

# Digitaalsete kaamerate analüüs 1994-2007

Elis Käär

17/12/2020

## Andmestiku kirjeldus

Andmestikuks on valitud Kaggle'ist 1038 digitaalse kaamera kohta käiv andmestik (<https://www.kaggle.com/crawford/1000-cameras-dataset>), mis kirjeldab ära 13 kaamerate kohta käivat tunnust. Lisaks olema-solevatele tunnustele on lisatud juurde ka firma tunnus, mis oli lihtsasti eraldatav mudeli nimest, ning ka tootjariigi tunnus.

Olulisemad tunnused on: hind, mudel, mudeli väljalaskeaasta, tootjariik, maksimaalne lahutusvõime(PPI), makro fookuskaugus, telesuum (mm), kaal(koos patareidega).

Kahjuks ei ole enamiku tunnuste puhul teada täpset mõõtühikut, seepärast pole neid graafikutel ka kirjas.

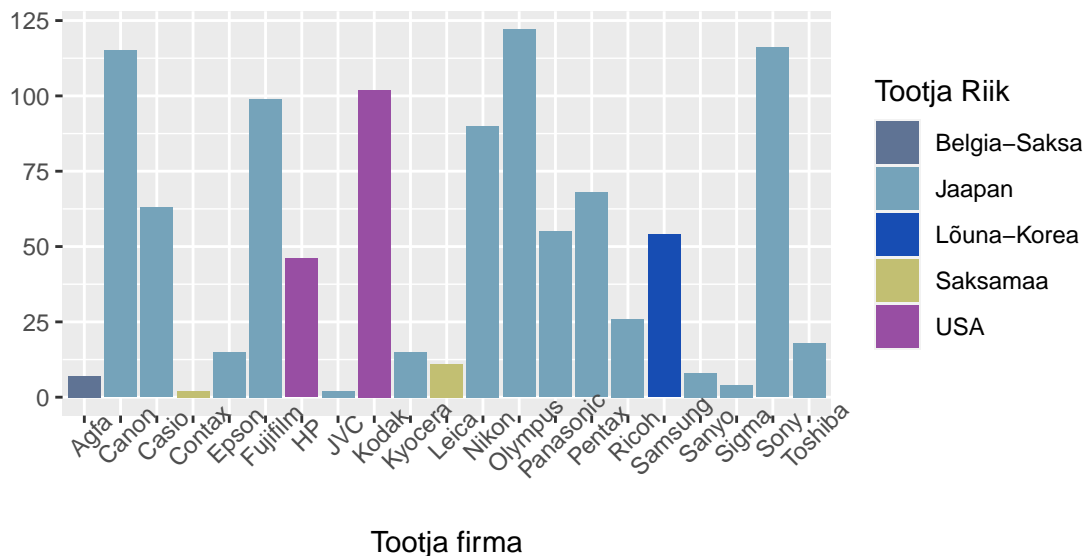
Järgnevas tabelis on näha andmestikus kirjeldatud tootjafirmad, mis on eraldatud kaamera mudelinimest

	sagedus
Contax	2
JVC	2
Sigma	4
Agfa	7
Sanyo	8
Leica	11
Epson	15
Kyocera	15
Toshiba	18
Ricoh	26
HP	46
Samsung	54
Panasonic	55
Casio	63
Pentax	68
Nikon	90
Fujifilm	99
Kodak	102
Canon	115
Sony	116
Olympus	122

