

Tartu Ülikool  
Matemaatika ja statistika instituut

# **Riikide võrdlus CIA World Factbook näitajate abil**

LTMS.00.016. Rakendustarkvara: R

Ilmar Uduste

<b>Sissejuhatus</b>	<b>3</b>
<b>Hüpoteesid</b>	<b>4</b>
<b>Andmete eeltöötlus</b>	<b>5</b>
<b>Analüüs</b>	<b>5</b>
Hüpotees nr 1. Populatsiooni kirjaoskuse seos sündimusega	7
Hüpotees nr 2. SKT seos telefonide arvuga elaniku kohta	8
Hüpotees nr 3 ja 4. Rändesaldo	9
Hüpotees nr 5. SKT ja teenindussektoris töötavate inimeste osakaalu seos	10
<b>Kokkuvõte</b>	<b>11</b>
LISA 1. Riikide jaotumine maailmajagudesse	12

# Sissejuhatus

Õppeaine LTMS.00.016. Rakendustarkvara: R raames oli lõpuprojekti jaoks tarvis uurida mingisugust andmestikku, pannes paika mingid hüpoteesid ja siis analüüsi käigus neid kinnitada või ümber lükata.

Otsustasin uurida riikide näitajaid CIA (*Central Intelligence Agency*) World Factbookist ning andmestik oli kättesaadav andmeteadust propageerivalt veebilehelt Kaggle.com. Andmed ise pärinevad aastast 2017 (World Factbook, Central Intelligence Agency, USA, 2017, <https://www.kaggle.com/fernandol/countries-of-the-world>).

Andmestik sisaldab endas 227 riigi andmeid ja tabelis väljatoodud näitajad (mida oli 20 veergu) olid järgmised:

- **Country** (Riigi nimi),
- **Region** (maailmajagu),
- **Population** (rahvastiku suurus),
- **AreaSqMi** (Riigi pindala ruutkilomeetrites),
- **PopDensityPerSqMi** (rahvastiku tihedus ruutkilomeetri kohta),
- **CoastlineCoastAreaRatio** (ranniku ja rannikuala suhe),
- **NetMigration** (migrantide arv 1000 elaniku kohta, rändesaldo),
- **InfantMortalityPer1000** (imikute suremus 1000 sünnituse kohta),
- **GDPPerCapita** (SKT elaniku kohta),
- **Literacy** (kirjaoskus),
- **PhonesPer1000** (telefonide arv 1000 elaniku kohta),
- **Arable, Crops, Other, Climate**,
- **Birthrate** (sündimus 1000 elaniku kohta),
- **Deathrate** (suremus 1000 elaniku kohta),
- **Agriculture** (põllumajandussektoris töötavate inimeste osakaal),
- **Industry** (tööstussektoris töötavate inimeste osakaal),
- **Service** (teenindussektoris töötavate inimeste osakaal).

Ülaltoodud näitajate põhjal formuleerisin viis hüpoteesi, mis on välja toodud järgmises peatükis.

# Hüpoteesid

Antud töö käigus uurisin püstitasin ja uurisin järgmisi hüpoteese:

## **1. Mida suurem on kirjaoskus, seda väiksem on sündimus.**

Üldiselt on sündimus ülimalt suur just arengumaades, kus on keskmisest madalam kirjaoskus, ning arenenud riikides, kus on ülimalt kõrge kirjaoskus (peaaegu 100%), on alatihti probleeme madala sündimusega.

## **2. Mida suurem on SKT, seda rohkem on inimestel telefone.**

Mida suurem on SKT, seda tõenäolisemalt on riigi elanikel piisavalt finantsilisi vahendeid, et endale telefon soetada.

## **3. Suurim rändesaldo on Euroopa riikide seas.**

See mõte on tingitud Euroopas suhteliselt soodsate elamistingimustest ja ümbritsevates maailmajagudes toimuvatest konfliktidest ja ebasoodsatest elamistingimustest.

## **4. Väikseim rändesaldo on riikides, kus on suur osakaal põllumajandussektoris töötavatel inimestel.**

Inimesed üritavad asuda elama sinna, kus on soodsamad elamistingimused, ning tihtipeale on suure põllumajandussektoris töötavate inimeste osakaaluga riikides just ebasoodsad elamistingimused. Sellest võiks siis olla tingitud nende riikide negatiivne rändesaldo.

