

TARTU ÜLIKOOL
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Agnes Annilo
Patsientide analüüs
Projekt

Õppeaine: Rakendustarkvara R

Juhendaja: Kaur Lumiste

Tartu 2019

Andmete kirjeldus

Uuritavad objektid ja tunnused

Uuritavateks objektideks on 1339 Ameerika Ühendriikide elanikku. Tegemist ei ole reaalse andmestikuga, vaid see on simuleeritud Ühendriikide avaldatud statistika põhjal õppematerjaliks. Kuna andmestik on mõeldud algajatele andmeteaduse õppijatele, et õppida järel dama tunnuste põhjal nende ravikulu, ei ole andmed ilmselt tegelikule Ühendriikide elanikkonnale vastavad, olgugi et andmestik on loodud reaalse statistika põhjal.

Uuritavaid tunnuseid on seitse: "age"- inimese vanus aastates, "sex"- sugu, "bmi"- kehamassiindeks (kg/m^2), "children"- lapsi (laste arv, kes on oma vanema ravikindlustusel), "smoker"- kas suitsetaja (yes/no), "region"- isiku elamispiirkond Ühendriikides (northeast-kirre, southeast-kagu, southwest-edel, northwest-loe), "charges"- ravikindlustuse esitatud ravikulud dollarites (ilmselt kogu isiku kindlustusaja vältel ehk hetkest kui nad ei olnud enam vanematest sõltuvad).

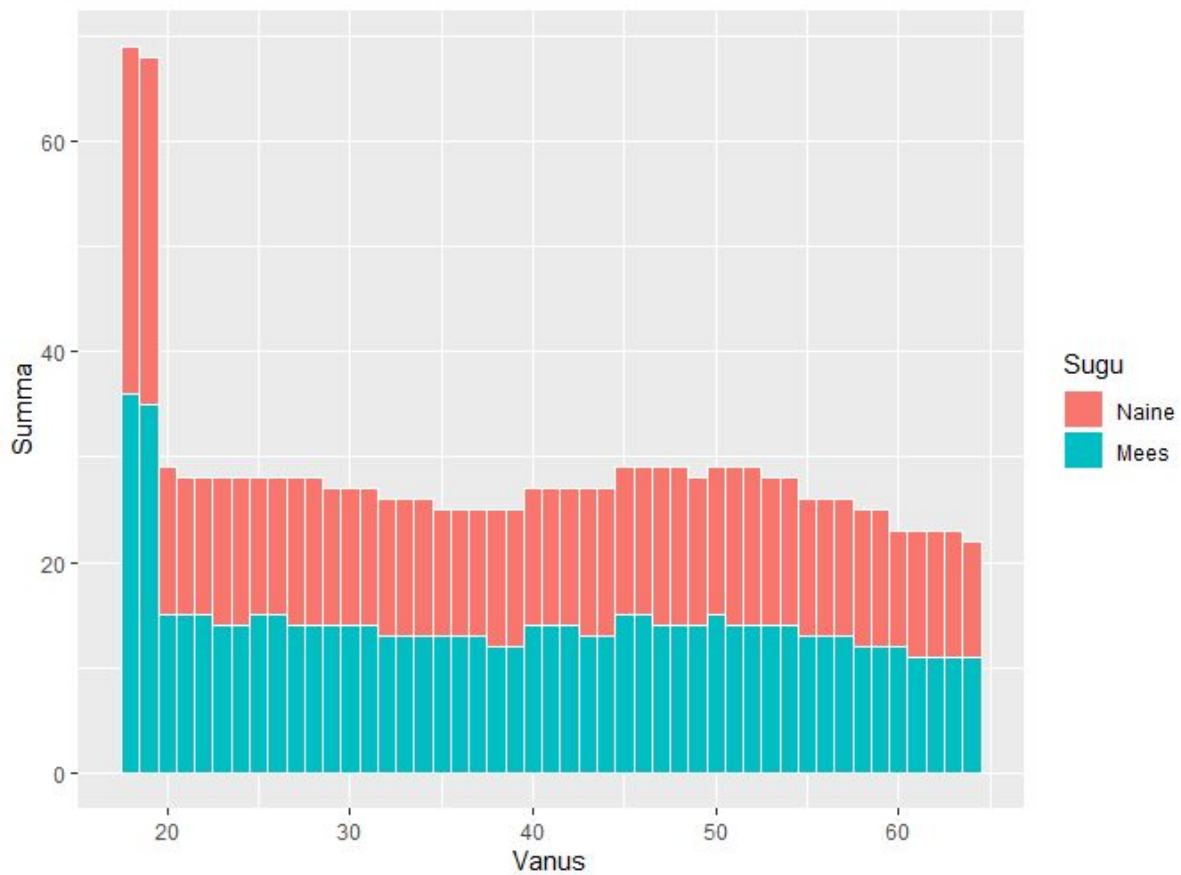
Analüüs

Üldine ülevaade andmetest

Andmestikus on inimesed vanuses 18-64. Vähim kehamassiindeks on 15.96 (18 aastane mees, mittediagnostiseeritud ja lasteta), suurim 53.13 (samuti 18 aastane mees, mittediagnostiseeritud ja lasteta). Uuritavate seas on suurim laste arv viis last, vähim ravikulu 1122 dollarit ja suurim 63770 dollarit. Uurin tunnuste mõju teistele tunnustele.

