

TALLINNA ÜLIKOOL
INFORMAATIKA INSTITUUT

**Avatud lähtekoodiga tarkvarale üleminekuvõimaluste hindamine
andmeanalüüsi kursuse näitel**

Magistritöö

Autor: Kristo Talmar

Juhendaja: Kairi Osula

Autor: ,, ,,2014

Juhendaja: ,, ,,2014

Instituudi direktor: ,, ,,2014

Tallinn 2014

Autorideklaratsioon

Deklareerin, et käesolev magistritöö on minu töö tulemus ja seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....

(kuupäev)

.....

(autor)

SISUKORD

SISSEJUHATUS.....	5
Teema valiku põhjendus ja aktuaalsus	5
Probleemi sõnastus.....	6
Töö eesmärgid.....	7
Töö struktuur.....	7
1. AVATUD LÄHTEKOODIGA TARKVARA.....	9
1.1 Ajaloost.....	9
1.2 Avatud lähtekoodiga tarkvara mõiste.....	10
1.3 Firmaomane versus vaba lähtekoodiga tarkvara	13
1.4 Avatud lähtekoodiga tarkvara hariduses	16
1.4.1 Avatud lähtekoodiga tarkvara üldhariduskoolides.....	18
1.4.2 Avatud lähtekoodiga tarkvara kõrghariduses.....	19
2. TARKVARA VALIK	21
2.1 Uue tarkvara sobivuse hindamine	21
2.1.1 Majanduslikud aspektid.....	21
2.1.2 Tehnilised aspektid.....	22
2.1.3 Halduslikud aspektid	23
2.1.4 Uue tarkvara juurutamine.....	23
2.2 Statistika programmide ülevaade	25
2.2.1 Võrreldavad tarkvarad.....	25
2.2.2 Ülevaade statistika programmidest	25
2.2.3 Statistika tarkvarade populaarsus	29
2.3 Vaba lähtekoodiga tarkvara „R“ ülevaade	33
3. EMPIIRILINE UURING	35

3.1. Üliõpilaste seas läbiviidud uuring	35
3.1.1. Uuringu taust	35
3.1.2. Uuringu tulemused	37
3.2. Ekspert hinnang	46
KOKKUVÕTE	50
SUMMARY	51
KASUTATUD MÕISTED	52
KASUTATUD KIRJANDUS	56
LISAD	59
LISA 1. Kursuseprogrammid IFI7041 ja IFI7044	59
LISA 2. Internetilehel Kdnuggets läbi viidud uuring	63
LISA 3. R-i installeerimise ja kodutöö juhend	65
LISA 4. Magistrant Lauriga läbiviidud intervjuu	73

SISSEJUHATUS

Teema valiku põhjendus ja aktuaalsus

Free software contributes to human knowledge, while non-free software does not. Universities should therefore encourage free software for the sake of advancing human knowledge, just as they should encourage scientists and scholars to publish their work.

Richard M. Stallman(Stallman, 1992)

Vaba tarkvara üks eestvedajaid Richard Stallman on tabavalt sõnastanud paralleeli vaba tarkvara ja ülikoolide peamise funktsiooni vahel, milleks on teaduse edendamine(Tartu Ülikooli seadus, 16.02.1995). Selle seisukoha veendunud pooldajana leian, et Eesti ülikoolides on vaba tarkvaraga seonduva mõttemaailma poole veel palju minna. Oma töös uurin võimalusi, kuidas seda teha statistilise analüüsi kursuse raames.

Antud teema on põhjalikumaks uurimiseks välja pakutud Tallinna Ülikooli Informaatika instituudi poolt. Teema vajab uurimist, kuna hetkeolukord põhjustab ebamugavusi andmeanalüüsi kursusel osalejatele. Sellise järelduse saab teha andmeanalüüsi kursuste tagasisidet analüüsides, isiklikust kogemusest kursuse läbinuna ning eriala seminari kursusel kogutud tagasisidest antud töö tutvustusele, millest ilmnes teema aktuaalsus ka kursusekaaslaste hulgas. Tallinna Ülikoolis kasutatakse andmeanalüüsi õpetamiseks programmi IBM SPSS Statistics. Antud programm on firmaomane tarkvara (IBM) ja kuigi õpilastele tehakse ostul tihti soodustusi, on programmi isiklikuks tarbeks soetamine suhteliselt kallis. Tallinna Ülikool on soetanud vajalikud litsentsid, et õpilased saaksid kooli arvutiklassides tarkvara tasuta kasutada. Kodutööde tegemiseks on võimalik alla laadida tasuta prooviversioon, mis laseb programmi kasutada 14 päeva. Seetõttu on andmeanalüüsi kursuste ülesehitusel koondatud koduste tööde tegemiseks ette nähtud aeg kursuse lõppu ning teoreetiliselt peaks kogu koduse praktilise töö ära tegema 14 päeva jooksul. Kui kursust saaks samaväärselt läbi viia ka vaba lähtekoodiga tarkvaral, muudaks see õppimise mugavamaks ja teadmiste kontrollimise paindlikumaks.

Aktuaalne on teema ka Eestis laiemalt. Aastast 2017 tõstab Microsoft järk-järgult kahe aasta jooksul oma toodete hinda Eesti koolidele. Hinnatõus peaks esialgsete andmete kohaselt olema umbes kümnekordne. Paljud omavalitsused on andnud märku, et sellise hinnatõusu

puhul ei ole neil enam võimalik koolidele tarkvara soetada. Alternatiiv tasulisele tarkvarale on tegelikult olemas, kuid seda kardetakse vähese kogemustepagasi tõttu. Koolitamine on kindlasti aeganõudev, kuid selle alustamise edasilükkamiseks või alternatiivide leidmiseks ei pruugi varsti enam aega olla. Alg-, kesk-, ja kõrghariduse järk-järguline liikumine vaba tarkvara kasutamise poole hoiaks kokku maksumaksja raha ning arendaks ka vaba ja teadmisi jagavat ühiskonda. Just sellist keskkonda peaksime püüdma koolides luua.

Probleemi sõnastus

Vaba tarkvara levik on Eesti üldhariduskoolides hoo sisse saanud. Lisaks riigi toele (LibreOffice projekti rahastamine) on ka paljud üldhariduskoolide informaatikaõpetajad iseseisvalt otsustanud vaba tarkvara kasutamise kasuks. Tihti on stiimuliks just majanduslik kasu koolile, kuid kindlasti ka:

- jagamise kultuur, inimlike väärtuste õpetamine;
- oluliselt madalamad kulud haridusele;
- praktiliselt olematud kulud õpilastele;
- sunnitud tarkvarapiraatluse välistamine;
- ligipääs tohutule hulgale intellektuaalsele materjalile - hea võimalus programmeerijate koolitamiseks, keda Eestile väga vaja on;
- võrdsemad võimalused (Kebbinau, 2014).

Madalamatest haridusastmetest pealekasvav trend kasutada vaba tarkvara, peaks edasi kanduma ka ülikoolidesse, et neid väärtusi edendada.

Andmeanalüüsi kursusel osalenute tagasisidest tuli välja, et üliõpilastel tekkis probleeme programmi SPSS allalaadimise ja installeerimisega. Samuti leiti, et kuna tegemist on küllaltki keerulise ja kalli programmiga, peaks see olema harjutamiseks kasutatav pikemalt kui kaks nädalat. Üliõpilastel tekkis probleeme eksamiks ettevalmistumisel, kuna kodutööde jaoks installitud programm oli oma prooviaja ületanud. Seetõttu sooviti kodutööde lahendamiseks ja aine läbimiseks ka vabavariat alternatiivi.

Käesolevat magistritööd ja selle eesmärgi tutvustaval erialaseminaril sain kursusel osalenutelt teada, et hätta jäädi programmi hankimisega ning leiti, et allalaadimise ja installimise peale kulus liiga palju aega ja energiat. Kõik tagasisidet andnud üliõpilased leidsid, et teema on nende jaoks aktuaalne ning vajadus parema lahenduse järgi on olemas.

Töö eesmärgid

Tutvustada avatud lähtekoodiga tarkvara, selle kasutust, levikut ja ajalugu. Kirjeldada tulevikuperspektiive ja uurida, kas ja miks avatud lähtekoodiga tarkvara kõrgharidussektoris kasutada.

Uurida, millistest etappidest koosneb uue tarkvara valiku protsess, milliseid aspekte peaks silmas pidama ning millised peaksid olema protsessi etapid ja järjekord. Koguda infot erinevate andmeanalüüsi tarkvarade kohta ja neid omavahel võrrelda. Pakkuda välja alternatiivne andmeanalüüsi tarkvara firmaomasele SPSS-ile.

Anda ülevaade ja tutvustada väljavalitud vaba lähtekoodiga programmi. Koostada juhend programmi paigaldamise ja kasutamise kohta. Testida programmi kasutusmugavust tudengite testgrupis. Koostada rahuloluuuring ning analüüsida tudengite tagasisidet.

Uurida kogunud kasutajatelt valituks osutunud programmi kasutamise kogemuste kohta. Selgitada, kas parimaks osutunud programmiga hetkel Tallinna Ülikoolis või mujal kõrgharidussektoris kursuseid läbi viiakse, ja millised on kasutajate kogemused.

Töö struktuur

Magistritöö koosneb neljast peatükist. Eraldiseisvateks osadeks on kasutatud mõisted ja lisad.

Sissejuhatav peatükk annab ülevaate teema valiku põhjendusest ja aktuaalsusest, probleemi sõnastusest, töö eesmärgist ning töö struktuurist.

Esimeses peatükis antakse ülevaade avatud lähtekoodiga tarkvarast, selle ajaloost ning levikust haridusvaldkonnas. Võrreldakse firmaomase ja avatud lähtekoodiga tarkvara positiivseid ja negatiivseid külgi.

Teises peatükis antakse ülevaade, millistel alustel hinnata tarkvara sobivust asutusele ning millised on uue tarkvara juurutamise etapid. Võrdlen erinevaid andmeanalüüsi programme. Valin välja sobivaima vabavaralise alternatiivi firmaomasele tarkvarale lähtuvalt andmeanalüüsi kursuse vajadustest. Koostan juhendi, milles selgitan, kuidas alternatiivse tarkvaraga on võimalik samu analüüse läbi viia. Tutvustan lähemalt programmi R, mida pean parimaks alternatiiviks tasulisele andmeanalüüsi tarkvarale. Samuti tutvustan R-i graafilist kasutajaliidest ja tööd hõlbustavaid lisapakette.

Kolmandas peatükis antakse ülevaade läbiviidud uuringust ning analüüsitakse tulemusi. Samuti tehakse kokkuvõtte R-i spetsialistiga tehtud intervjuust ning R-iga kursuseid läbiviinud õppejõu arvamusest.

1. AVATUD LÄHTEKOODIGA TARKVARA

Käesolevas peatükis antakse ülevaade avatud lähtekoodiga tarkvarast, selle ajaloost ning levikust haridusvaldkonnas. Võrreldakse firmaomase ja avatud lähtekoodiga tarkvara positiivseid ja negatiivseid külgi.

1.1 Ajaloost

Oli aeg, kui avatud lähtekoodiga tarkvara peeti vaid veidriku ja nohikute maailma kuuluvaks, ning see eksisteeris tavainimestele läbipääsmatute barrikaadide taga. Esimeseks suureks barjääriks oli arusaamatu kõnepruuk. Kasutajad, kes ei saanud aru käskudest nagu „*configure, make, sudo make install*“- jäeti tihti mängust välja. Teiseks takistuseks oli asjaolu, et enamik avatud koodiga programme nagu võrgu- ja mailiserverid, olid mõeldud kasutamiseks serverites ja võrgu administraatoritele. Selliseid vajadusi ja oskusi tavalistel arvutikasutajatel polnud. Isegi rakendused, millel oleks võinud olla laialdasem kasutajaskond, nagu tekstiredaktorid, olid tihti nii keerulised, et nendega töötamiseks paistis olevat vaja spetsialisti kraadi. Veel üheks tõkkeks oli esialgu positiivsena paistev asjaolu, et vaba lähtekoodiga tarkvara kirjutati täiesti tasuta Unixi-sugustele operatsioonisüsteemidele. Hiljem muutus see uueks takistuseks, kuna keskmise arvutikasutaja jaoks tekkis vale arusaam, et tasuta tarkvaral pole kindlasti nii lihtsat kasutajaliidest ja tehnilist tuge nagu kommertstarkvaral. Nii jäigi avatud koodiga maailm tavainimestele pikaks ajaks kinniseks. (Wiley, 2006)

Arvutiteajastu algusest saadik on eksisteerinud grupp riist- ja tarkvaraarendajaid, keda peetakse pigem entusiastideks kui professionaalideks. Neid on paelunud võimalus arendada süsteeme koostöös teistega. Tavaliselt on selleks väga vaba koostöövorm, kus ei ole kindlaid kohustusi või piiranguid. Esialgne nn jaosvara liikumine kujutas endast programmeerijate poolt loodud programmide levitamist erinevate võrkude kaudu. Sel ajal tähendas võrk tuttavate inimeste võrgustikku. Jagati diskette ja vastutasuks oodati vabatahtlikke rahalisi annetusi. Interneti laialdasem kasutuselevõtt tappis sedasorti jaosvara levitamise ja tasustamise korra, kuna programmide levitamine muutus väga lihtsaks. Programme oli nii lihtne internetist tasuta alla laadida, et keerulisema osa, ehk maksmiseni, enam ei jõutudki. (Hopkinson, 2007)

1.2 Avatud lähtekoodiga tarkvara mõiste

Olenevalt isiku vaatepunktist on „Avatud lähtekoodiga tarkvara“ (Open Source Software) mõistel tänaseks mitmeid erinevaid tähendusi.

Rangelt võttes tähendab avatud lähtekood, et:

- 1) lähtekoodi on teistel kasutajatel võimalik kasutada, vaadata ja muuta;
- 2) kõik võivad seda tasuta edasi levitada ilma tarkvara omanikule autori- või litsentsitasusid maksmata. (Hopkinson, 2007)

Vastukaaluks, ostetud tarkvara lähtekoodi tavaliselt ei jagata ega tehta täielikult kättesaadavaks, ja enamasti ei tohi kasutajad tarkvara kopeerida ega edasi levitada. (Hopkinson, 2007)

Terminit „avatud lähtekood“ kasutatakse tihti teatud tarkvara arendamismudeli tutvustamiseks. Selleks on detsentraliseeritud lähenemine, mida kasutavad tuntud projektid nagu Linux ja Apache. Nende toodete arendamisse on kaasatud ülemaailmne programmeerijate kogukond, kes annavad enda panuse tarkvara arendamisse ja hooldusesse. Selliseks tegevuseks on panustajatel palju erinevaid motiive. (Hopkinson, 2007)

Ühes spektri otsas on vabatahtlikud, kellele ei maksta, ja kelle motivatsiooniks on keeruliste probleemide lahendamisest saadav mõnu ja tunnustus. Prestiiž programmeerijate ja kasutajate kogukondades võib olla väga suureks tõukejõuks. Samuti tahe välja töötada midagi keerulist isiklikuks tarbeks. (Hopkinson, 2007)

Teiseks grupiks on programmeerijaid, kes on palgatud lisama kohalikele vajadustele vastavaid uuendusi ja panustama avatud lähtekoodiga tarkvara arengusse. Tööandjateks on sel juhul firmad või asutused, mis on teinud otsuse mingis valdkonnas üle minna avatud lähtekoodiga tarkvarale. (Hopkinson, 2007)

Kolmandaks on firmad, mis on otsustanud toetada vaba tarkvara arendamist, sest nad müüvad teenuseid, mis on selle tarkvaraga seotud. Selliste firmade hulka kuulub ka IBM, mis arendab avatud lähtekoodiga tarkvara tsentraliseeritud ja kontrollitud keskkonnas. Tulu teenivad nad pakkudes hooldust ja muid väärtust lisavaid teenuseid, mis on antud tarkvaraga seotud. (Hopkinson, 2007)

Viimased arengud aga toovad avatud koodiga tooteid järjest rohkem tavakasutajateni. Lisaks vabavaralisele operatsioonisüsteemile Linux on tänased vaba lähtekoodiga tarkvarad enamasti ühilduvad ka firmaomaste operatsioonisüsteemidega Microsoft Windows (edaspidi Windows) ja Mac OS X. Selliste rakenduste paigaldamine on sama lihtne kui kommertstarkvara paigaldamine ning ka kasutamine ei ole enam keerulisem. Paljudele enamlevinud kasutusalaadele on tänaseks kommertstarkvarale vaba lähtekoodiga tarkvara hulgas alternatiiv olemas. Näiteks Microsoft Office-i asemel saab kasutada LibreOffice-it, mis ühildub Microsofti tarkvaraga suurepäraselt. Microsoft Internet Explorerile on heaks alternatiiviks nii Firefox kui Google Chrome, mille funktsionaalsus on vähemalt samaväärne. Microsoft Outlook-i asemel võime kasutada Thunderbird-i jne. Teksti redigeerimine, tabelitöötlus, internetis surfamine, e-kirjadega tegelemine ja sõpradega vestlemine võtabki tõenäoliselt suurema osa tavalise arvutikasutaja ajast. Kõiki neid tegevusi saab teha väga mugavalt avatud lähtekoodiga tarkvaral.(Wiley, 2006)

Lisaks lähtekoodi avalikustamisele on avatud lähtekoodiga tarkvarale seatud *Open Source Initiative*-i (OSI) poolt veel mitmeid detailseid nõudmisi, et programmi saaks lugeda avatud lähtekoodiga tarkvaraks. OSI poolt on defineeritud 10 punkti, millele peab tarkvara vastama:(The Open Source Initiative, 2014)

- 1) Vabalt levitatav – litsents ei tohi piirata ühelgi isikul tarkvara müümist või ära andmist.
- 2) Lähtekood – programmiga peab kaasas olema lähtekood ja peab olema lubatud lähtekoodi jagamine. Kui lähtekoodi kaasas pole, peab olema hästi kättesaadav võimalus lähtekoodi hankimiseks. Eelistatult internetist tasuta allalaadimise võimalus.
- 3) Tuletatud teosed – litsents peab lubama tuletatud teoseid ja koodi muutmist ning nende edasiarenduste jagamist samadel litsentsitingimustel nagu originaaltootel.
- 4) Autori koodi terviklikkus – litsents võib piirata lähtekoodi levitamist muudetud kujul vaid siis kui litsents lubab koodiga koos nn paikasid, mis on mõeldud programmi muutmiseks selle koostamise ajal.
- 5) Isikute või gruppide diskrimineerimise keeld – litsents ei tohi diskrimineerida ühtki inimest ega gruppi.
- 6) Tegevusvaldkondade diskrimineerimise keeld – litsents ei tohi piirata programmi kasutamist mingis tegevusvaldkonnas. Näiteks ei tohi olla seatud piiranguid programmi kasutamiseks ärilistel eesmärkidel.

- 7) Litsentsi levitamine – õigused, mis on programmile sätestatud peavad kehtima ka kõigile neile, kellele programmi on jagatud, ilma lisa litsentsita.
- 8) Litsents ei tohi olla tootespetsiifiline – programmi kasutusõigused ei tohi sõltuda sellest, kas programm on suurema programmikogumi osa. Kui programm on sellest kogumist välja võetud, peaksid selle kasutajatele jääma samad õigused, mis suurema programmikogumi kasutajatele.
- 9) Litsents ei tohi piirata muud tarkvara – litsents ei tohi seada piiranguid muule tarkvarale, mida jagatakse litsentseeritud tarkvaraga koos.
- 10) Litsents peab olema tehnoloogiliselt neutraalne – ükski litsentsi säte ei tohi keelata mingi kindla tehnoloogiaga koostöötamist või kellegi poolt toodetud tarkvaraga ühildumist.(The Open Source Initiative, 2014)

Avatud lähtekoodiga tarkvara peetakse järjest enam võrdseks alternatiiviks firmaomasele tarkvarale ja see mõjutab märkimisväärselt kommertsettevõtete ärikultuuri. Unix (Linux), tuntum avatud lähtekoodiga operatsioonisüsteem, on kasutusel 28% ettevõtete serverite arvutites. Apache serveri tarkvara on kasutusel 67% (W3Techs, 2014) kõikidest veebilehekülgedest. MySQL-il, tuntuimal avatud lähtekoodiga andmebaaside tarkvaral, on üle kuue miljoni aktiivse installatsiooni üle maailma. Suurkorporatsioonid nagu IBM, HP, SUN Microsystems ja Intel on hakanud Linuxi toetajateks. Riistvara tootjad müüvad masinaid, millel on Linuxi tarkvara ning pakuvad sellele ka tugiteenust.(Hopkinson, 2007)

Muidugi ei tohi unustada ka puuduseid, mis avatud lähtekoodiga tarkvara kasutamine võib kaasa tuua.

Esiteks, kuna tarkvara jagatakse tasuta, ei pruugi olla võimalik rahastada kõiki uuendusi, mida kasutajad vajavad, just siis, kui neid vajatakse. Kuid seda juhtub tihti ka tasulise tarkvara puhul.(Hopkinson, 2007)

Teiseks, enamik varajastest avatud koodiga tarkvaradest arendati tehnikateadlikelt inimestelt tehnikateadlikele inimestele. Nende projektide puhul ei peetud esmatähtsaks lõppkasutajate kasutusmugavust nagu on tänapäeva arendustes kombeks. Tänapäevaks on sellistest probleemidest hakatud üle saama, heaks näiteks on siin Firefox-i ja Mozilla Thunderbird-i uuemad versioonid.(Hopkinson, 2007)

1.3 Firmaomane versus vaba lähtekoodiga tarkvara

Tabelis 1 on välja toodud firmaomase tarkvara ning vaba lähtekoodiga tarkvara võrdlus. Välja on toodud tarkvara hind, teeninduse ja kasutajatoe kirjeldus, turvalisus, kasutusmugavus, kättesaadavus ning tarkvara usaldusväärsus.