## **5. Mida suurem on SKT, seda suurem on teenindussektoris töötavate inimeste osakaal.**

See mõte tuleb seosest, et tihtipeale on arenenumates riikides väiksem osakaal rahvastikust töötamas põllumajandussektoris ja rohkem kontoris, mis üldiselt liigitub teenindussektori alla.

Ülalolevate hüpoteeside paikapidavust uurin andmestikuga järgmises peatükis.

# Andmete eeltöötlus

Andmestikku uurides ilmnes probleem, et kõik andmed kasutasid kümnendkoha eraldajana sümbolit “,” (koma), samuti väljade eraldamiseks “,” ning andmepunktide eristamiseks jutumärke, kuigi programmeerimiskeel R vaikumisi tunnistab kümnendkoha eraldajana sümbolit “.” (punkt). Sellest tingituna oli tarvis andmestiku arvud teisendada sobivale kujule - seda protsessi on kujutatud projektiga kaasas käivas R failis.

Probleeme tekkis ka andmestikus kasutatavate maailmajagude klassifikatsiooniga - maailmajagused oli esialgu liiga palju (kokku 10), et erinevaid maailmajagused oleks olnud võimalik arusaadavalt kujutada/grupeerida graafikul. Seega pidi vähendama maailmajagude arvu (lõpuks 6 maailmajaod peale) ning selle jaoks koondasin teatud maailmajaod kokku üheks suuremaks maailmajaoks (nt regioonid “WESTERN EUROPE”, “EASTERN EUROPE” ja “BALTICS” läksid kõik Euroopa maailmajaod alla). Andmestikus tekitasin riigi üldisema regiooni kuvamiseks uue veeru *RegionGeneral*, mille jaotamise algoritmi loogikat on selgitatud tabelis 1.

Tabel 1. Riikide jaotamine maailmajagudesse.

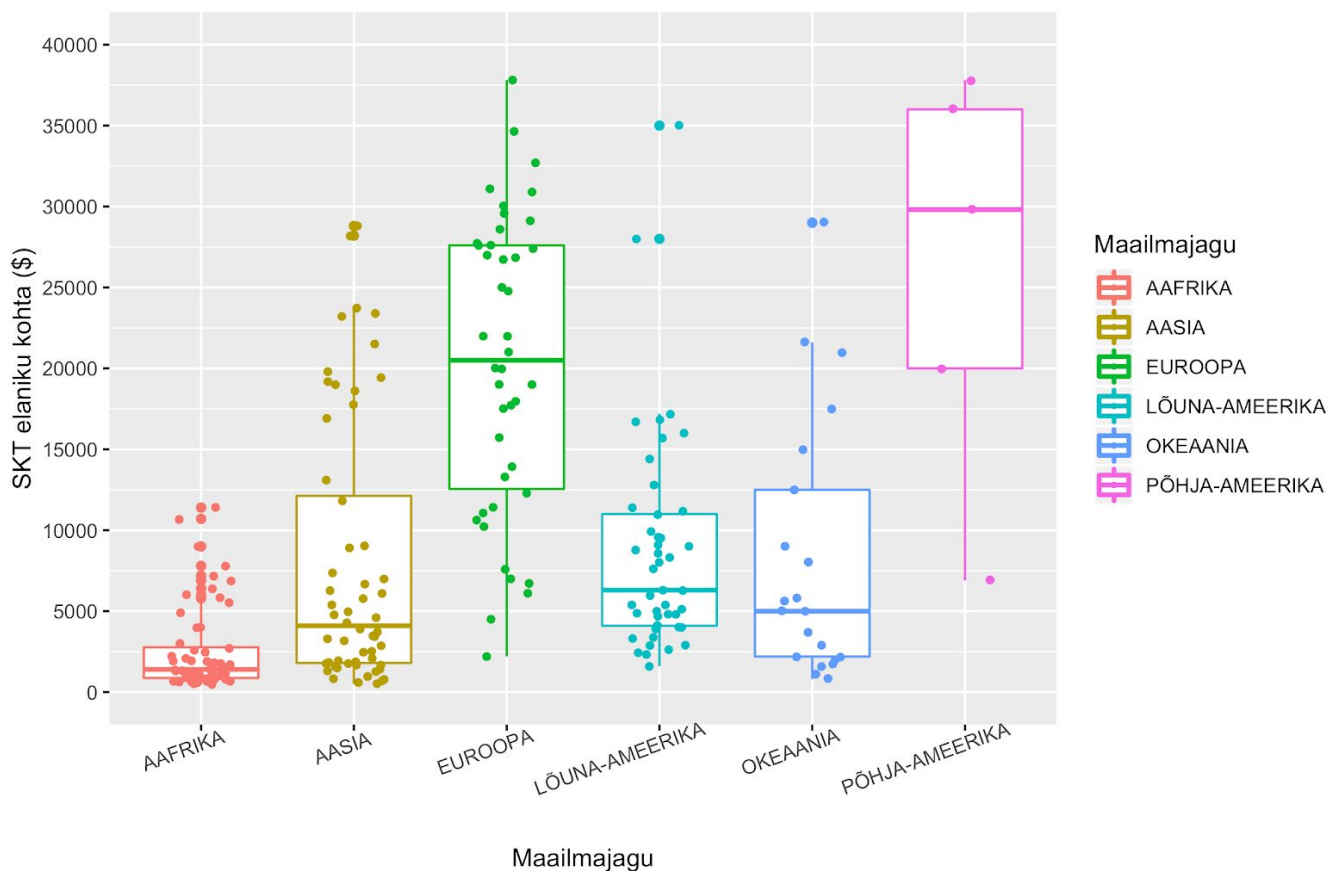
Andmestiku maailmajaod	Töös kasutatav maailmajagu	Riikide arv
Northern Africa Sub-Saharan Africa	Aafrika	57
Asia (Ex. Near East) C.W. of Ind. States (CIS) Near East	Aasia	56
Baltics Eastern Europe Western Europe	Euroopa	43
Latin Amer. & Carib	Lõuna-Ameerika	45
Northern America	Põhja-Ameerika	21
Oceania	Okeaania	5