Vanus

Andmestikus on kõige enam alla 20-aastaseid, ülejäänud vanuste esindajaid on andmestikus ühtlaselt 30-ringis. Keskmine vanus on 39.21. Ka soo jaotumine vanuseti on üsna võrdne. (Joonis 1)



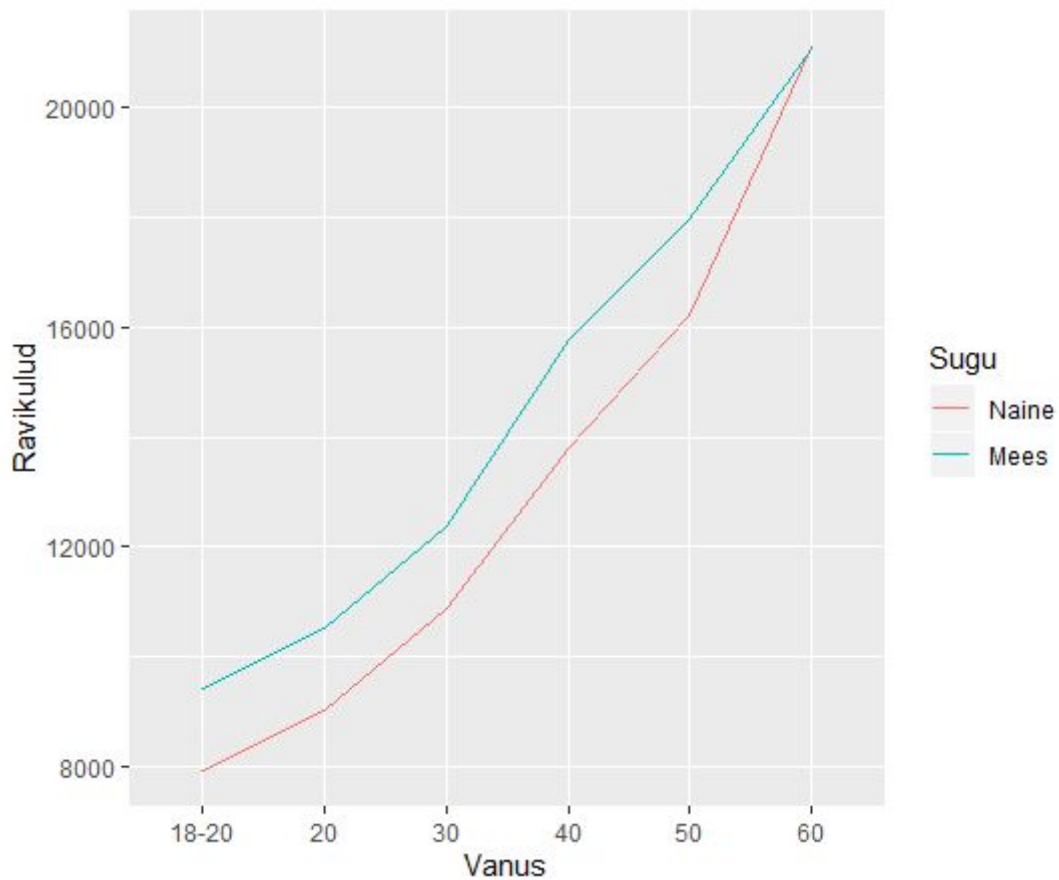
Joonis 1. Andmestikus olevate inimeste arv vanuseti

Kui jaotada inimesed 10 aastastesse vanuseklassidesse, on samuti selgelt näha, et vanemaid inimesi on uuringus üsna vähe, samas alla 20-aastaseid väga palju, 18-20 ehk kahe aasta sisse jääb 166 inimest, 10-aastastes klassides on inimesi vaid umbes poole rohkem. (Tabel 1)

Tabel 1. Inimeste koguarv andmestikus vanuseklassiti

| Vanus | 18-20 | 20 | 30 | 40 | 50 | 60 |
|-------|-------|-----|-----|-----|-----|----|
| Summa | 166 | 278 | 257 | 281 | 265 | 91 |

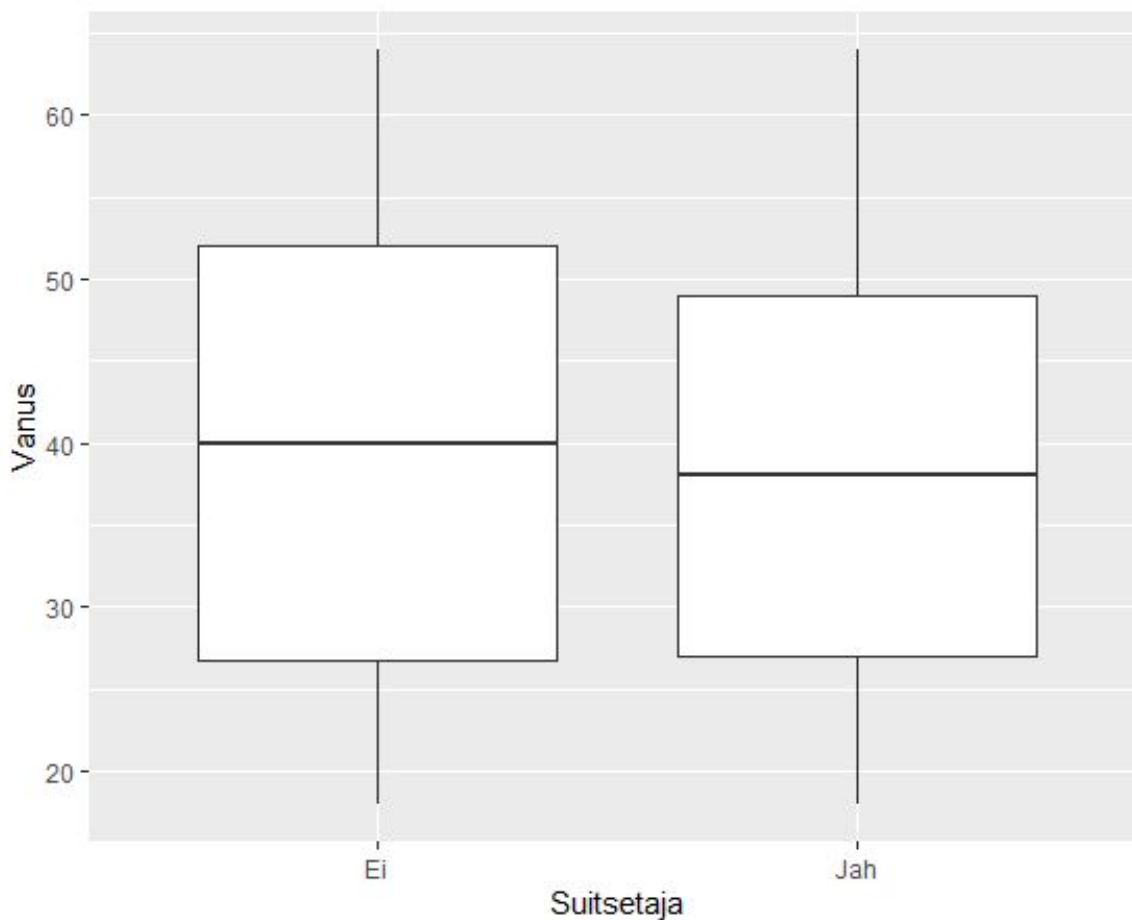
Vanuse suurenedes kasvavad keskmised ravikulud järsult, samas on see järsk kasv ühtlane, sarnanedes peaaegu lineaarsele kasvule. Meeste ravikulud on naiste omadest kõrgemad, ühtlustudes vaid üle 60 aastaste seas. (Joonis 2)