Omadused	Firmaomane tarkvara	Vaba Lähtekoodiga Tarkvara (VLT)
Hind	Varieerub mõnest eurost kuni mitmesaja tuhande euronni olenevalt süsteemi keerukusest. Tarkvara hinna moodustavad tarkvara baastasu, integratsioon, teenused ja igaaastane litsentsi/toe tasu. Mõnede jaoks võib selline tasu olla takistavaks, kuid firmade väitel maksab klient järgmiste teenuste eest: kohandatud toode usaldusväärset tootjalt, mis tagab kõrgema turvalisuse ja funktsionaalsuse, pideva ja jätkuva arendustegevuse. Treeningute ja toe pakkumise ning väiksemad nõudmised tehnilise taibu osas.	VLT on tasuta. Organisatsioonid võivad seda tüüpi tarkvara kasutada nii kaua kui soovivad. Puuduvad kuutasud, aktiveerimise tasud või uuendamise kulud. Raha eest on võimalik paljudele turul olevatele programmidele osta toe teenust kuid vabavara kogukond pakub enamasti tasuta abi.
Teenindus ja tugi	Teenindus on firmaomase tarkvara suurim eelis. Tasulise tarkvara pakujad meelitavad kasutajaid, kelle tehniline võimekus ei ole kõige parem ning pakuvad neile alalist tuge. Kui toote kasutusjuhendist ei piisa ja kasutaja takerdub mõnda probleemi, on olemas koht, kuhu helistada ja nõu küsida. Mõneti maandatakse kasutajate riski kui töötatakse firmadega, kes vastutavad oma toote eest ja omavad töötajaid, kellel on väga põhjalikud teadmised nende toote kohta. Kuna teenindus ja tugi on üks suurimaid põhjuseid, miks firmaomane tarkvara valitakse VLT asemel, siis on toote pakujad järjest rohkem hakanud võistlema toodete toe turul. Kui kasutajate tugi on tasemel, on ka firma toodete menu oluliselt suurem. Samas võib suurfirmade nagu Microsoft jaoks jääda	Teenindus on üks põhilisi probleeme vaba koodiga tarkvara puhul. Siin loodetakse võrgus olevale kogukonnale, mis pakub õppimise võimalusi foorumitest ja blogidest. Kuigi on olemas väga suur, lojaalne ja hingega asja juures olev kogukond, kelle poole saab pöörduda, nõuab selline teabe hankimine kasutajalt mõningaid põhioskusi, et tagasisidet mõista ja kasutada. Vahel on veaotsing kiirem kui firmaomase tarkvara puhul, kuid kasutajad, kes teevad päringu, peaksid suutma enda probleemi võimalikult täpselt kirjeldada, muidu ei ole ka tagasiside edasiviiv. Eestis on võimalik abi paluda ka Vaba Tarkvara Liidu liikmetelt, nõu antakse nii

	<p>tähelepanuta väikekliendid igapäevaste printeri või tarkvara muredega. Rohkem ressursse suunatakse ikkagi lepingulistele suurklientidele, kes maksavad palju litsentsitasusid.</p>	<p>tarkvara kui riistvara küsimustes. http://alvatal.ee/liikmed.htm</p>
Turvalisus	<p>Turvalisuse seisukohalt on käimas pidev debatt parema lahenduse osas. Paljusid firmaomaseid tarkvarasid, mis on tehtud firmaomastele operatsioonisüsteemidele, peetakse vähem turvalisteks kui VLT-d. Kuid firmaomase tarkvara lõpp-produkti peetakse turvaliseks kuna seda arendatakse kontrollitud keskkonnas ühtse eesmärgiga tiimi poolt. Lähtekoodi, mis on põhjalikult kontrollitud, saavad näha ja muuta ainult tiimi liikmed. See peaks elimineerima riski, et tarkvara sisaldab „tagaust“ või Trooja hobust ning vähendab erinevate vigade ja probleemide arvu tarkvaras (seda juhul kui tootja on täiesti usaldusväärne ja ei jäta endale meelega tarkvara ülevõtmise võimalust).</p>	<p>Vaba lähtekoodiga operatsioonisüsteeme peetakse kõige turvalisemateks. Klassikaliseks näiteks on siin Linux. Vabalt kättesaadava lähtekoodi puhul on kõik turvavead kõigile näha ja nende avastamine ja lahendamine käib kiiremini kui firmaomase tarkvara puhul.</p>
Kasutusmugavus	<p>Firmaomase tarkvara puhul on arendusprotsessi enamasti kaasatud ka kasutusmugavuse testijad. Tarkvara on tavaliselt tehtud kindlale kasutajategrupile ja seetõttu rohkem kohandatud. Kasutusmugavust hinnatakse enamasti üsna kõrgeks. Tarkvaraga on kaasas ka põhjalikud kasutusjuhendid. See võimaldab kiiremat väljaõpet ja õppimiskõveral liikumist. Lisatoena pakutakse seminare, suunatud koolitusi ja kasutajatuge, mis lubab tarkvarast viimast võtta.</p>	<p>Varasemaid VLT-sid kritiseeriti nende kehva kasutusmugavuse tõttu. Tavaliselt ei vaata tarkvara üle kasutusmugavuse eksperdid ja see ei rahulda suuremat osa arvutikasutajaid. Tänapäeval testib VLT-d kogukond ja kasutusmugavuse arengusse aidatakse kaasa ülemaailmselt. Suuremate projektide arengusse on kaasatud ka spetsiaalselt kasutusmugavust parendavad eksperdid (Ubuntu, Firefox jne) Samuti on VLT tänaseks väga hästi dokumenteeritud.</p>

Kätte- saadavus	Tarkvara on kättesaadav nende arendaja firma kaudu, kellel on õigused tarkvara muutmiseks ja levitamiseks. Enamasti on võimalik prooviversioone tasuta allalaadida ja testida.	Tarkvara on internetis tasuta saadaval. Mõned VLT-d on rendatud ka piiratud firmaomaseks tarkvaraks, kus pakutakse kasutajatuge seitsmel päeval nädalas ööpäevaringselt, internetikogukonna küsimustele vastamiseks ja arendajatega suhtlemiseks.
Usaldus- väärsus	Firmaomane tarkvara rendatakse välja tootjapoolsete meeskondade poolt nii, et müüki jõuavad ainult valmis tooted. Kuna mittekooskõlastatud muudatused ei satu lõpptootesse, siis peaks lõpp-produkt olema kõrge usaldusväärusega.	VLT-d peetakse usaldusväärseks kuna probleemide lahendamiseks kaasatakse kogukond. Kasutajad, mitte ainult ei edasta veateateid, vaid ka parandavad neid ja saavad parandused arendajatele.

Tabel 1.(N. Pankaja, 2013)

Tabelist 1. näeme, et mõlemal tarkvara liigil on omad plussid ja miinused. Firmaomasel tarkvaral on tihti parem kasutajatugi, mis kajastub toote hinnas. Samuti peetakse paremaks firmaomase tarkvara kasutusmugavust. VLT puhul on eelisteks kättesaadavus, hind ja turvalisus. Usaldusväärsus oleneb pigem konkreetsest tootest, mitte liigist.

1.4 Avatud lähtekoodiga tarkvara hariduses

Üleüldise vaba lähtekoodiga tarkvara leviku taustal on see endast mitmeti märku andnud ka hariduse vallas. Mõjutusi on kõige rohkem näha suures tarkvara hulgas, mis on suunatud haridusasutustele ja õpetajatele. Avatumat lähenemist on lisaks tarkvara kasutamisele näha ka väljaspool arvuteid. Mitmete e-kursuste materjale jagatakse avalikult, luuakse avalikke uurimustööde hoidlaid ja elektroonilisi avaldamise foorumeid.(Wiley, Open source, openness, and higher education, 2006)

Haridusasutustel on järjest laienev valik kasutatava tarkvara osas. Õppimise ja kursuse juhtimise süsteemid nagu WebCT ja Blackboard saab nüüd asendada alternatiividega nagu Sakai ja Moodle.(Wiley, Open source, openness, and higher education, 2006)

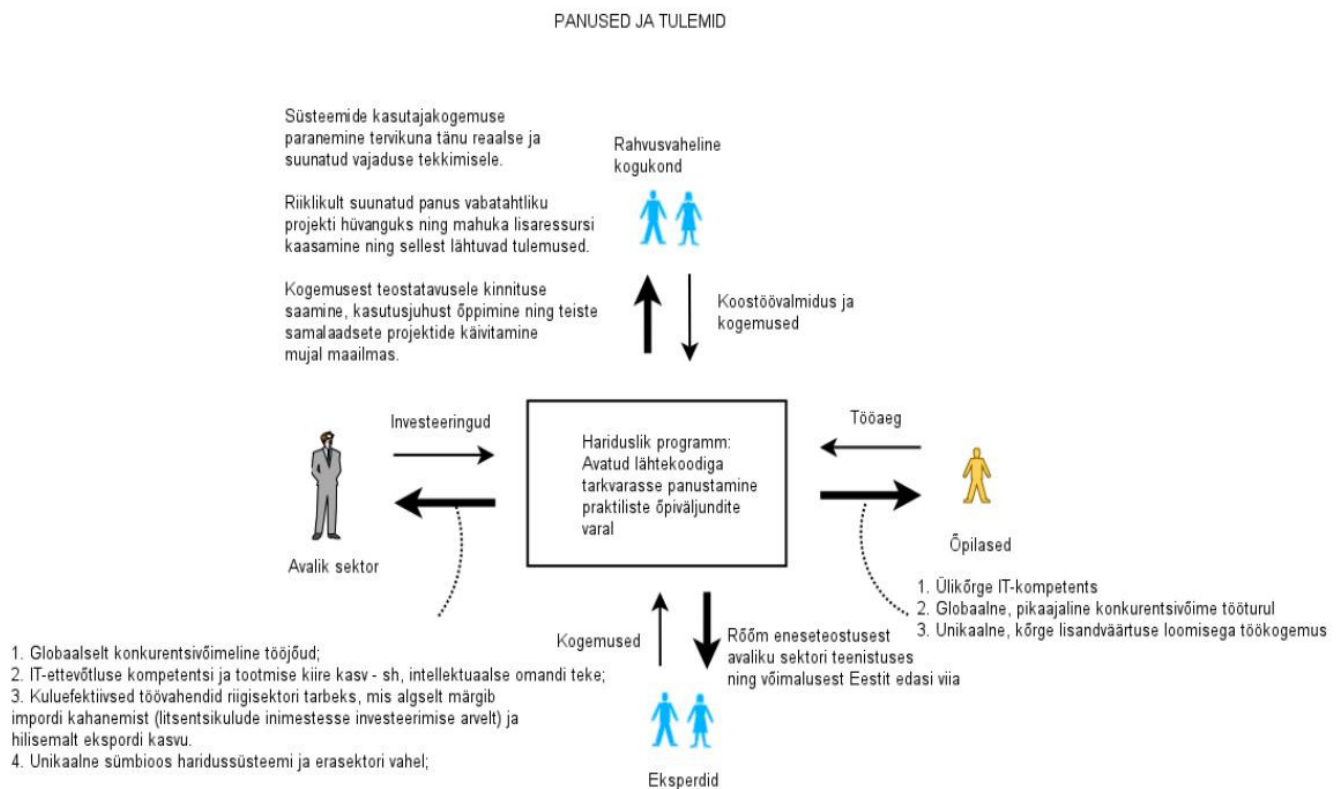
Hariduspõhise tarkvara loomise potentsiaal olemasoleva vaba lähtekoodiga tarkvara baasil lubab mitmeid uusi võimalusi. Vaba lähtekoodiga tarkvara on enda olemuselt väga kohanemisvõimeline ja tänu sellele saab tarkvara kasutada mitmetes valdkondades. Näiteks fotogalerii tarkvara, mis on loodud piltide jagamiseks ja hindamiseks võib kohandada esseede, kursusetööde, CAD failide (joonestustarkvara failid) jagamiseks ja hindamiseks. Tarkvara, mis on mõeldud interaktiivse multimeedia esitluste loomiseks võib kohandada esitlemaks bakterite kasvu visuaalselt või nõudluse ja pakkumise vahekorra illustreerimiseks majanduse loengutes. Kuigi tava tarkvara muutmine haridusasutustes kasutatavaks tarkvaraks on alati olnud võimalik, siis vaba lähtekoodiga tarkvara väga paindlik disain on ideaalseks vahendiks see võimalus ka realselt ellu rakendada.(Wiley, Open source, openness, and higher education, 2006)

Vaba lähtekoodi filosoofia on avaldanud ülikoolides palju laiemat mõju kui ainult tarkvara litsentseerimise valdkonnas. Muutunud on ka viisid, kuidas teaduskonnad jagavad teadustööde tulemusi ja õppematerjale. Üle 2500 vaba kursuse on saadaval rohkem kui 200-st ülikoolist üle maailma(Tom Caswell, 2008). See number on aasta-aastalt suurenenud. Suurema osa nendest kursustest pakub välja MIT OpenCourseWare (Massachusettsi Tehnoloogiainstituudi vabade kursuste leht). Kuid vabade kursustega väljatulemine on tänaseks populaarne kõikjal maailmas.(Wiley, Open source, openness, and higher education, 2006)

Miks kasutada vabavara hariduses? Vabavara aitab ehitada tugevamat kogukonda läbi inimväärtuste õpetamise. Jagamise kultuur ja millegi tegemine üldise heaolu nimel, mitte

koheste tasu saamiseks, õpetab lastele varakult häid väärtusi. Üksteise aitamine ja toetamine läbi võimaluse legaalselt tarkvara jagada vähendaks oluliselt piraatlust. Koolid võivad firmaomase tarkvara küll tasuta saada ja õpilaste käsutusse anda, kuid peale kooli lõpetamist tuleb tarkvara ikkagi osta. Vabavara kasutamine aitab säästa niigi nappi maksumaksja raha ja võimaldab seda investeerida Eesti majandusse. Riik ei koolitaks maksumaksja raha eest enam välismaistele äriettevõtetele kliente, mille tulemusena voolaks taas raha riigist välja. Kasutajad väldiksid tootjalukkusid (*vendor lock-in*). (Laugasson, 2014)

Joonisel 1. on kujutatud vaba lähtekoodiga tarkvara projektis osalejate panused ja saadavad tulemid.



Autorid: vabatarkvara kogukonna aktiivsed tegijad

Joonis 1. Vabavara arendamisesse panustajad ja saadud kasu. (Laugasson, 2014)

Microsofti hinnatõus

Maailmapanga andmeil (The World Bank) on Eesti aastal 2014 kõrge sissetulekuga maa: 12 616 € ja rohkem aastas – see summa suureneb tasapisi, sellest tulenevalt ei saa Microsoft pakkuda enam programmi „Partners in Learning“ (PiL) hüvesid – soodsamaid hindu

litsentside rentimisel. Hinnatõus toimus 01.07.2011 kuid veidi enne seda pakuti 3-aastast püsilepingut, mille enamus koole ka valis – 3 aastat saab täis 30.06.2014. Tähtaega on tänaseks pikendatud kuni 30.06.2017 – selle ajani on võimalik üle minna vabatarkvarale, mille on välja öelnud ka Haridus- ja Teadusministeerium. Tegelik hinnatõus on kümnetes kordades suurem kui avalikult räägitud ning sisaldab tehnilisi nüansse, mis võivad hinda veelgi tõsta. (Laugasson, 2014)

1.4.1 Avatud lähtekoodiga tarkvara üldhariduskoolides

Tänaseks on ka Eestis mitmed üldhariduskoolid ja lasteaiad võtnud kasutusele vabatarkvara.

Mõned nendest on:

- Kivimäe Põhikool
- Rapla Vesiroosi Gümnaasium
- Tallinna LV Haridusameti pilootprojekt
- Tallinna Mustamäe Gümnaasium
- Tallinna Mahtra Gümnaasium
- Merivälja Kool
- Tallinna Mustamäe I Lasteaed-Algkool
- Tallinna Mesimummu Lasteaed

Kindlasti on tehtud esimesi katsetusi ka teistes Eesti koolides ja täielikku uuringut ei ole veel läbi viidud. (Laugasson, 2014)

Tallinna linn tellis Ernst & Youngilt uuringu ja analüüsi, mille eesmärgiks oli välja selgitada kui palju hoiaks linn kokku vabatarkvarale üleminekul. Kui Tallinna linnavõim paigaldaks oma koolidesse-lasteaedadesse Microsoft Office-i asemel vabavaralise kontoritarkvara (LibreOffice), siis oleks võimalik viie aastaga säästa 490 000 eurot litsentsitasu. Kui aga vahetada välja ka MS Windowsi operatsioonisüsteem Linuxi vastu, siis võiks viie aastaga kokku hoida 700 000 eurot. Vabatarkvara kasutamise võimalusi uurib Tallinna linn eelpool väljatoodud Microsofti hinnatõusu tõttu. Kui täna tuleb ühe arvuti Windowsi operatsioonisüsteemi ja Office-i kasutamise eest maksta kokku alla kolme euro, siis kolme aasta pärast juba üle 35 euro. Peale kulude rehkendamise katsetas linn käesoleva aasta algusest alates kolmes koolis ja kahes lasteaias vabatarkvara kasutuselevõttu ka praktikas. Ametlik kokkuvõte valmib selle kooliaasta lõpuks, kuid testgrupi ülemineku põhjal saab projektile ka algse hinna määrata. Tallinna haridusasutuste 6000 arvuti kohta oleks see 93 300 eurot. See summa oleks võidetud litsentsitasusid arvesse võttes üsna väike. (Tammeorg, 2014)

Merivälja kool on üks viiest asutusest, kus üleminekut katsetatakse. Koolidirektor Kaja Laanmäe nentis, et üleminek on olnud tülikas. Probleeme on olnud klassides kasutatavate interaktiivsete tahvlitega ja printeritega. Laanmäe sõnul on kooli töötajatel tunne, justkui oleksid nad tehnoloogilise arenguga ajas tagasi läinud ja arvutite kasutamine võtab nüüd rohkem aega. Tagasi Microsofti tooteid kasutama ei plaanita sellele vaatamata minna. Vabataarkvara on niikuinii kõikide tulevik, leiab Laanmäe. (Tammeorg, 2014)

Täna sel päeval sõltub iga kolmas Tallinna koolide arvutikasutaja Microsofti toeta jäänud Windows XP-st. 538 vastaja hulgast kasutab 185 (38%) vastanud ikka Windows XP-d. Vaid kuus (1%) vastajat kasutab Linuxit. Natuke parem on olukord kontoritarkvara osas, kus Microsofti toeta jäänud Office2003 kasutajaid oli 48 (9%). LibreOffice-i ja OpenOffice-i kasutajaid oli kokku 23 (4%). (Tammeorg, 2014)

1.4.2 Avatud lähtekoodiga tarkvara kõrghariduses

Avatud lähtekoodiga tarkvara juured viivad meid tagasi ülikoolidesse. Paljud täna tuntud kommertstarkvarad, mis on ülikoolides kasutusel, said alguse teistest ülikoolidest ja nende vajadustest. Täna maailmas tuntud *Blackboard* (virtuaalse õppimise keskkond) tarkvara arendati välja Ameerika ülikoolis Cornell enda õpilaste tarvis ja nende poolt. Cisco arvutivõrkude süsteemid tehti valmis Standfordi ülikooli võrgundustarkvara aluseks võttes. Enamik raamatukogude automatiseerimise süsteeme on alguse saanud ülikoolidest ja arendatud ühistegevusena. Parimad näited on siin Dortmundi (DOBIS/LIBIS toetajaks oli ka IBM kelle miniarvutitel süsteem töötas), Virginia (DYNIX ja VTLS) ja Birminghami (BLCMP) ülikoolide raamatukogusüsteemid. (Hopkinson, 2007)

Digitaalsete raamatukogude arenduse eestvedajateks on samuti ülikoolid, nende seas ka Tallinna Ülikool. Greenstone digitaalraamatukogu tarkvara on arendatud Waikato ülikooli poolt, see on kättesaadavaks tehtud tänu UNESCO sponsorelusele. (Hopkinson, 2007)

Varem oli tavaks, et ülikoolid, kes olid enda tarbeks välja töötanud tarkvara, millel oli väärtus ka teiste koolide jaoks, müüsid selle edasi firmadele või tegid ise erafirma, et toodet edasi toetada. Täna on läbi „Avatud Koodi“ liikumise võimalik toodet edasi arendada läbi koostöö ja arendamise jagamise. Kõrghariduse spetsiifilisi vajadusi arvesse võttes on juba mitmeid tarkvarapakette arendatud läbi *Open Source Initiative*-i (Avatud Lähtekoodi Algatus).

Viimase kümne aasta jooksul on erinev arvutitarkvara muutunud kõrgema hariduse lahutamatuks osaks. 1990-ndatel hakkasid õpilased ja teaduskonnad tõsisemalt kasutama

õppetöös e-posti. Tänapäevaseks on erinev tarkvara kõigis õpinguvaldkondades igapäevaseks kaaslaseks. Mõned kasutusel olevad tarkvarad on näiteks ettevõtte ressursside planeerimise süsteemid, mida kasutatakse ülikoolide administreerimiseks, ülikoolis korraldatavate küsitluste läbiviimiseks tehtud küsitlus-vastamis platvormid, virtuaalsed õppekeskkonnad, veebipõhised kursused, teadustööde avaldamiseks loodud keskkonnad jpm. Ülikoolilinnakute arvutivõrgud peavad toetama suurtes kogustes kohalikke ja väliseid kasutajaid, tagama kõrge turvalisuse taseme ja vastama suuremale tarkvara nõudlusele kui enamik suurfirmasid. Nende hulgas on linnaku portaalid, ÕIS (õppeinfosüsteem) ja koolide jaoks kohandatud partnervõrgud (peer-to-peer network) failide jagamiseks.(Hopkinson, 2007)

Kui erinevad teaduskonnad MIT-s, Utah'i, Bostoni ja teistes ülikoolides kutsuti avaldama oma õppekavasid, loengumärkmeid, kodutöid ja muid õppematerjale, siis üllataval kombel palusid paljud endale lisaega, et materjale korrastada, üle vaadata ja redigeerida. Nad olid ettevaatlikud, sest liikumine avatuse suunas viib õpetamise otse akadeemilise maailma südamesse – ollakse avatud kvaliteeti tõstva spetsialistide hinnangu suhtes. Selline avatus annab materjalide ülevaatamiseks võimaluse palju rohkematele ning muudab hariduse läbipaistvamaks. Kõik huvirühmad nagu õpilaste vanemad, vilistlased, tööandjad, ülikooli sisseastujad jne, saavad hea ülevaate ülikooli õppetegevusest.(Wiley, Testimony to the Secretary of Education's Commission on the Future of Higher Education., 2006)

2. TARKVARA VALIK

Käesolevas peatükis antakse ülevaade, millistel alustel hinnata tarkvara sobivust asutusele ning millised on uue tarkvara juurutamise etapid. Võrreldakse erinevaid statistilise analüüsi programme ja valitakse välja sobivaim vaba lähtekoodiga tarkvara lähtuvalt andmeanalüüsi kursuse vajadustest. Seejärel antakse ülevaade, kuidas erinevate tarkvarade abil viia läbi lihtsamat statistilist analüüsi. Koostan juhendi, milles selgitan, kuidas alternatiivse tarkvaraga on võimalik samu analüüsi läbi viia.