Täielik nimekiri maailmajagudest ja nendes asuvatest riikidest leidub Lisas 1.

Põhjalikum andmete analüüs asub järgmises peatükis.

# Analüüs

Enne hüpoteeside uurimist oleks tarvis uurida andmeid üldisemalt. Suurt huvi pakub riikide SKT-de jaotus maailmas ning selle kuvamiseks konstrueerisin *Joonise 1*.



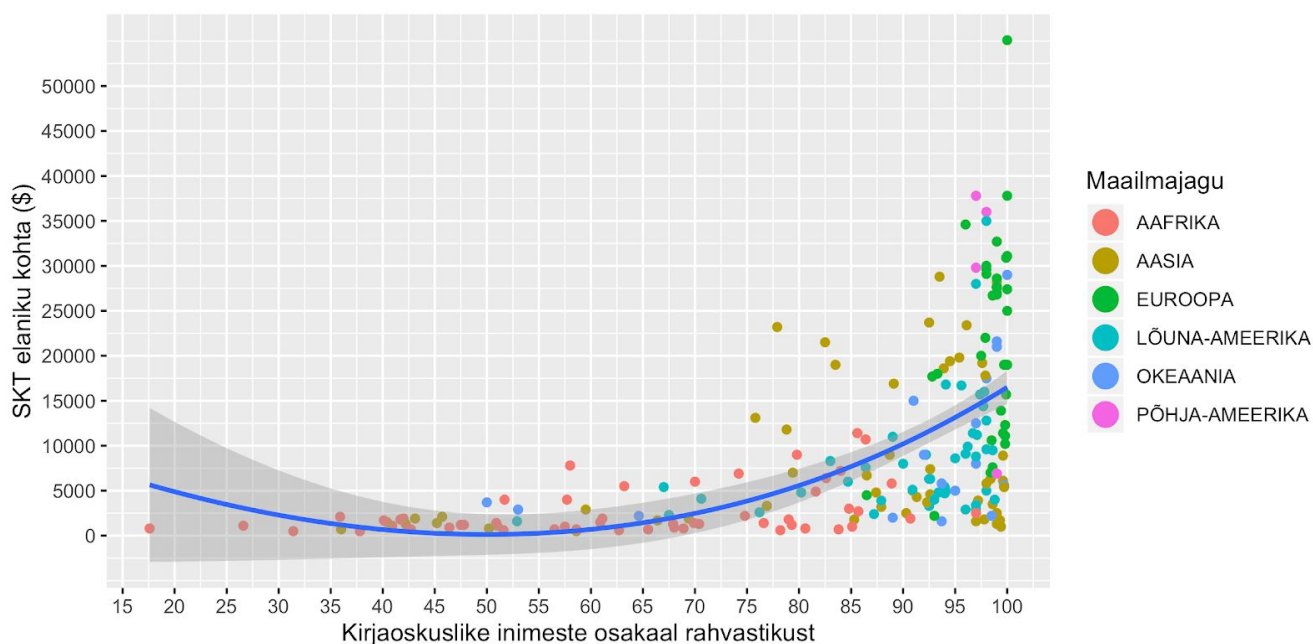
*Joonis 1.* Riikide sisemajanduse kogutoodangu jaotus suuruse ja maailmajao järgi.