Joonis 2. Ravikulud sõltuvalt vanusest ja soost

Suitsetamine

Suitsetajaid on kokku 274, mitesuitsetajaid 1064, seejuures keskmise suitsetaja vanus on 40. On näha, et suitsetajatest 75% on nooremad kui 52 ning 25% nooremad kui 27. (Joonis 3)



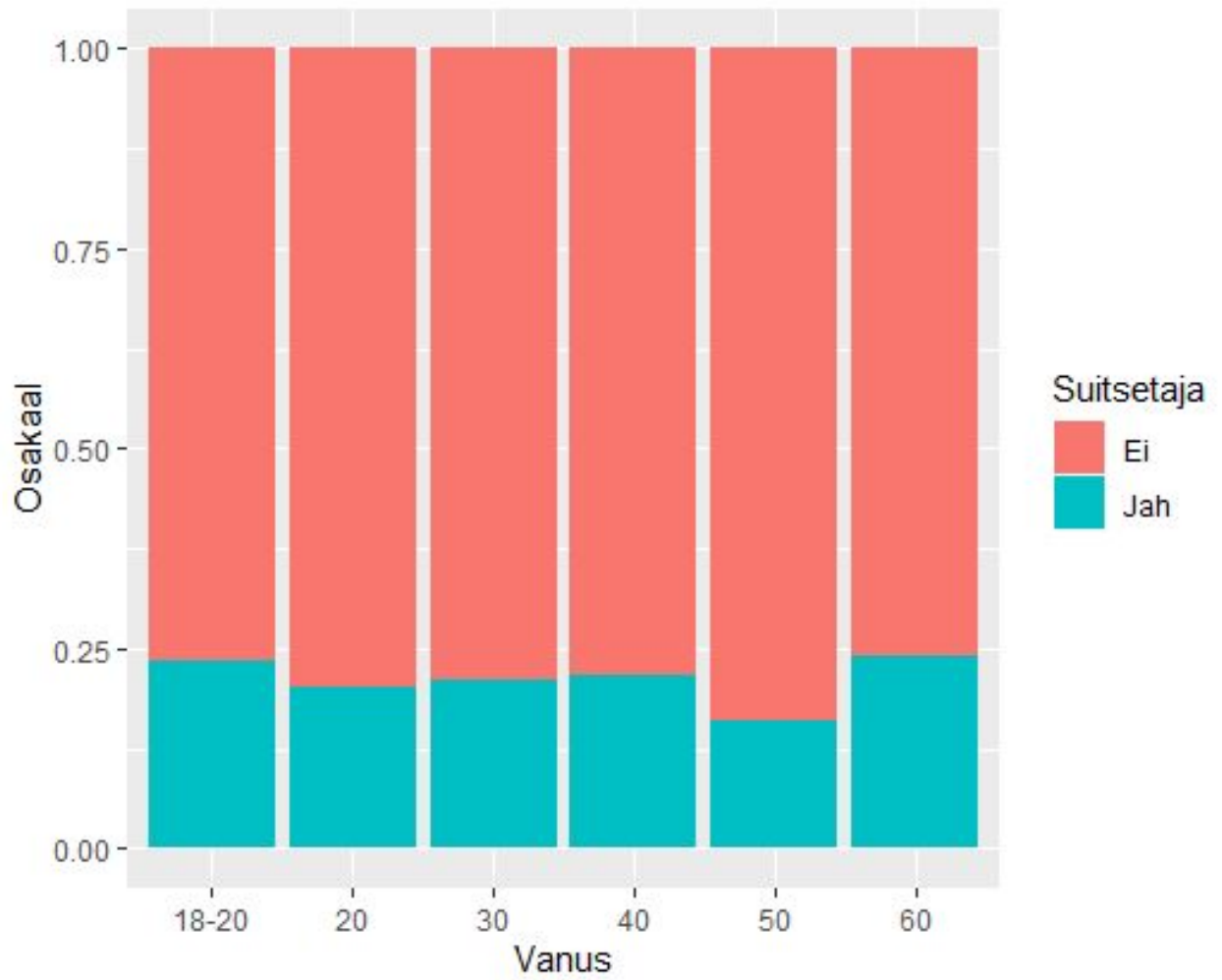
Joonis 3. Suitsetajate ja mitesuitsetajate vanuste karakteristikute võrdlus

Suitsetajate arv on sooti sarnane. Siiski suitsetavad naised vähem kui mehed, protsentuaalselt on naiste seas suitsetajaid 21%, meeste seas 30%. (Tabel 2)

