

TARTU ÜLIKOOL
MATEMAATIKA-INFORMAATIKATEADUSKOND
Arvutiteaduse instituut
Informaatika (2476) õppekava

Eerik Muuli

Korrelatsioonimaatriksi visualiseerimine

Bakalaureusetöö (9 EAP)

Juhendaja(d): Vambola Leping

Tartu 2015

Korrelatsioonimaatriksi visualiseerimine

Lühikokkuvõte:

Käesolevas bakalaureusetöös antakse ülevaade visualiseerimisest ja selle headest tavadest. Lõputöö kirjutamise käigus valmis ka rakendus, mis loob ning visualiseerib korrelatsioonimaatrikseid etteantud andmetest. Rakendus on kirjutatud programmeerimiskeeles Java ning visualiseerimisel on kasutatud teeki nimega JFreeChart.

Võtmesõnad:

Korrelatsioon, visualiseerimine, korrelatsioonimaatriks, Java, JFreeChart

Correlation matrix visualization

Abstract:

This Bachelor's Thesis gives an overview of visualization, its best practises. In the course of writing this thesis, an application was developed that creates and visualizes correlation matrices from given data. The application is written in Java with the help of a visualization library JFreeChart.

Keywords:

Correlation, visualization, correlation matrix, Java, JFreeChart

Sisukord

Sissejuhatus	5
1. Visualiseerimine.....	6
1.1 Mis on visualiseerimine?	6
1.2 Miks visualiseerida?	7
1.3 Kuidas visualiseerida?	9
1.4 Kuidas ei tohiks visualiseerida?	11
2. Korrelatsioonimaatriks ja selle visualiseerimise näiteid	13
2.1 Korrelatsiooni definitsioon	13
2.2 Lineaarse korrelatsiooni puudused	13
2.3 Kuidas Pearson'i korrelatsiooni arvutada?	14
2.4 Teisi korrelatsioone	15
2.5 Mis on korrelatsioonimaatriks?	16
2.6 Olemasolevad lahendused korrelatsioonimaatriksi visualiseerimiseks	16
2.6.1 Excel.....	16
2.6.2 R.....	17
2.6.3 Stata.....	18
3. Korrelatsioonide visualiseerimise näiterakendus	19
3.1 Tehnoloogiad	19
3.2 Peamised vaated	20
3.2.1 Faili avamise vaade.....	20
3.2.2 Maatriksi vaade	21
3.2.3 Diagrammi vaade	21
3.3 Rakenduse funktsionaalsus.....	22
3.4 Kasutusjuhend	22
3.4.1 Rakenduse käivitamine	22
3.4.2 Vajaminevad tehnoloogiad.....	22
3.4.3 Sisendfaili tingimused.....	23
3.4.4 Rakenduse sulgemine	23
3.4.5 Piltide salvestamine	23
3.5 Edasised arendused.....	23
Kokkuvõte	25
Kasutatud kirjandus.....	26
Lisad	28

I.	Terminid.....	28
II.	Rakendus, selle lähtekood ning testandmestik.....	29
III.	Litsents	30

Sissejuhatus

Visualiseerimiseks (ingl.k. *visualization*) nimetatakse tehnikat, mille abil edastatakse teavet optilisel kujul. Palju on selliseid valdkondi, mis tegelevad suurte andmemahtudega ning milles tuleb leida seoseid andmete vahel. Üheks selliseks iseloomustajaks on korrelatsioon (ingl.k. *correlation*), mis näitab erinevate tunnuste vahel esinevaid seoseid ning nende tugevust. Kui andmete hulgas on suured, on mugav need korrelatsioonid maatriksisse kanda nii arvuliste väärtustena kui ka silmale sõbralikumalt – diagrammide abil.

Käesoleva lõputöö eesmärgiks on luua rakendus, mis CSV formaadis sisendfailist loob visualiseeritud korrelatsioonimaatriksi. Rakendus arvutab tunnustevahelised korrelatsioonid, kannab need maatriksisse ning lisaks kujutab tunnustevahelisi seoseid ka graafiliselt diagrammidena. Rakendus võimaldab diagramme ning maatrikseid ka pildiformaadis salvestada. Rakendust saab kasutada erinevates valdkondades – oluline on vaid see, et sisendfaili andmed oleks arvulisel kujul.

Lõputöö esimene peatükk viib kurssi visualiseerimisega. Selles antakse ülevaade visualiseerimise tähtsamatest aspektidest ning tutvustatakse visualiseerimise häid tavasid ning harjumusi. Antakse visualiseerimise definitsioon ning põhjused, miks visualiseerimine vajalik on. Tuuakse välja efektiivsed visualiseerimise meetodid ning vastupidiselt – kuidas ei tohiks andmete visualiseerimist teostada.

Teises peatükis tehakse täpsemalt selgeks, et mis on korrelatsioon ning korrelatsioonimaatriks. Tutvustatakse erinevaid korrelatsioonitüüpe ning nende arvutamisevõimalusi. Lisaks tuuakse välja erinevad olemasolevad lahendused korrelatsioonimaatriksitest ning sellest, kuidas need on visualiseeritud.