Suurem osa riikidest jääb oma sisemajanduse kogutoodang elaniku kohta poolest vahemikku 0 - 10 000 \$ ning enamik sellesse vahemikku kuuluvatest riikidest asuvad kas Aafrikas või Aasias. Tasub mainida, et riikide SKT mediaan on 5550\$ ja keskmine 9690\$. Kõige suurem variatsioon piirkonnasiseselt tundub olevat kas Aasias või Euroopas, kuigi Aasias on suurem osa riike madalama SKT-ga. Võib aga kindlalt öelda, et antud maailmajagude jaotusega on Aafrika riigid kõige vähem rikkamad maailmas. Üldiselt tundub, et Euroopa ning teatud Aasia riigid on kõige rikkamad, ühtlasi on ka mõni Põhja-Ameerika, Lõuna-Ameerika ja Okeania riik üsnagi jõukas.

Järgnevalt uurin esialgselt püstitatud hüpoteese, alustades kirjaoskuse ning riigi jõukuse seosest.

## Hüpotees nr 1. Populatsiooni kirjaoskuse seos sündimusega

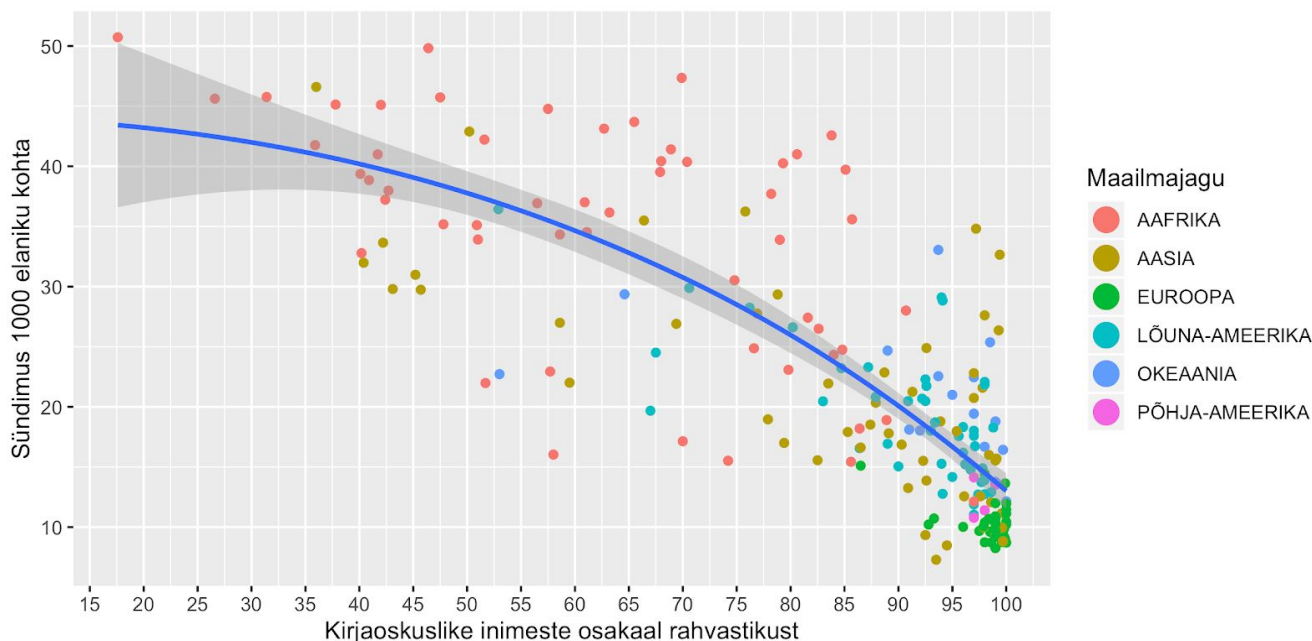
Esiialgu uurin, kuidas on riigi elanike kirjaoskus seotud riigi sisemajanduse kogutoodanguga, et täpsemalt saada aimu seosest ligikaudselt riigi elanike heaolu ja kirjaoskuse vahel. Joonisel 2 on näha korrelatsiooni ja üpriski suure tõenäosusega kehtib järgmine seos: mida suurem on kirjaoskus, seda suurem on SKT. Eelnev seos aga kehtib vaid teatud piirini, sest eksisteerib ka vaeseid riike, mille kirjaoskuslike elanike osakaal on ligilähedane 100%-le. Kõige rikkamatel riikidel on aga kindlalt üle 95% elanikkonnast kirjaoskuslikud.