Tabel 1: Firmade sagedustabel

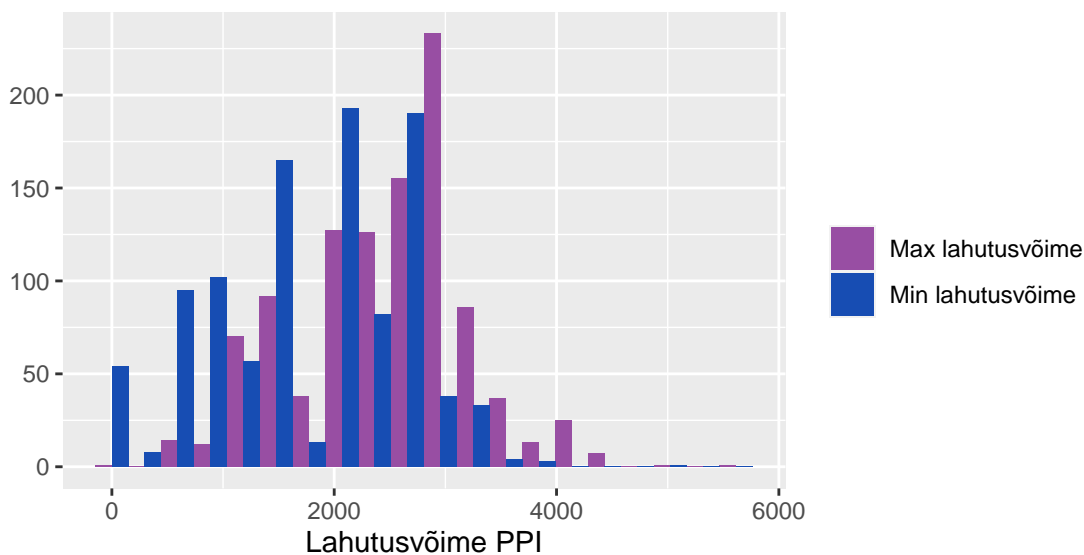
## Andmete esmane analüüs

Järgnevalt on teostatud andmete esmane visuaalne analüüs, kus on uuritud tunnuste jaotusi, sagedusi ja muid huvipakkuvaid statistikuid.



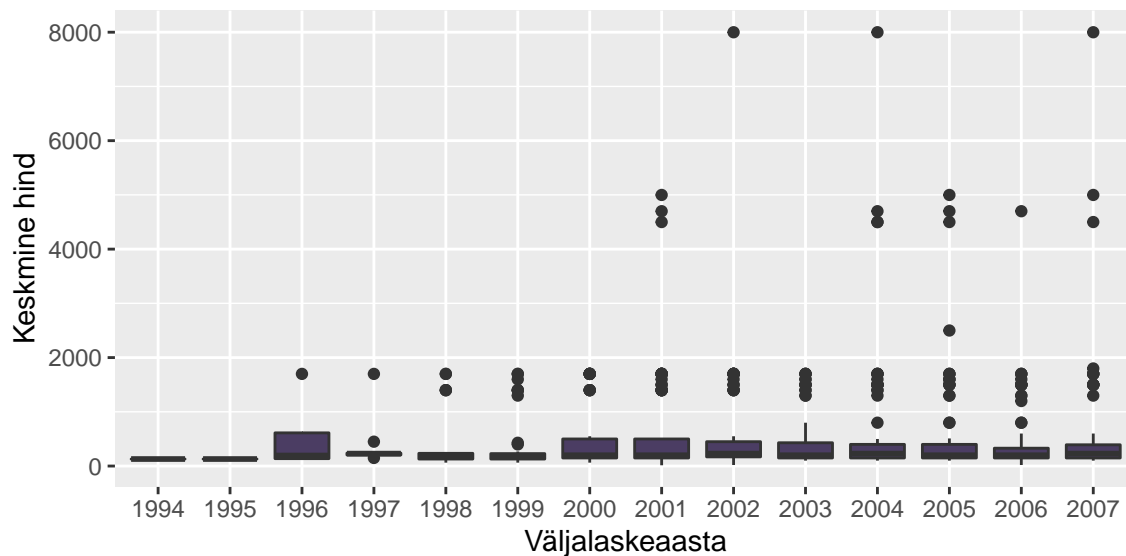
Joonis 1: Firmade sagedused tootjariikide järgi

Jooniselt 1 on näha, et tootjariikide seas on selgelt domineeriv Jaapan - selle riigi firmade poolt toodetud mudelid moodustavad andmestikust ligi 80%. Tuntuimad Jaapani kaameraid-tootvad firmad on Canon ja Nikon. Teisena tuleb järgi USA, kust on pärit kuulsad firmad Kodak ja HP. Lõuna-Koreast on pärit Samsung.



Joonis 2: Maksimaalse ja minimaalse lahutusvõime jaotus

Joonisel 2 on kujutatud maksimaalse ja minimaalse lahutusvõime jaotusi. Jaotused meenutavad sümmeetrilist normaaljaotust, kuid maksimaalse lahutusvõime jaotus on küllaltki raske vasaku sabaga, mis tähendab, et rohkem mudeleid on madalama maksimaalse lahutusvõimega.