2.1 Uue tarkvara sobivuse hindamine

Tarkvara hankimine ja paigaldamine algab selle sobivuse hindamisest praeguse infotehnoloogia süsteemiga. Koostöötamise võimet kasutatava tarkvaraga, operatsioonisüsteemiga ja muude rakendustega. Hinnang tuleks anda järgmistele punktidele:

- **majanduslikele aspektidele**, mis on seotud tarkvara soetamis- ning ülalpidamiskuludega;
- **tehnilistele aspektidele**, mis on seotud tarkvara paigaldamise, ühilduvuse ja uuendamisega;
- **halduslikele aspektidele**, mis on seotud kasutajate harjumuste ja töövoogude, aga ka asutuse eesmärkide ja poliitikaga. (Horm, 2012)

Hindamisel lähtutakse arvestusest, et tarkvara hankimise puhul on võimalikud kolm varianti:

1. Uue originaaltarkvara tellimine
2. Olemasoleva tarkvara täienduse tellimine
3. Valmistarkvara hankimine (Pöldmaa H. , 2011)

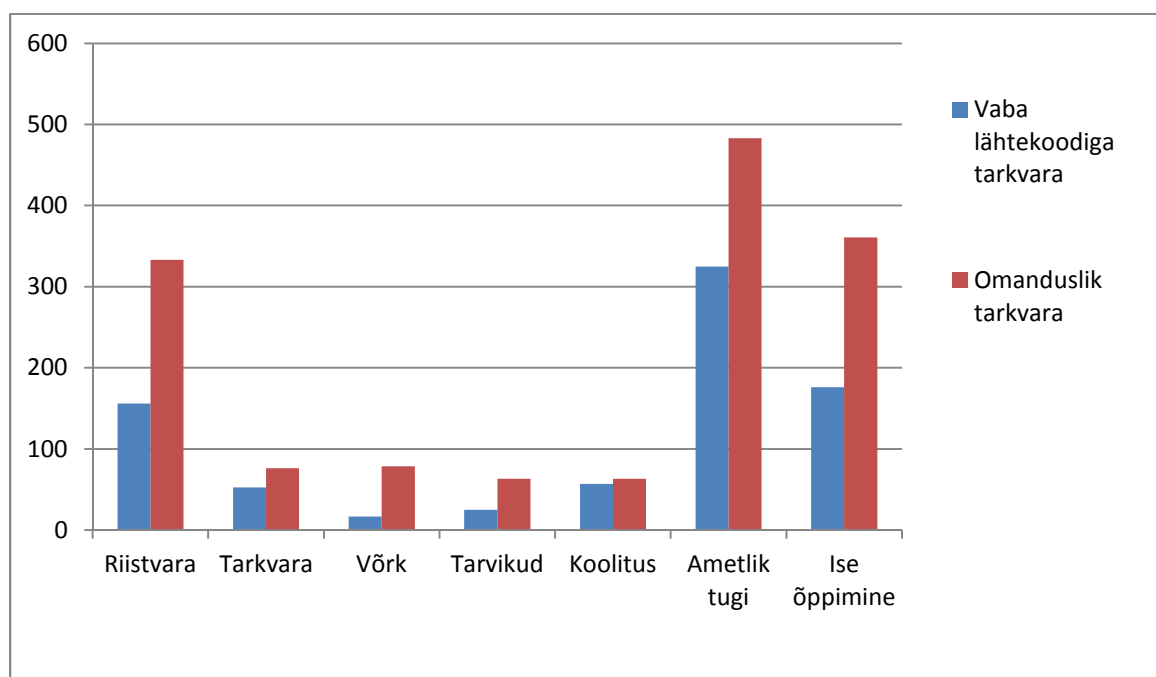
2.1.1 Majanduslikud aspektid

Tarkvara soetamise esmane hind ei tohiks kindlasti olla ainus asi, mida uue tarkvara valimisel silmas pidada, lisaks tuleks selgeks teha ka järgnevad kulud (Cornford, 2011):

- **Otsingukulud** – kulutused, mida tehakse tarkvara eeluuringule. Juurutamise võimalikkuse hinnang, olemasoleva süsteemiga koostoimimise analüüs.
- **Soetamiskulud** - tarkvara maksumus, firma vajadustele vastavaks kohandamine.

- **Integratsioonikulud** – migreerimise kulud (kasutajad ja andmed), väljaõppe kulud, protsesside muudatuse ja parimate tavade väljatöötamise ning kasutuselevõtu kulud.
- **Kasutuskulud** – tugiteenuste maksumus – majasisene, majaväline. Haldus- ja uuendamiskulud.
- **Väljumiskulud**- tarkvara elutsükli lõppedes tehtavad kulutused.(Cornford, 2011)

Suurbritannias koostatud uuringus vaadeldi, mille peale kulutavad koolid kõige rohkem raha IT valdkonnas. Joonisel 2 kujutatakse kogukulusid (TCO -*Total cost of ownership*) ühe arvuti kohta aastas Suurbritannias läbiviidud uuringu põhjal.



Joonis 2. Vaba- ja kommertstarkvara TCO (*Total cost of ownership*) kulu võrdlus (hind eurodes/ ühe masina kohta).(Becta, 2005)

Jooniselt näeme, et omanduslik tarkvara on kõikides kategooriates osutunud kulukamaks. Üllatuslikult on koolide jaoks suurimaks kuluartikliks tarkvarale toe ostmine.

2.1.2 Tehnilised aspektid

Enne tarkvara soetamist tuleks kindlaks teha tarkvara poolt esitatavad tehnilised nõuded. Tuleb kaaluda hangitava tarkvara sobivust juba olemasoleva riistvara, operatsioonisüsteemidega, andmebaasimootoritega ning muu olemasoleva baastarkvaraga.(Põldmaa H. , 2012)

Riistvaraplatvormide puhul tuleb vaadata, millistel riistvaraplatvormidel soovitud tarkvara toimib, ning kas see on asutuses juba kasutusel. Juhul, kui sellist riistvara asutuses kasutusel ei ole, tuleks uurida, kas ja mis tingimustel on võimalik seda hankida. Samuti on vajalik uurida kui palju maksab sellise riistvara kasutamine – selle haldus ja hooldus. (Põldmaa H. , 2012)

Hangitava tarkvara puhul tuleb samuti jälgida, millistel operatsioonisüsteemidel ta toimib. Enamik avatud tarkvara mudelit kasutavaid programme on küll olemas kõigile enamlevinud operatsioonisüsteemidele, kuid sageli on põhiliseks platvormiks valitud Linux ning Windows või Mac versioonid on kas tasulised või oluliselt väiksema funktsionaalsusega. (Põldmaa H. , 2012)

2.1.3 Halduslikud aspektid

Kommertstarkvara pakub IT teenuste osutamiseks terviklikku paketti, sisaldades enamikku baasteenuseid, mida taristu käigushoidmine ja haldamine vajab. Tarkvara komponendid on loodud ja testitud koos töötama. Niisugusel terviklahendusel on sageli keskne administreerimisliides, mille abil saab muudatusi rakendada tsentraalselt kõikidele seadistatavatele teenustele. See muudab võrguhalduri töö lihtsamaks ja efektiivsemaks. (Horm, 2012)

Vabavaralise tarkvara puhul tuleb tervikpildi komplekteerimiseks pisut rohkem ise vaeva näha. Kuna valik on suur, siis tuleb jälgida, et tooted ühilduksid ning nende administreerimine oleks võimalikult lihtsalt teostatav.

Tarkvara valikul tuleks kindlasti lähtuda eesmärgist, mitte kasutatava tarkvara põhjal seada eesmärgi. Tarkvara on abivahend eesmärgi saavutamiseks. (Horm, 2012)

2.1.4 Uue tarkvara juurutamine

Järgnevas alapeatükis on firma Asta ElecoSoft (asta ElecoSoft, 2014) poolt koostatud tarkvara juurutamise mudeli põhjal (Tabel 2.) kirjutatud lahti kõik protsessi etapid.

Uue tarkvara juurutamisel ja valimisel tuleks liikuda edasi samm sammult, see on interaktiivne protsess. Kõige olulisem on, et protsessi lõpuks oleks juurutatud tarkvara, mis vastab algsetele eesmärkidele ning selle kasutamine on tehtud koolituste ja abimaterjalidega selgeks kõigile kasutajatele.

Kõigepealt peaks välja selgitama organisatsiooni vajadused, mida soovitakse uue tarkvara kasutuselevõttuga saavutada. Selleks konsulteeritakse programmi põhiliste kasutajatega ning pannakse paika, mis funktsioonid peaksid uuel tarkvaral olema. Samuti peaks tegema kindlaks laiem ringkonna, kes uut programmi edaspidi kasutama hakkavad, et ka nende arvamused ja vajadused kirja panna.

Kui kasutajaskonnaga on konsulteeritud, peaks kirja panema süsteemi nõuded ning konfiguratsiooni. Vastavalt vajadustele otsitakse välja sobivad tarkvara lahendused ning tehakse tarkvara võrdlus. Selles punktis peaks kirjeldama ka veakindluse ja vajadusel taasteplaanid.

Suuremate projektide puhul, kuhu on kaasatud ka serverid ja muud võrguseadmed peaks enne installeerimise alustamist kõik süsteemid konfigureerima ja vigadeta tööle saama ning seejärel kasutajatele paigaldama hakkama.

Seejärel saab programmi paigaldada, kas kõigile kasutajatele või testgrupile. Nüüd peaks läbi viima ka esimesed koolitused kasutajate ja administraatorite hulgas.

Kui süsteem on mõnda aega (3-4 nädalat) töötanud, peaks tegema esimesed kokkuvõtted ning selgitama välja üleminekuga ilmnenud probleemid ja lisa koolituse vajadused.

1. Vajaduste selgitamine	<ul style="list-style-type: none"> Hinnata organisatsiooni vajadusi, mida soovitakse saavutada, kes seda kasutama hakkavad ja kuidas.
2. Rakendamise planeerimine	<ul style="list-style-type: none"> Kirjeldatakse süsteemi konfiguratsioon ja esitatakse nõuded.
3. Projekti ülesseadmine	<ul style="list-style-type: none"> Projekti konfigureerimine ja esitatud nõuetele vastavusse seadmine.
4. Paigaldamine	<ul style="list-style-type: none"> Programmi installeerimine organisatsiooni IT-süsteemidele. Õppejõudude, administraatorite koolitus.
5. Koolitus	<ul style="list-style-type: none"> Kasutajate koolitus
6. Paigaldusjärgne kontroll	<ul style="list-style-type: none"> Süsteemi rahuloluuring, muudatuste vajadused, koolituste vajadused.

Tabel 2. Uue tarkvara kasutuselevõtmine (asta ElecoSoft, 2014)

2.2 Statistika programmide ülevaade

Käesolevas alapeatükis annan lühikese ülevaate vaba lähtekoodiga andmeanalüüsi tarkvaradest ning valin töö eesmärkidest lähtuvalt välja kõige sobivama.

2.2.1 Võrreldavad tarkvarad

Statistilise analüüsi tarkvarade nimekiri on küllaltki pikk, seetõttu tegin enne programmide põhjalikumat võrdlemist esialgse valiku, millega jätkata. Võrdlusesse valisin kõik tuntumad programmid, mis suudavad teha statistilise analüüsi kursusel esmavajalikke analüüse ning on kasutatavad nii Windowsi, Maci kui ka Linuxi operatsioonisüsteemidel.

Lähtuvalt andmeanalüüsi kursuste teemadest, mis on ära toodud töö lõpus LISAS 1, valisin võrdlemiseks järgmised meetodid:

- Kirjeldavastatistikaanalüüsid (miinimum, maksimum, keskväärtus, mediaan, standard hälve jne)
- Vahemikhinnang
- T-test (sõltuvad valimid)
- T-test (sõltumatud valimid)
- Hii-ruut test
- Kruskal-Wallise test
- Tabelid
- Diagrammid
- ANOVA

Need analüüsimetodid ja tulemuste kujutamise võimalused sean otsitavale tarkvarale miinimumnõueteks. Sellest nimistust midagi välja jättes kannataks tarkvara tõttu kursuse programmi sisu.

2.2.2 Ülevaade statistika programmidest

Tabelis 3. jagasin võimalusterohkemad statistika tarkvarad kahte gruppi võttes aluseks nende litsentsi tüübid.

Firmaomane tarkvara	Microsoft Excel 2013, Maple, MATLAB + Statistics Toolbox, Mathematica, Nvivo, RATS, SAS, S-Plus, JMP, Sysdat, SigmaXL, SPSS, Stata
---------------------	--

Üldine Avalik Litsents (GNU General Public License)/ AGPL3 (Sofa)	AdaMSoft, ScaViS, GNU PSPP (edaspidi PSPP), R + erinevad lisa versioonid, Sage, SOFA Statistics
--	---

Tabel 3. Andmeanalüüsi tarkvarade jagunemine litsentsi alusel.(Wikipedia)

Lisaksin nimekirjale veel kolm programmi, mis ei toeta Linuxi ja Maci operatsioonisüsteeme, kuid on Windowsi kasutajatele piisavalt võimaluste rohked vahendid, et need ära märkida (kõik järgnevad on firmaomased tarkvarad).

- Minitab (puudub linuxi ja Maci tugi)
- SHAZAM (puudub linuxi ja Maci tugi)
- STATISTICA (puudub linuxi ja Maci tugi)

Vabatarkvaradest jäid esialgu valikusse:

- ADaMSoft
- SCaViS
- PSPP
- R + erinevad lisa versioonid
- Sage
- SOFA Statistics

AdaMSoft – vaba lähtekoodiga tarkvara, mis on kirjutatud Javas. Kasutatakse andmeteanalüüsiks ja andmekaevandamiseks. Programm on mitmekeelne ja töötab kõigil suurematel platvormidel, mis toetavad Javat. (ADaMSoft, 2014)

ScaViS - on keskkond andmeanalüüsiks ja andmete visualiseerimiseks, mis on mõeldud teadlastele, inseneridele ja õpilastele. Programm sisaldab ja kasutaja saab juurde lisada palju erinevaid vaba lähtekoodiga tarkvarapakette. ScaVis-i saab kasutada kõikjal, kus analüüsitakse suuri andmemahte või kaevandatakse andmeid (loodusteadustes, inseneriteadustes, finantsturgude modelleerimisel ja analüüsimisel). ScaVis töötab igal platvormil, kuhu on paigaldatud Java. (SCaViS)

PSPP – on programm statistiliseks andmeanalüüsiks. See on tasuta asendus firmaomasele SPSS tarkvarale. PSPP on tehtud välja nägemiselt võimalikult sarnaseks SPSS-ile kuid ei oma

samapalju funktsionaalsust. PSPP programmil on nii graafiline kasutajaliides kui ka käsurealt kasutamise võimalus. (GNU Operating System, 2014)

R – on tasuta programmeerimiskeel ja tarkvara statistiliseks andmeanalüüsiks ning graafiliseks kujutamiseks. R töötab kõigil suurematel operatsioonisüsteemidel. R-i keel on välja kasvanud keelest S, mille arendamist alustas Bell Laboratories juba 1975-1976 aastal. R-iga töötamine käib käsurealt, kuid erinevate lisapakettide paigaldamisel saab programmis kasutada ka graafilisest kasutajaliidest. (R-project, 2014)

Sage - vaba lähtekoodiga matemaatika tarkvara, mis on ehitatud mitmete teiste vaba lähtekoodiga tarkvarapakettide (NumPy, SciPy, matplotlib, SymPy, Maxima, GAP, FLINT, R jpt) peale. Kasutajaliideseks on tekstidokumendi aken internetilehitsejas või käsurida. Kasutades tekstidokumenti, ühendab Sage kohalikku arvutisse paigaldatud programmiga või Sage serverisse. (Sage)

SOFA Statistics – on kasutajasõbralik statistika-, analüüsi- ja aruandlusprogramm. Programm on tasuta ning paneb rõhku heale kasutusmugavusele, kiirele õppimisele ja hästi vormistatud väljundile. SOFA võimaldab analüüsi tulemusi kuvada hästi kujundatud vormingus, mida saab kergelt teistega jagada (Excel, PDF). SOFA aitab kasutajal programmi tundma õppida töö käigus. Nuppude ja funktsioonide kirjeldused on kirjeldatud graafilises kasutajaliidises. (Statistics Open For All)

Lähtuvalt statistika kursuse vajadustest võrdlen järgmises tabelis vabatarkvaralisi statistika programme ning hetkel kasutuses olevat SPSS-i. Lisasin võrdluse ka Excel 2013 kuna Tallinna Ülikooli õpilastele on kohustuslikus aineks „Arvuti töövahendina“, mis baseerub Microsofti kontoritarkvaral, ning seetõttu peaks kõikidel tudengitel olema algteadmised Excelist. Täpselt sama funktsionaalsus on ka vabavaralises programmis LibreOffice Calc, kus ei pea analüüside tegemiseks aktiveerima lisamoodulit nagu MS Excelis vaid saab kohe arvutama asuda.

	ADaMSoft	SCaViS	PSPP	R + erinevad lisa versioonid	Sage*	SOFA Statistic s	SPSS	Excel 2013 + Analysis ToolPak
Kirjeldav statistika (Me, Mo, s jne)	Jah (mitte graafilisest kasutajaliid esest)	Ei	Jah	Jah	Jah (mitte graafilisest kasutajaliid esest)	Jah	Jah	Jah
Tabelid (sagedus- ja risttabelid)	Jah	Jah	Jah (mitte objektina)	Jah	Jah	Jah	Jah	Jah
Diagrammi d	Jah	Jah	Jah	Jah	Jah	Jah	Jah	Jah
Korrelatsio on (Pearson, Spearman, Cramer)	Jah (mitte graafilisest kasutajaliid esest)	Ei	Jah (piiratud)	Jah	Jah (mitte graafilisest kasutajaliid esest)	Jah	Jah	Jah (Cramer-i funktsioon i peab valemina sisestama)
T-test	Jah (mitte graafilisest kasutajaliid esest)	Ei	Jah	Jah	Jah	Jah	Jah	Jah
Hii-ruut test	Jah (mitte graafilisest kasutajaliid esest)	Ei	Jah	Jah	Jah (mitte graafiliselt kasutajaliid esest)	Jah	Jah	Jah
ANOVA	Jah	Jah	Jah	Jah	Jah	Jah (ainult ühesuun aline)	Jah	Jah

Tabel 4. Statistika tarkvarade võrdlus

*-Windowsi keskkonnas võimalik töötada kas võrgus või VirtualBoxi kasutades.

Koostatud tabelist näeme, et tasuta tarkvarast vastavad etteantud tingimustele kõige paremini programmid R ja SOFA Statistics. SPSS-iga kõige sarnasem on programm PSPP, kuid sellel ei ole kõiki põhjaliku andmeanalüüsi läbiviimiseks vajaminevaid funktsioone. ADaMSoft on küll graafilise liidesega ja hõlpsasti kasutatav, kuid mõned vajalikud funktsioonid on kasutatavad käsurealt või scripti kirjutades (kasutatavad on ka programmis R koostatud scriptid). SCAViS ei vasta miinimumnõuetele, ehk puudub võimalus teha mitmeid andmeanalüüsi kursusel vajalikke analüüse. Sage kasutamine on Windowsi keskkonnas hetkel kasutajatele keerulisem, kuna see ei tööta eraldi programmina, vaid vajab Windowsi keskkonnas töötamiseks lisaprogrammi VirtualBox (või mõnd sarnast virtualiseerimise tarkvara), mis laseb Sage käivitada virtuaalmasinas Linuxi keskkonnas.

Excelil on lihtsamate analüüside tegemise võimalik aktiveerida lisamoodulid „*Analysis ToolPak*“ ja „*Analysis ToolPak VBA*“, mis teevad kasutaja jaoks kirjeldava ning ka üldistava statistika enamkasutatavad analüüsid üsna lihtsaks. Võrreldes spetsiaalsete analüüsi-programmidega on tudengitel Exceliga enamjaolt ka varasem kokkupuude olemas. Kogutud andmed on tihti juba Excelis või on neid sinna lihtne eksportida. Exceli asendamiseks spetsiaalse analüüsi tarkvaraga tekibki vajadus alles suurte andmemahtude ja komplekssemate analüüside puhul. Oma töös proovin tudengitele välja pakkuda võimalikult suure funktsionaalsusega analüüsi tarkvara, millega saaks ka peale kooli tööelus jätkata. Komplekssemate analüüside ja suurte andmemahtude jaoks ei ole Excel mõeldud ega sobilik. Samuti ei vasta Excel vaba lähtekoodiga tarkvara kriteeriumitele.