Joonis 2. SKT sõltuvus kirjaoskuslike inimeste osakaalust riigiti ja maailmajaoti.

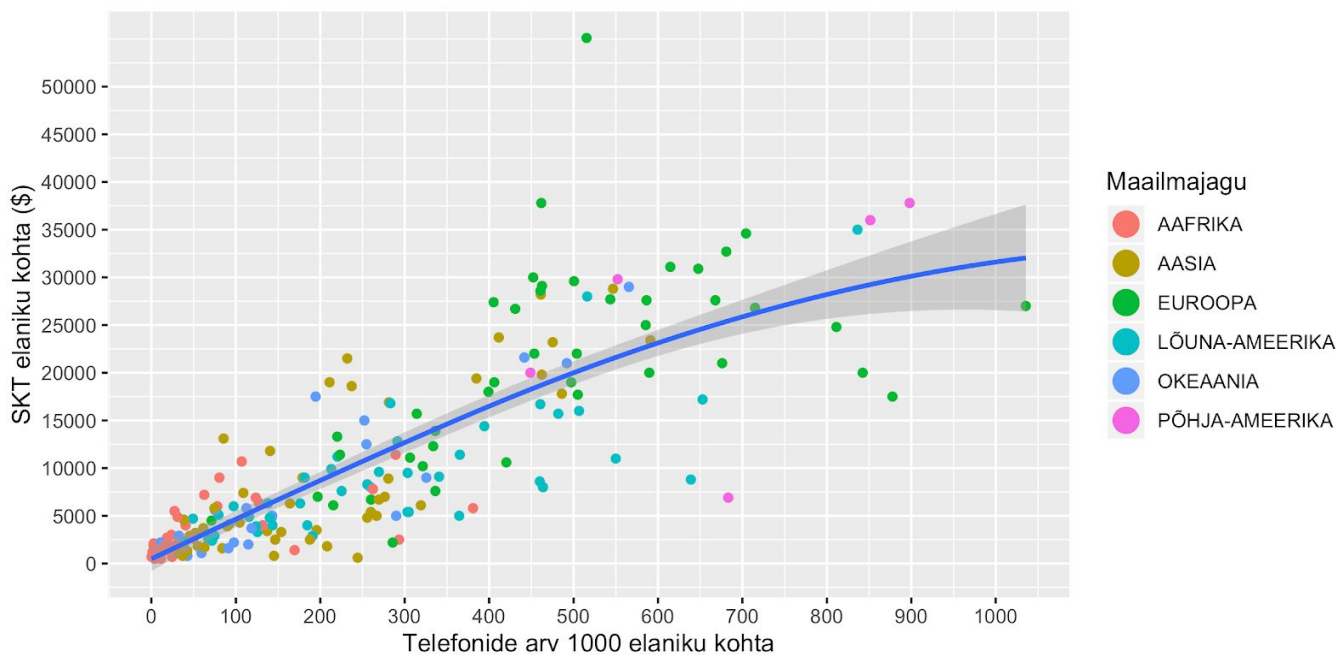
Joonisel 3 on välja toodud sündimuse seos riikide populatsioonide kirjaoskusega ning sealt on võimalik välja lugeda mitut asjaolu. Esiteks, madala kirjaoskusega riigid on enamasti kas Aafrikas või Aasias. Teiseks, Euroopa riigid, kus on enamjaolt suure kirjaoskusega populatsioonid, on maailmas kõige väiksem sündimus. Kõige väiksema kirjaoskusega riigid on samuti ka kõige suurema sündimusega maailmas. Kirjaoskuse mediaan riigiti on 92,5% ja keskmine 82,8%.

Võib väita, et esialgne hüpotees kehtib. **Mida suurem on kirjaoskus, seda väiksem on sündimus.**



Joonis 3. Sündimuse sõltuvus kirjaoskuslike inimeste osakaalust riigiti ja maailmajaoti.

