

# Jalgpallikoondiste omavahelised kohtumised aastast 1872 kuni 2019

Autor: Joonas Salu

## Sissejuhatus

Projektitöös vaadeldakse jalgpalli rahvuskoondiste omavahelisi kohtumisi ning seati eesmärgiks uurida kolme seose olemasolu:

- Kui palju mõjutab mängu toimumise aeg (kalendrikuu) mängus löödud väravate arvu?
- Kui palju mõjutab mängu tulemust koduväljakul mängimine (kas koduväljak annab eelise)?
- Kuidas on ajas muutunud ühes mängus löödavate väravate arv?

## Andmestiku kirjeldus

Vaadeldav andmestik on saadud veebilehelt <https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>. Vaadeldud on jalgpallikoondiste kohtumisi aastast 1872 kuni 2019. Kokku on andmestikus 41540 kohtumist ehk kõigi selles ajavahemikus toimunud rahvuskoondiste mängude info. Algandmestikus on esitatud 9 tunnust, millele lisati autori poolt üks algandmete põhjal arvatav tunnus (mängus löödud väravate summa).

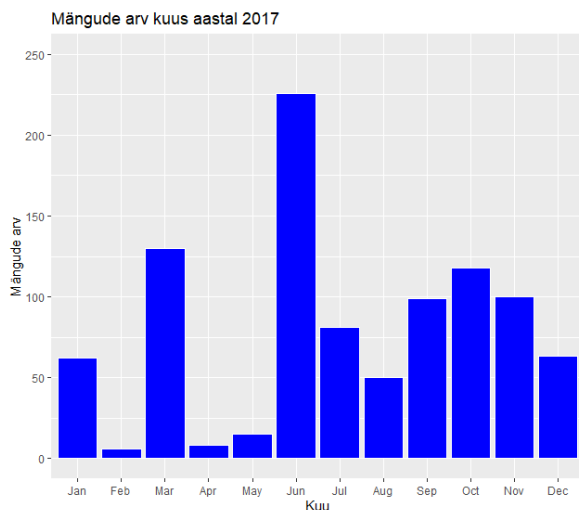
Andmestikus on igal mängul 10 tunnust:

- date (kuupäev) – diskreetne tunnus
- home\_team (kodumeeskond) – tekstiline tunnus
- away\_team (külalimeeskond) – tekstiline tunnus
- home\_score (kodumeeskonna väravate arv antud mängus) – täisarvuline tunnus
- away\_score (külalimeeskonna väravate arv antud mängus) – täisarvuline tunnus
- tournament (võistlus, mille raames mäng toimus) – tekstiline tunnus
- city (linn, kus mäng toimus) – tekstiline tunnus
- country (riik, kus mäng toimus) – tekstiline tunnus
- neutral (kas mäng toimus neutraalsel pinnal) – nominaalne tunnus (tõeväärtus)
- goals\_in\_game (mängus löödud väravate arv) – täisarvuline tunnus

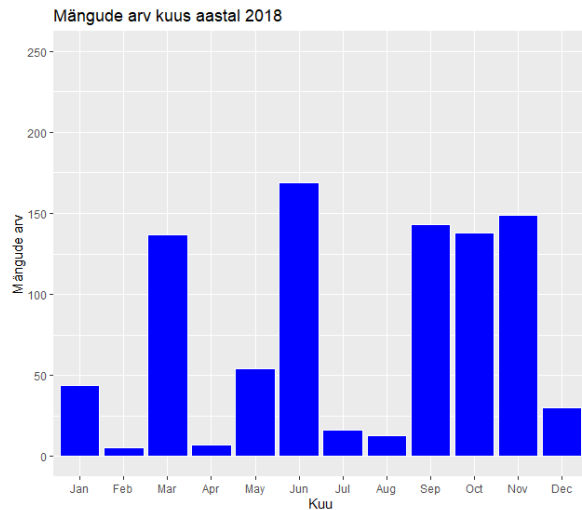
## Andmete analüüs

Kui palju mõjutab mängu toimumise aeg (kalendrikuu) mängus löödud väravate arvu?

Esimesena uuriti seost mängu toimumise aja (kalendrikuu) ja mängus löödud väravate arvu vahel. Analüüsiks võeti kaks järjestikust aastat (2017, 2018).