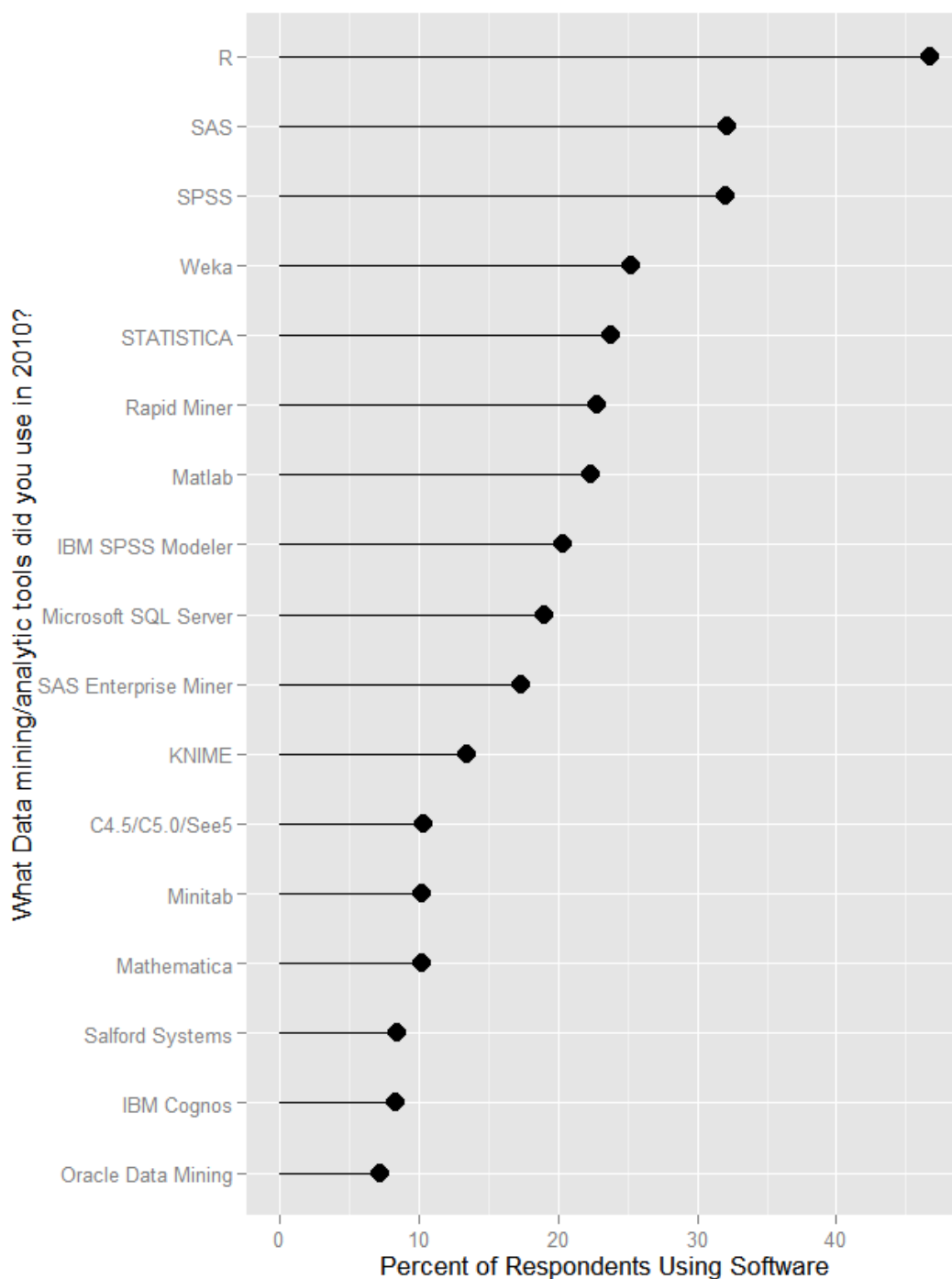
Edaspidises võrdluses võime välja jätta programmid SCAViS ja PSPP, mis ei suuda kõiki vajalikke funktsioone täita. Sage ei ole kasutatav Windowsi keskkonnas ning ei vasta seetõttu esitatud tingimustele. ADaMSoft-iga saab küll kõiki analüüse teha, kuid selleks peab kasutama käske käsurealt. Käsud saab kirjutada programmi R süntaksis, kuid samad asjad on R programmis graafiliselt tehtavad. Seega pole ADaMSoft-i edasine analüüsimine otstarbekas.

2.2.3 Statistika tarkvarade populaarsus

Tarkvara, mille abil õpilastele statistilise analüüsi kursusel andmete töötlust ja analüüsimist õpetatakse, peaks olema kasutatav ka edaspidises elus. Seetõttu peaksid olema eelistatud kõige laialdasemalt kasutatavad programmid. Hea oleks, kui õpilane saaks kasutada olemasoleval või tulevasel töökohal sama tarkvara, mida õpingute jooksul. Tasub jälgida ka

trende, mis näitavad, kas tarkvara kasutamine on kasvamas või kahanemas, et õppida selgeks kõige parem ja laialdasemalt kasutatav tarkvara.

2011. aastal on läbi viidud uuring, mida peetakse suurimaks analüütikute, andmekaevanduse ja andmete teaduses. Uuringu nimi on „Rexer's Data Miner Survey“(Rexer, 2011). 2010. aasta uuringu tulemuste põhjal on kokku pandud järgnev joonis:




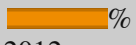











Joonis 3: Rexer's Data Miner Survey 2010 a. tulemused (Rexer, 2011)

Jooniselt näeme, et 2010 aastal kasutas vastanutest 47% oma andmeanalüüsi ja andmete kaevandamise ülesannete lahendamiseks vaba lähtekoodiga programmi R. Teist ja kolmandat kohta jagavad tabelis firmaomased tarkvarad SAS ja SPSS. Kui vaadata ka eelnevaid küsitluste tulemusi, siis on R-i turuosa kasv olnud pidev juba aastaid.

Jooniselt ei leia tarkvara Sofa tõenäoliselt seetõttu, et programm on mõeldud rohkem kodukasutajale ja keerukamate projektide puhul jääb selle funktsionaalsusest väheks.

Erinevus klassikalise andmeanalüüsi ja andmete kaevandamise tarkvara vahel on rohkem turunduse mõiste kui erinevus analüütilises võimekuses. Seetõttu võime uuringu tulemuste põhjal teha järeldusi ka enim kasutatava andmete analüüsi tarkvara kohta.

Aastal 2013 viidi samalaadne uurimus läbi ka andmete kaevandamise lehel Kdnuggets, olen siin väja toonud uuringu kümme enimkasutatud tarkvara (Tabel täispikkuses LISAS 2):

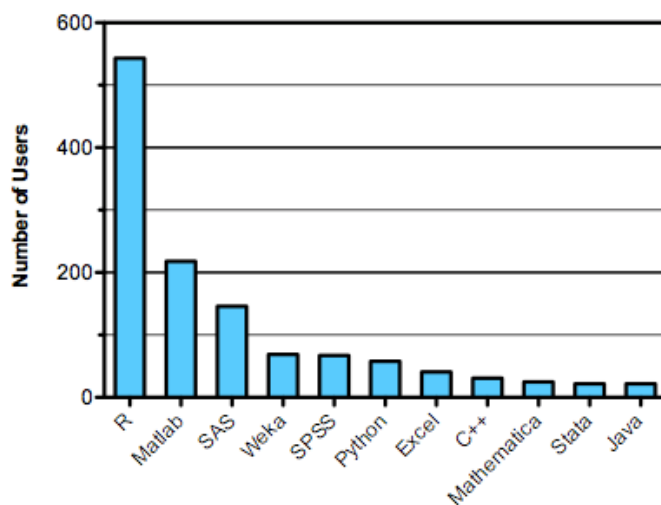
Millist analüüsi, andmekavanduse, „Big Data“ tarkvara sa kasutasid reaalse projekti tegemisel viimase 12 kuu jooksul? [1880 vastajat]	
Legend: Punane: tasuta/ vabatarkvara programmid	 % kasutajaid 2013
Roheline: Kommerts tarkvara	 % kasutajaid 2012
Rapid-I RapidMiner/RapidAnalytics tasuta versioon(737), 30.9%	 39.2%  26.7%
R(704), 6.5%	 37.4%  30.7%
Excel(527), 0.9%	 28.0%  29.8%
Weka / Pentaho(269), 5.6%	 14.3%  14.8%
Python + numpy/scipy/pandas/iPython... pakettidega(250), 0%	 13.3%  14.9%
Rapid-I RapidAnalytics/RapidMiner firmaomane versioon(225), 52.4%	 12.0%

SAS(202), 2.0%	10.7%
	12.7%
MATLAB(186), 1.6%	9.9%
	10.0%
StatSoft Statistica(170), 45.9%	9.0%
	14.0%
IBM SPSS Statistics(164), 1.8%	8.7%
	7.8%

Tabel 5: KDnuggets 2013 a. uuring.(KDnuggets, 2013)

Tabelist näeme, et esikohal on RapidMiner/RapidAnalytics free edition, mis on eelkõige mõeldud tööks andmebaasidega ja seetõttu meie nõudmistele ei vasta. Teisel kohal on R, mille kasutatavus on 2013 aastal vastajate seas tõusnud 2012 aastaga võrreldes 6,9%.

Üheks tarkvara elujõulisuse mõõdupuuks võiks võtta reaalse kasutatavuse võistlusel, kus iga osaleja valib endale meelepärase programmi. Kaggle.com on interneti lehekülg, mis sponsoreerib andmeanalüüsi võistlusi. Kasutajad saavad sinna postitada andmeanalüüsi ülesandeid koos rahasummaga, mida ollakse nõus maksma ülesande kõige paremini lahendanud inimesele või meeskonnale. Aastal 2011 tehtud uuringus küsiti osalejatelt, millist tarkvara probleemi lahendamiseks kasutati. Vastused on kujutatud joonisel 3.



Joonis 4: Andmeanalüüsi võistlusel kasutatav tarkvara (2011).(Muenchen, 2012)

Jooniselt on näha, et tarkvara „R“ juhib teiste programmide ees märkimisväärselt. Erinevate ülesannete võitjate seas on „R“-i edu veelgi suurem, üle 50% võitjatest kasutas just seda tarkvara.

Töö ja töötajate otsijatega tegelevate interneti lehekülgede põhjal tehtud uuring näitab aga muid tendentse. Tööandjate poolt on kõige enam nõutud tarkvara oskus SAS, sellele järgneb SPSS ja kolmandal kohal on R. R-i turuosa on küll viimastel aastatel kiirelt tõusnud, kuid SAS on siiski kindlalt esikohal. (Muenchen, 2012)

Eelneva põhjal võib järeldada, et igas valdkonnas on oma kindlad liidrid. Suurte korporatsioonide seas kasutatakse rohkem tarkvara SAS ja SPSS, Põhja-Ameerika ülikoolides kasutatakse palju Statat. Tallinna Ülikoolis on kindel liider SPSS. Kõigis valdkondades on juhtivate tarkvarades seas ka vabavaraline tarkvara R.

Eelnevate uuringute põhjal võib väita, et vaba lähtekoodiga tarkvarade hulgast on R ainuke, mis pakub tõsist konkurentsi kommertstarkvarale ning mille õppimisest oleks suure tõenäosusega õpilastele kasu ka tulevikus.

Sofa programmiga on suhteliselt kerge teha lihtsamaid analüüse, diagramme ja graafikuid, kuid suurte andmemahtude ja keerukamate ülesannetega jääb antud programm hätta. Seetõttu ei leia Sofat ka ühestki spetsialistide hulgas korraldatud kasutatavuse küsitluste vastuste seast.

Koostasid R-i installeerimiseks ja kirjeldava statistika ülesannete lahendamiseks juhendi (LISAS 4), mille põhjal Tallinna Ülikooli üks Andmeanalüüsi kursus oma koduse töö tegi.

2.3 Vaba lähtekoodiga tarkvara „R“ ülevaade

Käesolevas peatükis tutvustan lähemalt programmi R, mida pean parimaks alternatiiviks tasulisele andmeanalüüsi tarkvarale. LISAS 4. on leitav juhend programmi R allalaadimiseks, installeerimiseks ja esmaste kirjeldava statistika ülesannete lahendamiseks.

R on keel ja keskkond statistilise andmetöötluse ja graafilise esitamise jaoks. R pakub mitmesuguseid statistilisi (lineaarsed ja mittelineaarsed mudelid, klassikalised statistilised testid, aegridade analüüs, liigitus, rühmitamine jne) ja graafilisi tehnikaid suurte andmemahtudega töötamiseks ja on erinevate liidrite abil laiendatav ning paindlik. (R-project, 2014)

R on saadaval vaba tarkvarana *Free Software Foundationi GNU General Public License*-i tingimustel avatud lähtekoodina. See töötab paljude erinevate UNIX platvormide peal ning ka Windowsi ja Mac-i keskkonnas. (R-project, 2014)

Terminit „keskkond“ kasutatakse teadlikult ilmestamaks R-i kui täielikult planeeritud ja järjepidevat süsteemi, mida saab vajadusel laiendada ning muuta. Mitte konkreetset ja paindumatut programmi nagu andmeanalüüsi tarkvarad tihti on. (R-project, 2014)

R-i keskkond on ehitatud arvuti keele ümber, see lubab kasutajatel lisada funktsionaalsust defineerides uusi funktsioone. Suur osa süsteemist on kirjutatud R keeles, mis teeb algoritmiliste valikute tegemise kasutajatele lihtsalt jälgitavaks. (R-project, 2014)

R-i installeerimisel on kasutajal võimalus kasutada käsurea liidest (CLI – *command line interface*), mis on soovitatav kasutajaliides kogenud kasutajatele paindlikkuse ja käskude otse kasutamise tõttu. Käsurea kasutamine vajab väga head R keele tundmist, algajatele võib see hirmutavana mõjuda. Graafilise kasutajaliidesega on tavaliselt õppimiskõver lühem kui käsurealt töötamisel. Kuid põhjalikumad teadmised analüüsides saab käsurealt töötades, seal on käsud lihtsalt salvestatavad ja uuesti kasutatavad. R-i graafiline kasutajaliides leiab enim kasutust väheste kogemustega kasutajate seas ja õpetajate hulgas, kes peavad lühikese ajaga õpilastele programmi lihtsamad funktsioonid selgeks tegema. (R-project, 2014)

3. EMPIIRILINE UURING

Käesolevas peatükis esitatakse kasutajauuringu tulemused. Saamaks tagasisidet õppeprotsessis osalevatelt eri osapooltelt viidi antud magistritöö raames läbi kahetasandiline uuring: uuringusse kaasati lisaks üliõpilastele ka eksperdid, kes on programmi R kasutanud nii õpetamis- kui ka õppimisprotsessis.

Üliõpilaste seas läbiviidud uuringu eesmärgiks oli saada tagasisidet programmi R kasutamise kohta kirjeldava statistilise analüüsimeetodi rakendamisel. Lisaks uuriti, kui keeruliseks peeti tarkvara installeerimist ning seda, millist abi vajaksid üliõpilased nimetatud programmiga töötamisel. Kuna üliõpilased olid uuringu läbiviimise ajaks osalenud neljas andmeanalüüsi praktikumis ning kasutanud programmi SPSS, paluti vastajatel võrrelda tarkvarade kasutusmugavust ja funktsionaalsust.

Ekspertidena kaasati uuringusse kaks programmi R kasutajat:

Magistrant Lauri, kasutab programmi R aktiivselt enda igapäevatoos. Ta on osalenud kolmel andmeanalüüsi kursusel, mis kõik on baseerunud programmi SPSS kasutamisel. Viimase kursuse raames eelistas ta kasutada andmete analüüsimiseks programmi R.

Õppejõud Märk Möls, kes on kasutanud R-i erinevate kursuste läbiviimiseks juba üle 15 aasta. Ka praegu kasutab ta R-i statistika ja matemaatika kursuste läbiviimisel nii bakalaureuse- kui magistriõppes. Varasemalt ka arstiteaduskonna doktoriõppes.

3.1. Üliõpilaste seas läbiviidud uuring

Käesolevas alapeatükis annan ülevaate üliõpilaste seas läbiviidud uuringu sisust ja tulemustest.

3.1.1. Uuringu taust

Uuringus osales 45 esimese kursuse üliõpilast, kelle põhierialaks ei ole informaatika (ei eeldata kõrgemaid arvutikasutusoskusi). Õppetööväliselt ei olnud statistika tarkvaraga kokku puutunud 37 (82,2%) vastajat. Programmi SPSS oli eelnevalt kasutanud 5 üliõpilast.

Vastajate keskmine hinnang enda arvutikasutus-oskusele 10 palli süsteemis oli 5,7 (SD=1,9). Kõik uuringus osalenud üliõpilased olid kursuse “Andmeanalüüs” kuulajad ning neil oli rühmatöö ajaks toimunud neli programmi SPSS baasil seminari (vt. Tabel 6.).

1. seminar	Sissejuhatus ainesse. Ülevaade kursuse korraldusest. SPSS for Windows. Tunnuste defineerimine, andmete sisestamine.
2. seminar	Põhilised andme- ja failiteisendused: sorteerimine, arvutamine, andmete jagamine gruppidesse, selekteerimine. Andmete transport SPSS-i.
3. seminar	Andmestiku korrastamine (analüüsiks ettevalmistus): puuduvate väärtuste analüüs ja asendamise võimalused, sisestusvigade otsimine, skaalade pööramine. Andmete tüübid.
4. seminar	Ühemõõtmeline analüüs: andmete esitamine tekstina, sagedustabelid. Tulemuste redigeerimine ja tõlgendamine. Kirjeldavate arvnäitajate arvutamine, tulemuste tõlgendamine.
Is.töö (1) 10p.	Rühmatöö, rühmas 3-4 üliõpilast – miniuuringu läbiviimine. Esmane analüüs. Tähtaeg 13.03.2014. Tähtajast hiljem saadetud töö maksimumpunktid 6p.

Tabel 6. Andmeanalüüsi kursuseprogramm (LISA 1.)

Aine „Andmeanalüüs“ kuulajad pidid rühmatööna läbi viima 5-7 küsimusega miniuuringu ning andma ülevaate saadud tulemustest kasutades esmase analüüsi vahendeid (sagedustabelid, kirjeldavad arvnäitajad). Rühmatöö toimus kolme- kuni neljaliikmelistes gruppides. Rühmatöö täistekst on toodud töö lisas nr.4.

Andmeanalüüsi kursusel osalenute seas läbiviidud uuringu küsimustik jagunes kolmeks osaks. Esimesed kümme küsimust hindasid süsteemi kasutatavust *System usability scale* (SUS) küsimustiku põhjal. SUS on lihtne kümnest küsimusest koosnev ja Likerti skaalal põhinev küsimustik, mis annab üldise ülevaate kasutatavusest subjektiivsete hinnangute põhjal. Selle hindamissüsteemi töötas juba 1986. aastal välja John Brooke ning see on tänapäeval laialdaselt kasutusel ja annab usaldusväärset tagasisidet hoolimata valimi suurusest (Stetson, 2004). Nimetatud hindamissüsteemi on väga hea kasutada just vähese kogemusega või programmi esmakasutajate hulgas, kuna see ei nõua vastamiseks programmi kohta põhjalikke teadmisi. Vastuste põhjal saab välja arvutada SUS punktid, mida saab võrrelda teiste sama süsteemiga hinnatud programmide tulemustega. (Measuring Usability With The System Usability Scale (SUS))

Küsimustiku teine osa koosnes rühmatööd puudutavatest küsimustest, mille kaudu uuriti, kas vastaja installeeris programmi või oli vaid tarkvara kasutaja rollis. Sellest tulenevalt jagunesid

ka tekkinud probleemid erinevalt ning erines aeg, mis kulus vastajal rühmatöö tegemiseks. Kuna üliõpilased omasid enne rühmatöö tegemist vähemalt nelja akadeemilise tunni pikkust kogemust SPSS-i kasutamisel, paluti neil ka kahte tarkvara omavahel võrrelda ning välja tuua mõlema programmi tugevused ja nõrkused.

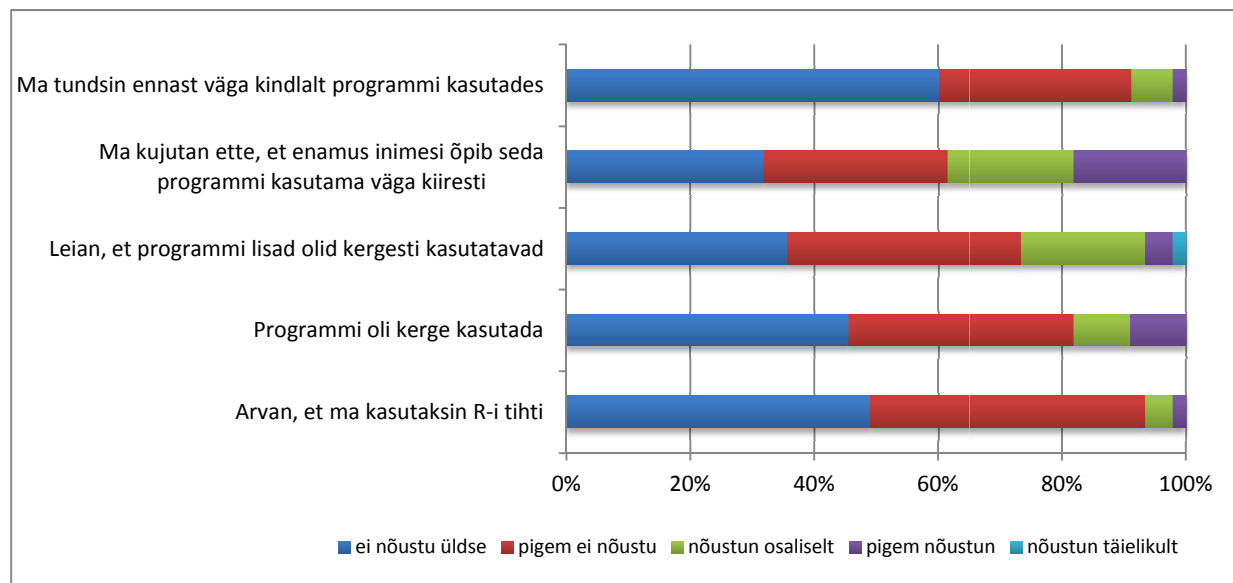
Küsimustiku kolmas osa kirjeldas vastajate tausta. Uuriti, milliseid statistika tarkvarasid on üliõpilased eelnevalt kasutanud ning milliseks hindavad nad enda üldist arvutikasutusoskust.

Uuringu läbiviimiseks kasutati veebipõhist rakendust *SurveyMonkey* (Surveymonkey).

3.1.2. Uuringu tulemused

Süsteemi kasutatavuse küsimused võib nende olemuselt jagada kahte rühma: pooled küsimused annavad ülevaate sellest, kui lihtne oli kasutajate arvates programmi kasutada, pooled sellest, milliste aspektidega oldi rahulolematud.