Joonis 3: Kaamerate hinnad väljalaskeaasta järgi 1994-2007

Hindade haare ulatub mõnest kümnest ühikust kuni ligi 8000-ni. Jooniselt 3 joonistub, et enamik kindlal aastal esitletud mudelite hindadest jäävad alla tuhande, kuid hajuvus on väga lai. Numbriliselt on seda kirjeldatud tabeliga nr 2.

	Aasta	Keskmine	Standardhälve	Mudelid(tk)
1	1996	554.00	765.27	4
2	1997	367.18	448.57	11
3	1998	382.75	502.35	32
4	1999	349.64	460.83	53
5	2000	482.44	541.53	61
6	2001	635.19	966.63	85
7	2002	552.37	953.25	89
8	2003	413.32	467.69	101
9	2004	507.67	969.08	141
10	2005	448.93	760.29	143
11	2006	367.93	517.97	153
12	2007	436.04	854.42	163

Tabel 2: Hindade statistikud tinglikult aastate järgi 1994-2007

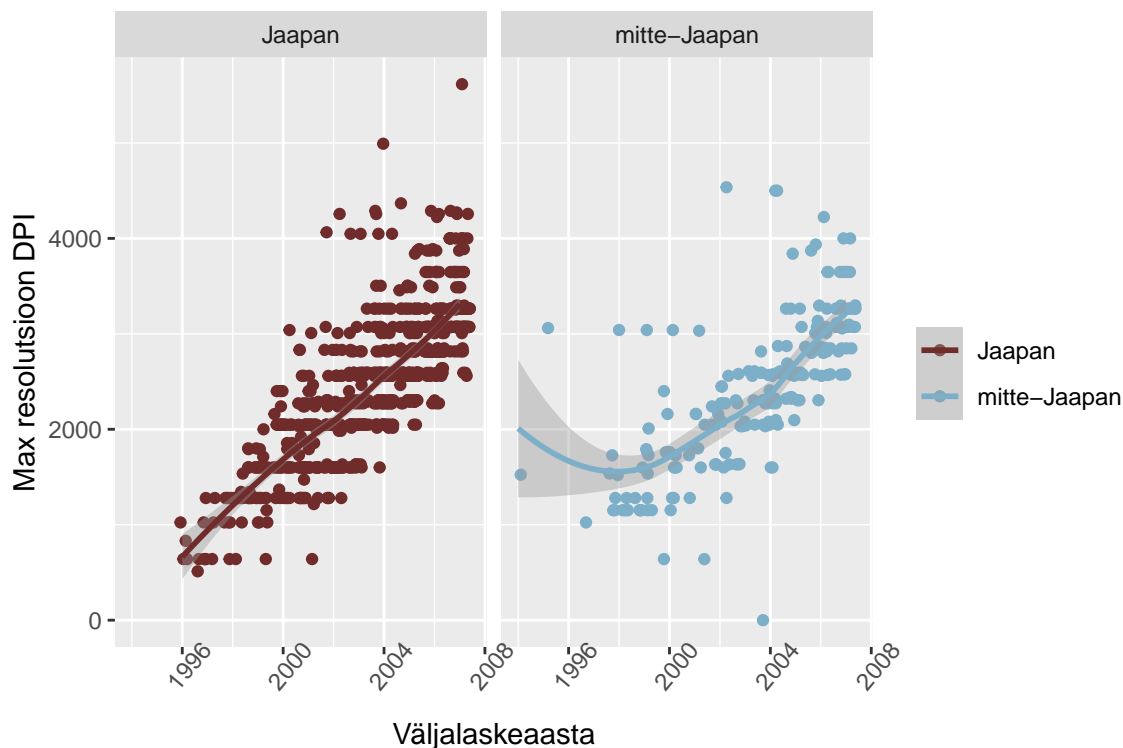
# Uurimisküsimused

Järgnevalt leian vastused neljale uurimisküsimusele:

1. Kas Jaapanis toodetud kaameratel (nt Nikon, Canon) toimus vaadeldavatel aastatel areng kiiremini võrreldes teistes riikides toodetud kaameratega? (Maksimaalne lahutusvõime)
2. Kas väljalaskeaasta kasvades langeb kaamera kaal?
3. Kas kaamera hinda mõjutab rohkem väiksem makrofookus või suurem suum?
4. Milliseid seoseid võib kaamera omaduste vahel leida?