## Hüpotees nr 2. SKT seos telefonide arvuga elaniku kohta



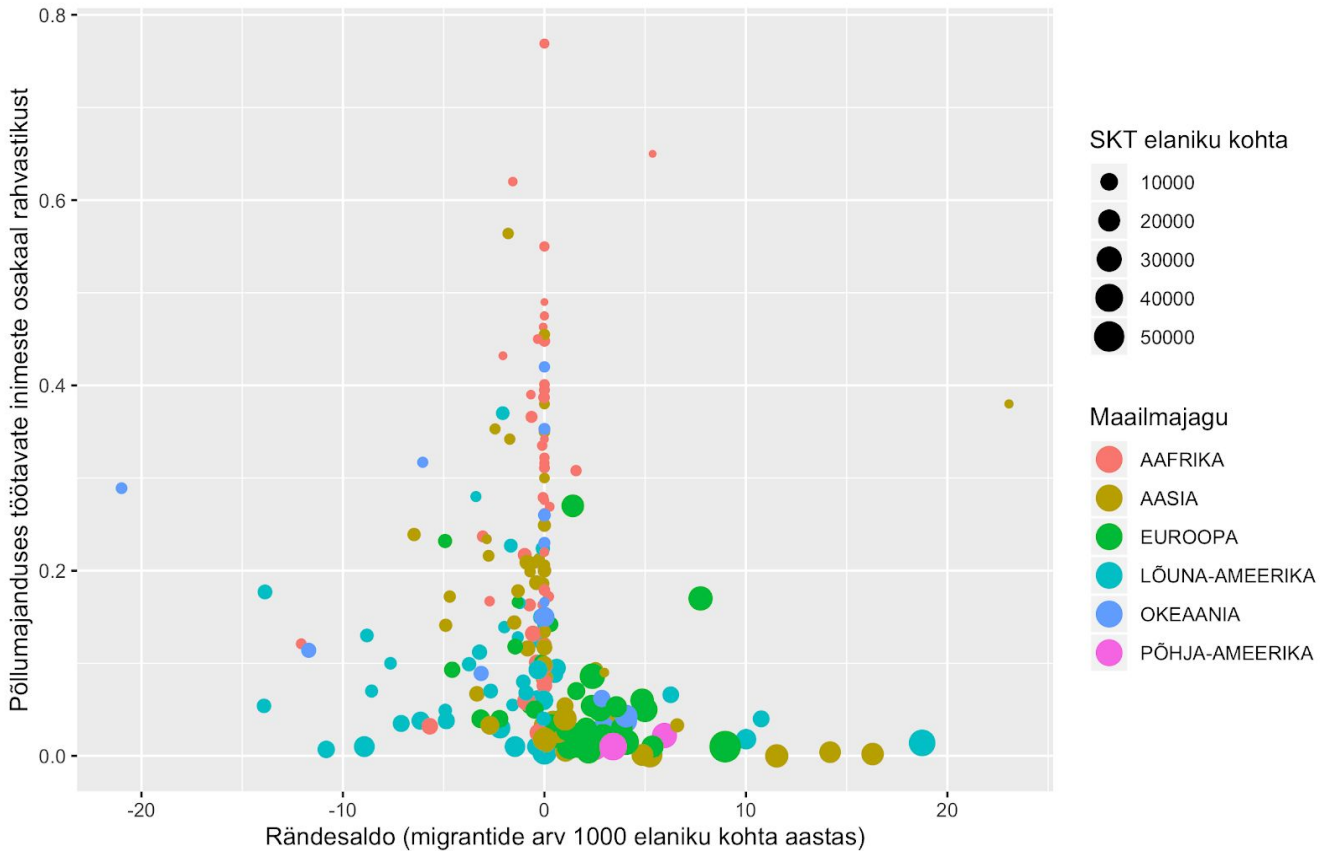
Joonis 4. SKT sõltuvus kirjaoskuslike inimeste osakaalust riigiti ja maailmajaoti.

Riigi finantsilist heaolu on võimalik ka mõõta telefonide arvuga, seda seost sümboliseerib Joonis 4. Taaskord on Aafrika riigid madalaima SKT ning telefonide arvuga elaniku kohta. Muude maailmajagudega on hajuvus suurem, kuid üldjoontes on Euroopa riikides ikka suurem SKT ning suurem telefonide arv elaniku kohta. Maailmas on riigiti telefonide arv 1000 elaniku kohta mediaan 176 ja keskmine 236. Graafikul on näha küll mõningaid riiki,



mis asuvad väljaspool standardhälvet, kuid võib kindlalt öelda, et varasemalt püstitatud seos kehtib. **Mida suurem on SKT, seda rohkem on inimestel telefone.**