Joonis 1. Mängude arv kuude lõikes aastal 2017



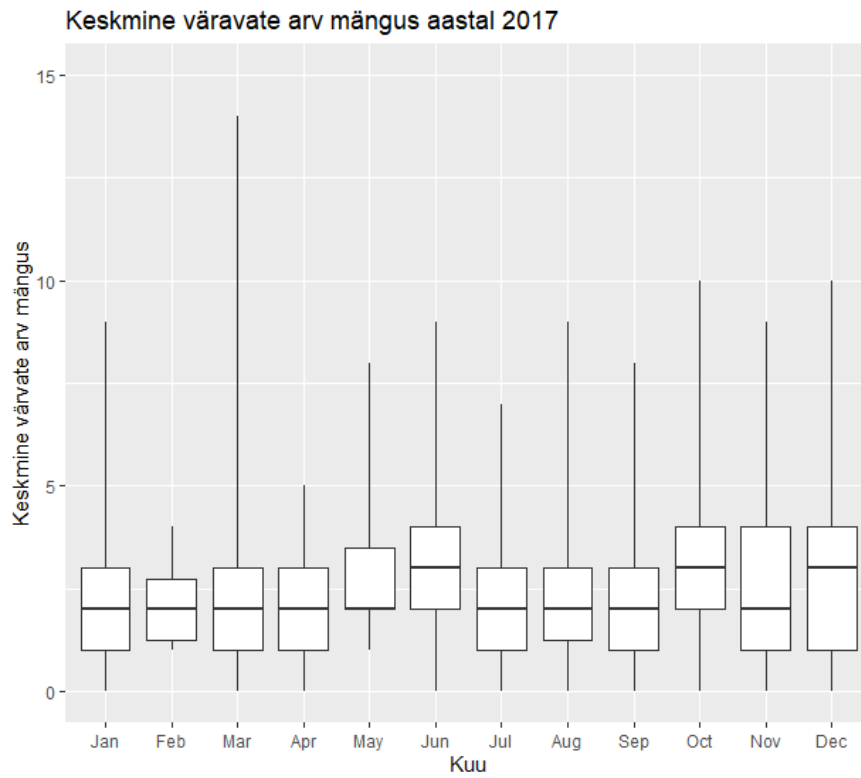
Joonis 2. Mängude arv kuude lõikes aastal 2018

Joonis 1 illustreerib toimunud mängude arvu kuude lõikes aastal 2017. Näeme, et enim toimus mängu juunis, selle põhjuseks on toimunud FIFA konföderatsioonide karikaturniir ja muud maailmameistrivõistlusteks valmistumisega seotud turniirid.

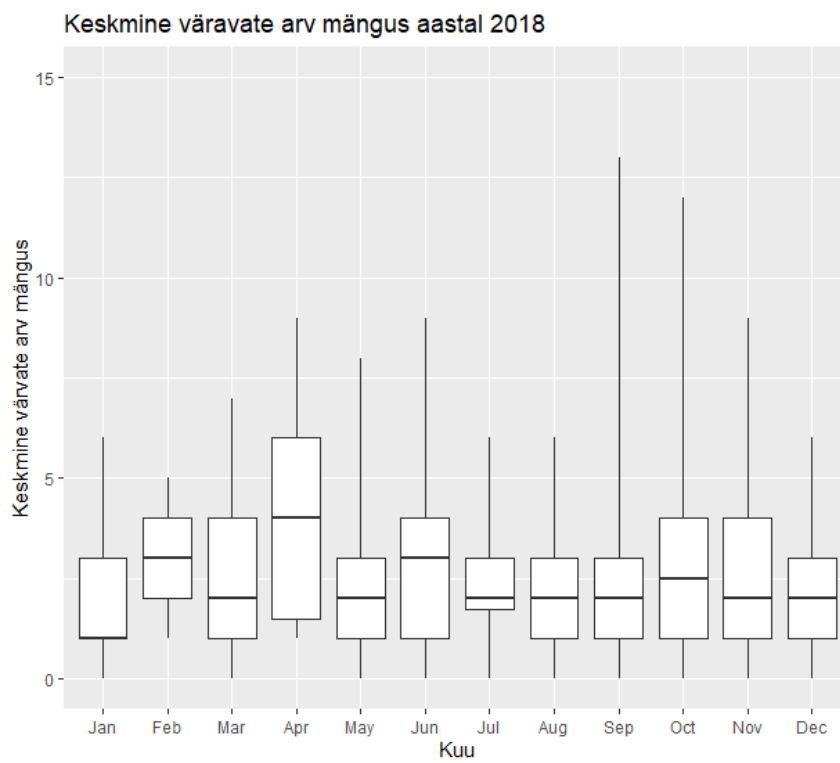
Joonis 2 illustreerib toimunud mängude arvu kuude lõikes aastal 2018. Näeme, et enim toimus mängu juunis, selle põhjuseks on toimunud jalgpalli maailmameistrivõistlused Venemaal.