Kolmas peatükk annab ülevaate selles lõputöös valminud rakendusest, selle valmimisest ning kasutatud tehnoloogiast. Kaetakse rakenduse funktsionaalsus ja lisatud on ka kasutusjuhend. Arutletakse ka rakenduse edasiste arenduste üle.

Lõputöö käigus valminud rakendus on loodud autori poolt programmeerimiskeeles Java. Visualiseerimiseks on kasutatud lisaks Java rohketele sisseehitatud teekidele teeki nimega JFreeChart, mis on valmistatud just diagrammide loomiseks.

1. Visualiseerimine

Õeldakse, et andmed on uue maailma valuuta ning läbi interneti toimub selle valuuta vahetus. Tarbimisühiskonnas elatakse päevast päeva selles andmetest tulvil maailmas. See on kõikjal, alustades toidu pakenditest, lõpetades igapäevaste uudistega, millest enamikuga pole tarbijatel mingit pistmist. Disaineritel tuleb pidevalt võidelda selle eest, et just nende disainitud tooted oleksid kõige pilkukõitvamad, sümpaatsemad ning arusaadavamad [1].

1.1 Mis on visualiseerimine?

Informatsiooni visualiseerimine selle arusaadavamaks edastamiseks ei ole hiljuti tekkinud valdkond. Esimeseks visualiseerimise vahendiks peetakse koopamaale, mis on tänaseks juba mitukümmend tuhat aastat vanad. Väga ammusteks vahenditeks saab lugeda veel hieroglüüfe, geomeetriat ning tehnilisi joonestamise meetodeid. Lisaks on juba tuhandete aastate jooksul kasutatud visualiseerimist jooniste ja kaartide tegemiseks [2].

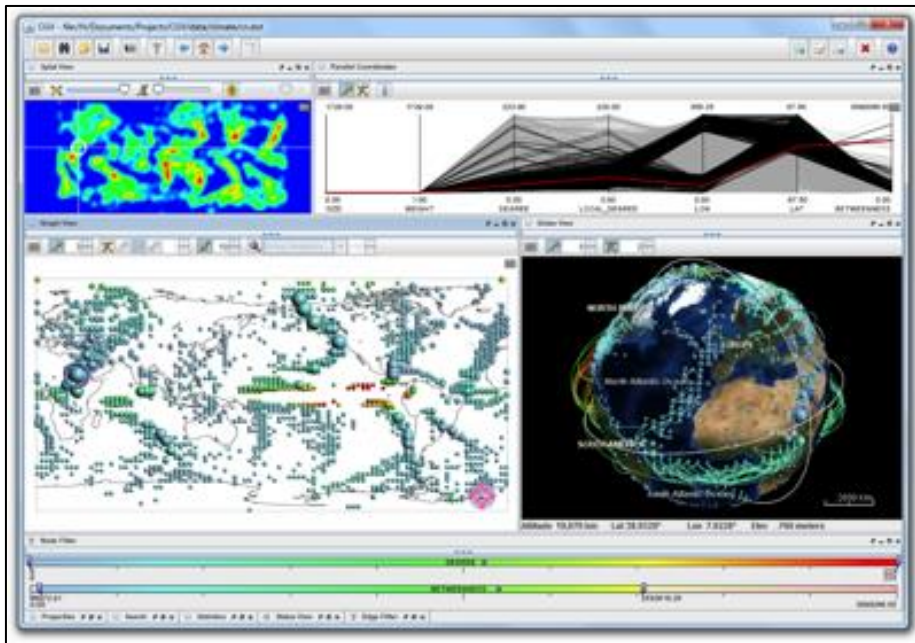
Üldisemas mõistes on visualiseerimine meetod, mille tulemusena on võimalik informatsiooni esitada näiteks piltide, diagrammide, graafikute, skeemidena või muul sarnasel moel. Tänu visuaalsele kujutamisele saab kasutaja mugavalt omastada infot, mis on näiteks arvuline, geograafiline või lihtsalt raskesti mõistetav. Graafiline esitus aitab probleemi või kitsaskohta paremini esile tuua, lihtsustada ning arusaadavamaks muuta [3][2].

Autor, ajakirjanik ja disainer David McCandless ütles enda TED konverentsil järgnevat: „Andmeid visualiseerides muudame need maastikuks, justkui informatsiooni kaardiks, mida saame uurida. Ning kui olla uppunud massilisse informatsiooni, võib informatsiooni kaardist abi olla.“ [1] Seejuures pole oluline, missuguse informatsiooniga on tegemist – visuaalne pool on igasuguse informatsiooni esitamisel ääretult oluline [3]. Hea visualiseerimine loob andmetest süsteemi ning aitab neid esitada vaatlejatele arusaadavalt, anda neile tähendusi ning näha seoseid [4].

