

Analiza utakmica Bundeslige od sezone 2014-15 do 01.02.2017.

Alen Murtic, Mario Slatinac

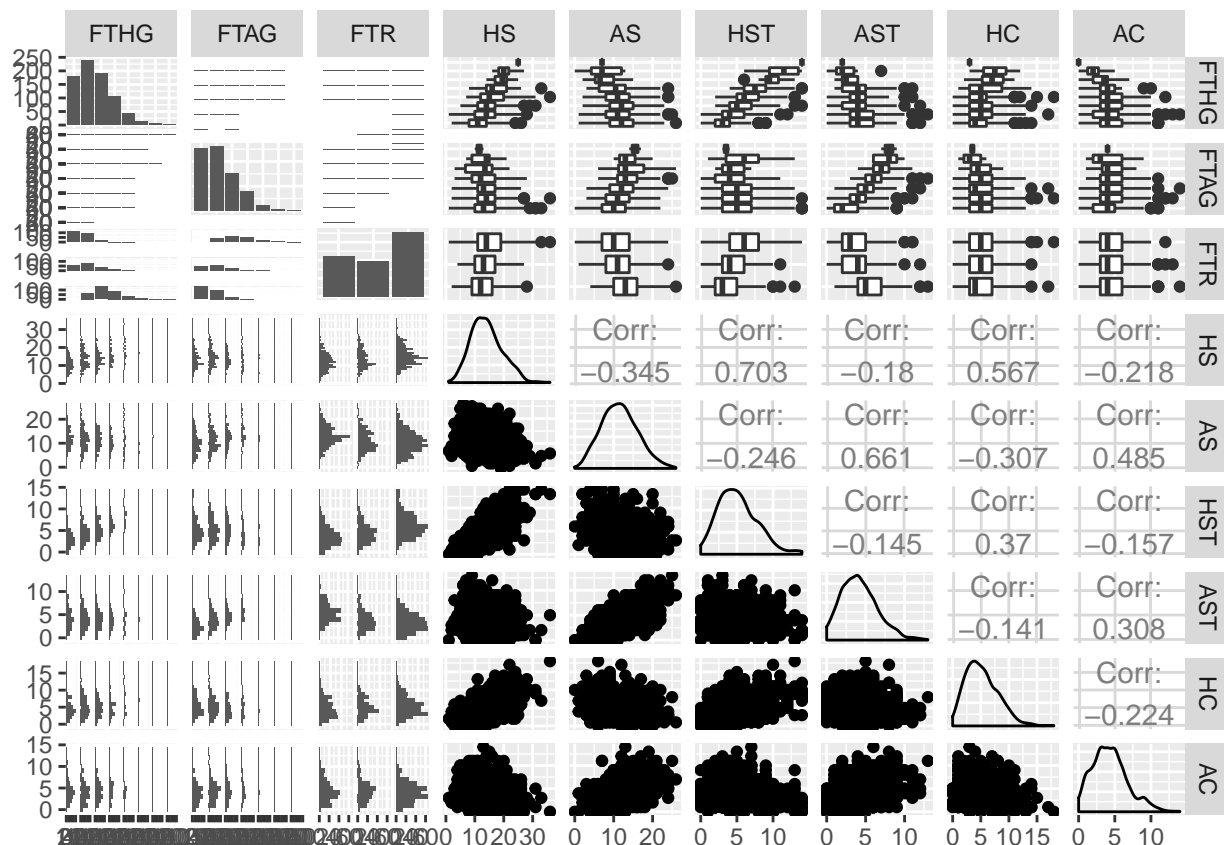
5 veljace 2017

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` 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.

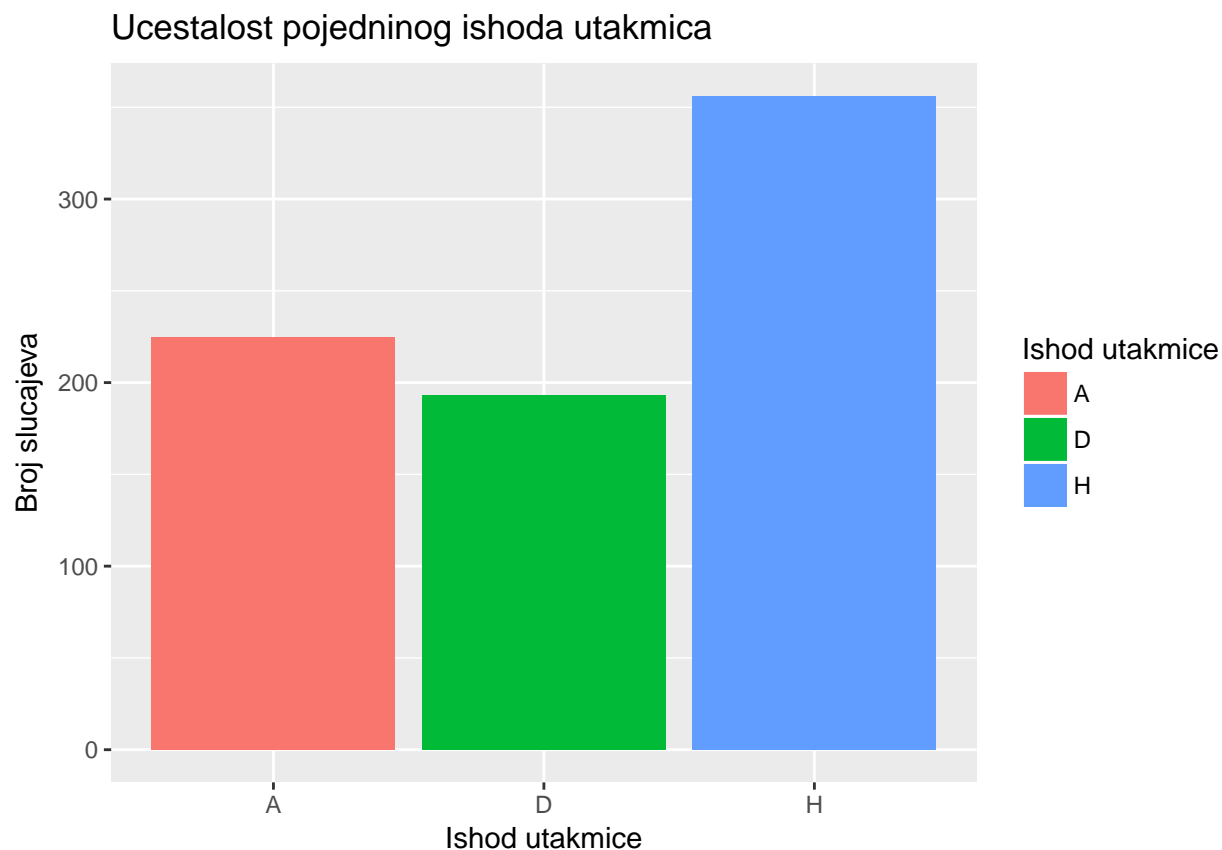
U ligi je u navedene tri sezone nastupilo 21 različitih klubova. Oni su:

```
## [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 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.

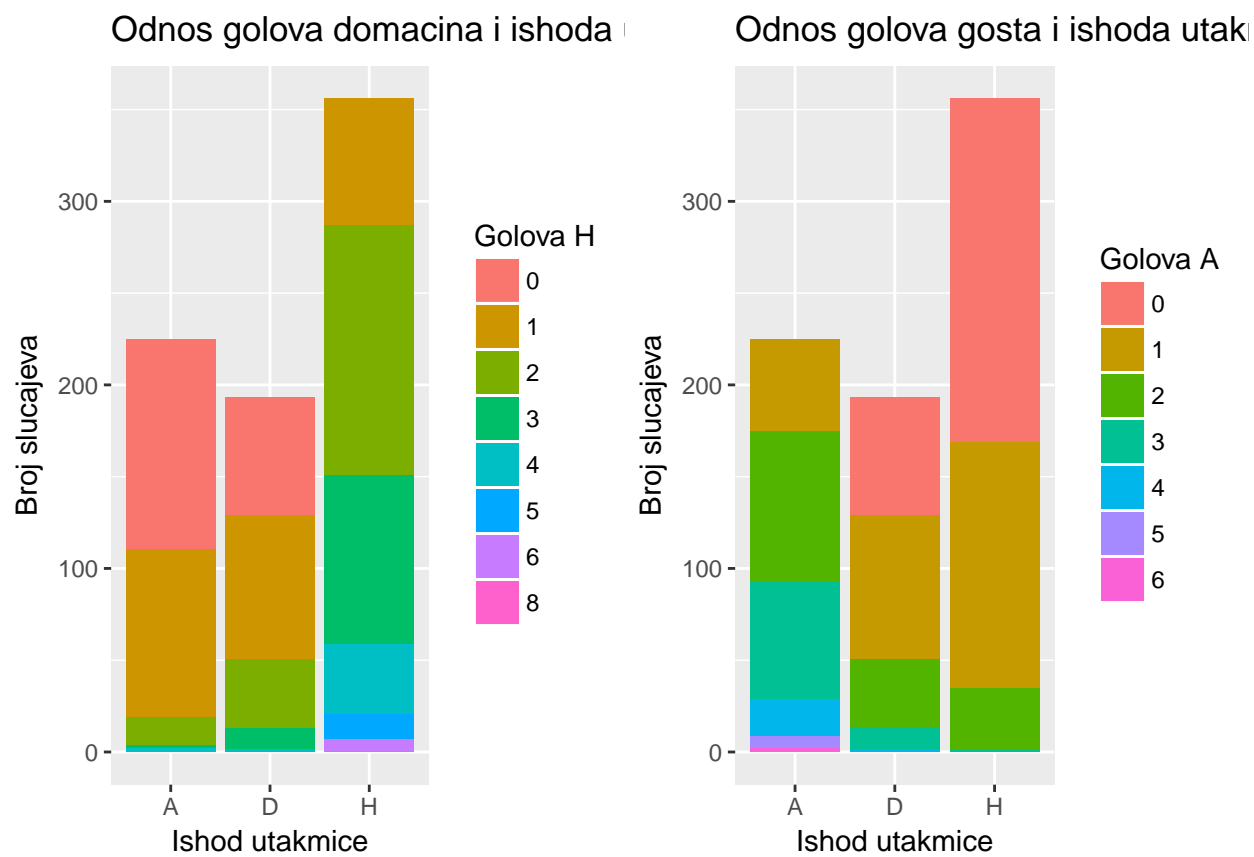
A predstavlja pobjedu gosta (away), D neodlučen rezultat (draw), a H pobjedu domaćina (home).



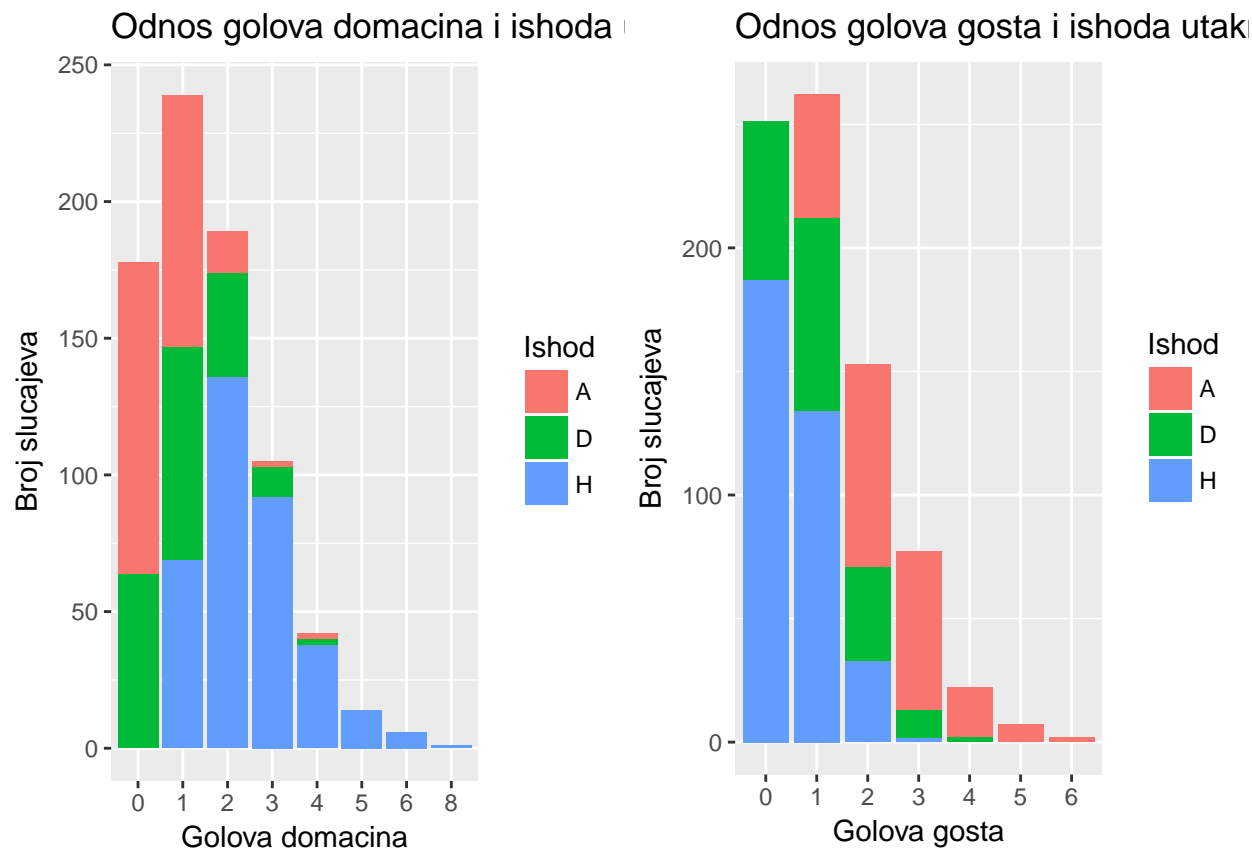
Nakon čega možemo pogledati konkretne brojke pojavljivanja pojedinog ishoda utakmice:

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

Na sljedećim grafovima vidimo broj domaćina/gosta za pojedini ishod utakmice:



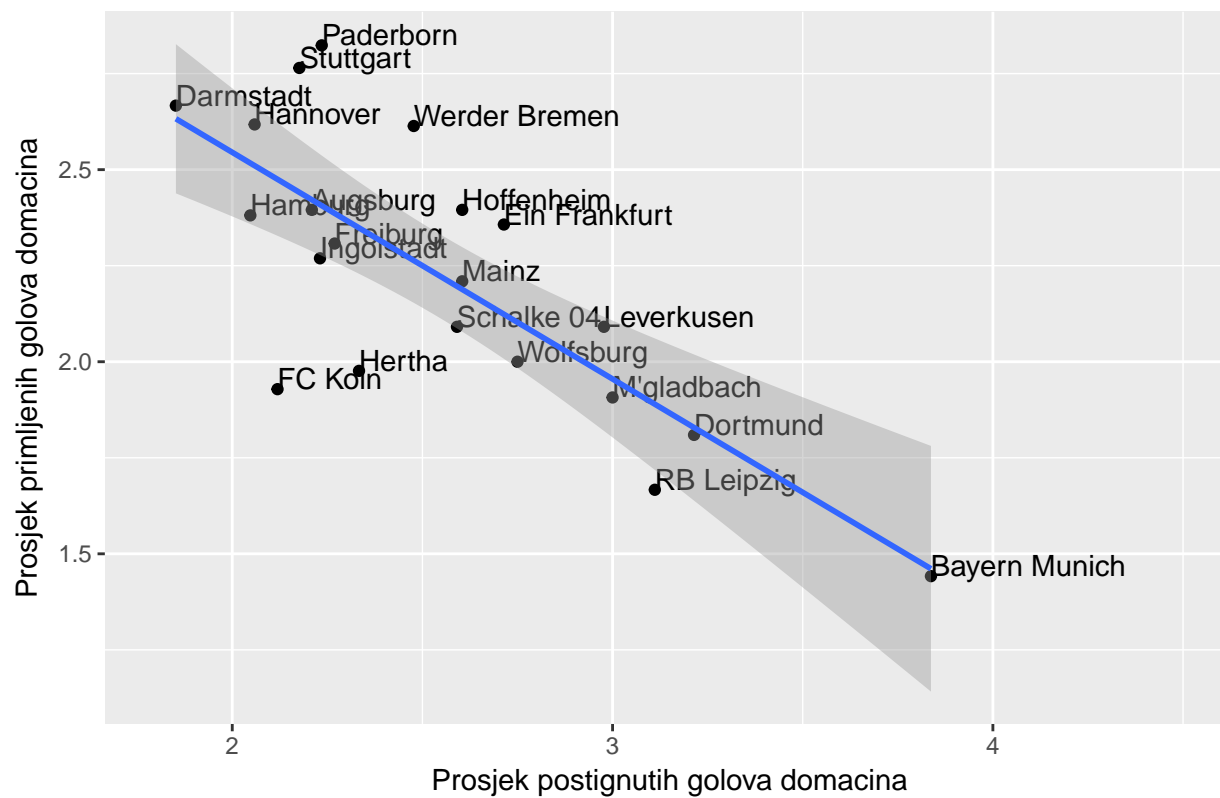
Iste podatke možemo pogledati na drukčiji način, tako da za pojedini broj golova domaćina ili gosta promatramo ishode:



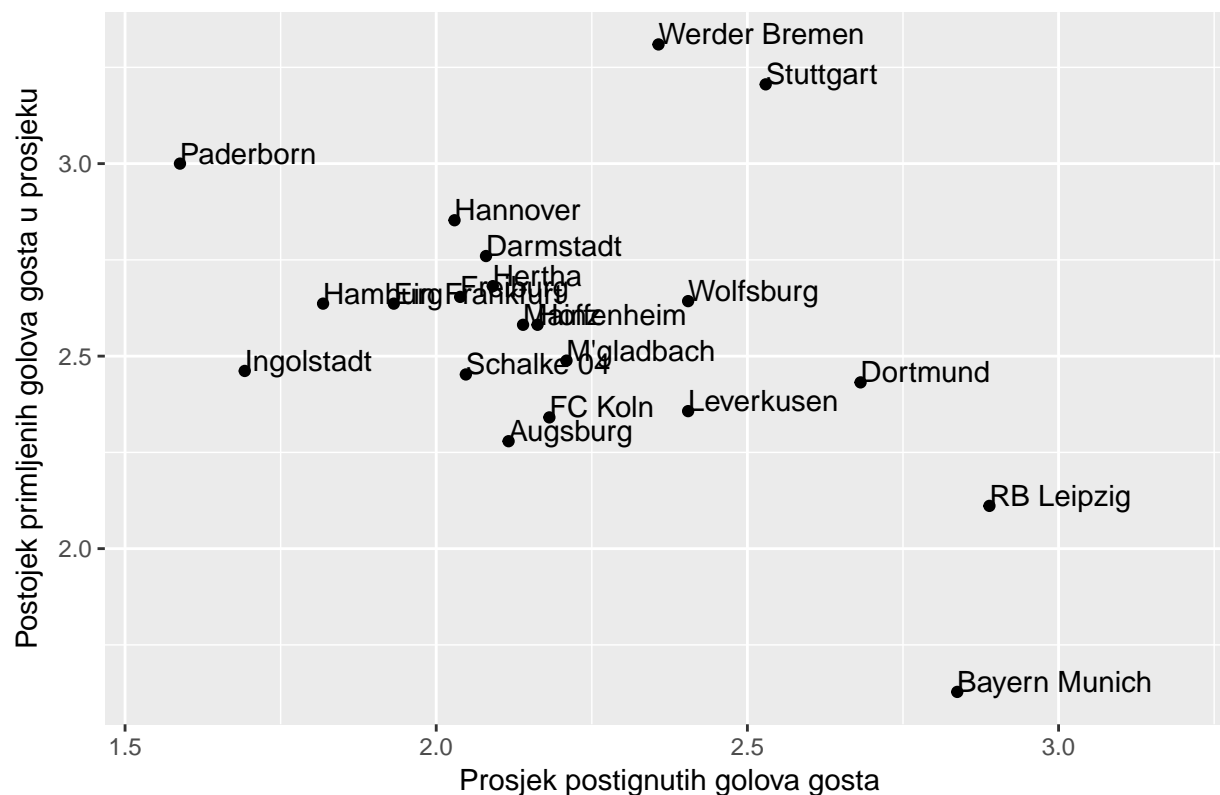
Prilično očita informacija na grafovima je da bilo koja, domaća ili gostujuća, momčad mora zabiti više od 1 gola da joj najvjerojatniji ishod ne bi bio poraz. I to semantički ima smisla jer je Bundesliga jedna od najatraktivnijih europskih liga po ukupnom broju golova.