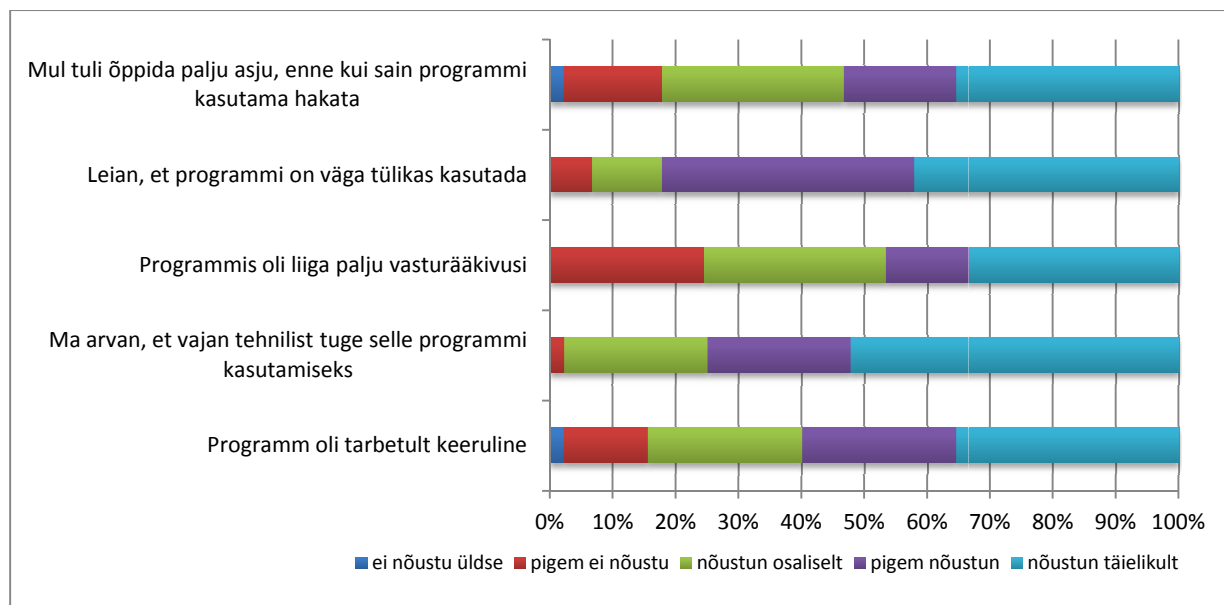
Jooniselt 5. näeme, et programmiga rahulolu (osaliselt, pigem või täielikult nõustumise) osakaal jääb kõikides väidetes alla 40%. Kõige enam (38,6%) väljendasid vastajad nõustumist väitega, et enamus inimesi õpib seda programmi kasutama väga kiiresti.



Joonis 5. Vastajate rahulolu programmiga

Programmiga rahulolu väljendavad väited olid kõik omavahel keskmises positiivses seoses. Näiteks, mida kindlamalt tundsid vastajad ennast programmi kasutamisel, seda rohkem oldi nõus, et programmi oli kerge kasutada ($\rho=0,59$; $p<0,01$). Kui kasutaja väitis, et ta kasutaks programmi R tihti, siis oli ta arvamusel, et enamik inimesi õpiks seda programmi kasutama väga kiiresti ($\rho=0,56$; $p<0,01$).

Programmi kasutamise rahulolematuse väidetega nõustuti märksa enam kui rahulolu väidetega. Kõige enam oldi nõus väidetega, et programmi oli väga tülikas kasutada ning programm oli tarbetult keeruline. Tehnilise toe vajalikkuse tõid välja peaaegu kõik vastajad (43).

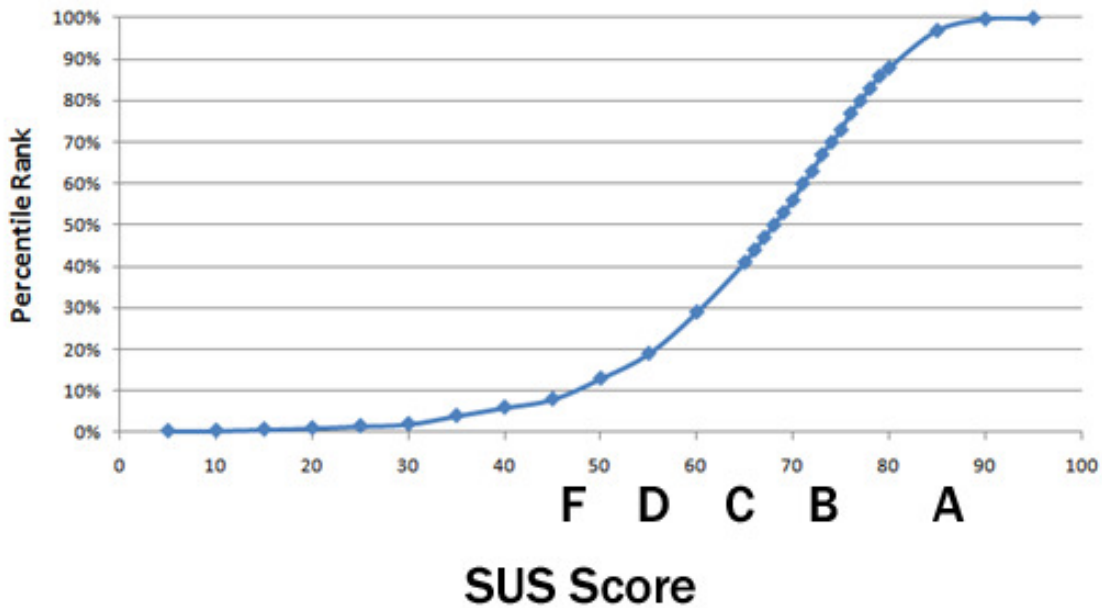


Joonis 6. Programmi puudused vastajate hinnangul

Rahulolematust kirjeldavad väited olid samuti omavahel keskmises positiivses seoses. Tugev samasuunaline seos esines vaid tehnilise toe vajalikkuse ja programmi kasutamise keerukuse vahel ($\rho=0,71$; $p<0,01$). Erandiks olid samuti väited “programm oli tarbetult keeruline” ning “mul tuli õppida palju asju, enne kui sain programmi kasutama hakata”, mis ei olnud seotud ($\rho=0,08$; $p=0,56$).

SUS punktide analüüs

SUS punkte kasutatakse programmi kasutatavuse hindamiseks. Keskmiseks SUS hindeks loetakse 68 punkti. Hinnangud, mis jäävad alla 51 punkti, saavad hindeks F (skaalal A...F, kus F on kõige madalam hinne). Hinne F tähistab alumist 15% tulemustest. Kui SUS tulemus on üle 80,3, siis on tõenäoline, et kasutajad soovitsid seda programmi ka oma sõpradele. SUS punktide hindamiskaala on välja toodud Joonisel 3.(Measuring Usability With The System Usability Scale (SUS))



Joonis 7. SUS punktide hindamisskaala (Measuring Usability With The System Usability Scale (SUS))

Punktide arvutamiseks kasutatakse järgmist valemit:

Küsimuste 1, 3, 5, 7 ja 9 vastused annavad punktideks tabelis oleva positsiooni -1. Vastused 2, 4, 6, 8 ja 10 annavad punktideks 5 miinus tabelis olev positsioon. Punktid arvutatakse kokku ja korrutatakse 2,5-ga.

Meie tabeli põhjal arvutatud punktid on järgmised:

- 1) 27 (kõikide vastajate punktid kokku) / 45 (vastajate arv) = $0,6$ (esimese küsimuse keskmine punktisumma)
- 2) $55/45=1,2$
- 3) $36/44=0,8$
- 4) $33/44=0,8$
- 5) $45/45=1$
- 6) $65/45=1,4$
- 7) $55/44=1,3$
- 8) $37/45=0,8$
- 9) $23/45=0,5$
- 10) $59/45=1,3$

Kokku $9,7 * 2,5 = 24,25$ (Kõikide vastuste kogusumma korrutatud 2,5-ga)

Tulemus 24,25 annab hindeks kindlalt F-i, mis näitab, et kasutajate arvates oli programmi raske installeerida ja kasutada. SUS küsimuste 4 ja 10 põhjal tuleb välja, et küsitlusele vastanud tudengid vajavad programmi kasutamiseks tehnilist abi või lisamaterjale.

Tabelis 7. on toodud SUS küsimustiku vastuste põhjal arvatud keskväärtused („ei nõustu üldse“ = 1 ja „olen täiesti nõus“ = 5). Väited on esitatud kahes blokis nii, et esimesena antakse ülevaade programmiga rahulolu kohta. Nende väidete puhul näitab kõrgem keskväärtus suuremat rahulolu. Kõige enam (M=2,25; SD=1,1,) jäid vastajad rahule 7. väitega "Ma kujutan ette, et enamus inimesi õpib seda programmi kasutama väga kiiresti". Olles selles blokis kõige kõrgema keskmisega väide, jääb see siiski alla skaala keskmise, mistõttu võib taaskord järeldada, et üldine rahulolu programmi kasutamisega oli madal. Teises blokis said kasutajad väljendada rahulolematust, kus suurem keskväärtus näitab suuremat rahulolematust. Kõige enam nõustuti sellega, et programmi R kasutajad vajaksid tuge selle programmi kasutamisel (M=4,25; SD=0,89) ning, et programmi on väga tülikas kasutada (M=4,18; SD=0,89).

Küsimused	Keskmine (M)	Standardhälve (SD)
1. Arvan, et ma kasutaksin R-i tihti	1,60	0,69
3. Programmi oli kerge kasutada	1,82	0,95
5. Leian, et programmi lisad olid kergesti kasutatavad	2	0,98
7. Ma kujutan ette, et enamus inimesi õpib seda programmi kasutama väga kiiresti	2,25	1,10
9. Ma tundsin ennast väga kindlalt programmi kasutades	1,51	0,73
2. Programm oli tarbetult keeruline	3,78	1,15
4. Ma arvan, et vajan tehnilist tuge selle programmi kasutamiseks	4,25	0,89
6. Programmis oli liiga palju vasturääkivusi	3,56	1,20

8. Leian, et programmi on väga tülikas kasutada	4,18	0,89
10. Mul tuli õppida palju asju, enne kui sain programmi kasutama hakata	3,69	1,18

Tabel 7. Kasutatavuse hinnangute keskvärtus ja standardhälve.

Järgnevalt antakse ülevaade rühmatöös erinevaid rolle täitnud üliõpilaste vastustest.

Roll rühmatöös	N	SUS punktide summa
Installeeris tarkvara	29	19,50
Kasutas tarkvara	6	37,75
Osales rühmatöös, kuid ei installeerinud ega kasutanud tarkvara	10	33,50

Tabel 8. Rühmatöös erinevaid rolle täitnud üliõpilaste vastused.

Programmi ise installeerinud tudengid said SUS punkti summaks 19,5, mis on vastajate keskmisest tulemusest märgatavalt madalam. Punkti summadest ja vastustest võime järeldada, et programmi installeerimine põhjustas kasutajatele kõige rohkem raskusi. Kasutamine ning ülesannete lahendamine oli juba mõnevõrra lihtsam. Kõigi kolme grupi punktid jäid alla 51 punkti, mis tähendab hinnet F, ehk väga madalat kasutusmugavust.

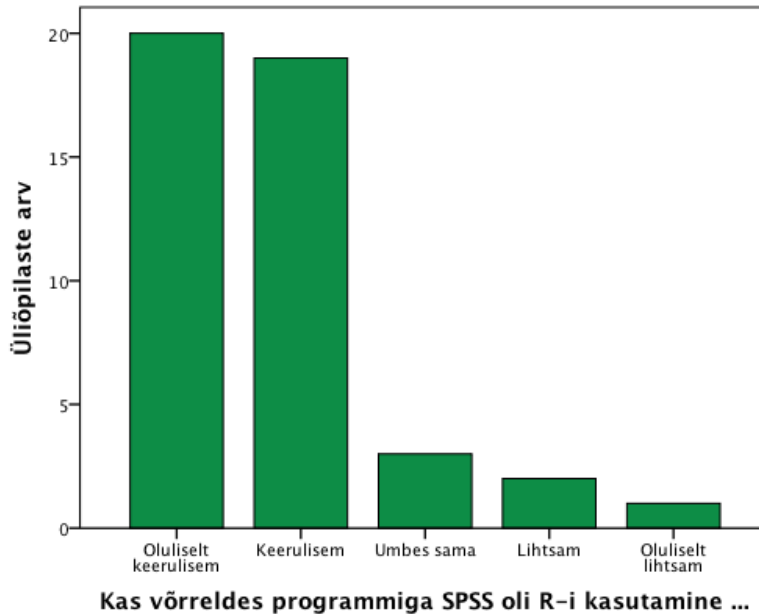
Arvutikasutusoskust said üliõpilased hinnata 10-palli skaalal, kus 1 tähistas algajat arvutikasutajat ning 10 eksperti. Hinnangute võrdlemiseks moodustati järgmised grupid:

Hinnang arvutikasutusoskusele	N	SUS punktide summa
1-4	11	28,30
5-7	26	25,75
8-10	8	17

Tabel 9. Hinnang arvutikasutusoskusele

Esimese grupi vastajad said SUS punkti summaks 28,3, mis tähendab, et tudengid, kes pidasid enda arvutikasutusoskust madalaks, hindasid programmi kasutatavust kõrgemalt kui uuringus osalenud kokku (24,25). Kolmandasse gruppi kuuluvad õpilased pidasid enda arvutikasutusoskust kõige paremaks. Nende SUS punkti summa oli 17, mis näitab, et üllatuslikult pidasid programmi kõige halvemini kasutatavaks just kõige paremate arvutiteadmistega tudengid.

Võrreldes vastajate hinnanguid erinevate programmide kasutamise kohta, peeti programmi R kasutamist võrreldes SPSS-ga keerulisemaks (Joonis 8). Lausa 39 üliõpilast (86%) pidas R-i kasutamist keerulisemaks või oluliselt keerulisemaks. Eelmistest punktidest võime küll järeldada, et suuremat raskust valmistab programmi installeerimine kui kasutamine, kuid vastajate üldhinnang on tugevalt SPSS-i poole kaldu.

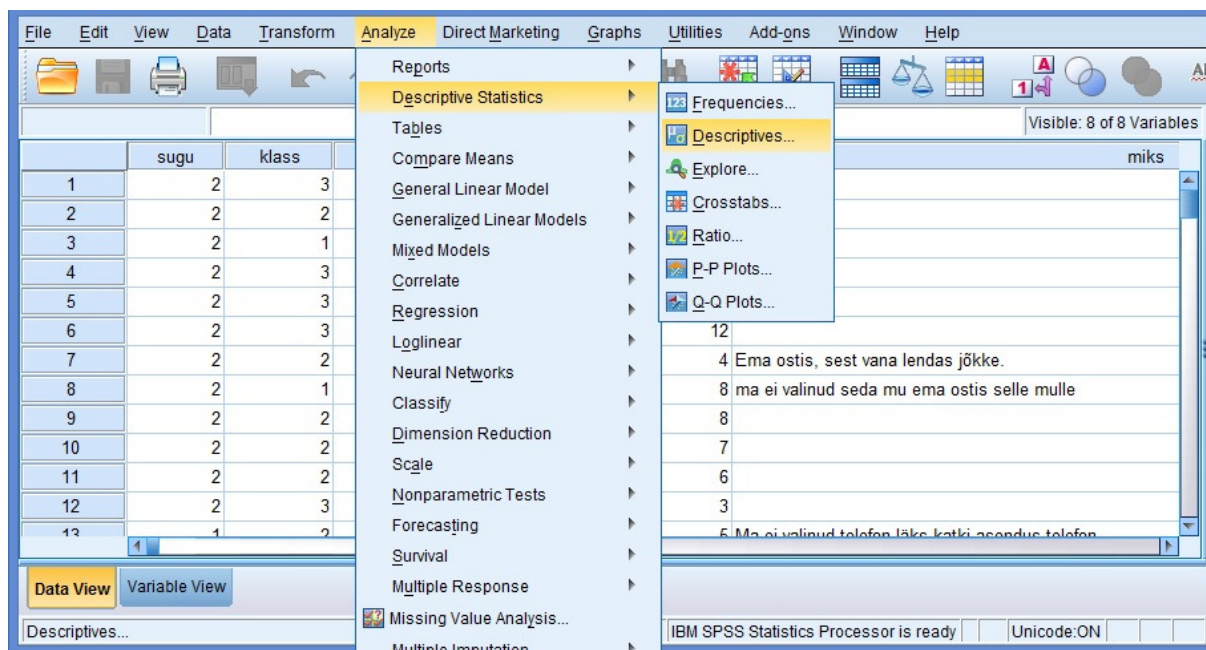


Joonis 8. Kas võrreldes programmiga SPSS oli R-i kasutamine ...

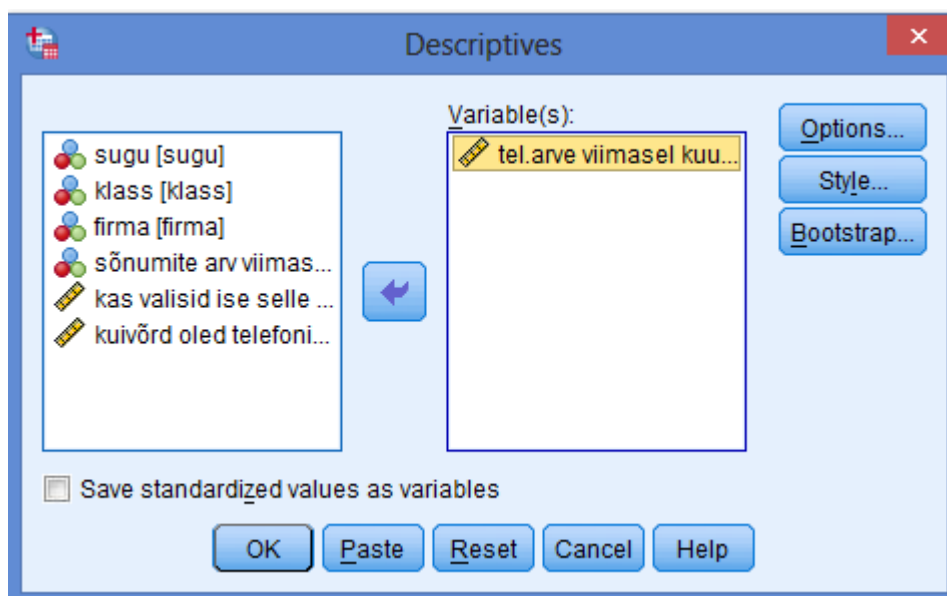
Tudeng, kes ei pea statistilise analüüsi programmi kasutama enda igapäevatööks vaid kursuse koduste tööde lahendamiseks, ei pea kõige olulisemaks programmi suutlikust, vaid just lihtsust ja õppimiskõvera pikkust. Kui SPSS-i prooviversioon internetist peale erinevate tingimustega nõustumist ja enese kasutajaks registreerimist lõpuks kätte saada, siis on installeerimine suuresti „Next“ klahvi korduv vajutamine. Kuid R vajab töö lihtsustamiseks ka lisapakette, mille allalaadimise ja installeerimisega tudengid hätta jäid. Kui programmid on lõppkasutajale ettevalmistatud (kooli arvutiklass), siis on kasutamine paljuski väga sarnane.

Järgnevalt on toodud võrdluseks üliõpilaste rühmatöö ühe ülesande (kirjeldavate arvnäitajate arvutamine) lahendamine nii programmiga R kui ka SPSS.

Kirjeldavate arvnäitajate arvutamine programmiga SPSS.



Joonis 9: SPSS-i kirjeldava analüüsi vaate avamine



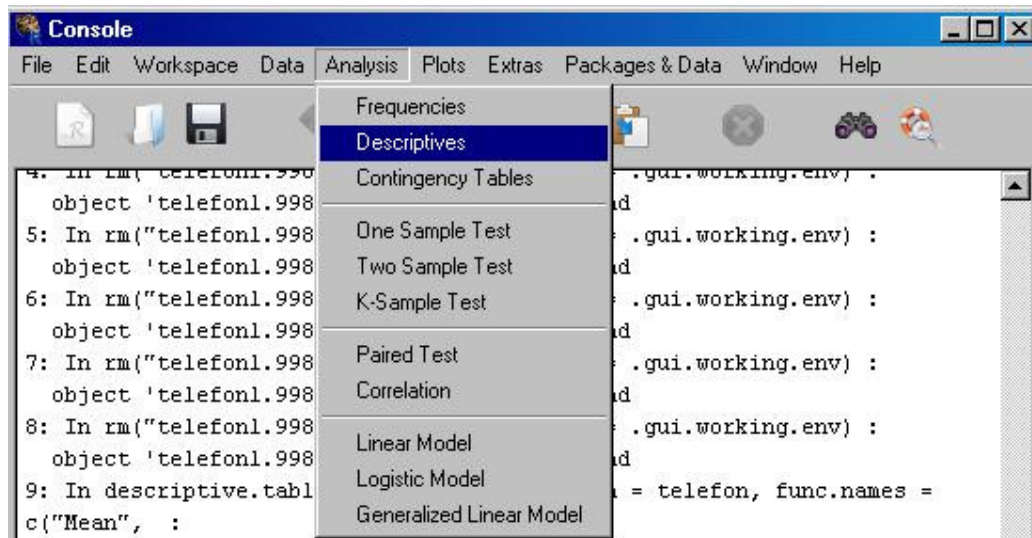
Joonis 10: SPSS-i kirjeldava analüüsi muutuja määramine.