Diagrammidelt on näha, et suur mängude hulk langeb märtsi, juunisse ja sügisperioodi, seega kõige usaldusväärsemad andmed uuritavas küsimuses saame just neid kuude tulemusi vaadelda.

Järgnevalt uuriti andmeid karp-vurrud diagrammiga, millelt näeme väravate arvu varieeruvust, kvartiilvahemikke ja mängudes löödud väravate arvu mediaani.



Joonis 3. Keskmine löödud väravate arv mängus kuude lõikes 2017.aastal



Joonis 4. Keskmine löödud väravate arv mängus kuude lõikes 2018. aastal

2017. aastal ei ole üldist keskmiste tulemuste suurt kõikumust – üheksal kuul on mediaan ühesugune ja ülejäänud kolmel kuul omakorda väga sarnane. 2018. aastal on eristuvaid kuid rohkem, kuid üldpilt on sarnane 2017. aastaga.

Eristuvad need kuud, kus on vähe mänge, siis on iga mängu tulemusel suurem mõju keskmise arvutamisel. Kõige enam eristub 2018. aasta aprill.

Teine eristumisi põhjustav tegur on suurvõistluste esinemine vaadeldavas kuus, see tõstab mediaan tulemust, sest neis mängudes tuleb sageli ette vajadus selgitada välja võitja, milleks kasutatakse lisaaega.

2017. ja 2018. aasta jooksul mängudes löödud väravate arvu ja mängu toimumise kuu vahel seos puudub või on väga nõrk, selgub koostatud karp-vurru diagrammidest, mille paljude kuude jaotised on väga sarnased, eriti 2017. aastal, mil ei olnud maailmameistrivõistlusi, millest saamegi teha vastava järelduse.

Täiendavalt arvutati samade andmete kohta Spearmani korrelatsioonikordaja, mille väärtuseks saadi ligikaudu -0,07, mis näitab samuti andmete vahel seose puudumist.

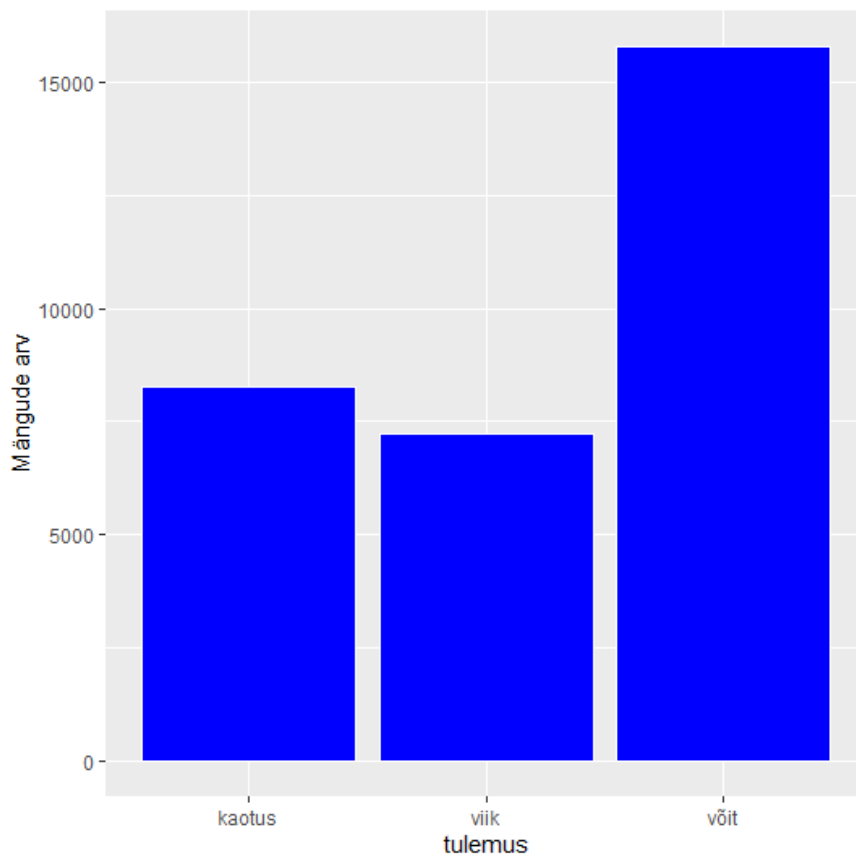
## Kui palju mõjutab mängu tulemust koduväljakul mängimine (kas koduväljak annab eelise)?