## Kas Jaapanis toodetud kaameratel toimus vaadeldavatel aastatel areng kiiremini võrreldes teistes riikides toodetud kaameratega?

Antud küsimuses on arengu all mõeldud maksimaalse lahutusvõime kasvu.



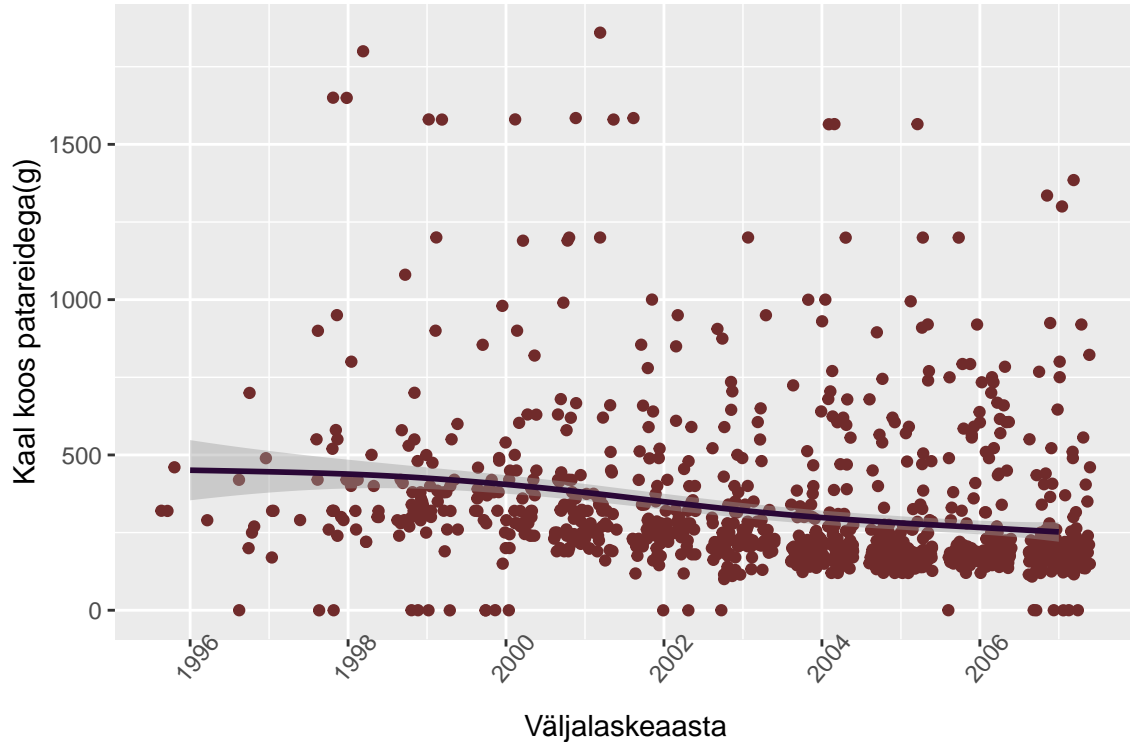
Joonis 4: Jaapani tootjate ja mitte-jaapani tootjate areng üle aastate 1994-2007

Jooniselt 4 võib näha, et Jaapanis toodetud mudelite areng on suhteliselt lineaarne üle kogu vaadeldava perioodi - sama ei saa öelda aga Jaapanist väljaspool toodetud mudelite kohta. Siinkohal kasutati mõlema kategooria jaoks Spearmani korrelatsiooni ning saadi, et mõlema kategooria puhul on tulemused statistiliselt olulised. Tulemused on esitatud Tabelis 3.

Nagu näha tabelist, siis Jaapanis ja väljaspool on aja möödudes areng toimunud sama kiiresti, erinevus on peaaegu et märkamatu.

	Riik	Spearman
1	Japan	0.83
2	Non-Japan	0.80

Tabel 3: Spearmani korrelatsioonikordajad Jaapanis ning väljaspool toodetud mudelite arengu põhjal



Joonis 5: Kaamera kaalu langemine üle aastate 1996-2007

### Kas väljalaskeaasta kasvades langeb kaamera kaal?

Jooniselt 5 on märgata, et üldine trend on tõepoolest negatiivne - see on rohkem väljajoonistunud graafiku alumises pooles, kus on kajastatud kaamerad, mis kaaluvad alla 600g.