Descriptive Statistics

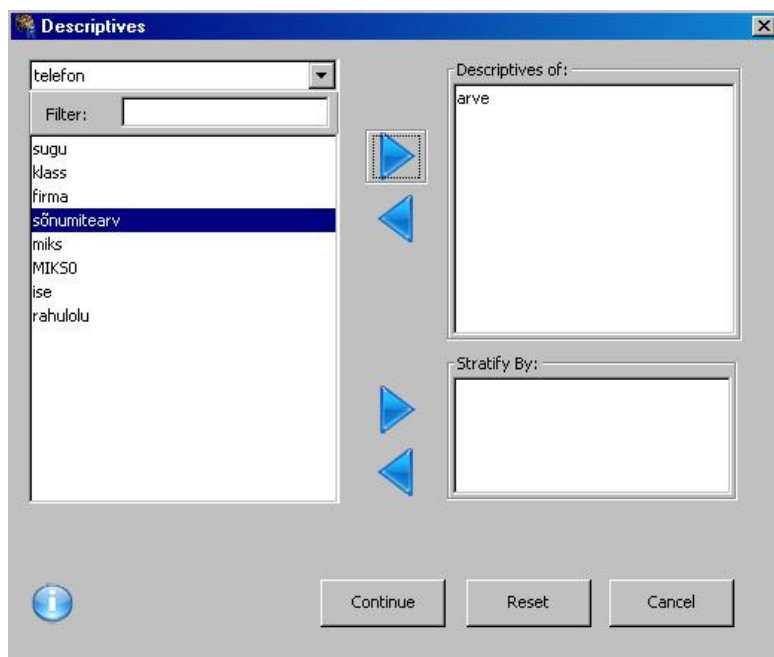
	N	Minimum	Maximum	Mean	Std. Deviation
tel.arve viimasel kuul	102	2,36	28,22	13,0085	5,42348
Valid N (listwise)	102				

Tabel 10: SPSS-i vastus

Kirjeldavate arvnäitajate arvutamine programmis R.



Joonis 11: R-i kirjeldava analüüsi vaate avamine



Joonis 12: R-i kirjeldava analüüsi muutuja määramine.

Descriptive Statistics

	Mean	St. Deviation	Valid N	Median	Minimum	Maximum
arve	13.01	5.42	102	12.35	2.36	28.22

Joonis 13: R-i vastus

Küsimusele „Kui kaua võttis aega R-i installeerimine ja ülesannete lahendamine?“ said kasutajad vastata vabateksti vormis. Vastused algasid 15 minutist ja lõppesid mitme päeva või ebaõnnestumisega (võttes arvesse, et kõik grupid koduse töö esitasid, siis keegi igast grupist suutis siiski programmi tööle saada ja seda kasutada). Kui installeerimise aega oli mitmel korral mõõdetud päevades, siis lahendamise peale oli enamasti kulunud tund või vähem.

Küsimusele „Kirjuta lühidalt, mis sulle meeldis ja mis häiris programmide R ja SPSS juures“ said kasutajad vastata jällegi vabatekstina. Vastustest ilmnes, et enim valmistas probleeme R-i lisade allalaadimine ja kuna õiged „linnukesed“ jäid tegemata, siis pidi igal programmi käivitamisel lisad uuesti laadima. Samuti oli probleeme vastuste ilusa vormistamisega (kuna selleks vajaliku lisamooduli installeerimisega jäädi hätta). Ka andmete eksportimine jättis soovida.

SPSS-i peeti lihtsamaks, selgemaks ja loogilisemaks. Tudengid olid seda varasematel kursustel kasutanud ja omasid suuremat kogemust. Meeldis ilus kasutajaliides ja head andmete eksportimise võimalused.

Küsimusele „Kui sul ei õnnestunud programmi R tööle saada, siis kirjelda protsessi, mis sa tegid ja millist abi oleksid veel oodanud.“ said kasutajad samuti vastata oma sõnadega. Selgus, et põhiliseks komistuskiviks sai R-i lisa „*DeducerRichOutbut*“, mida on vaja vastuste paremaks vormindamiseks.

Juhendi koostaja arvutisse õnnestus see lisa ilma probleemideta installeerida, kuid paljudel kasutajatel see ei õnnestunud. Siin aitas juhendi muudatus, mis aitas küll mitmel grupil installeerimise edukalt lõpuni viia, kuid vastused esitati kehva vorminguga tabelites. Abi saadi põhiliselt e-kirjade teel minult kui juhendi koostajalt, kuid see ei ole ilmselgelt jätkusuutlik tarkvara kasutajatoe süsteem. Veel leiti abi internetist ja kursusekaaslastelt, kellel oli sarnaseid probleeme tekkinud. Täiendavat abi üldiselt vajalikuks ei peetud.

Analüüsi kokkuvõte

Kasutajate üldine hinnang programmi kasutusmugavusele oli väga madal. Kriitilise hinnangu andmise peamiseks põhjuseks oli tarkvara installeerimine ja töökeskkonna ettevalmistamine. Tarkvara edasine kasutamine oli juba mõnevõrra lihtsam ja ülesannete lahendamisele kulunud aeg lühem. Installeerimise juures valmistas suurimaid probleeme lisapakettide laadimine ja kasutamine.

Selliste probleemide edaspidiseks vältimiseks peaks tudengitele tegema R-i kasutamise kohta ühe sissejuhatava praktikumi. Samuti võib mõelda graafilise liidese väljajätmisele, et algajad kasutajad ei peaks tegelema lisapakettide laadimisega. Lisamaterjalidena peaks looma R-i sissejuhatava praktikumi materjalid. Aluseks võib võtta Märt Mölsi poolt koostatud R-i sissejuhatuse, mis on kasutusel Tartu Ülikooli Biomeetria kursusel <http://www-1.ms.ut.ee/mart/biomeetria2013/praks1.pdf> (Möls, 2014).

3.2. Eksperthinnang

Analüüsimaks vabavaraliste programmide kasutamise tugevusi ja nõrkuseid õppeprotsessis, kaasati uuringusse ka kaks eksperti, kes omavad valikus parimaks osutunud programmi R kasutuskogemust.

R tugevustena nimetati järgmisi aspekte:

- R pakub laia ning paindlikku analüüsimeetodite valikut
- R on tasuta kättesaadav
- Andmefailide transport teiste programmidega on lihtne
- R-i sisend-väljund voog on kasutajale lihtsamini mõistetav kuna väljundis on täpselt see, mida küsiti ja ei midagi rohkemat. Kasutaja peab suutma küsimused õigesti koostada ja oskama analüüsi tulemusi tõlgendada.
- R kogukond on toetav ning abistab foorumites meelega.

Kui võrrelda eelpool toodud tugevuste korral programme R ning SPSS, siis analüüsimeetodite ning tulemi võrdluses väga suurt erinevust ei ole. Küll on aga R-il kindel eelis hinnas ning toe võimalustes.

Negatiivse poole pealt märgiti, et R-i õppimiskõver on küllaltki pikk ka kogenud programmeerija jaoks. Selgus, et reeglina ei kasuta R-iga igapäevaselt töötav inimene graafilist kasutajaliidest vaid pigem programmeerimiskeelt ja skripte. Õppimiskõver on ka kogenud programmeerijale pikk, mis teeb algajale antud programmiga uue aine õppimise raskeks.

Töö alustamise lihtsustamiseks välja pakutud *RStudio* (RStudio) töö autori hinnangul kasutajatele suurt võitu ei anna. Selleks, et töötada graafilise liidesega, peab algul ikkagi pakette lisama ning siis avama uue JGR akna, kus töö toimuks. *RStudio* konsool jääks aga tahaplaanile, mis tähendab kasutaja jaoks veel üht lisaakent ja segadust. Kui võtta suund, et õpetamist alustatakse käsurealt, siis oleks *RStudio* kindlasti kasulik, kuid aine omandamise seisukohast läheks programmi õpetamisele liiga palju aega.

Kui diagrammide koostamine on üldiselt õppijatele sisuliselt hästi mõistetav osa, siis R-i kasutamine graafikute koostamisel võib algajatele osutada üsna keeruliseks, kuna võimalusi on mitmeid. Töö kirjutamisel konsulteeritud ekspert leiab, et R-iga saab graafikuid hästi koostada ja lisapakettide abiga pole selleks teist programmi vaja kasutada.

Professor Märt Mölsi arvates on väga oluliseks vaheks see, kas tudeng on eelnevalt kokku puutunud programmeerimisega või käsurealt arvutile käskude andmisega. Programmeerimisega realselt kokkupuudet omavatel tudengitel (ükskõik millise programmeerimiskeelega) ei tekita analüüsi läbiviimine R-is raskuseid. Samas on Möls siiski neile läbiviidud kursuste raames esimeses praktikumis tutvustanud programmi R ning selle loogika aluseid.

Tudengid, kes küll otseselt ei ole programmeerinud, kuid kellel on siiski kogemus käsurealt programmide käivitamise kohta, kohanevad programmiga R samuti suhteliselt kergesti. Sissejuhatuses tuleb siiski kulutada natuke kauem aega, harjumiseks kulub umbes 2 praktikumi.

Neil tudengitel, kellel puudub varasem kokkupuude programmeerimisega või käsurealt käivitavate töövahenditega, on õppimiskõver pikem - nende jaoks on tegemist põhjaliku mõttemaailma muutusega ja selline muutus võtab paratamatult aega. Näiteks arstiteaduskonna bakalaureuseõppes või filoloogidele suunatud statistikakursuses Möls R-i õppetöös pigem ei kasutaks.

R-i kasutamise kohta erinevate eelteadmistega õppijate korral, ütleb Möls järgmist:

„Minu arvates on R-i täiesti edukalt võimalik õppetöös kasutada, kuid tuleb arvestada teatava täiendava ajakuluga R keele veidruste omandamiseks (ajakadu umbes 2 praktikumi).

Samas võidavad need tudengid, kes hilisemas elus rohkelt statistilist analüüsi tegema peavad, märkimisväärselt (R on jätkuvalt minu arvates praegu kättesaadavatest statistikaprogrammidest - ükskõik kas vabavaralised või tasulised - parim töövahend, kuigi ühe või teise konkreetse meetodi puhul võib mõni teine programm - SAS, SPSS, Lisrel vms - paremaks osutuda). Need, kelle kokkupuude statistikaga piirdub pigem teiste inimeste poolt tehtud statistiliste analüüside lugemise/tõlgendamise/tarbimisega - nende jaoks on R-i õppimiseks kulunud aeg aga kahjuks üsna kasutult raisatud aeg.“

Märt Mölsi kursuste tagasiside on olnud väga positiivne (viimasel korral oli keskmine hinne lausa 4,73), mis näitab, et R-i kasutamine õppetöös ei ole isenesest takistuseks head vastuvõttu leidva kursuse tegemisel (peab märkima, et enamike kuulajate jaoks polnud tegemist kohustusliku ainega, mistõttu võis olla ka üliõpilaste motivatsioon ainet õppida kõrgem). Samas oli tagasisides mitmel korral märgitud, et kodutöö tegemine R-iga võttis palju aega ja oli keeruline. Pikemaajaline tagasiside Mölsi kursustele, kus kasutatakse R-i, on olnud positiivne ning kuulajatel on olnud programmi õppimisest kasu ka edaspidises elus.

Eduka kursuse läbiviimiseks annab Möls õppejõududele järgmised soovitusel:

„Õppejõud peab ise uskuma, et see töövahend, mida ta tudengitele tutvustab, on parim võimalikest. Mina tõepoolest usun, et R on praegusel hetkel parim töökeskkond statistilise analüüsi tegemiseks. Sestap on mul ka kerge anda tudengitele edasi tunnet, et ma tutvustan/õpetan neile midagi väga head. Kui ma ise arvaksin, et R on lihtsalt odav aseaine tõeliselt heale töövahendile - vaesuse tõttu pealesunnitud lahendus - siis segaks selline õppejõu eelhäälestus ka vastava õppeaine edukat õpetamist tuntavalt.“

Programmi R poolt pakutavate andmeanalüüsi meetodite mitmekesisus võib nii algajale õppejõule kui ka õppijale kasu asemel kahju tooma hakata. Möls selgitab seda järgmiselt: *„R võib osutuda õppejõule lõksuderohkeks töövahendiks. Paljude analüüside korral võib valida kas äärmiselt lihtsa ja elegantse lahenduse, mis on aga paraku spetsiifiline just selle analüüsi/graafiku/olukorra jaoks ja võib leida ka veidi pikema tee lahenduse, mis seevastu on aga kergesti üldistatav ja kasutatav ka teiste analüüside/olukordade korral. Kui kasutada neid veidi pikemaid ja kohmakamaid, aga paremini üldistatavaid lahendusi, saab tudengitele minu arvates kiiresti kätte anda oskuse toime tulla vägagi erinevate olukordadega. Väiksema kogemusega õppejõud võib aga kergesti vaimustuda leitud elegantsest lahendusest, mis aga paraku ei pruugi tudengit eriti aidata veidikene muutunud olukorras/teistsuguse ülesande korral.“*

Kokkuvõttes on eksperdid ühel meelel, et R-i õppimiskõver võib programmeerimiskaugetele tudengitele olla pikk. Aine õpetamisel R-iga tuleb arvestada täiendava ajakuluga programmi tundmaõppimiseks. Kuid mõlemad toetasid R-i kasutamist õppetöös. Programmi kasutamist soovitati alustada käsurealt, sest hilisem töö käib ilma graafilise liideseta ja nii välditakse ümberõppe vajadust. Hind ja võimalusterohkus teevad R-ist vahendi, mille õppimine tuleb tudengitele kindlasti kasuks.

KOKKUVÕTE

Käesolev magistritöö juhtis tähelepanu probleemidele, mis on tingitud tasuliste programmide kasutamisest õppeasutustes. Andis ülevaate avatud lähtekoodiga tarkvara ajaloost, arengust ning haridusvaldkonnas kasutamisest. Alates kõige madalamatest haridusastmetest tuleks võtta arvesse, et aine läbiviimiseks kasutatavad programmid peavad olema kättesaadavad kõigile õpilastele, hoolimata nende majanduslikust olukorrast.

Sissejuhatuses püstitatud eesmärgid said töö käigus täidetud. Kogutud andmete ja analüüsi tulemusena selgitasin enda hinnangul välja parima vaba lähtekoodiga alternatiivi praegu kasutusel olevale firmaomasele tarkvarale SPSS. Erinevad spetsialistide seas läbiviidud uuringud näitavad, et R on statistikute ja analüütikute seas üks populaarsemaid ja hinnatumaid tööriistu. Seda ka võrdluses firmaomaste programmidega. Sain väljavalitud programmi testida ka tudengite peal reaalses õpikeskkonnas ja samade ülesannetega, mida siiani on kasutatud. Kasutamisele järgnenud küsitlusest ning vastuste analüüsist sai palju väärt informatsiooni. Tudengite hinnang programmile ja juhendile ei olnud positiivne. Järgmisel katsel tuleb sisse viia töös väljatoodud muudatused ning tudengite tagasisidet uuesti analüüsida. Esimesel katsel sai proovitud võimalikult lähedast varianti SPSS-ile, lisades R-ile graafilise kasutajaliidese. R-i kogenud kasutajad soovitasid edaspidi kulutada rohkem aega programmiga tutvumisele ja alustada tööd kohe käsurealt.

Spetsialistide hinnangul on R tarkvara, mis on väga võimekas ja mida tasub õppida, kuid välja tuuakse, et programmi töö selgekssaamine võtab aega. Õppimiskõver on pikem kui kasutajad ei ole kokku puutunud programmeerimise või käsurealt programmide käivitamisega. Kogenud kasutajad ei tööta R-is graafilise kasutajaliidese ja soovivad sellist lähenemist ka tudengitele. See omakorda teeb õpikõvera pikemaks kuid suurendab programmi tööst ja põhimõtetest arusaamist.

Evaluating the Transition to Open-Source Software on the Data Analysis Course Case Study

Master's Thesis

Kristo Talmar

SUMMARY

This thesis draws attention to the problems caused by the use of proprietary software in educational institutions. It gives an overview of the history and development of open-source software and its use in educational systems. It should be taken into account that the software used in all levels of education must be accessible to all students, regardless of their economic situation.

The stated objectives in introduction were met during the work. With the collected data and its analysis I chose the best open-source alternative to the currently used proprietary software SPSS. Different studies conducted among statistician and analyst specialist show that R is one of the most popular and valued tool for this kind of work. That is also true when we compare it with proprietary softwares. I had the opportunity to test the program on students in their real learning environment and with the same assignments they would have done with SPSS. From the answers and analysis of the follow-up survey I was able to collect much good information to use in the future of this project. Students' assessment of the program and instructions were not positive. On the next attempt the changes suggested in my work should be made and students' feedback should be re-analyzed. On the first attempt the R was made as similar to SPSS as possible by adding graphical user interface. Experienced users of R advise to spend more time to the introduction of the program and start using it from command line right from the beginning.

In the assessment of R specialist the software is very capable and worth the input time. They also state that learning curve of R might be long if the person hasn't had experience with programming. Experienced users do not work in the graphical user interface and recommend that new users start using the program from command line. This will make the learning curve longer but will give better understanding of the principles and program work.

KASUTATUD MÕISTED

Vaba tarkvara - (free software), FSF'i (Free Software Foundation) poolt levitatav avatud koodiga tarkvara, mida kasutajatel on õigus paremaks muuta ja levitada kas tasuta või tasu eest. Tingimuseks on seejuures, et muudatused oleksid korralikult dokumenteeritud ja lähtekood jääks avatuks.

Vaba tarkvara on tarkvara, mida kasutajal on vabadus käitada, kopeerida, uurida, muuta ja täiendada. Olulised on siinkohal neli vabadust:

- * Vabadus käitada programmi mis tahes eesmärgil
- * Vabadus uurida, kuidas programm töötab ning seda soovikorral oma vajadustele vastavaks muuta (eeldab ligipääsu lähtekoodile)
- * Vabadus jaotada koopiaid nii, et ka Sinu naaber saab sellest abi
- * Vabadus jaotada enda poolt muudetud versioone teistele. Tehes seda võimaluse saada kasu Sinu tehtud muudatustest (eeldab ligipääsu lähtekoodile)

Kui kasutajal on kõik need vabadused, siis on tegemist vaba tarkvaraga.

<http://www.heiki.org/post/Vaba-tarkvara-avatud-lahtekood-ja-vabavara.aspx> - Heiki Tähis

<http://www.vallaste.ee/> - vabavara, freesoftware

Tasuta tarkvara (freeware) priivara - autoriõigusega kaitstud tarkvara, mida autor lubab tasuta kasutada kas kõigil soovijatel või teatud kasutajate rühmal, näit. haridusasutustel. Priivara kasutamine oma isiklikuks tarbeks on vaba, kuid seda ei tohi edasi müüa (levitada kommertseesmärkidel). Tavaliselt kirjutab autor täpselt ette, mida konkreetse priivaraga tohib teha ja mida mitte.

Priivara on ühtlasi ka omanduslik tarkvara

<http://www.vallaste.ee/> - priivara, freeware

<http://viki.pingviin.org/Priivara>

Omanduslik tarkvara - tarkvara, mis on saadaval tasuta (ilma raha maksmata), kuid on omanduslik, kuna kasutajatel ei ole vabadust seda kasutada, kopeerida, uurida, muuta ja levitada välja arvatud tarkvara looja lubatud juhtudel. Tegemist on tavaliselt suletud lähtekoodiga ning tasulise tarkvaraga ja binaarfailide dekompileerimine ei ole lubatud. Võidakse jagada ka priivara kujul ehk siis küll tasuta ent mitmete piirangutega - näiteks tohib kasutada vaid kodus ja mitte kuskil mujal; ei tohi levitada, täiendada jne. Ostes omandusliku tarkvara litsentsi kaasnevad sellega tavaliselt mitmed piirangud. Näiteks tohib seda paigaldada vaid ühele (või teatud arvu) arvutile, seda ei tohi edasi levitada, ei tohi dekompileerida (see on tavaliselt ka takistatud), ei tohi muuta ega muudetud versiooni levitada.

Omandusliku tarkvara ostmisega kaasneb õigus kasutada litsentsilepingus lubatud tingimustel kuid tarkvara jääb endiselt selle loojate omandusse - sealt ka nimetus "omanduslik tarkvara".

http://viki.pingviin.org/Omanduslik_tarkvara - omanduslik tarkvara

Vaba Tarkvara Fond - FSF (Free Software Foundation) - vabavara loomisele ja levitamisele pühendunud organisatsioon. Fondi põhitegevuseks on Richard Stallmani poolt 1984.a. asutatud GNU projekti toetamine. Projekti aluseks on Stallmani seisukoht, et informatsioon kuulub rahvale ja nii peab ka tarkvara olema kõigile vabalt kättesaadav.

FSF egiidi all loodud tarkvara levitatakse GNU litsentsi (GNU General Public Licence) alusel. Tarkvara on küll tasuta saadaval, kuid kasutajaid kutsutakse üles toetama FSF tegevust tarkvara levitamisteenuste eest tasumisega ja/või vabatahtlike annetustega. Need ongi FSF peamised sissetulekuallikad. FSF-i tegevust toetavad ka paljud tuntud riistvara tootjad.

<http://www.vallaste.ee/> - vaba tarkvara fond, FSF

Avatud lähtekood (Open source) - Üldjuhul nimetatakse avatud lähtekoodiks mistahes programmi, mille lähtekood on tehtud programmeerijatele ja kasutajatele kättesaadavaks nii kasutamiseks kui muutmiseks. Patenttarkvara (omandiõigusega kaitstud tarkvara) tootjad üldiselt ei avalda lähtekoode. Avatud lähtekoodiga tarkvara töötatakse välja ja arendatakse koostöös avalikkusega ning see on saadaval tasuta

<http://www.vallaste.ee/> – open source

Üldine Avalik Litsents (GNU General Public License) - FSF (Free Software Foundation'i) tarkvara ja GNU projekti litsents, mis garanteerib kasutajatele vabaduse seda tarkvara levitada ja modifitseerida. Sõna "free" tähendab siin nimelt "vaba", mitte "tasuta", sest FSF võtab uue tarkvara avaldamisel selle eest tasu.

<http://www.vallaste.ee> – GNU GPL (GNU General Public License)

Jaosvara, ühiskasutusega tarkvara (shareware) - algselt nimetati jaosvaraks sellist tarkvara, mida jaotati n.ö. aumeeste mängu põhimõttel. Autor lubas jaosvara teatud aja vältel tasuta proovida ja kui see teile meeldis, siis paluti annetada autorile väike rahasumma.

Nüüdseks on jaosvara mõiste sisu muutunud. Seda saab endiselt tasuta alla laadida, kuid kasutusaeg on piiratud (nn. prooviaeg, näit. 1 kuu) või on osa funktsioone blokeeritud. Kui soovite jaosvara kasutada ka pärast prooviaja lõppu või aktiveerida kõik funktsioonid, siis tuleb tasuda nõutud summa. Kui olete maksnud, registreeritakse teid jaosvara pakkuja juures ning te hakkate regulaarselt saama kõnealuse tarkvara täiendusi, uusi versioone ja uudiskirju.