### Hüpotees nr 3 ja 4. Rändesaldo



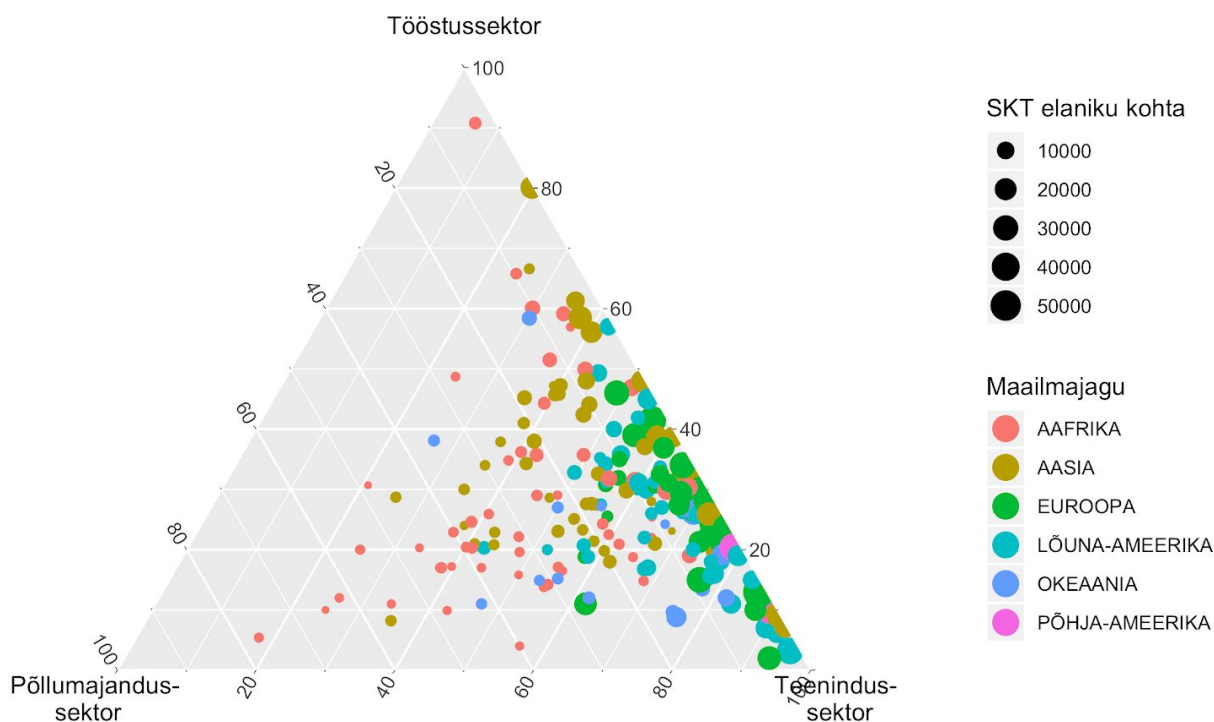
Joonis 5. Rändesaldo sõltuvus põllumajandussektoris töötavate inimeste osakaalust riigiti ja maailmajaoti, kusjuures andmepunkti suurus viitab SKT suurusele elaniku kohta.

Kui vaadata rändesaldot maailmajao järgi (Joonis 5), siis tundub, et Euroopa riigid on terve klastrina teistest maailmajagudest kõrgema rändesaldoga, isegi kui suurima rändesaldoga on mõned Lõuna-Ameerika ja Aasia riigid. Hüpotees nr 3 leiab kinnitust. **Suurim rändesaldo on Euroopa riikide seas.**

**Küll aga ei ole võimalik kinnitada hüpoteesi nr 4.** Põllumajandussektoris töötavate inimeste osakaal on eriti suur just Aafrika ning teatud Aasia riikides, kuid nendel riikidel on rändesaldo suhtes vigane info. Andmestikus on teatud riikidel rändesaldo väli puudu ja paljudel on see väärtusega "0". On üpriski vähetõenäoline, et nii suurel hulgal Aafrika riikidel on rändesaldo täpselt 0, kuid vale jätta need andmepunktid välja, sest nad on siiski andmestikus mingisuguse väärtusega. Seetõttu on vastavad andmepunktid Joonisel 5 esindatud.

## Hüpotees nr 5. SKT ja teenindussektoris töötavate inimeste osakaalu seos

Kuna riigi töösektoreid jagatakse primaarseks (põllumajandus), sekundaarseks (tööstus) ja tertsiaarseks (teenindus) sektoriteks ning iga riigi sektorite osakaalude summad moodustavad terviku, siis on võimalik nendest töösektoritest teha kolmnurkdiagramm (Joonis 6).