Kaalu ja väljalaskeaasta vahel esineb tõepoolest statistiliselt oluline seos: Spearmani korrelatsioonikordaja väärtusega -0.409, mida võib lugeda keskmise tugevusega negatiivseks seoseks.

### Kas kaamera hinda mõjutab rohkem väiksem makrofookus või suurem suum?

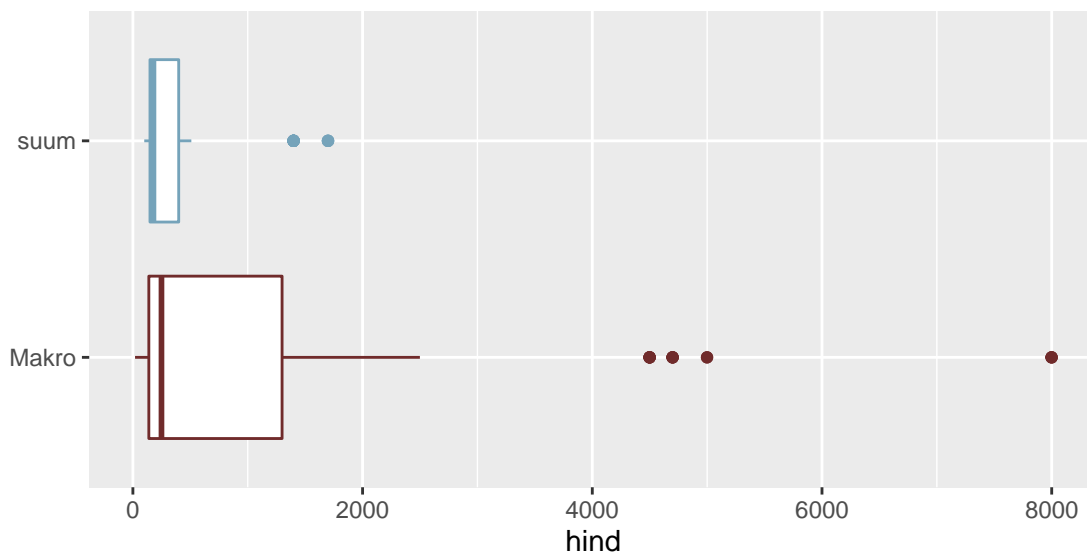
Siinkohal on võetud 200 kõige parema näitajaga vaatlust mõlema tunnuse jaoks, s.o 200 suurima telesuumiga kaamerat ja 200 väikseima makrofookuskaugusega kaamerat. Nendest on saadud 2 sõltuvat valimit. Hinna erinevuse testimiseks sõnastame ühepoolsed hüpoteesid ning sooritame t-testi valitud olulisuse nivool  $\alpha = 0.05$ :

$H_0$ : Keskmise hind parema makrofookusega < keskmine hind parema telesuumiga

$H_1$ : Keskmise hind parema makrofookusega > keskmine hind parema telesuumiga

T-testi tulemusena saame p väärtuseks väga väikese arvu,  $p < \alpha$ . Saame vastu võtta alternatiivse hüpoteesi, et keskmine hind parema makrofookusega on kõrgem kui keskmine hind parema telesuumiga.

Joonisel 6 on näha ka vastavat karpdiagrammi. Tuleb ka välja, et keskmised hinnad erinevad üle 500 ühiku võrra.