Teiseks uuriti seose olemasolu mängu tulemuse ja koduväljakul mängimise vahel. Andmestik (1872-2019) on mänge, mis on mängitud ühe meeskonna koduväljakul või neutraalsel väljakul (kumbki meeskond ei mängi koduväljakul). Uuriti ainult mänge, kus mäng toimus ühe võistkonna koduväljakul, neid mänge oli kokku 31289.

home_team	Mänge
Brazil	568
Argentina	548
Mexico	513
Germany	506
England	498
Sweden	486
France	484
South Korea	465
Hungary	453
Italy	443

away_team	Mänge
Uruguay	541
Sweden	530
England	511
Hungary	475
Paraguay	453
Germany	443
Argentina	432
Poland	431
Zambia	418
Finland	414

Tabel 1. Top 10 enim kodumänge mänginud riigid ja top 10 enim võõrsilmänge mänginud riigid



Joonis 5. Kodumeeskonnana mängitud mängude tulemused

Jooniselt 5 on näha, et koduväljakul saavutab meeskond võidu pisut rohkem kui pooltel juhtudel ja kaotab vaid veidi rohkem kui veerand mängudest. Seega on kodumängud kindlasti eelistatud, seepärast on olulistel mängudel enamasti kaks mängu – üks kummagi võistkonna koduväljakul. Suurvõistlustel puhul mängivad enamus võistkondi neutraalsel väljakul, mis välistab koduväljaku eelise. Mis on koduväljaku eelise põhjuseks, saab vaid oletada.

Täiendavalt vaadeldi mängus löödud väravate arvu ja siin on ilmselt üks olulisi koduväljaku eelise põhjuseid, koduväljakut tuntakse väikseima detailini ja seal õnnestub rohkem väravaid lüüa. Kodumeeskondade keskmine mängus löödud väravate arv 1,78, aga võõrsil mänginud meeskondade oma on 1,12 väravat mängus, see on väga suur erinevus, kui teada, et keskmiselt lüüakse mängus 2-3 väravat.

