by(bl, HomeTeam, FTR) %>% summarise(Count = n()) -> homeTeamResults

ggplot(homeTeamResults, aes(x = FTR, y = HomeTeam)) + geom_label(aes(label = Count, color = Count, size
```



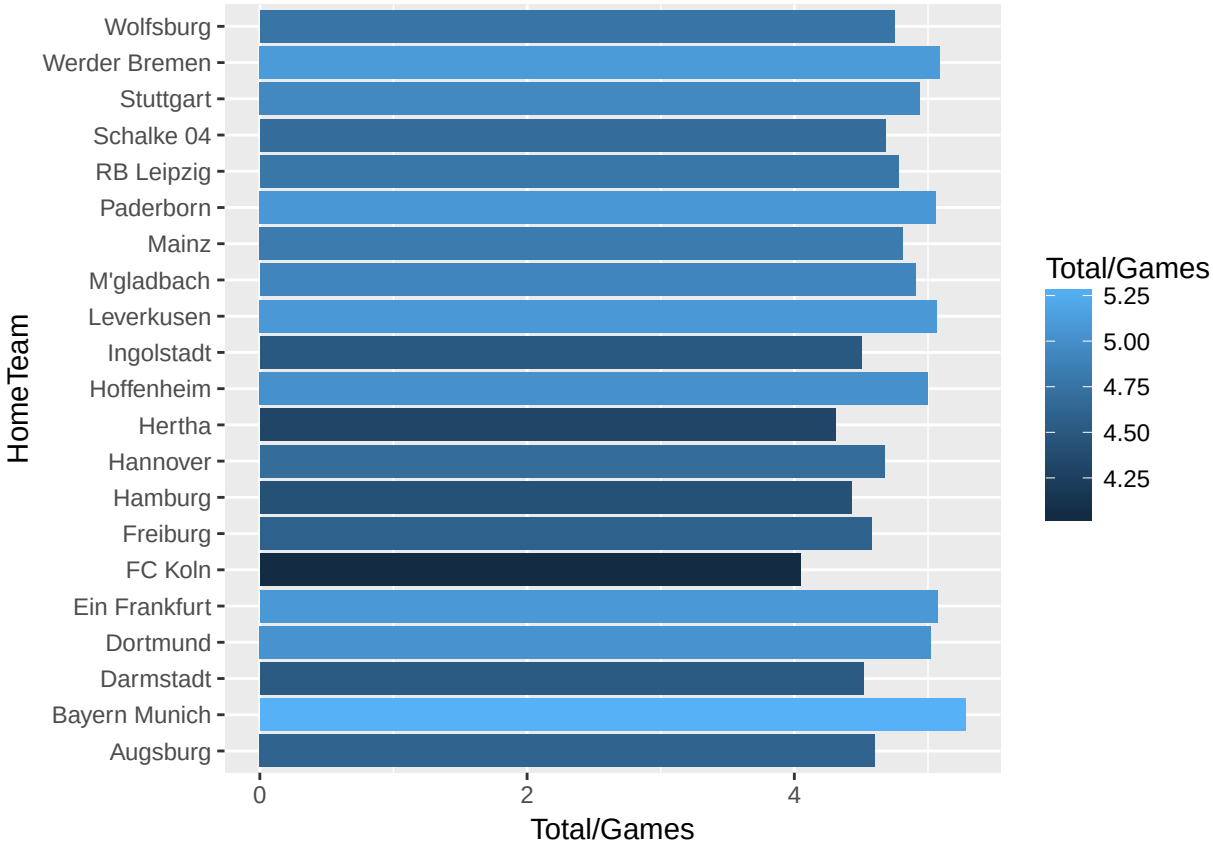