Joonis 6: Telesuumi ja makrofookuskauguse karpdiagramm

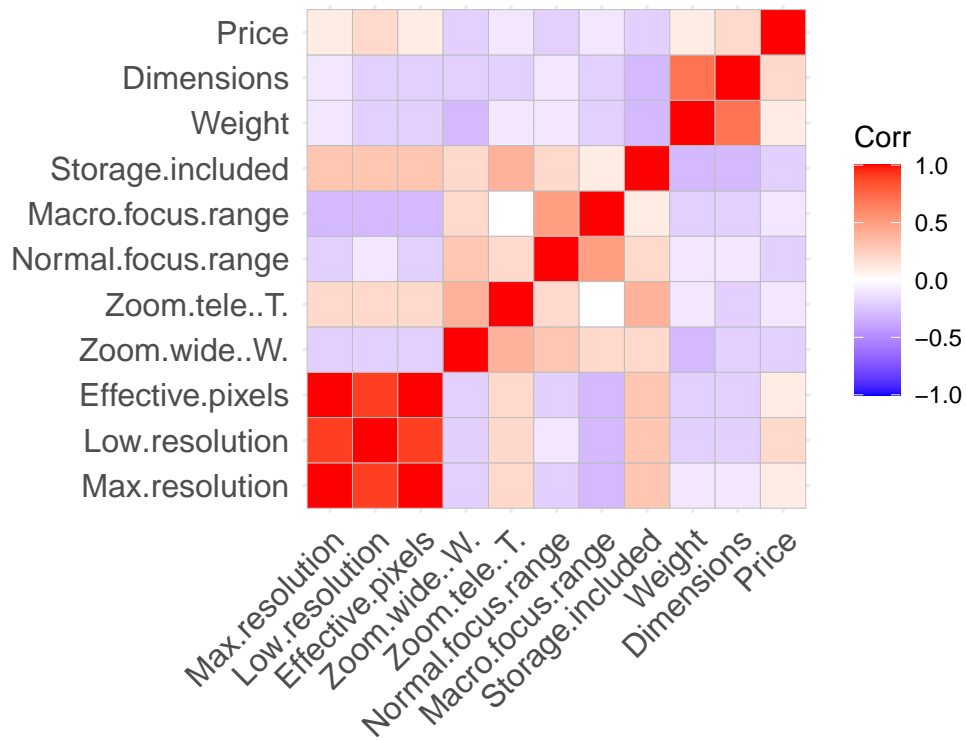
### Milliseid seoseid võib kaamera omaduste vahel leida?

Kaamera omaduste seas on defineeritud 11 huvitavat arvulist näitajat. Eesmärk on teada saada, milliste omaduste vahel on kõige tugevamad monotoonsed seosed (jättes välja triviaalsed seosed nt efektiivsete pikslite ja maksimaalse lahutusvõime vahel ning kaalu ja mõõtmete vahel)

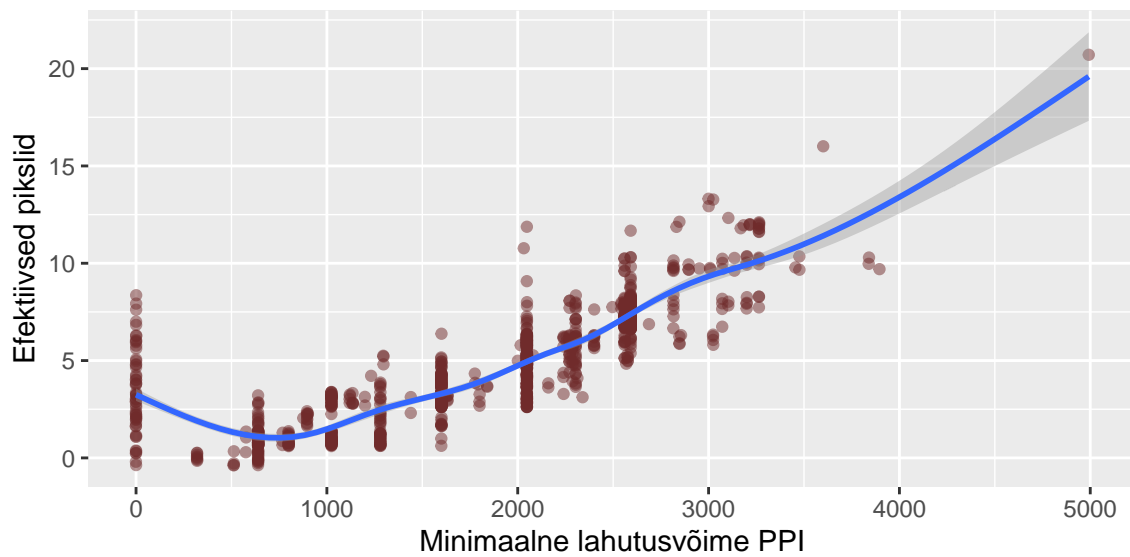
Selleks, et vaadelda seoseid kõik-võimalike arvuliste kaamera tunnuste vahel, saame konstrueerida Spearmani korrelatsioonimaatriksi

Sellest maatriksist pakuvad meil huvi 2 seost - kuidas täpsemalt käitub madalaim lahutusvõime seoses efektiivsete pikslitega ja kuidas käitub maksimaalne lahutusvõime sõltuvuses makrofookuskaugusega. Joonistame välja vastavad graafikud.