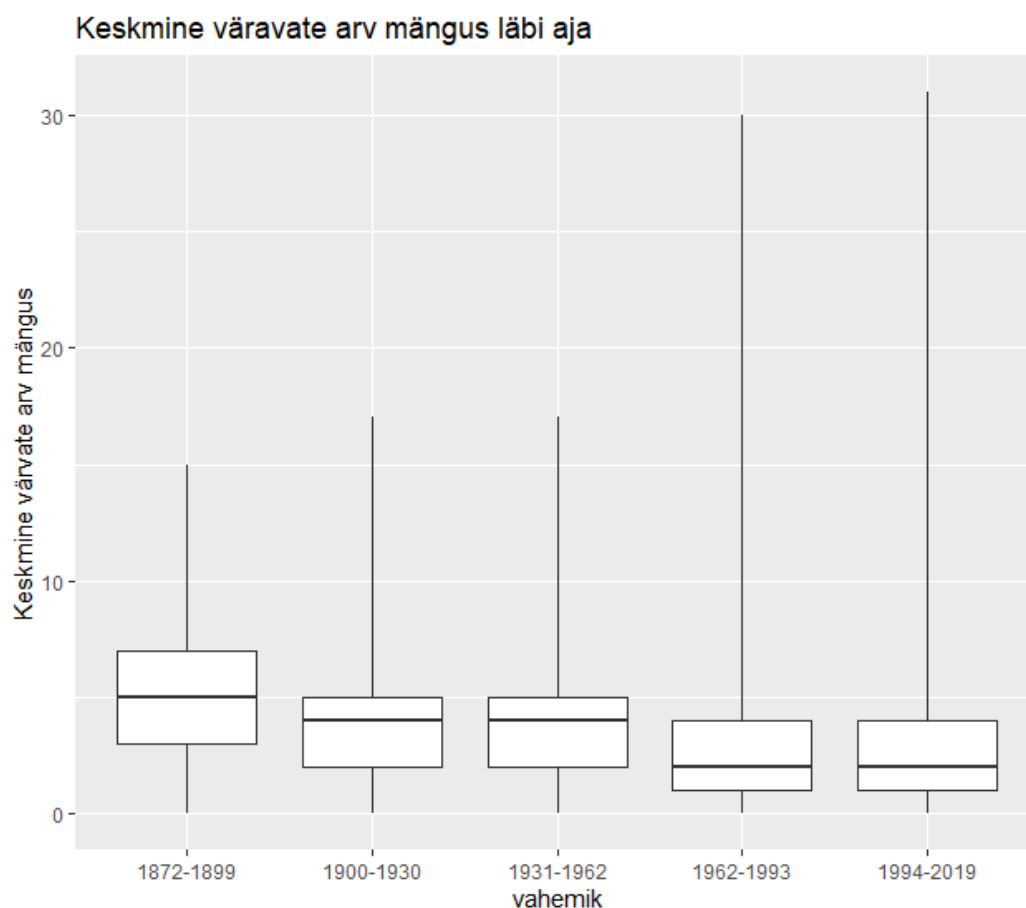
Kuidas on muutunud ühes mängus löödavate väravate arv?

Kolmandaks uuriti, kuidas on ajas muutunud mängudes löödud väravate arv. Uuritavad mängud jagati viieks võrdseks perioodiks. Perioodide piirid on võimaluse korral seotud reeglistikus toimunud muudatuste rakendamisega.

	vahemik	minimaalne_väravatearv	keskmine_väravatearv	mediaan	maksimaalne_väravatearv
1	1872-1899	0	5.195489	5	15
2	1900-1930	0	3.951679	4	17
3	1931-1962	0	4.058109	4	18
4	1962-1993	0	2.760293	2	30
5	1994-2019	0	2.758777	2	31

Tabelid 2. erinevate ajaperioodid väravate tabelid.

Tabelitest ja karp-vurrud diagrammilt on näha, et keskmine löödud väravate arv on ajas langenud, mis oli eeldatav ja tuleneb reeglitesse tehtud muudatustest. Ootamatu tulemus on see, et maksimaalne väravate arv ühe mängus on ajas tõusnud. Selle muutuse põhjuseks võib olla mängijate taseme suurem diferentseerumine, st meeskonnad on muutnud rohkem ebavõrdseks ehk halvima ja parima meeskonna erinevus on suurenenud ajas.



Joonis 6. Keskmiste väravate arv mängus läbi aja

## Kokkuvõte

Projektitöös uuritud kolm küsimust olid autorile huvipakkuvad ja otsitud seosed leidsid kahel juhul kinnitust ja kolmas seos puudus.

Koduväljaku eelisest räägitakse palju, aga et see nii tugevalt tulemustes ilmnes, oli ootamatu. Põnev oleks uurida, millest see tuleneb.

Seose puudumine mängus löödud väravate ja mängu toimumise aja vahel oli üllatuslik, see näitab, et mängijad on väga heas vormis aastaringselt ja tulemus ei sõltu hooajast, samuti ei ole väga suuri muutuseid oluliste võistlustega seose, seega ei toimu ka olulist vormi ajastamist, nagu paljude teiste spordialade puhul.

Ajas väravate arvu vähenemine oli ootuspärane, ometi on tänapäevased mängud väga huvipakkuvad oma kiiruse ja tehnilise osavusega. Väravate arvu vähenemise põhjuseks on tugevnenud kaitsemäng.