Tabel 2. Suitsetajate jaotumine sooti

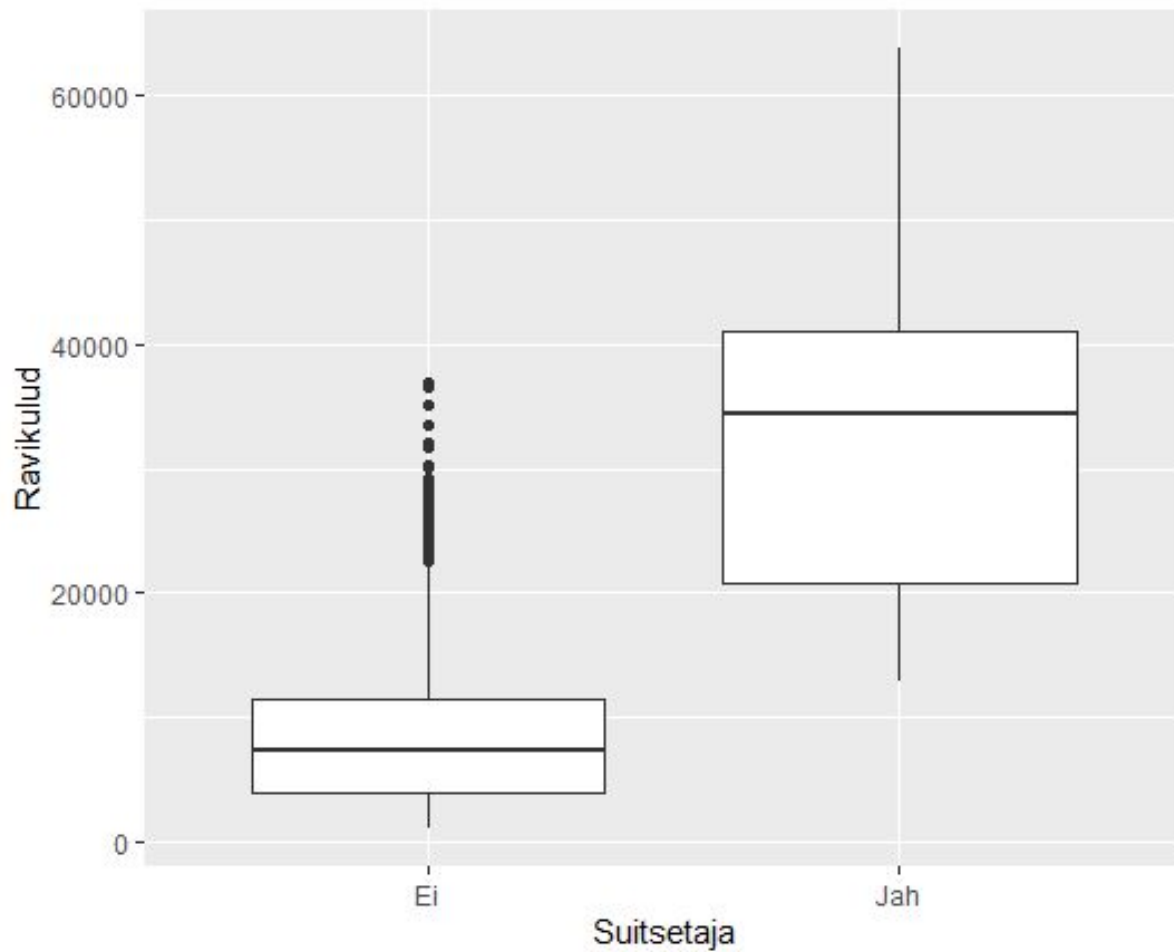
| Sugu/Suitsetaja | Ei | Jah |
|-----------------|-----|-----|
| Naine | 547 | 115 |
| Mees | 517 | 159 |

Suitsetajate osakaalust on näha, et vanuseklassiti suitsetajate arv suuresti ei erine. Langus toimub vaid 50-59 aastaste seas. (Joonis 4)



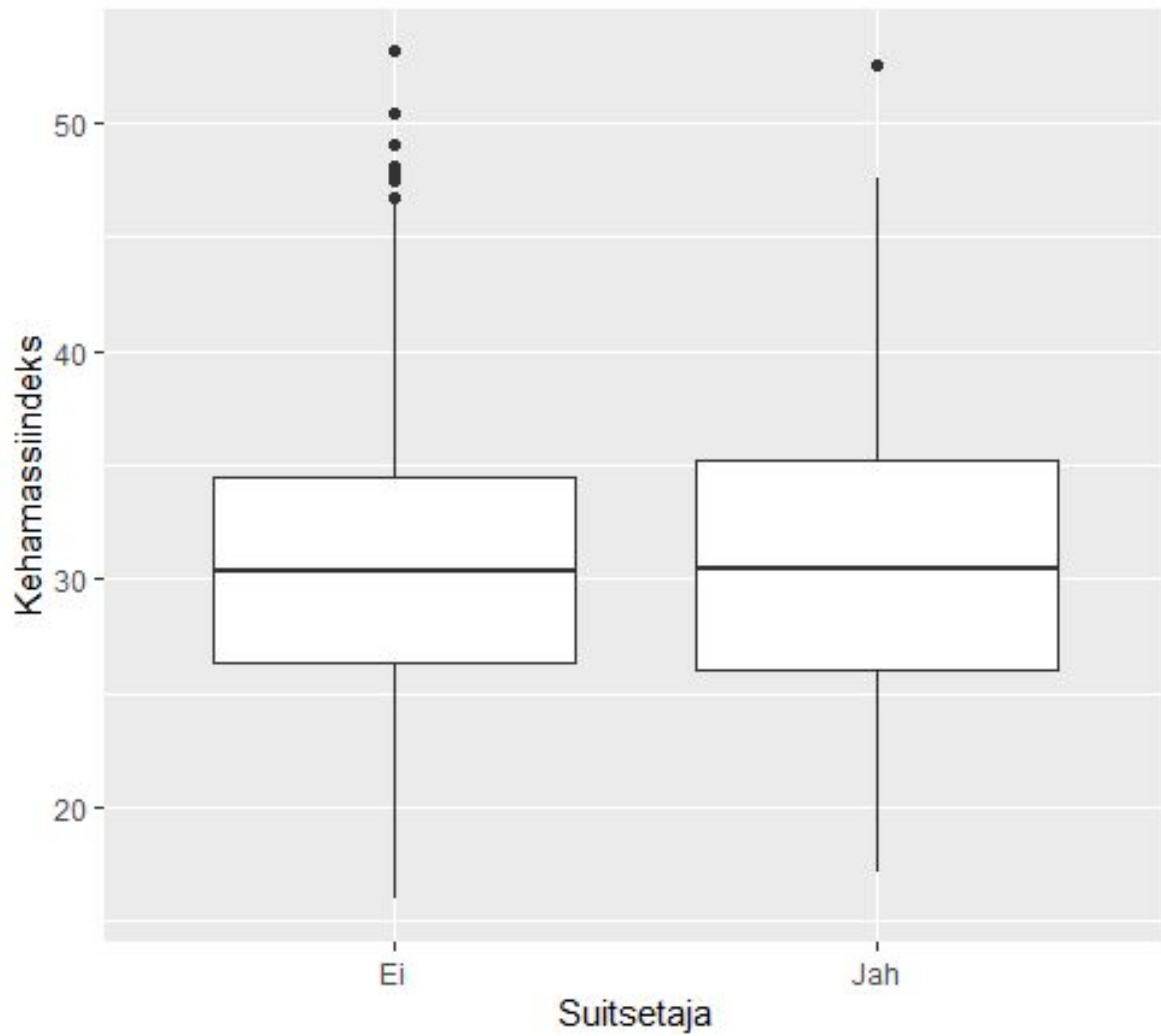
Joonis 4. Suitsetajate osakaal vanuseklassides