No, znamo da sve momčadi u projekciji ne sudjeluju jednako te sljedeća dva grafa prikazuju kako momčad zabija/prima golove kad je domaćin, odnosno gost:

Odnos zabijenih i primljenih golova kad je momcad domacin



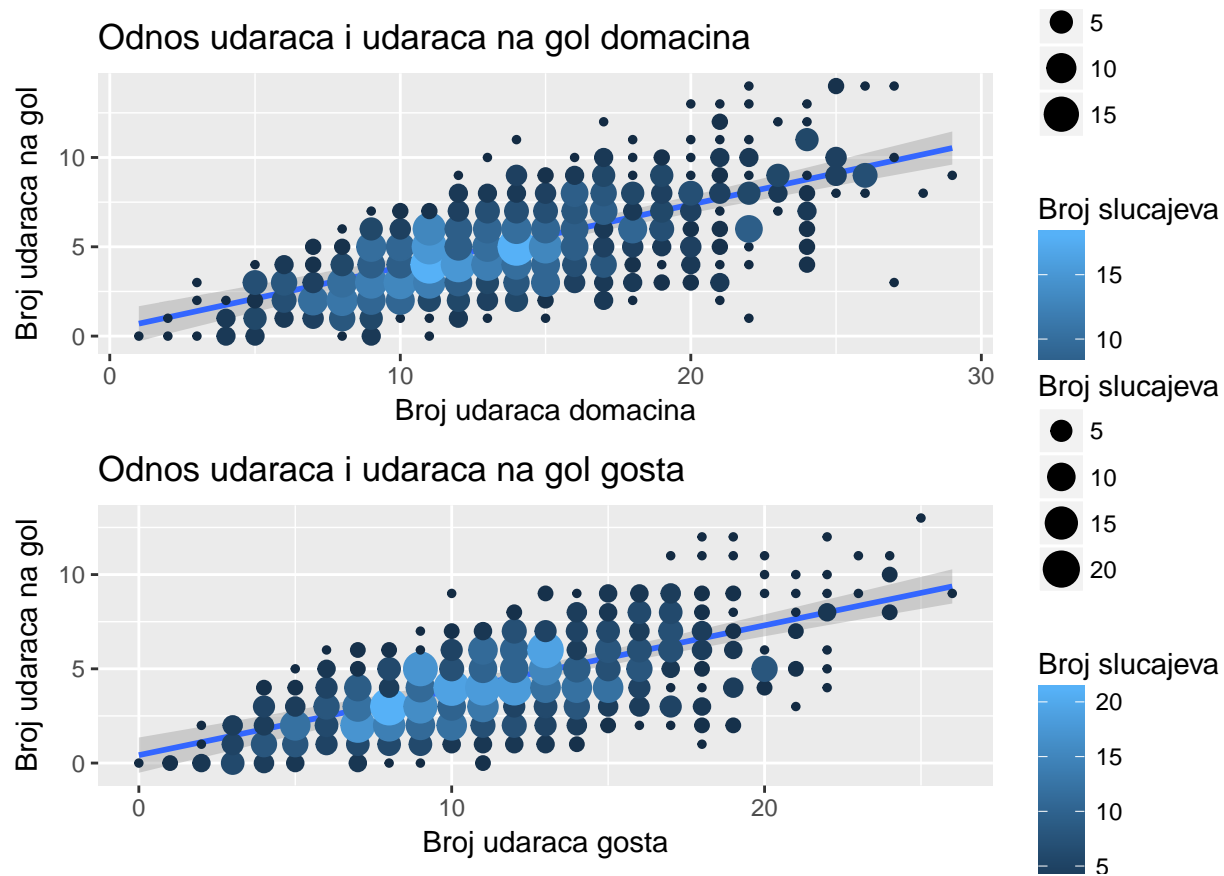
Odnos zabijenih i primljenih golova kad je momcad gost



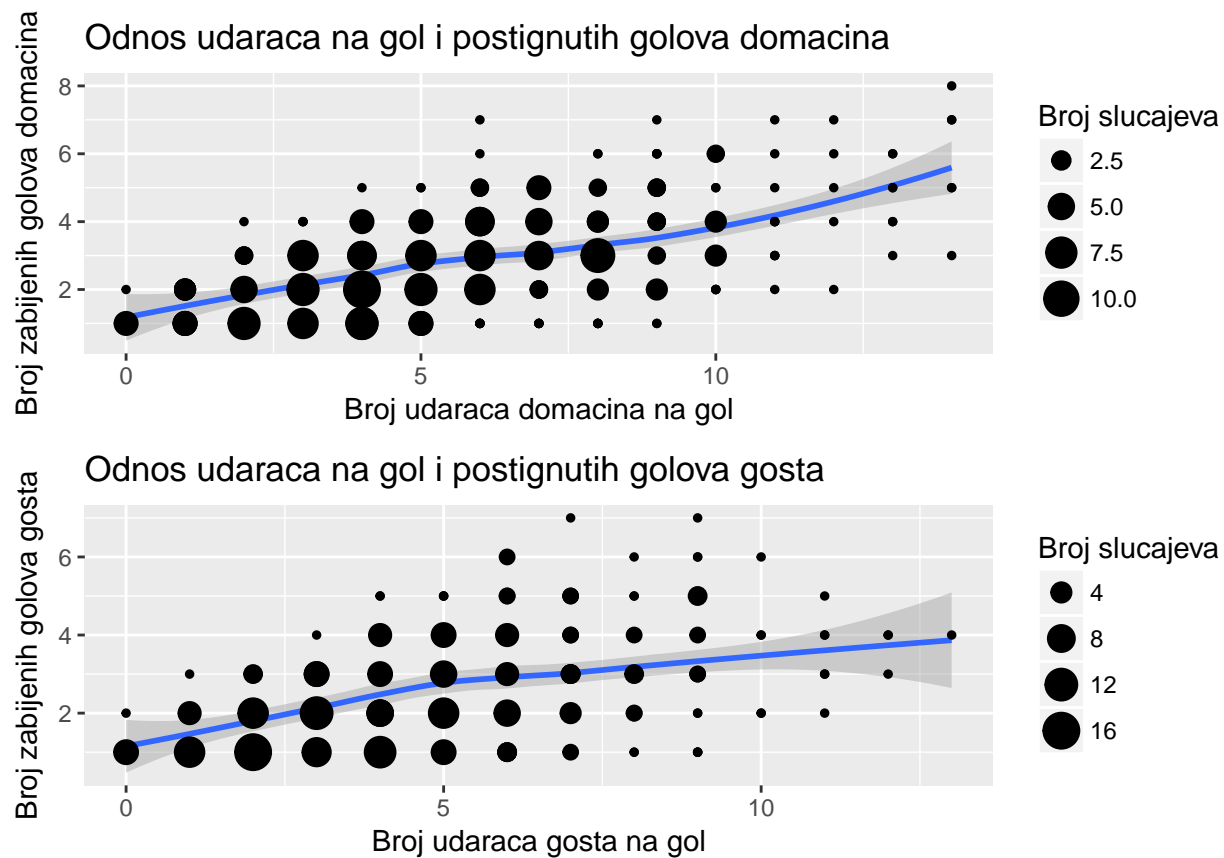
Iz njih je očito da je Bayern klasa za sebe, prima najmanje i zabija najviše golova u oba slučaja. Zanimljivo je da se golovi momčadi na domaćem terenu ponašaju praktički linearno, uopće nisam planirao nacrtati pravac preko podataka, ali nakon pogleda na graf, bilo je očito da tako mora biti. S druge strane, golovi momčadi u gostima su nabijeni u sredinu grafa, a outlieri su 3 najbolje momčadi (Bayern, Dortmund i Leipzig - mali uzorak za RBL), 