Jaosvara on odav, sest selle on tavaliselt kirjutanud eraisikust programmeerija või mõni väike firma ja seda pakutakse üle Interneti otse klientidele. Jaosvara erineb avalikust tarkvarast (public domain software) selle poolest, et jaosvara autoriõigused on kaitstud, s.t. seda ei tohi oma nime all edasi müüa

<http://www.vallaste.ee> – Jaosvara (shareware)

Avatud Lähtekoodi Algatus (OSI - Open Source Initiative) - Organisatsioon, mis on pühendunud avatud lähtekoodiga tarkvara levitamistingimuste definitsiooni haldamisele ja edendamisele ühiskonna hüvanguks

<http://www.vallaste.ee> – OSI (Open Source Initiative)

Partnervõrk (P2P Peer-to-Peer network) - Partnervõrk, võrdõigusvõrk arvutivõrkude puhul selline võrk, kus kõik laua- ja sülearvutid töötavad samaaegselt nii klientide kui serveritena ja jagavad oma faile kõigi teiste võrgus olevate kasutajatega. Võrdõigusvõrke kasutatakse sageli väikestes büroodes, kus pole mõtet üles seada spetsiaalset failiserverit. Harilikult tehakse igal

masinal kõigile ligipääsetavaks ainult teatud kataloogid, mitte kogu kõvaketas.

Internetis tähendab P2P ajutist internetis moodustatud võrku, mis võimaldab sama võrgustamisprogrammi kasutaval arvutikasutajate rühmal üksteisega ühenduses olla ja omada vahetut juurdepääsu teiste rühmaliikmete arvutite kõvaketastele. Firmsid loodavad kasutada P2P tarkvara selleks, et nende töötajad saaksid omavahel faile vahetada ilma, et oleks vaja raha kulutada tsentraalse serveri peale ja et firmsid saaksid omavahel informatsiooni vahetada ilma vahendajata.

<http://www.vallaste.ee> – partnervõrk (P2P Peer-to-Peer network)

Firmaomane (proprietary) - Omandiõigusega kaitstud toode või süsteem. Firmaomase tarkvara lähtekoodi ei avaldata.

<http://www.vallaste.ee> – Firmaomane (**proprietary**)

Omamise kogukulud (TCO, Total Cost of Ownership) - on termin personaalarvuti omamisega seotud summaarsete kulude kohta. See sisaldab arvuti- ja tarkvara ostuhinda, riist- ja tarkvara värskenduste hinda, hoolduskulusid, koolituskulusid, seisakutest tekkivat kahjumit jne. Hinnanguliselt ületab TCO ligikaudu 3 kuni 4 korda arvuti ostuhinna.

<http://www.vallaste.ee/> - Omamise kogukulud (TCO, Total Cost of Ownership)

KASUTATUD KIRJANDUS

- ADaMSoft. (2014). Kasutamise kuupäev: 26.04.2014. a., allikas
<http://adamssoft.sourceforge.net/>
- asta ElecoSoft. (2014). Kasutamise kuupäev: 11.04.2014. a., allikas Implementation:
<http://www.astadev.com/services/implementation/>
- Becta. (2005). Open Source Software in Schools.
- Cornford, M. S. (2011). *Total Cost of Ownership of Open Source Software*. London: LSE Research Online.
- Fuchs, I. H. (2008). Challenges and Opportunities of Open Source in Higher Education. *The Tower and The Cloud*, 150-158.
- GNU Operating System. (2014). Kasutamise kuupäev: 28.04.2014. a., allikas GNU PSPP:
<https://www.gnu.org/software/pspp/>
- Hopkinson, A. J. (2007). Information Technologies in Education in the 21st Century. (A. j. Hopkinson, Toim.) *Open source software in higher education*.
- Horm, I. (2012). *Vaba tarkvara kasutamine võrguteenuste osutamisel Tallinna*. Tallinna Tehnikaülikool.
- KDnuggets. (2013). KDnuggets Annual Software Poll. Kasutamise kuupäev: 15.1.2014. a., allikas <http://www.kdnuggets.com/2013/06/kdnuggets-annual-software-poll-rapidminer-r-vie-for-first-place.html>
- Kebbinau, K. (2014). Vaba tarkvara - mis ja milleks? Tallinn. Kasutamise kuupäev: 16.1.2014. a., allikas
https://www.dropbox.com/sh/nedmt0f0bnxh5bz/ug6RFCWB0Q/Vaba_tarkvaramis_ja_milleks_Kalle-Kebbinau.pdf
- Laugasson, E. (2014). Vabavara Eesti koolidesse. (lk 30). Hariduse infotehnoloogia sihtasutus. ALVATAL-i juhatuse liige, TLÜ doktoranti, Eesti Informaatikaõpetajate Seltsi juhatuse liiget.

- Measuring Usability With The System Usability Scale (SUS)*. (kuupäev puudub). Kasutamise kuupäev: 30.03.2014. a., allikas Measuring Usability:
<http://www.measuringusability.com/sus.php>
- Muenchen, R. A. (2012). The Popularity of Data Analysis Software. Kasutamise kuupäev: 9.12.2013. a., allikas <http://r4stats.com/articles/popularity/>
- Möls, M. (2014). *Biomeetri bioloogidele*. Kasutamise kuupäev: 15.04.2014. a., allikas Sissejuhatus R-i: <http://www-1.ms.ut.ee/mart/biomeetria2013/praks1.pdf>
- N. Pankaja, M. R. (2013). Proprietary software versus Open Source Software for Education. *American Journal of Engineering Research (AJER)*, 2(7), 124-139.
- Pöldmaa, H. (2011). Tarkvara hindamise põhimõtted. Tallinn: www.riso.ee. Kasutamise kuupäev: 20.03.2014. a.
- Pöldmaa, H. (2012). Soovitused vabavara kasutamiseks riigiasutustes. Kasutamise kuupäev: 19.11.2013. a., allikas <http://www.riso.ee/et/koosvoime/tarkvara>
- Rexer. (2011). Rexer's Annual Data Miner Survey. Kasutamise kuupäev: 4.12.2013. a., allikas <http://www.rexeranalytics.com/Data-Miner-Survey-Results-2011.html>
- R-project. (2014). *R, Introduction to*. Kasutamise kuupäev: 14.04.2014. a., allikas R: <http://www.r-project.org/>
- RStudio*. (2014). Kasutamise kuupäev: 20.03.2014. a., allikas Welcome to RStudio: <https://www.rstudio.com/>
- Sage*. (2014). Kasutamise kuupäev: 27.04.2014. a., allikas <http://www.sagemath.org/>
- SCaViS*. (2014). Kasutamise kuupäev: 27.04.2014. a., allikas <http://jwork.org/scavis/>
- Stallman, R. M. (1992). Free Software, Free Society: Selected Essays of Richard M. Stallman. (J. Gay, Toim.) *Releasing Free Software if You Work at a University*.
- Statistics Open For All*. (2014). Kasutamise kuupäev: 27.04.2014. a., allikas SOFA: <http://www.sofastatistics.com/home.php>
- Stetson, T. s. (2004). A Comparison of Questionnaires for Assessing Website Usability. *UPA 2004 Konverents*.

- SurveyMonkey*. (2014). Kasutamise kuupäev: 30.03.2014. a., allikas
<https://www.surveymonkey.com/>
- Tammeorg, T. (22.04.2014. a.). Tallinn säästaks Linuxiga 700 000 eurot. *EestiPäevaleht*.
- Tartu Ülikooli seadus*. (16.02.1995). RTI 1995, 23, 333.
- The Open Source Initiative*. (2014). Kasutamise kuupäev: 30.10.2013. a., allikas
<http://opensource.org/osd>
- The World Bank*. (2014). Kasutamise kuupäev: 25.04.2014. a., allikas
http://data.worldbank.org/about/country-classifications/country-and-lending-groups#High_income
- Tom Caswell, S. H. (2008). International Review of Research in Open and Distance Learning. *Open Educational Resources: Enabling universal education*, 9(1).
- W3Techs*. (2014). Kasutamise kuupäev: 23.11.2013. a., allikas World Wide Web Technology Surveys: http://w3techs.com/technologies/overview/operating_system/all
- Wikipedia*. (2014). Kasutamise kuupäev: 26.04.2014. a., allikas Comparison of statistical packages: http://en.wikipedia.org/wiki/Comparison_of_statistical_packages
- Wiley, D. (2006). Open source, openness, and higher education. *Innovate Journal of Online Education*, 3(1).
- Wiley, D. (2006). Testimony to the Secretary of Education's Commission on the Future of Higher Education. Kasutamise kuupäev: 16.11.2013. a., allikas
<http://www.ed.gov/about/bdscomm/list/hiedfuture/3rd-meeting/wiley.pdf>
- Wiley, D. (2006). Testimony to the Secretary of Education's Commission on the Future of Higher Education. Kasutamise kuupäev: 23.11.2013. a., allikas
<http://www.ed.gov/about/bdscomm/list/hiedfuture/3rd-meeting/wiley.pdf>

LISAD

LISA 1. Kursuseprogrammid IFI7041 ja IFI7044























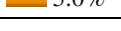

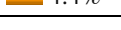


Kursuseprogramm IFI7041		ANDMEANALÜÜS: statistiline andmestik ja kirjeldav statistika	
Maht: 4 EAP	Kontaktundide maht: 20 tundi (loeng 4, praktikum 16)	Õppesemester: S	Eksam
Eesmärk:		Toetada andmete kogumiseks, töötlemiseks ning lihtsama statistilise analüüsi läbiviimiseks vajaminevate teoreetiliste teadmiste ja praktiliste oskuste omandamist. Tutvustada statistikapaketi kasutusvõimalusi andmete töötlemisel ning statistilisel analüüsil. Arendada teadmiste ja oskuste praktilise rakendamise-kogemuse kujunemist, mis võimaldab teha iseseisvalt otsustusi sobiva(te) analüüsimeetodi(te) valikuks ning analüüsi käigus saadud tulemuste korrektseks esitamiseks.	
Aine lühikirjeldus: (sh iseseisva töö sisu kirjeldus vastavuses iseseisva töö mahule)		Statistiline andmestik, selle kogumine ning töötlemine. Statistiliste tunnuste tüübid. Kirjeldav statistika e. andmete kokkuvõtu- ning esitlusmeetodid: erinevad tabelid, diagrammid ja arvnäitajad. Seoste kirjeldamine: seosekordajad ning risttabelid. Sobiva analüüsi- või esitlusmeetodi valik. Mitme valikuvõimalusega küsimuste analüüs. Kursuse põhiosa koosneb seminari tüüpi loengutest ja praktikumidest, kus üliõpilastelt eeldatakse aktiivset kaasamõtet ja -töötamist. Lisaks tuleb igal üliõpilasel teha mahukas iseseisv töö, mis koosneb õppejõu poolt ette antud praktilistest andmeanalüüsi ülesannetest kogu läbitud materjali ulatuses. Kasutatavad andmestikud võivad olla kas õppejõu poolt ette antud või üliõpilaste poolt mingi teise aine raames kogutud (nende kasutamine tuleb õppejõuga eelnevalt kooskõlastada). Hinde saamiseks tuleb sooritada ka kirjalik test.	
Õpiväljundid:		Koostab korrektse ülesehitusega andmestiku. Püstitab andmetest ning eesmärkidest lähtuvalt statistilist analüüsi eeldavaid küsimusi. Selgitab käsitletud kirjeldava statistika meetodite olemust, teab nende rakendamise tingimusi ning oskab analüüsi tulemusi korrektselt tõlgendada. Eristab andmete/tunnuste tüüpe ning valib vastavalt andmete tüübile ning andmete kohta esitatud küsimuse sisule sobiva analüüsi meetodi (käsitletud meetodite piires). Kasutab juhendmaterjali abiga vastavat tarkvara lihtsama andmetöötluse ja -analüüsi läbiviimiseks.	








Hindamise meetodid:	Eksam (hindeline). Hinne kujuneb iseseisva töö (50% lõpphindest) ja avatud küsimustega testi (50% lõpphindest) koondtulemuste põhjal. Kirjalikku testi hinnatakse õppekorralduse eeskirjas toodud alustel. Iseseisvad tööd tuleb praktikumides etteantud tähtajaks esitada õp. Taivo Tuulingule. Teoreetilise osa test sooritatakse kursuse lõpus õp. Kairi Osulale.
Õppejõud:	lekt. Kairi Osula, õp. Taivo Tuuling
Ingliskeelne nimetus:	Data Analysis: Descriptive Statistics
Eeldusained:	Arvutikasutuse elementaarskused IFI6001 mahus
Kohustuslik kirjandus:	Loengumaterjalid
Asendus kirjandus: (üliõpilase poolt läbi töötatava kirjanduse loetelu, mis katab	Niglas, K. Videoloengud andmeanalüüsist (saadaval informaatika instituudis) Niglas, K. Statistika loengumaterjale (http://www.tlu.ee/~katrin/) Niglas, K. (2007) Andmeanalüüs statistikapaketi SPSS 14.00 abil. Tallinn, TLÜ. Hiob, K. (1995) Matemaatiline statistika. Algekursus koolidele, Tallinn

<i>Kursuseprogramm – IFI7044</i>		ANDMEANALÜÜS: ÜLDISTAV STATISTIKA	
4 EAP	Kontakttundide maht: 18	Õppesemester: S	Eksam
Eesmärk:		Luu võimalused erinevuste ja seoste statistilist usaldusväärsust peegeldava analüüsi läbiviimiseks kasutatavate teoreetiliste teadmiste ja praktiliste oskuste omandamiseks. Tutvustada tuntud statistikapaketi SPSS võimalusi üldistava statistika põhimeetodite kasutamisel. Aidata kaasa teadmiste ja oskuste ning praktilise rakendamiskogemuse kujunemisele, mis võimaldab õppijal teha iseseisvalt otsustusi sobiva(te) analüüsimeetodi(te) valikuks ning analüüsi tulemusi korrektselt tõlgendada.	
Aine lühikirjeldus: (sh iseseisva töö sisu kirjeldus vastavuses iseseisva töö mahule)		<p>Teemad:</p> <ul style="list-style-type: none"> • Üldkogum ja valim. Normaaljaotuse idee. • Parameetrilised ja mitteparameetrilised analüüsimeetodid. • Statistiline üldistamine: t-test, χ^2-test, dispersioonanalüüs, Kruskal-Wallise test; seosekordajate statistiline olulisus. • Sobiva analüüsi- ja esitlusmeetodi valik. <p>Korraldus: Kursuse sooritamiseks vajalik töö maht on 104 (4x26) tundi. Sellest kontakttundidena toimuvad seminarid (16 tundi) ja eksam (2 tundi). Peale seminare on üliõpilased kohustatud kinnistama õpitu iseseisva õppimise teel ning lahendama 3 kodust ülesannet (vastavalt teemale, kas personaalsed või rühmatööd). Eksami läbimiseks tuleb sooritada kirjalik avatud küsimustega test. Kursuse hinne kujuneb iseseisvate tööde ja kirjaliku testi koondtulemuse põhjal. Vt. täpsemalt: „Eksami hindamiskriteeriumid“</p>	
Õpiväljundid:		<ul style="list-style-type: none"> • Omab kogemust andmetest lähtuvate ning statistilist üldistamist eeldavate küsimuste püstitamiseks. • Mõistab käsitletud üldistava statistika meetodite olemust, teab nende rakendamise tingimusi ning oskab analüüsi tulemusi korrektselt tõlgendada. • Oskab eristada andmete/tunnuste tüüpe ning valida vastavalt andmete tüübile ning andmete kohta esitatud küsimuse sisule sobivad analüüsi meetodid (käsitletud meetodite piires). • Oskab juhendmaterjali abiga kasutada vastavat tarkvara andmetötluse ja 	

	<p>üldistava statistika meetoditel põhineva analüüsi läbiviimiseks.</p>
<p>Hindamismeetodid:</p>	<p>Eksam (hindeline). Kursuse hinne kujuneb iseseisva töö ja avatud küsimustega testi koondtulemuste põhjal. Positiivse hinde saamiseks on vajalik saada nii testi kui ka iseseisva töö tulemuseks vähemalt 51%. Hindamisele pääsevad kõik ainele registreeritud tudengid vaatamata sellele, kas ja kui palju on nad õppetöö kontakttundides osalenud. Vt täpsemalt kursuseprogrammi alalõik „Hindamiskriteeriumid“.</p>

LISA 2. Internetilehel Kdnuggets läbi viidud uuring

Millist analüüsi, andmekeevanduse, „Big Data“ tarkvara sa kasutasid reaalse projekti tegemisel viimase 12 kuu jooksul? [1880 vastajat]	
Legend: Punane: tasuta/ vabatarkvara programmid	 % kasutajaid 2013
Roheline: Kommerts tarkvara	 % kasutajaid 2012
Rapid-I RapidMiner/RapidAnalytics tasuta versioon(737), 30.9%	 39.2%  26.7%
R(704), 6.5%	 37.4%  30.7%
Excel(527), 0.9%	 28.0%  29.8%
Weka / Pentaho(269), 5.6%	 14.3%  14.8%
Python + numpy/scipy/pandas/iPython... pakettidega(250), 0%	 13.3%  14.9%
Rapid-I RapidAnalytics/RapidMiner firmaomane versioon(225), 52.4%	 12.0%
SAS(202), 2.0%	 10.7%  12.7%
MATLAB(186), 1.6%	 9.9%  10.0%
StatSoft Statistica(170), 45.9%	 9.0%  14.0%
IBM SPSS Statistics(164), 1.8%	 8.7%  7.8%
Microsoft SQL Server(131), 1.5%	 7.0%  5.0%
Tableau(118), 0%	 6.3%  4.4%
IBM SPSS Modeler(114), 6.1%	 6.1%  6.8%

KNIME tasuta versioon (110), 1.8%	 5.9%  21.8%
SAS Enterprise Miner (110), 0%	 5.9%  5.8%
Rattle (84), 0%	 4.5%
JMP (77), 7.8%	 4.1%  4.0%

LISA 3. R-i installeerimise ja kodutöö juhend.

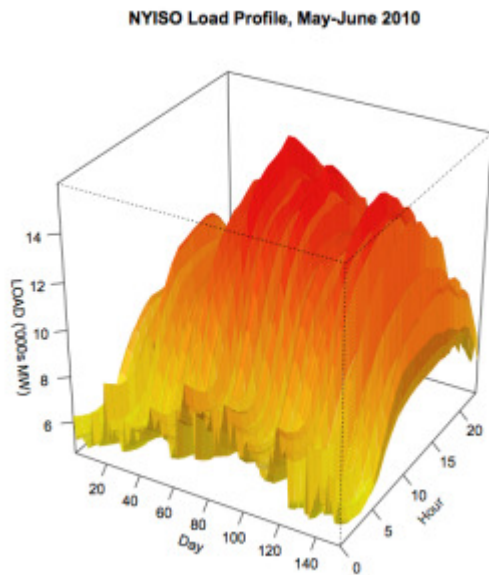
Ülesande kirjeldus

1.1. Andmete kogumine

- Püstitage/sõnastage uurimisprobleem
- Moodustage selle probleemi kirjeldmiseks/uurimiseks viiest küsimusest koosnev küsitlus. Loodavad küsimused peavad olema järgmist tüüpi: 1 binaarne tunnus, 1 järjestustunnus, 2 intervalltunnust, 1 nimitunnus.
- Kasutades veebipõhiseid andmekogumisvahendeid, koostage elektrooniline küsitlus ning koguge kaasõpilaste käest andmeid. Andmete kogumiseks kasutame FB vastavat andmeanalüüsi gruppi (postitage oma küsitlus, määrake vastamisaeg ja vastake teiste postitatud küsitlustele). Selles ülesandes olge palun väga aktiivsed ja mõistvad. Samas ma arvan, et teie miniuuringud tulevad nii põnevatel teemadel, et teil on huvitav nendele küsimustele vastata!
- Vastuseid võiks olla vähemalt 40

Tarkvara R installeerimine

Andmete analüüsimiseks kasutame vabavara R, mis on tasuta andmete analüüsi tarkvara. R-il on väga hea ja laialdane dokumentatsioon, mis on internetis lihtsalt leitav ja tasuta kasutatav. Samuti on R-i kasutajate kogukond väga suur ja abivalmis. R-i kasutajate arv tõuseb kiiresti ja uuendusi ning lisapakette avaldatakse peaaegu iga päev. R-iga on võimalik teha väga huvitavaid ja ilusaid graafikuid:



PS. Pane tähele, et teil on vaja tarkvara installeerida vaid ühte arvutisse. Analüüsimiseks kogunege grupiga selle arvuti juurde;)

- R-i installimiseks klikki ühel järgmistest viidetest:
 Windows: <http://cran.r-project.org/bin/windows/base/>
 Mac: <http://cran.r-project.org/bin/macosx/>
 Linux: <http://cran.r-project.org/bin/linux/>
- Tarkvara installeerimiseks nõustu kõikide vaikesätetega ja vajuta *Next* -i kuni installeerimine jõuab lõpule.
- Töö lihtsustamiseks on vaja installida R-ile mõned lisa paketid (lisade eelduseks on, et arvutis on olemas Java. Java saab vajadusel allalaadida siit: <http://java.com/en/download/index.jsp>). Lisadeks on JGR ja Deducer. JGR on R-i graafiline kasutajaliides, mis on kirjutatud Javas (J-Java G-graafiline liides R-ile). JGR on loodud, et R-i õppimine oleks lihtsam ja kiirem, koodi kirjutamise asemel saab andmeid töödelda ja analüüsida visuaalsete menüüde ja nuppude abil.

Deducer on samuti R-i graafiline liides kuid lisab JGR-ile veel funktsionaalsust. Näiteks Exceli sarnase andmete vaate ja menüü enimkasutatavate statistiliste analüüside jaoks.