Jooniselt on näha, et üldiselt mitesuitsetajate ravikulud on drastiliselt väiksemad, kuid samas on ka mitmeid mitesuitsetajaid, kelle ravikulud on üldisest kõrgemad. Keskmine ravikulu mitesuitsetajatele jääb alla 10 000, suitsetajatel on keskmine üle 30 000. (Joonis 5)



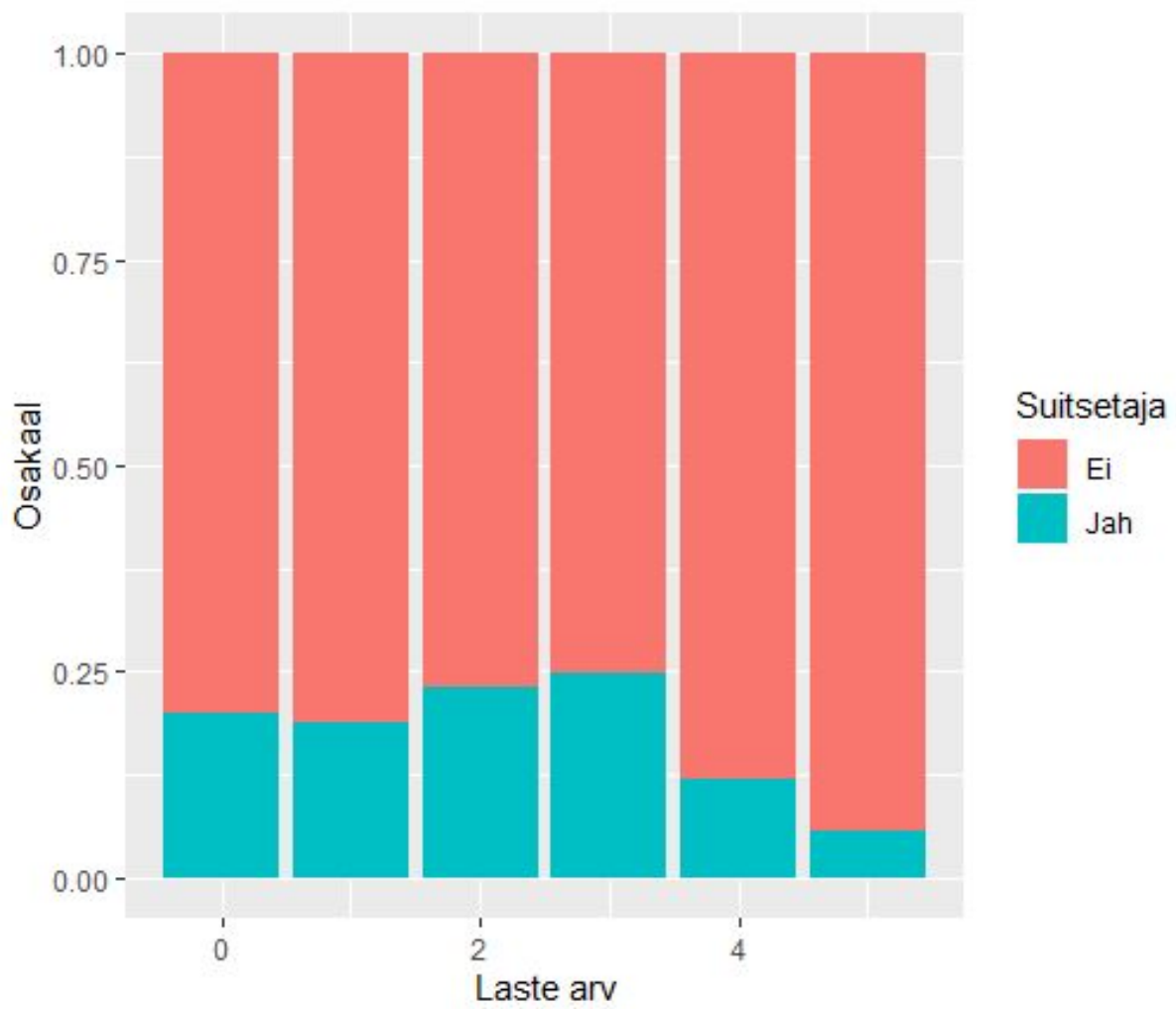
Joonis 5. Suitsetamise mõjud ravikuludele

Suitsetajate ning mittedsuitsetajate kehamassiindeksid teineteisest suurel määral ei erine.
(Joonis 6)