2 momčadi koje su ispale u drugu ligu (Stuttgart i Paderborn) te jedna koja godinama zabija i prima puno golova (Werder).

Analiza podataka o udarcima i kornerima

Na ovom grafu prikazan je odnos ukupnog broja udaraca i broja udaraca na gol za domaćine, odnosno goste. Eliminirani su podaci kad momčad ima više od 30 udaraca (bilo ih je nekoliko), s ciljem što usporedljivijeg skaliranja. Na grafu vidimo da domaćini na gol pucaju nešto češće, ali značajno preciznije.

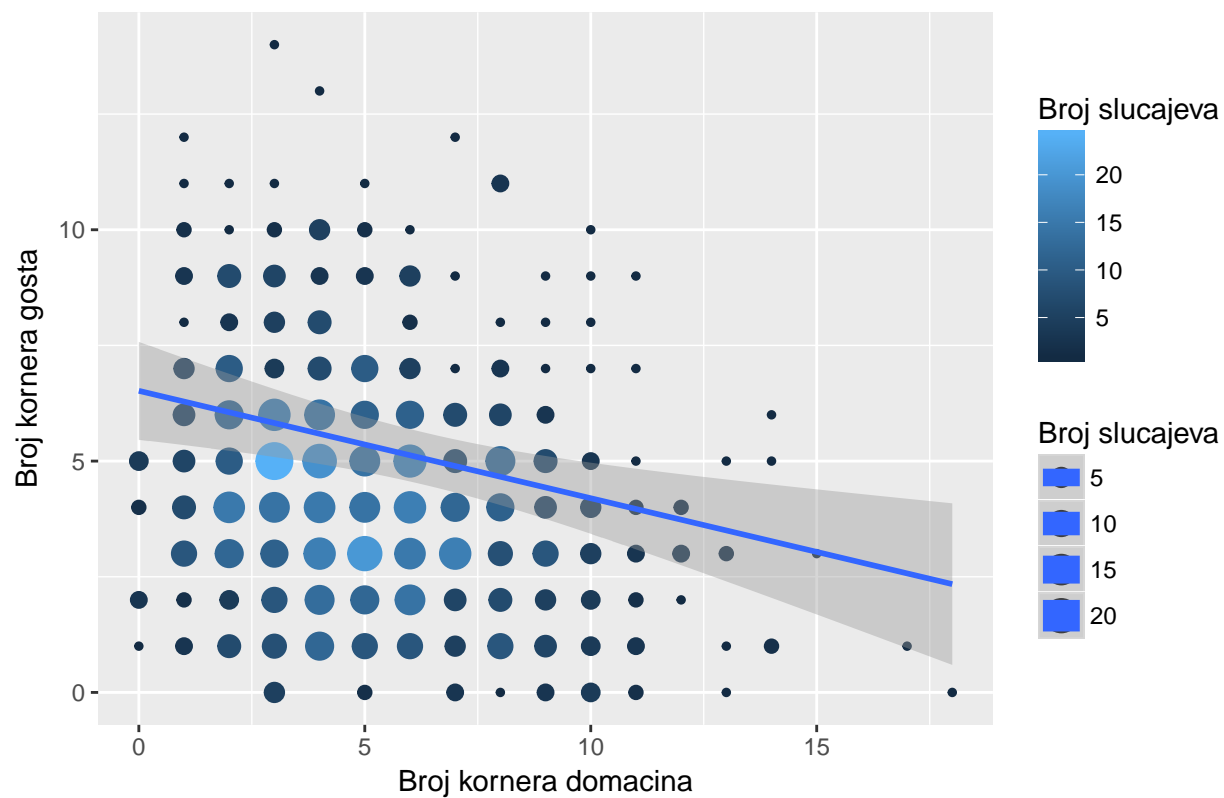


Sljedeći graf prikazuje odnos broja udaraca na gol domaćina/gosta te postignutih golova iste momčadi. Primjetno je da u realizaciji prilika nema nikakve razlike, jedino u dolasku do istih ako pogledamo desne strane grafa budući da se rijetko događa da gost uputi više od 5 udaraca na gol.