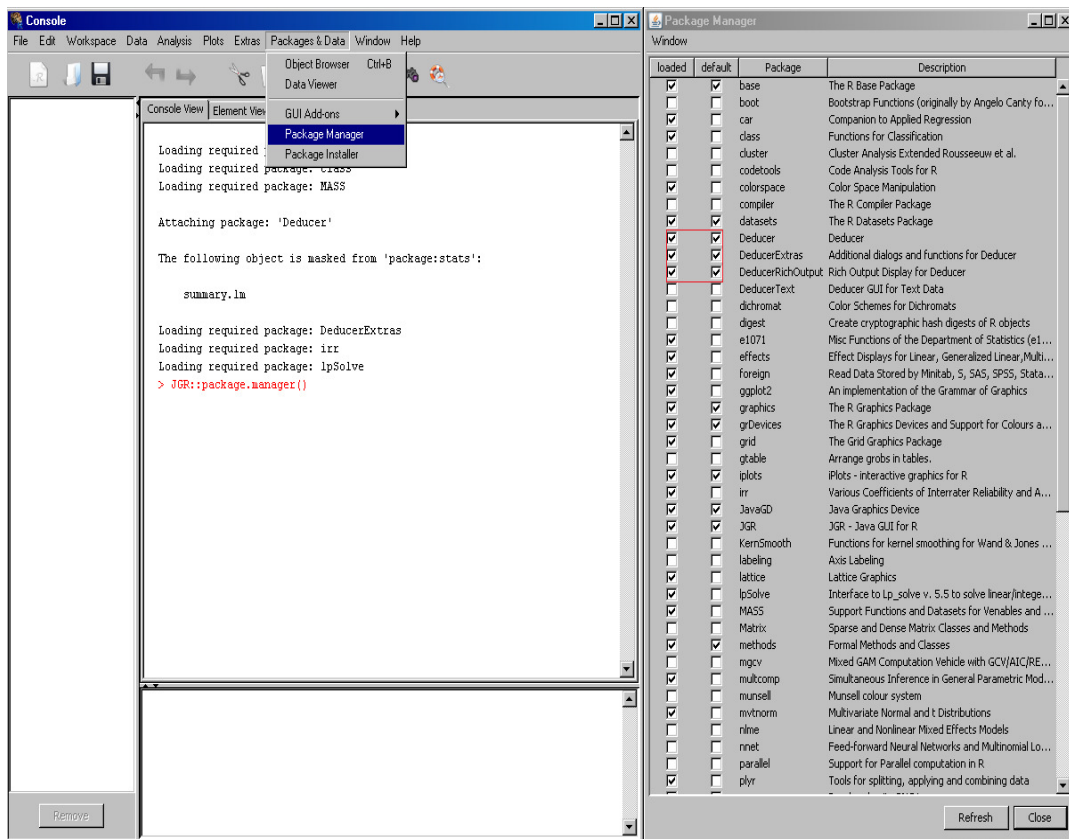
- JGR-i ja Deduceri installimiseks ava R ja kopeeri konsooli aknasse järgnev tekst (sult küsitakse millisest serverist failid allalaetakse. Mõistlik on valida Eestile mõni lähedalasuv riik):

```
install.packages(c("JGR", "Deducer", "DeducerExtras"))
```

```
library(JGR)
```

```
JGR()
```

Selle käsu peale installitakse vajalikud paketid ning avatakse JGR konsool (kui ei avata, vajuta Enter klahvi). Konsooli aknas tuleks valida Packages&Data->Package Manager ja sealt teha linnukesed Deducer, DeducerExtras ja DeducerRichOutbut ees olevatesse kastikestesse.



Edaspidi saab JGR-i käivitada allalaetud exe failist. Fail asub siin:

JGR allalaadimine:

Windows 32 bit: https://rforge.net/JGR/web-files/jgr-1_62.exe

Windows 64 bit: https://rforge.net/JGR/web-files/jgr-1_62-x64.exe

Linuxis tuleks R-i konsooli kopeerida järgmine käsk:

```
install.packages("JGR",dep=TRUE)
```

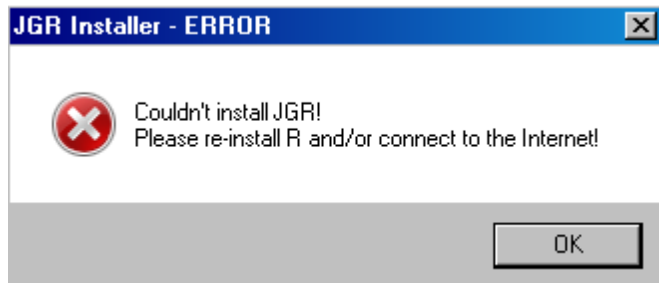
```
library(JGR)
```

```
JGR()
```

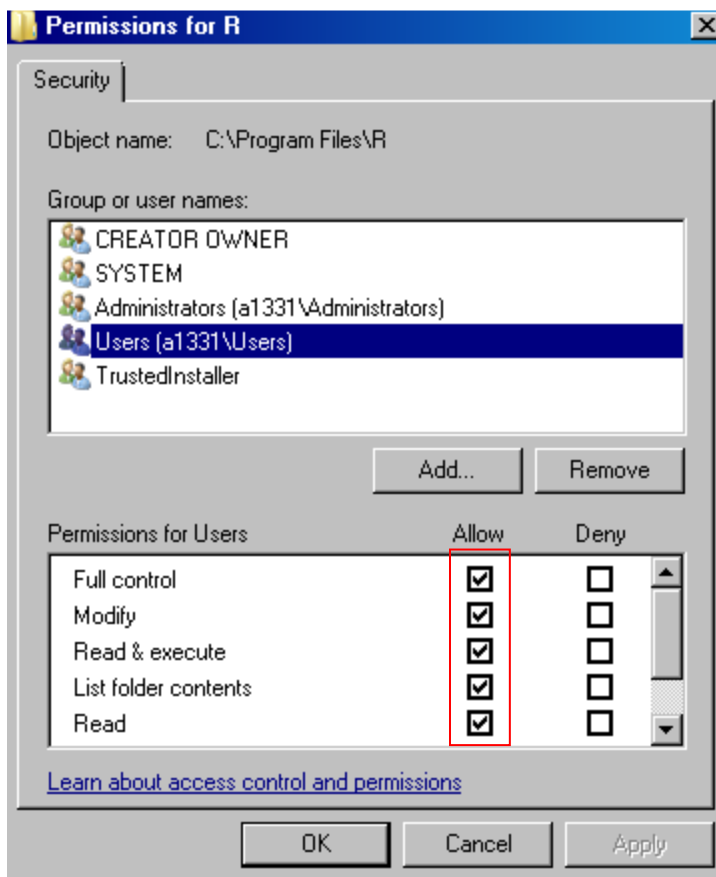
Mac: <https://rforge.net/JGR/files/JGR-1.8.dmg>

- Kui tekkis probleeme loe järgnevaid punkte

1) Kui installeerimisel tekkis alljärgnev viga,



siis peaks kaustale C:\Program Files\R andma enda kasutajale kõik õigused. Selleks ava „My Computer“ -> C ketas -> „Program Files“ ja vajuta kaustal „R“ parempoolset klahvi. Avanenud rippmenüüst vali „Properties“ -> Security -> Edit. Vali kasutajate nimekirjast enda kasutaja ja tee alumises aknas linnuke kasti „Full control“, kui enda kasutajat nimekirjas ei ole, saab selle „Add“ nupuga lisada. Nüüd võib vajutada Apply ja aknad sulgeda.



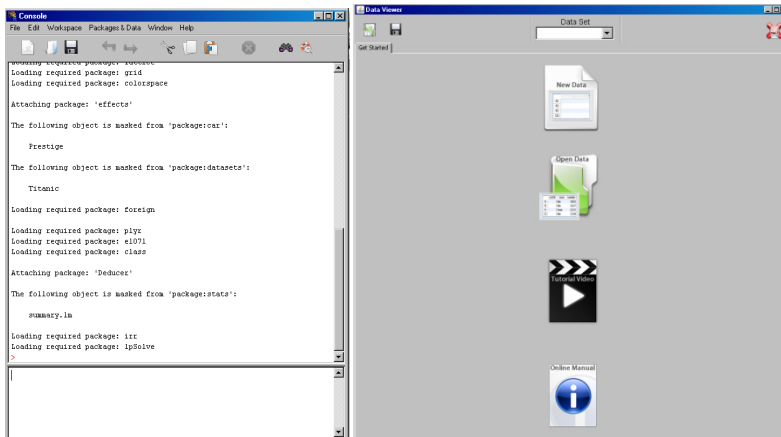
2) Kui JGR-i avamisel ei avane Data Viewer või tekib konsooli aknas mõni muu viga (Võib juhtuda uuemate Windowsite 64 bitisel versioonil), siis avage R ja laadige sealt käsitsi JGR()

konsool R-> Packages -> Load Package... -> JGR. Siis sisestage konsooli käsk JGR() ja vajutage enter. Avaneb JGR-i aken koos kõigi lisadega.

Andmestiku avamine R-s

- Ava andmestik

Avades JGR-i, avaneb lisaks konsooli aknale ka andmete vaatamise aken, seal saab andmeid sisestada või avada valmis andmestikke:



Konsool

Andmete vaade

- Vajutades „Open Data“ saab enda arvutist otsida ja avada varem valmis tehtud andmestik. Andmestik võib olla koostatud Excelis, SPSS-is, PSPP-s või paljudes muudes programmides.

Kirjeldavate arvnäitajate arvutamine

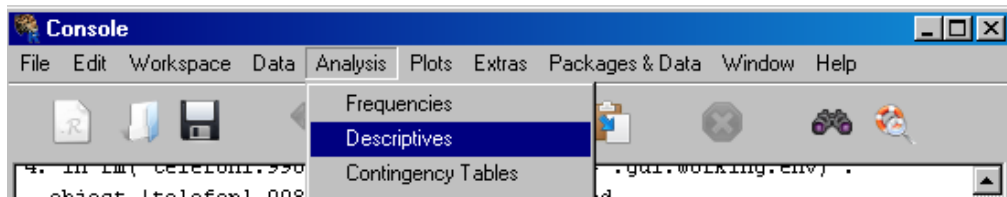
Kasutades R tarkvara, analüüsi kogutud andmeid järgmiselt:

- kirjelda ühe tunnuse väärtuseid kirjeldavate arvnäitajate kaudu. Too välja arvnäitajate väärtused (nii tabelina kui ka tekstis) ja interpreteeri saadud tulemusi.
- võrdle ühe tunnuse väärtuseid kahe (vabalt valitud/moodustatud) grupi lõikes. Esita taaskord tulemused nii tabelina kui ka tekstis. Kirjuta tulemusi avav järeldus.

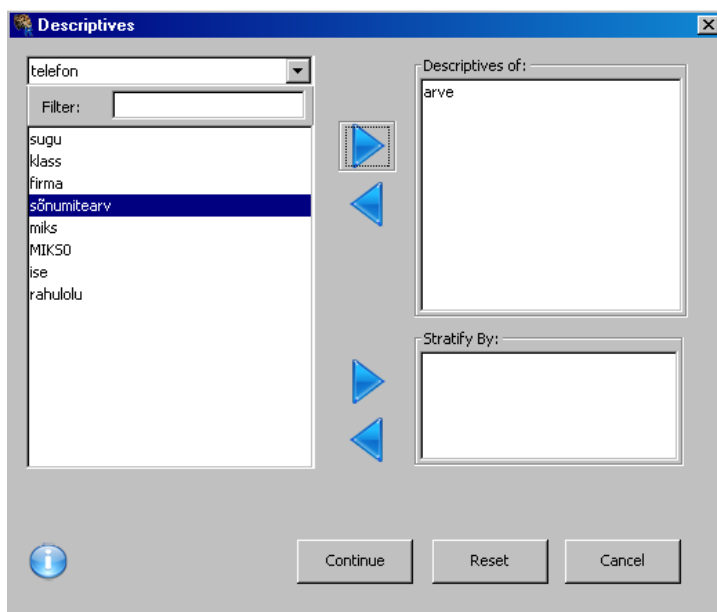
Kirjeldavate arvnäitajate arvutamise juhend

- Kui andmestik on avatud saab edasisi toiminguid teha konsooli aknas.

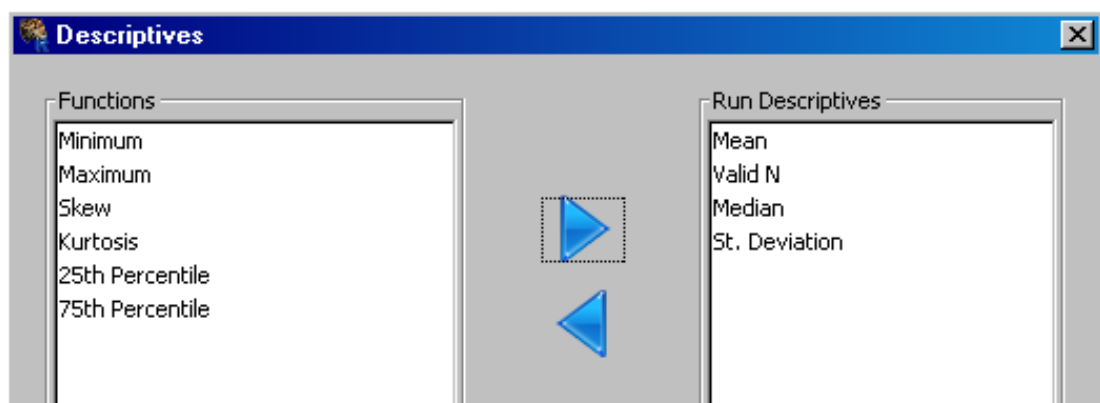
- Kirjeldavate arvnäitajate arvutamiseks vali: *Analysis/ Descriptives*



- Descriptives aknas määrake, millis(t)e tunnus(t)e väärtuste kohta arvnäitajaid arvutama hakatakse, viige vastav(ad) tunnus(ed) Descriptives of väljale.
- Lahter Stratify By: võimaldab jagada vastused mingi tunnuse järgi mitmeks osaks. Näites olevate andmete puhul võib vastused näiteks soo põhjal jagada kaheks, selleks tuleb ära märkida vasakus tulpas „sugu“ ja vajutada alumist paremale suunatud noolt.



- Peale andmete väljavalimist vajutage „Continue“ ning seejärel saab määrata arvnäitajad, mida arvutama hakatakse.



- Valige sobivad arvnäitajad ning klõpsake analüüsi tellimiseks *Run*. Tulemused kuvatakse konsooli aknas.
- Vastus kuvatakse sellisel kujul:

Descriptive Statistics

Variable: arve

	<i>sugu</i>	<i>Mean</i>	<i>St. Deviation</i>	<i>Valid N</i>	<i>Median</i>	<i>Minimum</i>	<i>Maximum</i>
1	poiss	13.47	6.29	37	13.22	2.36	28.22
2	tüdruk	12.75	4.90	65	12.14	3.29	23.90

Kuna vastus on HTML-i kujul, siis saab seda kopeerida ka Exceli tabelitesse. Jooniste ja diagrammide puhul on võimalik kohe salvestada PDF-i või mõnda muusse tuntud formaati.

LISA 4. Magistrant Lauriga läbiviidud intervjuu

Miks kasutate R-i? Kas valiku tegemisel tutvusite ka teiste võimalustega? Kas valikut kallutas R-i kasuks ka töökoht või mõni muu väline tegur? Millist programmi kasutaksite kui R välja jätta?

Huvi pärast ja kuna tegemist on rohkem programmeerimisvahendiga kui "point and click" süsteemiga (SPSS, Statistica). Ma ei valinud R-i mingite asjade hulgast - ma olin kasutanud SPSSi ja Statisticat (ja veel üht-teist) ja otsustasin R-i ka endale selgeks teha. Kui R-i ei kasutaks, siis ilmselt SPSS-i.

Kas teie arvates võiks ülikoolides statistilise analüüsi õpetamiseks kasutatav SPSS välja vahetada R-i vastu? Miks?

Võiks muidugi, aga eks seda tuleb väga ettevaatlikult ja läbimõeldult teha. Praegu juba on statistika kui selline väga segane tudengitele ja kui selle jaoks veel programmeerima peab, siis võib hariduse kvaliteet alla minna.

"Miks" on muidugi hea küsimus - R on tasuta, R on võimsam (rohkemate võimalustega), R-is käib töövoog teistmoodi (saad seda mida küsid, mitte meeletut kogust segast väljundit millest asju välja noppida). R-s salvestatud andmefailid on ka muudes programmides kasutatavad, jne, jne, jne.

Kas olete kasutanud ka SPSS-i? Millise programmi õppimine võttis kauem aega ja kumba algajale soovitaksid? Kumb on nende võrdluses paindlikum ja võimekam?

Olen kasutanud SPSSi, kasutan praegu ka paralleelselt R-ga mõnede tegevuste jaoks. Hea raamatu abil on SPSS täitsa hästi kasutatav. St. midagi väga selgeks ei saa, aga tehtud saab oma asjad ikka. R-i saab ka raamatu abil teha (Andy Fieldil on hea raamat, mul on see olemas), aga eks see vajab alguses omajagu rohkem õppimist ma arvan. Paindlikum ja võimekam on jällegi sõltuv sellest, et mis eesmärk on. Kõige parem on mõlemat teada ja tunda ja siis vastavalt vajadusele kasutada.

Millised on R-i kasutamist alustades suurimad takistused? Mida peaks oskama ja teadma, et õpikõver oleks võimalikult lühike?

Hea küsimus. Ma ei oska sellele vastata. Mul kui pikaajalises programmeerijas (23 aastat pidevat kogemust) oli päris keeruline R-i andmestruktuure panna oma mõtetes loogilisse süsteemi. Mul oli õppimiskõver ikka päris korralik, esimene kuu aega läks iga päev paar

tundi, et õppida data frame filtreerima ja sorteerima ja teisele kujule viima. Seega ma ei oska vastata sellele küsimusele. Ilmselt oleks kõige suurem kasu lihtsalt harimisest (teadmiste jagamisest ja küsimustele vastamisest) sel teemal, et mis see R on ja kuidas andmed sinna sisse lähevad ja sealt välja tulevad ja et tegelikult ei ole tegemist maagiaga.

Tudengite test grupile valmistas suuri raskusi R-i installeerimine ja lisapakettide lisamine. Kuidas teie arvates programmi esmane seadistamine võimalikult lihtsaks teha? Kas oleks lahenduseks valmis pakett kus mingid valitud lisad on juba kaasas või peaks kasutajatele püüdma kohe selgeks teha kuidas ükskõik millise lisapaketi allalaadimine ja kasutamine käib? *RStudio Server on üks võimalikest lahendustest. Sinna on siis kõik vajalik juba installitud ja kasutajad lähevad kõik oma RStudioga lihtsalt serveri külge ja ongi olemas. RStudio seest on ka lihtne neid lisapakette juurde tõmmata ja uuendada.*

Kas edasijõudnud kasutajana käib põhiline töö konsooli aknas või on kasutusel ka graafilinelides (Deducer vms). Kas graafilise kasutajaliidesega õppimise alustamine on mõistlik või peab selle nagunii keerulisemate ülesannete tekkides välja vahetama?

No ma kasutan ainult RStudiot, kus töö käib vaheldumisi konsoolis ja koodiaknas, st. üks on command-line ja teine on fail, kust saab ka rea kaupa asju käima lasta. Ma olen näinud ka screenshote graafilisest kasutajaliidesest, aga ise kasutanud pole, seega ei oska rohkem midagi selle kohta öelda.

Kust enda kogemuse põhjal kõige kiiremini R-i kohta abi leiad? Kas lihtsalt googeldades, R-i maililistist või mõnest R-iga seotud foorumist?

Googeldades täiesti kindlasti. "how to blablabla in r" ja tuleb.

Kas R-i kogukond on nii vastutulelik ja aktiivne nagu mitmelt poolt lugeda võib?

Minu arust on küll jah. Stackoverflow foorum on päris sõbralik, alati pakutakse mitu erinevat varianti soovitud tulemust saavutada ja tihtipeale on need kenasti jagunenud erineva selgituse sügavuse poolest - lihtne lahendus, põhjalikum lahendus, süsteemne arusaamine asjast.

Kas oskaja jaoks on R-iga graafikute ja jooniste tegemine lihtne ja mugav? Või eelistad saadud andmete põhjal jooniseid mõnes muus keskkonnas koostada (Excel).

Kui vaikumisi parameetrid sobivad, siis on lihtne ja mugav, aga eks see on igal pool samamoodi. Ning kuna iga vahendi puhul tuleb keerulisemate asjade jaoks juurde õppida, siis

minaurust on R selle jaoks päris hea. ggplot2 kodukal on hulgaliselt näiteid ja googeldades leiab kõikvõimalikke lahendusi ka väga keerukate ja huvitavate jooniste ja graafikute jaoks. Ma teen üldiselt suhteliselt vähe graafikuid, aga enamasti teen ikka R-s. Sellega on muidugi nii nagu ma eespool nimetasin - kui osata rohkem vahendeid, siis on valik suurem ja saab teha seal kus on efektiivsem. Näiteks kui keegi palub abi graafiku tegemisel ja tal on SPSSi fail, siis on muidugi lihtsam SPSS-s "graphic builder" lahti võtta ja asi valmis teha selle asemel, et andmeid R-i hakata konvertima ja seal graafikut ehitama.

Millised on kogenud kasutaja jaoks R-i paketid, mis kindlasti võiksid kasutusel olla?

Hmm. See oleneb ikka väga palju sellest, milliste andmetega ja mis eesmärgil tegeletakse. Ilmselt üks vajalikemaid asju on siiski andmete kuju muutmine, seega paketid nagu reshape, dplyr, dplyr2 (või mis iganes selle uue versiooni nimi oli) ja misiganes pakett seda data.table andmetüüpi pakub. Graafikute jaoks vaieldamatult ggplot2. Mul endal on tihtipeale loaditud ka "foreign", ehk siis see pakett, mis lubab lugeda ja kirjutada kõikvõimalikke eri formaadis andmefaile. Ülejäänu on juba selline, mis sõltub konkreetsest analüüsist ja tegevusest. Näiteks puhtalt machine learningut jooksutades pole ühtegi eelnimetatud vaja.

Milliseid R-i abimaterjale kasutate ja mida soovitaksid teistelegi?

Googeldamine hästi palju muidugi. Ise jälgib ka r-bloggers.com postitusi. Statistika õppimiseks (erinevalt siis väga paljust muust mida R-ga saab teha) on Andy Fieldi raamat hea - "Discovering Statistics using R".