Joonis 6. Suitsetamise mõju kehamassiindeksile

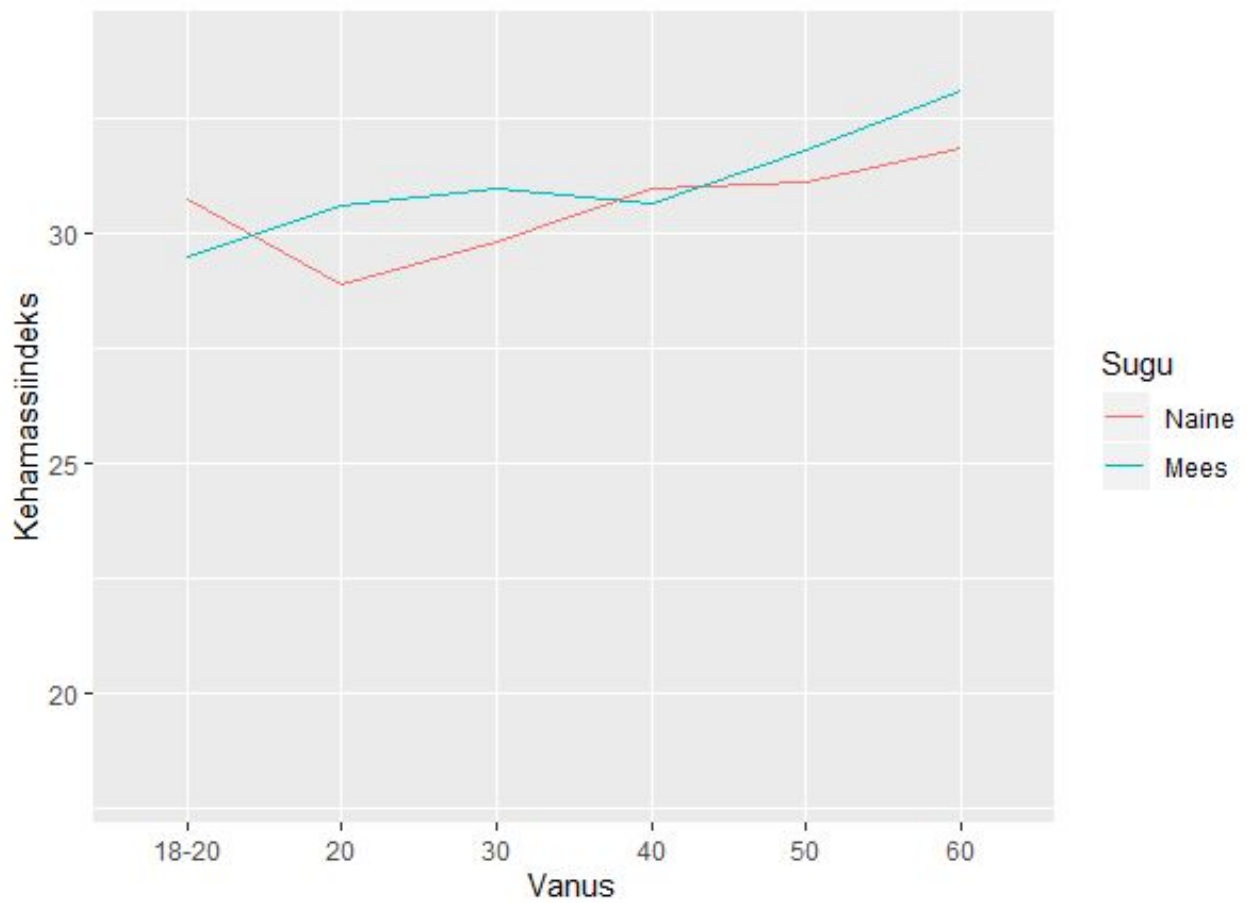
Samas on mittersuitsetajatel harva üle nelja lapse, kuigi see võib tuleneda ka sellest, et suitsetajaid on antud andmetes võrreldes mittersuitsetajatega kordades vähem. (Joonis 7)



Joonis 7. Suitsetamise mõju laste arvule

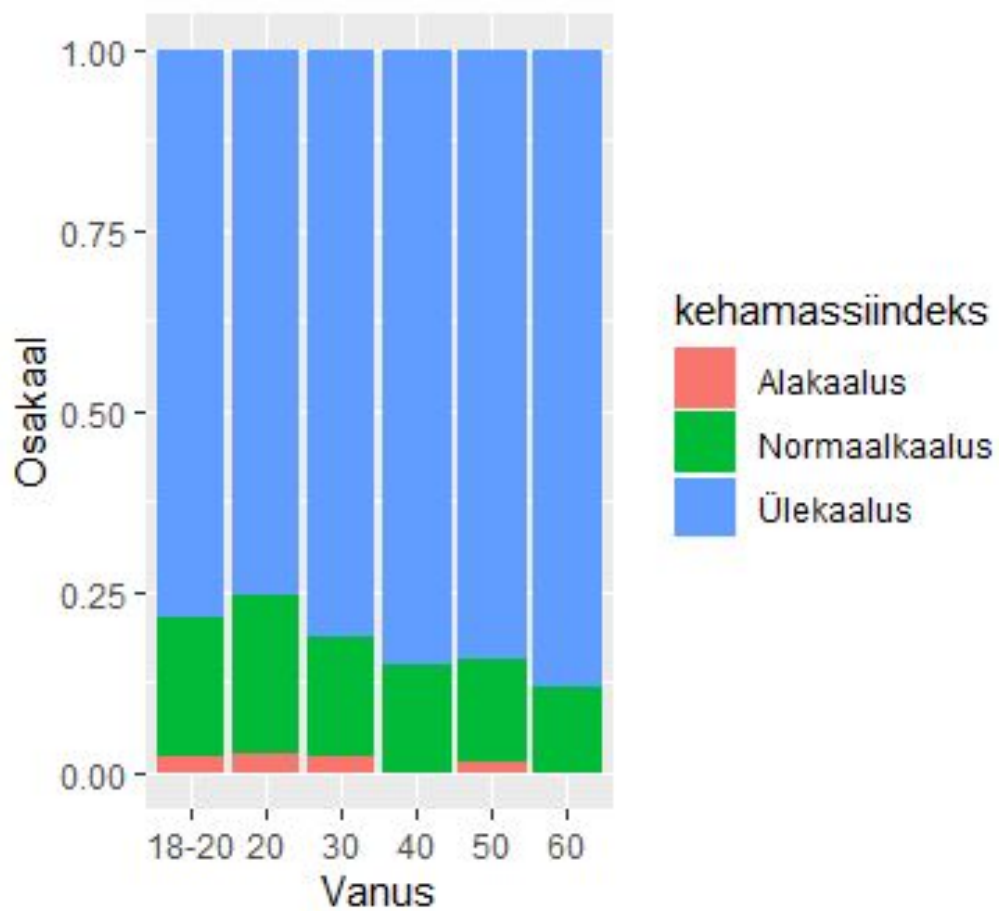
Kehamassiindeks

Naiste kehamassiindeks on üldiselt väiksem kui meestel, ometi on naiste kehamassiindeks kõrgem 18-20 aastastel ning 40-aastastel, kuigi meestel 40-aastates kehamassiindeks langeb. (Joonis 8)