Sljedeći graf prikazuje odnos broja kornera domaće i gostujuće momčadi, kao što možemo vidjeti, oni su obrnuto proporcionalni.

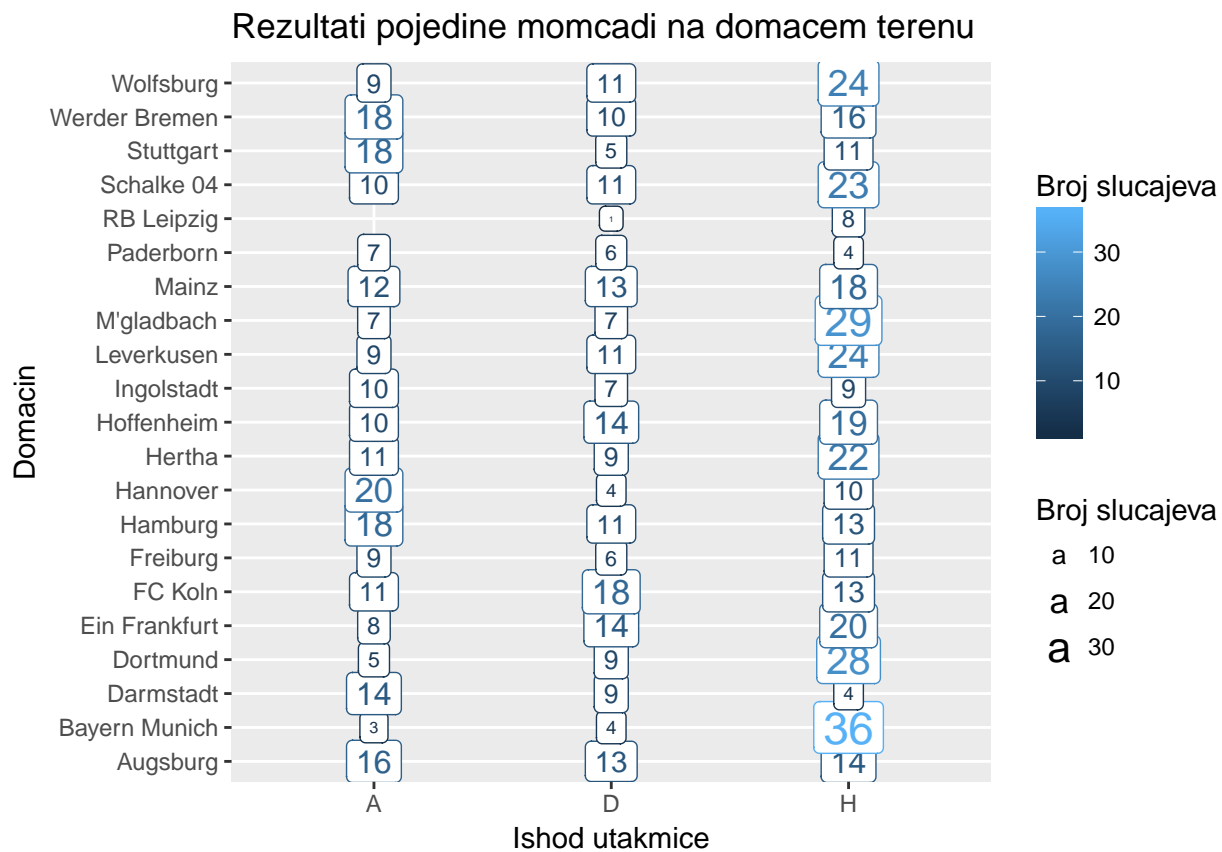
Odnos kornera domacina i gosta



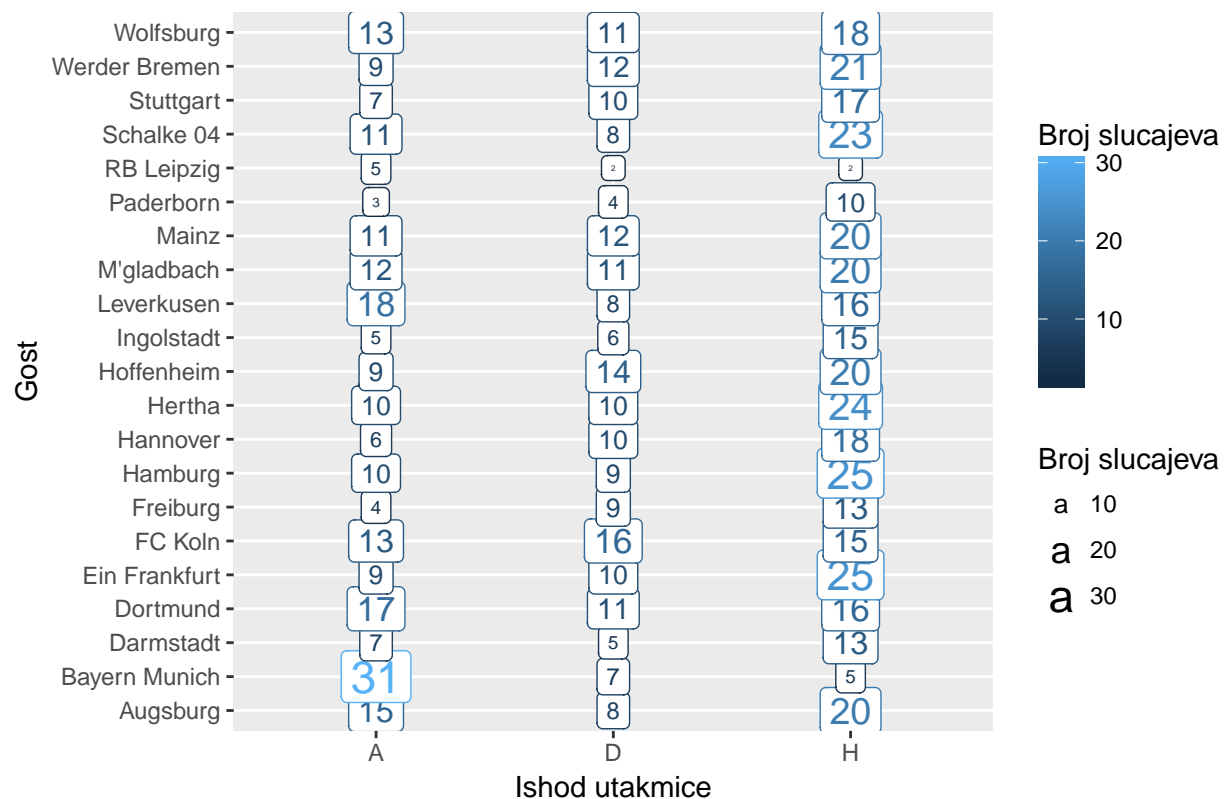
Korelacija domaćih i gostujućih kornera: -0.2728384

Ishodi utakmica svake momčadi

Naravno, najzanimljiviji aspekt analize sportskih podataka je predviđanje pobjednika. Budući da je ovo eksplorativna, a ne prediktivna analiza podataka, nećemo se baviti predviđanjem, no zanimljivo je pogledati koje su rezultate momčadi imale u postekle 2 i pol sezone.