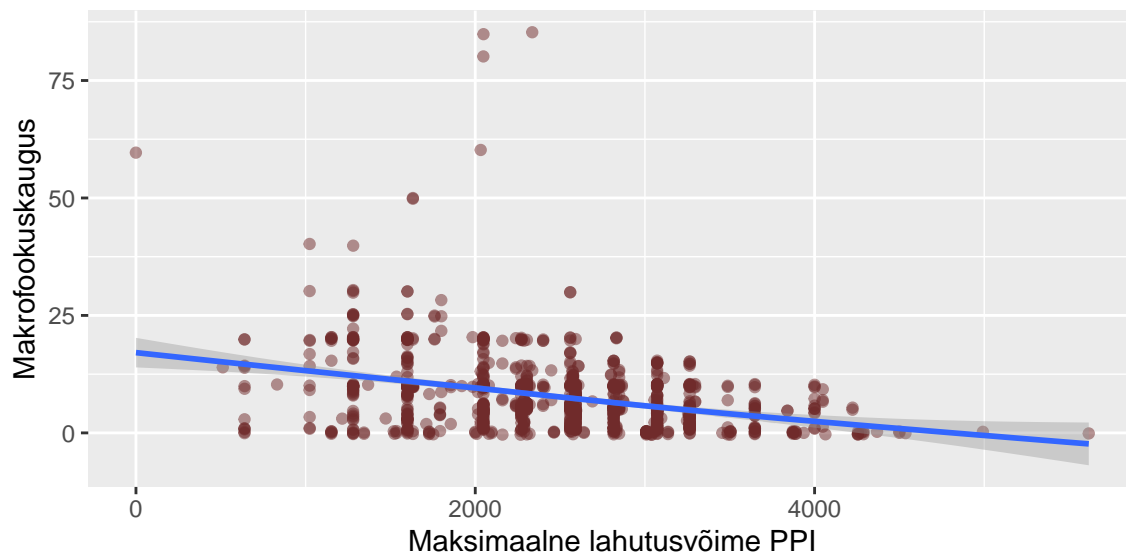
Huvitaval kombel näeme, et mida kõrgem on maksimaalne lahutusvõime, seda madalam on makrofookuskaugus. Seos ei tundu jooniselt väga tugev, kuid negatiivne trend on siiski näha.



Joonis 7: Korrelatsioonimaatriks 11 tunnuse vahel



Joonis 8: Madalaima lahutusvõime ja efektiivsete pikslite hajuvusgraafik



Joonis 9: Kõrgeima resolutsiooni ja makrofookuskauguse hajuvusgraafik

## Kokkuvõte

Projekti käigus sai uuritud 4 uurimisküsimust, millele leiti ka vastused. Esiteks tuli välja, et kuigi Jaapani firmad moodustavad enamiku kaamera-tootjatest, siis maksimaalse lahutusvõime arengu mõttes ei jää teistes riikides toodetud kaamerad maha. Erinevus on peaaegu et märkamatu.

Teiseks saime teada, et väljalaskeaasta kasvades väheneb kaamera kaal, sest nende kahe tunnuse vahel leidub keskmiselt tugev negatiivne korrelatsioon.

Kolmandaks uurisime kaamera hindade erinevust kahe tunnuse põhjal, Tuleb välja, et parema makrofookusega ja parema telesuumiga kaamerate keskmiste hindade vahe on enam kui 500 ühikut. Saime tulemuseks, et tõepoolest parema makrofookuskaugusega kaamerate hind on keskmiselt kõrgem kui parema telesuumiga kaamerate hind.

Neljandaks uurisime erinevaid seoseid kaamerate 11 arvulise omaduse vahel. Leidsime paar huvitavat esmapilgul mitte-triviaalset seost - Efektiivsete pikslite ja madala lahutusvõime vahel on väga tugev positiivne seos, samuti, et maksimaalse lahutusvõime ja makrofookuskauguse vahel on nõrk negatiivne seos.