Joonis 6. Kolmnurkdiagramm inimeste tööhõivest erinevate töösektorites (põllumajandus-, teenindus- ja tööstussektor), kusjuures andmepunkti suurus viitab SKT suurusele elaniku kohta ja andmepunkti värv viitab maailmajaole, kus konkreetne riik asub.

Kolme töösektori jaotuse uurimisel ilmneb asjaolu, et suurem osa rikkastest riikidest on minimaalse põllumajandussektori hõivega. Aasia riikides on tööstus suurema osatähtsusega kui mujal maailmajaos, aga Euroopas tundub teenindussektor kõige tähtsama kaaluga olevat. Aafrika riikide puhul on tegemist madala SKT-ga elaniku kohta, kuid kõrge hõivega kas põllumajandus- või tööstussektorites.

Riikides, kus on teenindussektori hõive suurem kui miski muu, on SKT elaniku kohta suurem kui kuskil mujal. Ehk hüpotees nr 5. leiab kinnitust. **Mida suurem on SKT, seda suurem on teenindussektoris töötavate inimeste osakaal.**

# Kokkuvõte

Käesoleva projekti raames uuriti riikide näitajaid, mida kogutakse igaaastaselt CIA (*Central Intelligence Agency*) poolt. Andmestik sisaldas endas 227 riigi andmeid ning andmetes esinevate näitajate põhjal pandi paika 5 hüpoteesi, millest 4 osutusid tõeseks.

Hüpoteesid olid järgmised:

1. Mida suurem on kirjaoskus, seda väiksem on sündimus. **Hüpotees osutus tõeseks.**
2. Mida suurem on SKT, seda rohkem on inimestel telefone. **Hüpotees osutus tõeseks.**
3. Suurim rändesaldo on Euroopa riikide seas. **Hüpotees osutus tõeseks.**
4. Väikseim rändesaldo on riikides, kus on suur osakaal põllumajandussektoris töötavatel inimestel. **Hüpoteesi polnud võimalik kinnitada.**
5. Mida suurem on SKT, seda suurem on teenindussektoris töötavate inimeste osakaal. **Hüpotees osutus tõeseks.**

Andmetöötlus ning andmete visualiseerimine tehti rakendustarkvaraga R ja teekidega *ggplot2* ning *ggtern*.

# LISA 1. Riikide jaotumine maailmajagudesse

**AAFRIKA:** Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Rep. Chad, Comoros, Congo, Dem. Rep. Congo, Repub. of the Cote d'Ivoire, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Saint Helena, Sao Tome & Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Western Sahara, Zambia, Zimbabwe

**AASIA:** Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Belarus, Bhutan, Brunei, Burma, Cambodia, China, Cyprus, East Timor, Gaza Strip, Georgia, Hong Kong, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, North Korea, South Korea, Kuwait, Kyrgyzstan, Laos, Lebanon, Macau, Malaysia, Maldives, Moldova, Mongolia, Nepal, Oman, Pakistan, Philippines, Qatar, Russia, Saudi Arabia, Singapore, Sri Lanka, Syria, Taiwan, Tajikistan, Thailand, Turkey, Turkmenistan, Ukraine, United Arab Emirates, Uzbekistan, Vietnam, West Bank, Yemen

**EUROOPA:** Albania, Andorra, Austria, Belgium, Bosnia & Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Faroe Islands, Finland, France, Germany, Gibraltar, Greece, Guernsey, Hungary, Iceland, Ireland, Isle of Man, Italy, Jersey, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Monaco, Netherlands, Norway, Poland, Portugal, Romania, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom

**LÕUNA-AMEERIKA:** Anguilla, Antigua & Barbuda, Argentina, Aruba, Bahamas, The Barbados, Belize, Bolivia, Brazil, British Virgin Is., Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Mexico, Montserrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts & Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad & Tobago, Turks & Caicos Is, Uruguay, Venezuela, Virgin Islands

**OKEAANIA:** American Samoa, Australia, Cook Islands, Fiji, French Polynesia, Guam, Kiribat, Marshall Islands, Micronesia, Fed., St. Nauru, New Caledonia, New Zealand, N. Mariana Islands, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu, Wallis and Futuna

**PÕHJA-AMEERIKA:** Bermuda, Canada, Greenland, St Pierre & Miquelon, United States