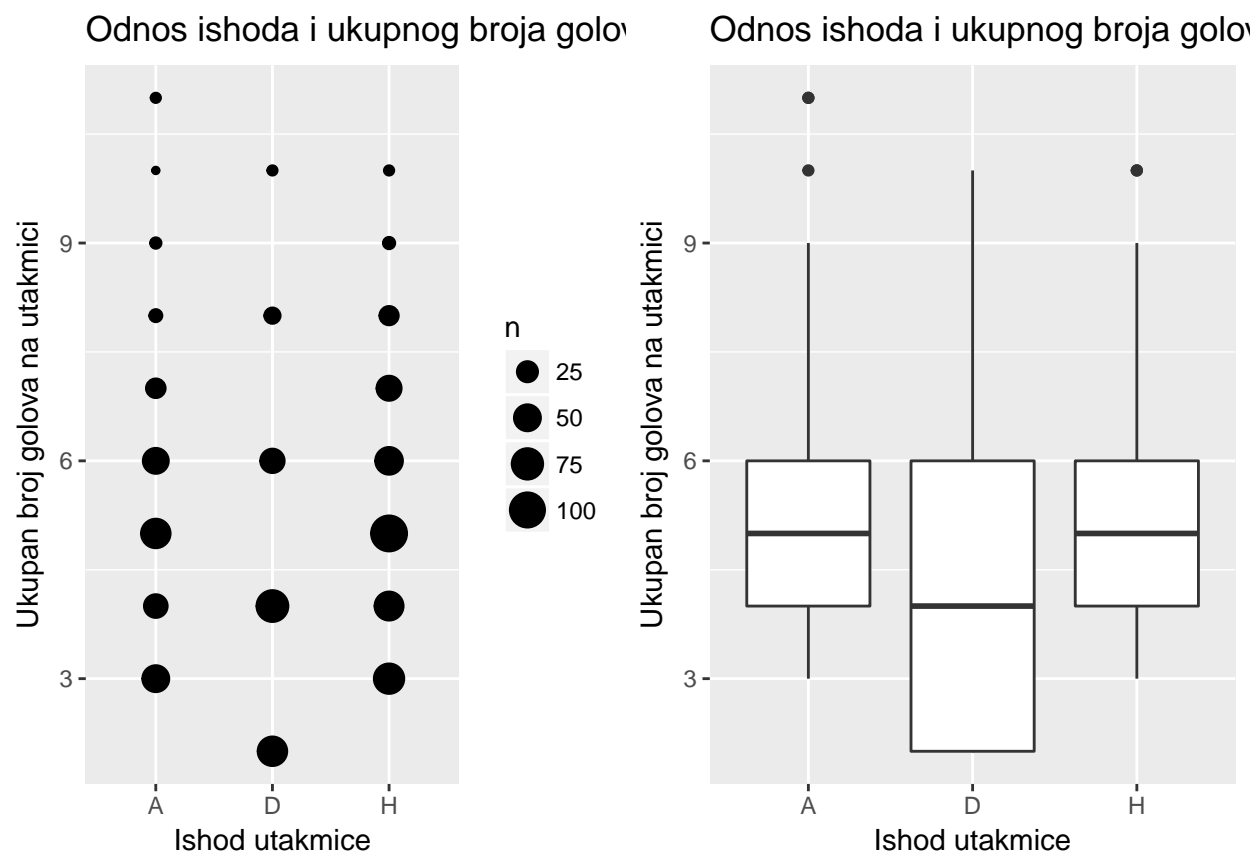


Rezultati pojedine momčadi na gostujućem terenu



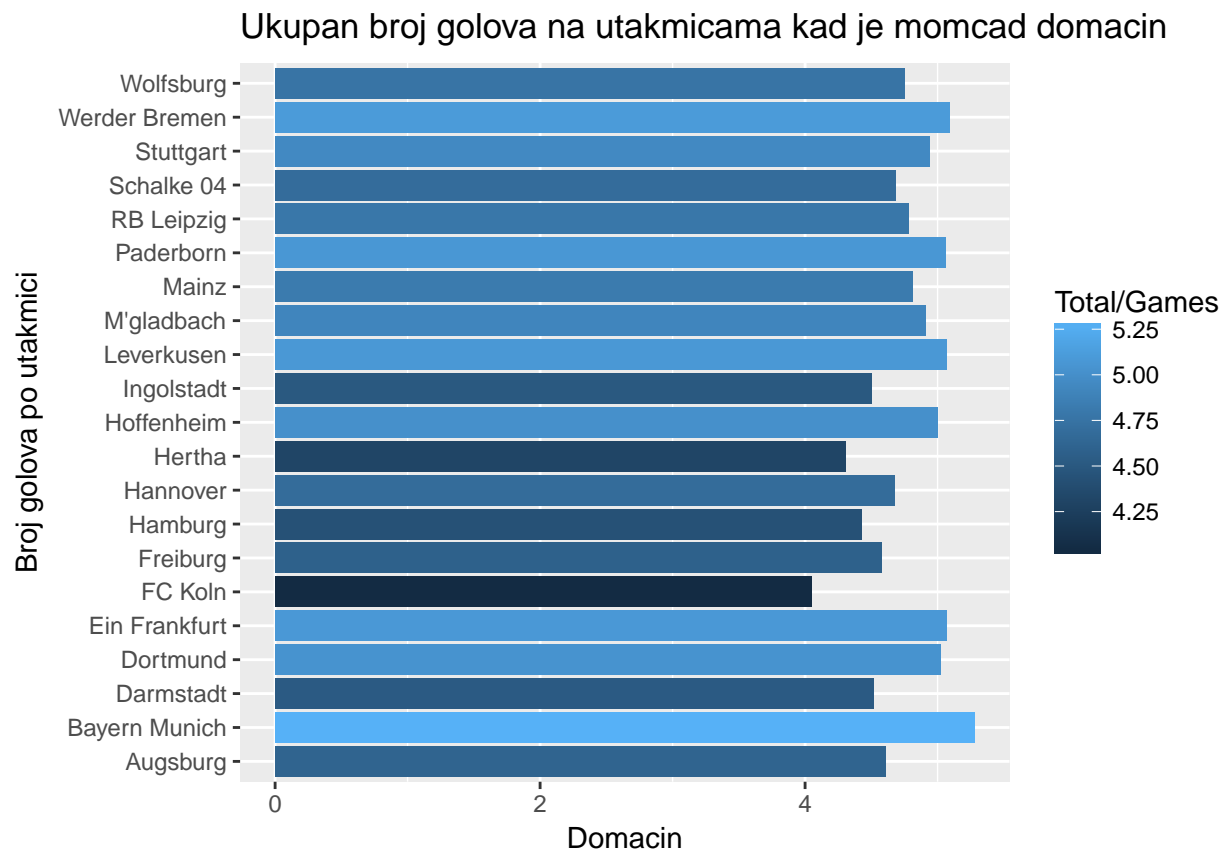
Iz estetike grafa očito je da je Bayern nadominantnija momčad na gostujućem terenu jer je većina protivnika podjednako loša na gostujućim terenima, iako se vidi razlika u kvaliteti momčadi prema čuvanju domaćeg terena. U ovom slučaju mali uzorak Leipziga znači da je svrstan među lošije momčadi.