```
group_by(bl, AwayTeam, FTR) %>% summarise(Count = n()) -> awayTeamResults

ggplot(awayTeamResults, aes(x = FTR, y = AwayTeam)) + geom_label(aes(label = Count, color = Count, size
```



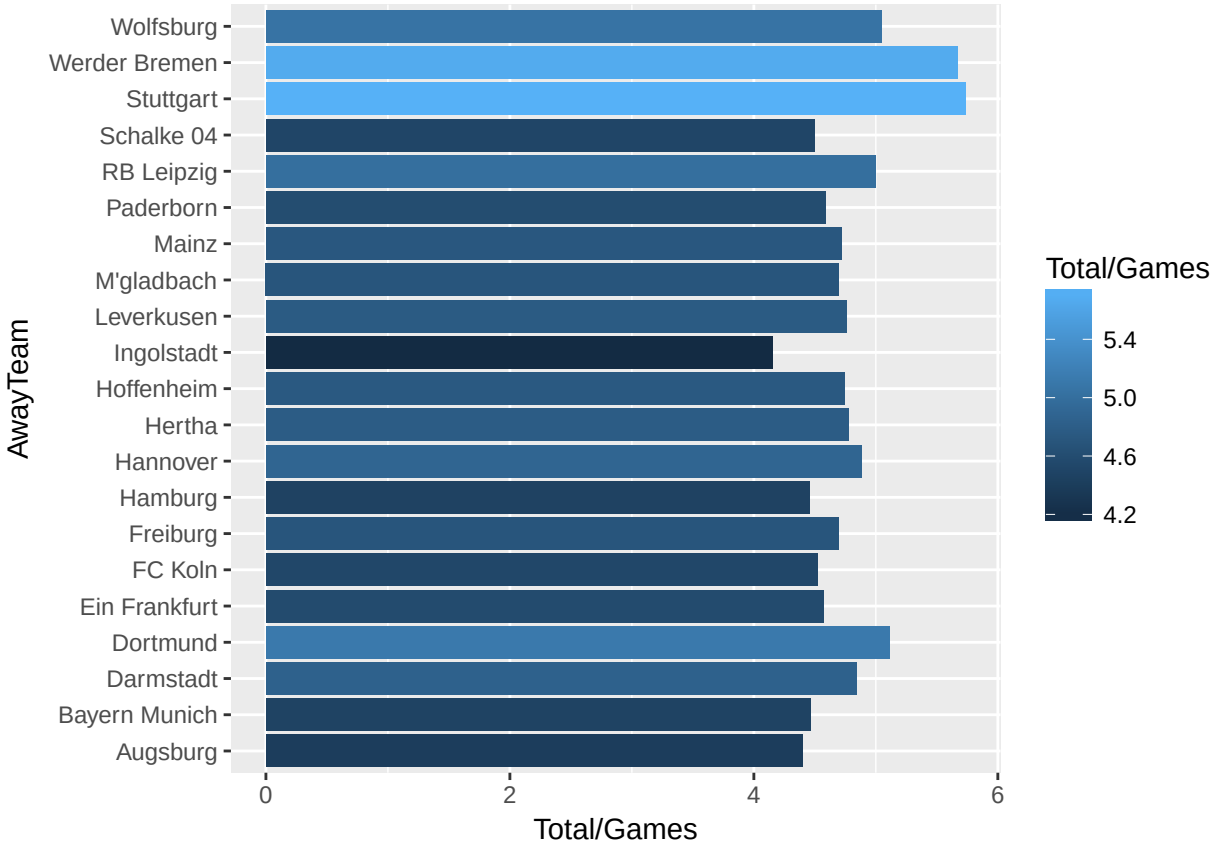

```
ggplot(bl, aes(x = FTR, y = FTHG + FTAG)) + geom_boxplot()
```




```
group_by(bl, HomeTeam) %>% summarise(Games = n(), Total = sum(FTHG + FTAG)) -> homeGoalsPerGame
ggplot(homeGoalsPerGame, aes(x = HomeTeam, y = Total/Games)) + geom_col(aes(fill = Total/Games)) + coor
```




```
group_by(bl, AwayTeam) %>% summarise(Games = n(), Total = sum(FTHG + FTAG)) -> awayGoalsPerGame
ggplot(awayGoalsPerGame, aes(x = AwayTeam, y = Total/Games)) + geom_col(aes(fill = Total/Games)) + coor
```


Wolfsburg
Werder Bremen
Stuttgart
Schalke 04
RB Leipzig
Paderborn
Mainz
M'gladbach
Leverkusen
Ingolstadt
Hoffenheim
Hertha
Hannover
Hamburg
Freiburg
FC Koln
Ein Frankfurt
Dortmund
Darmstadt
Bayern Munich
Augsburg
AwayTeam
0
2
4
6
Total/Games
Total/Games
5.4
5.0
4.6
4.2