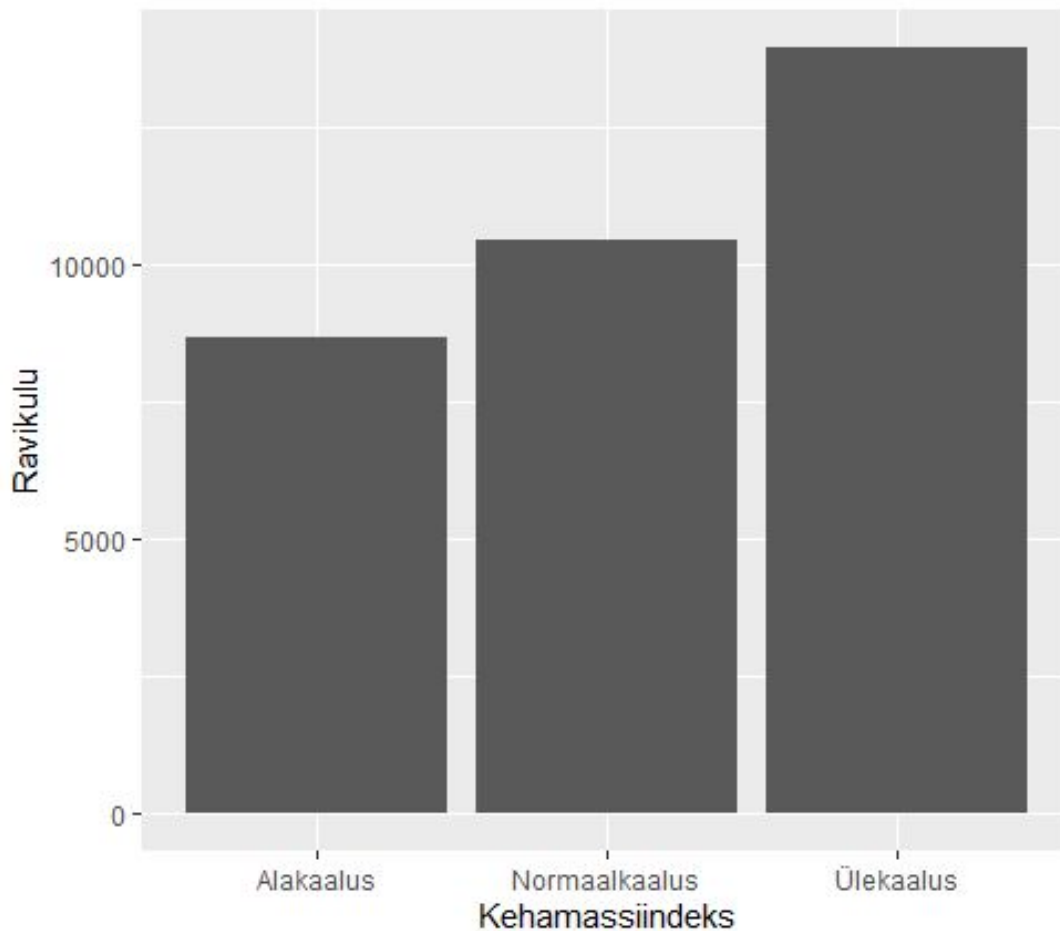
Joonis 8. Kehamassiindeks sõltuvalt vanusest ning soost

Alakaalus ning normaalkaalus isikuid on vähe, kuid ülekaalus inimesi on suur enamus. Vanuse suurenedes on rohkem ülekaalulisi, samas ka vähem, kui üldse, alakaalulisi. (Joonis 9)



Joonis 9. Kehamassiindeksi osakaal vanuseti

Keskmine ravikulu suureneb kehamassiindeksi kasvades (Joonis 10). Huvitaval kombel on alakaalus inimestel keskmiselt madalamad ravikulud kui normaalkaalulistel. Samas, eelnevatest joonistest (Joonis 9) on näha, et alakaalulised on pigem nooremad ning noorematel inimestel on madalamad ravikulud (Joonis 2). Seega nii suitsetamine (Joonis 5) kui ka kehamassiindeks mõjutavad ravikulu.



Joonis 10. Kehamassiindeksi mõju ravikuludele

Regioon

Tabelis on regioonid jaotatud Northeast, Northwest, Southeast ja Southwest. Seejuures keskmine ravikulu ei ole regiooniti ülemäära erinev, ka siis, kui jaotada regioonid North, South, East ja West. (Tabel 3)

Tabel 3. Ravikulud regiooniti

| Regioon | Ravikulu |
|-----------|----------|
| Southwest | 12346.94 |
| Southeast | 14735.41 |
| Northwest | 12417.5 |
| Northeast | 12406.38 |

Samuti on suurim kehamassiindeks on kagus (Southeast): 33.35599, kus on ka suurim ravikulu.