Sljedeći graf je izuzetno zanimljiv, on pokazuje da je ukupan broj golova u pobjedama domaćina i gosta podjednak, dok je broj golova u neodlučenim rezultatima manji. To je prikazano na dva načina, klasičnim prebrojavanjem i popularnim boxplotom.

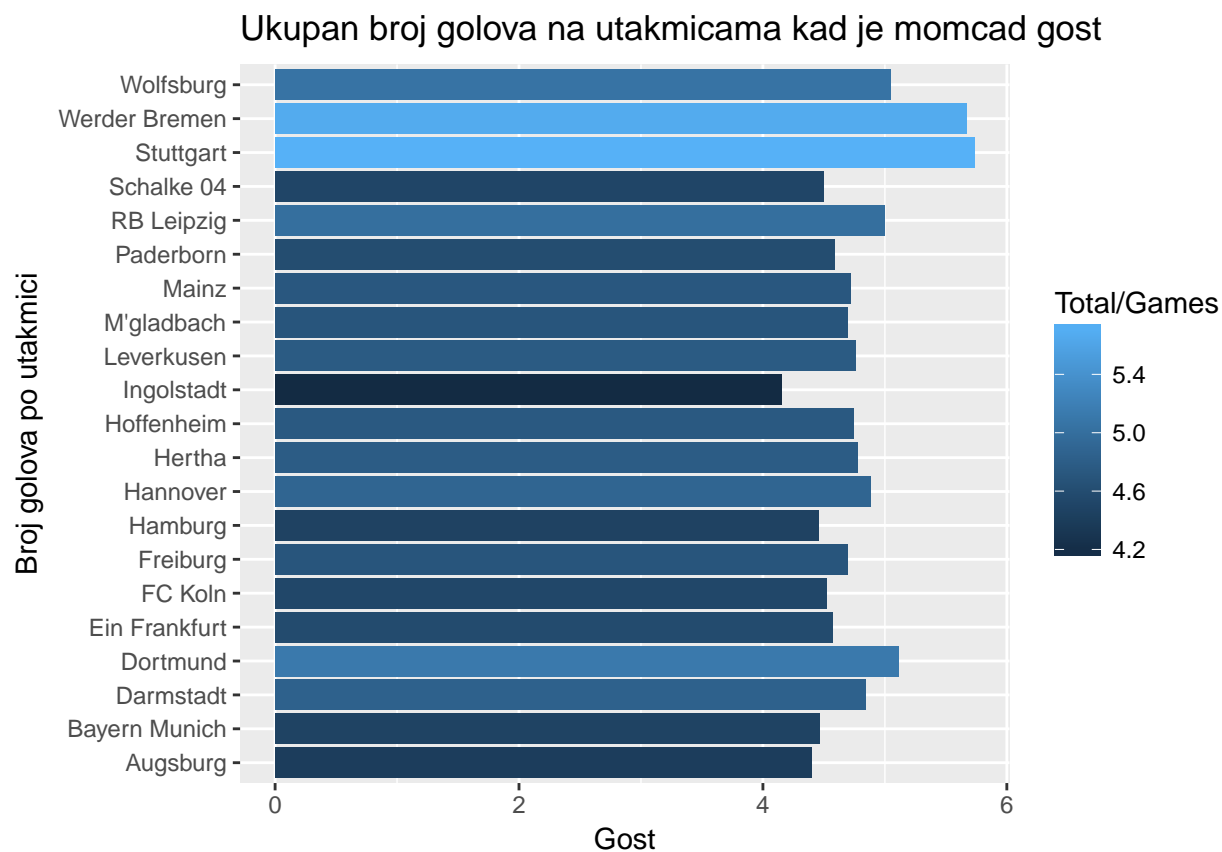


Odabir najzanimljivijih momčadi

Sljedeća dva grafa prikazat će ukupan broj golova na utakmicama kad je neka momčad domaćin/gost te će čitatelju pomoći u odabiru najvjerojatnije atraktivnih utakmica, tj. ako se na utakmicama dvaju momčadi uvijek zabija mnogo golova, vrijedi ih pogledati.



A zatim za gostujuće momčadi:



Odabiri su prilično očiti: Bayern kao domaćin te Werder i Stuttgart kao gosti garantiraju najviše golova. Činjenice potvrđuju takav odabir, Bayern je sezonu 2016/17 počeo 6:0 pobjedom protiv Werdera, sezonu prije je Stuttgart doma pobijedio 4:0, a Werder 5:0 te Stuttgart 2:0, a Werder 6:0 sezone 2014-15. To je ukupno 23 gola na 5 utakmica, tj. 4.6 gola po utakmici