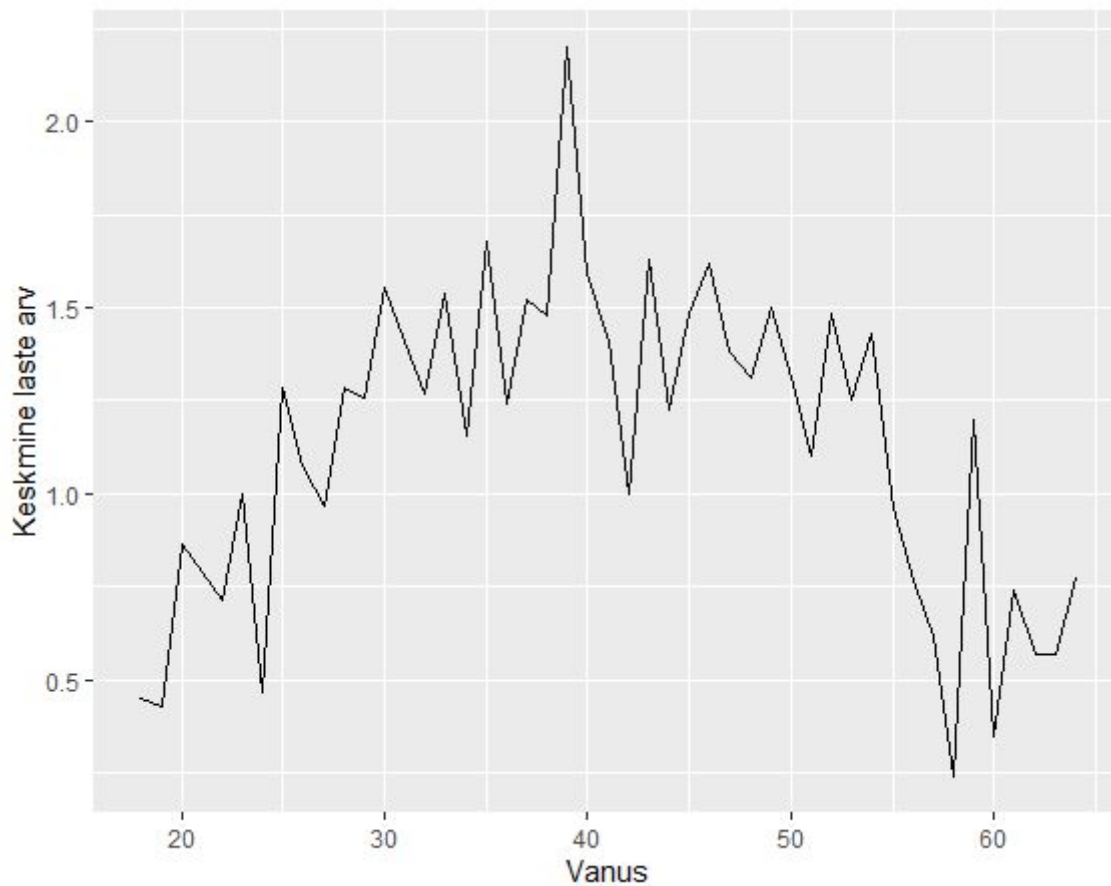
Siiski on kõige kõrgemad ravikulud Ühendriikide idaosas, seejärel lõunaosas. See klappib ka sellega, et kagus on suurim ravikulu. (Tabel 4)

Tabel 4. Ravikulud regiooniti

| Regioon | Ravikulu |
|---------|----------|
| North | 12911.22 |
| South | 13608.77 |
| East | 14109.53 |
| West | 12382.26 |

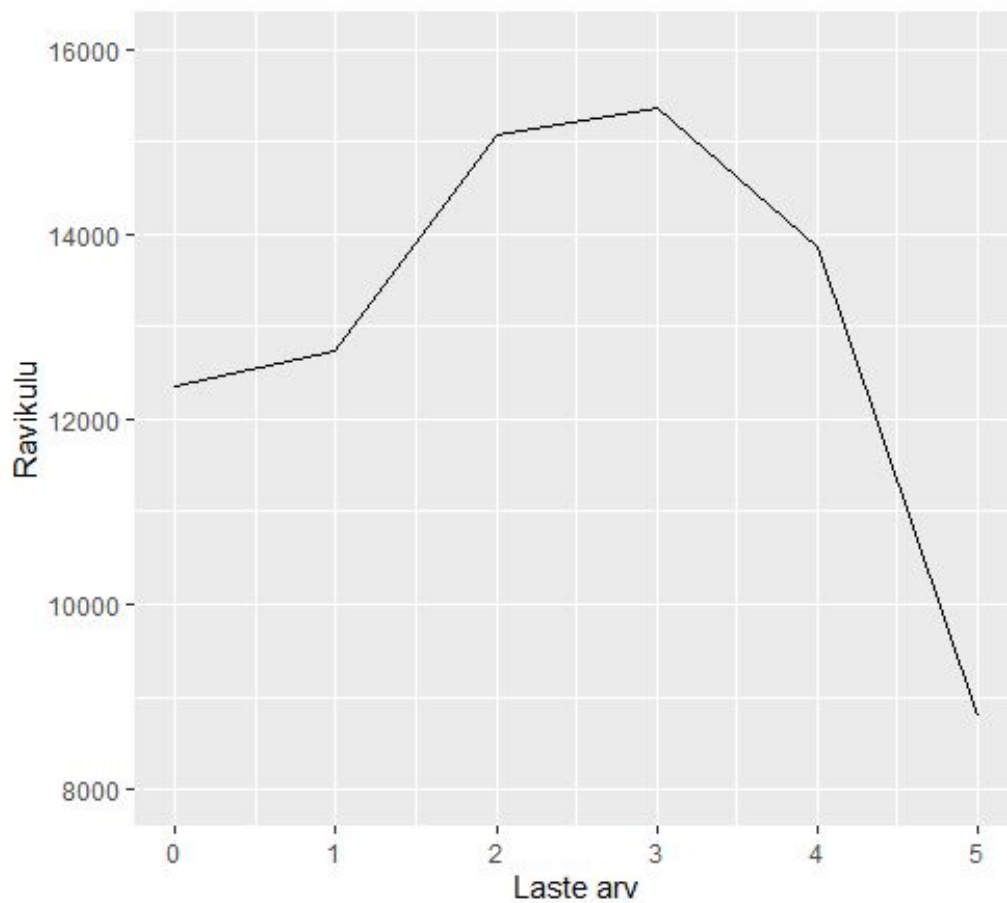
Laste arv

Keskmine laste arv on üks, ilmselt, kuna on palju inimesi, kellel ei ole lapsi ning ka palju neid, kelle lapsed enam neist ei sõltu. Ka laste moed vanuseti on enamasti nullid, seda ka 54-aastastest vanematel inimestel. Seevastu 54-aastastel inimestel on kõige tihedamini kolm, üks või null last. Noortel ja vanadel inimestel on ka keskmiselt vähem sõltuvaid lapsi. Huvitaval kombel on siiski ka üle 60 aasta vanuseid inimesi, kellel on endast sõltuvaid lapsi. (Joonis 11)



Joonis 11. Vanemast sõltuvate laste keskmine arv sõltuvalt vanusest

Kuna valimis on vaid välja toodud lapsed, kes on sama tervisekindlustuse poolt kaetud nagu nende vanem, on huvitav, et keskmine ravikulu kasvab laste arvuga kuni kolme lapseni, seejärel langeb järsult. Suurtes peredes viie lapsega on ravikulud mitmetuhande dollari võrra väiksemad kui vähemate lastega peredes, samas katab sama ravikindlustus kuute inimest. (Joonis 12)



Joonis 12. Ravikulu sõltuvalt laste arvust