Visualiseerimise peamised rakenduskohad on tänapäeval teaduses, meditsiinis, hariduses, projekteerimises ning interaktiivses multimeedias. Üheks tüüpilisemaks valdkonnaks on arvutigraafika. Just kiiret arvutigraafika arengut võib pidada visualiseerimise edu murdepunktiks. Ka animatsiooni areng soodustas suuresti visualiseerimise arengut [2].

Tänu visualiseerimisele sai võimalikuks teistsugune andmeid analüüsiv protsess, mida nimetatakse visuaalseks andmeanalüüsiks. See võimaldab uurijal saada süsteemi kohta

graafilist ülevaadet, mis võib osutada äärmiselt edasiviivaks. Sinna alla kuuluvad näiteks probleemid, mis on suuruse, keerukuse ja inim- ning masinanalüüsi koostöö puudumise tõttu võimatud [5]. Mõnedeks sellisteks valdkondadeks on näiteks füüsika, astronoomia, kliima uuringud ning meditsiin.



Joonis 1. Erinevate kliimamudelite simulatsioonide visuaalne toetus. Loodud rakendusega CGV (Coordinated Graph Visualization) [6].

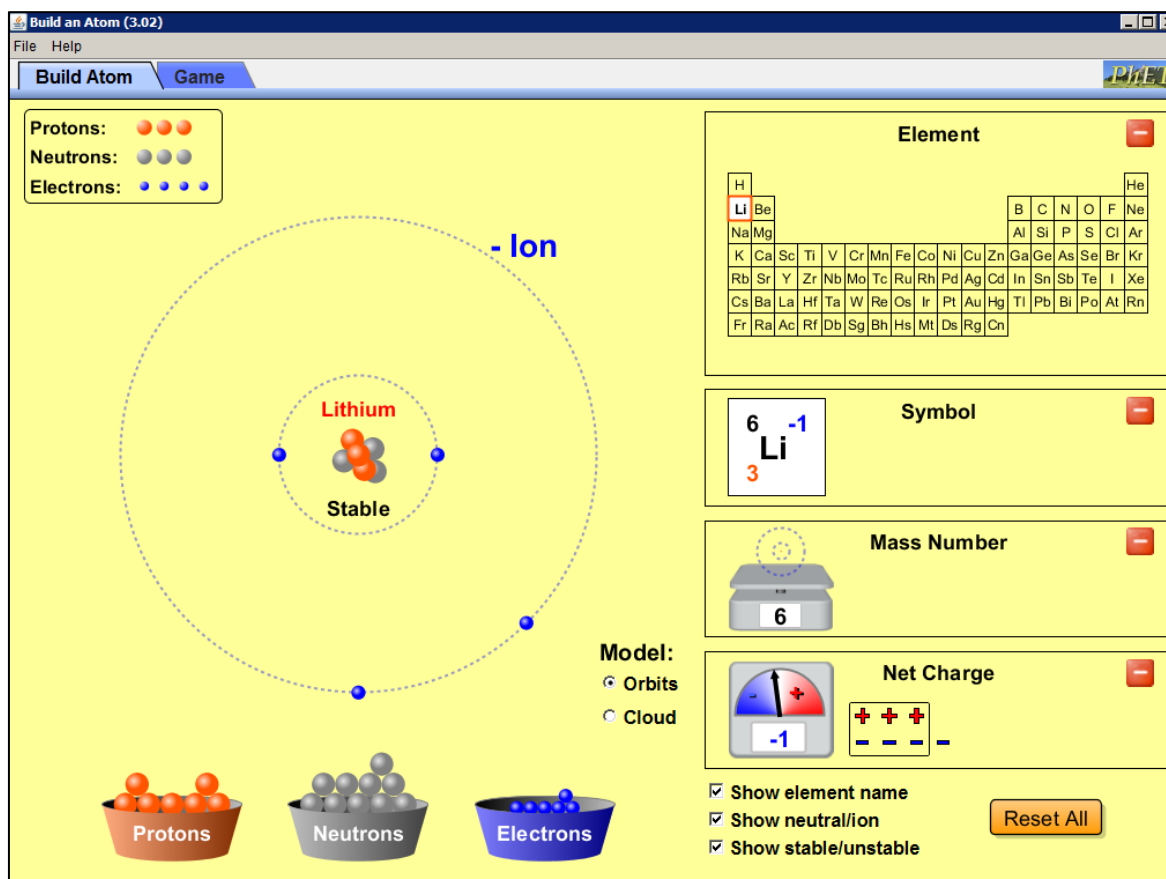
Veel üheks väga oluliseks visualiseerimise valdkonnaks võib pidada teaduslikku visualiseerimist. See on simulatsioonidest ning eksperimentidest kogutud andmete teisendus ning esitamine koos nende geomeetrilise ehitusega. Seda võib pidada nõnda oluliseks, kuna nähtuste ning katsete visualiseerimist on praktiseeritud teaduse arengu algusaegadest [2].

1.2 Miks visualiseerida?

Üks parimaid viise enda sõnumit, teadet või mistahes informatsiooni selgelt edastada, on seda visualiseerida, kuna tihti haarab see inimeste tähelepanu, ilma et nad seda ise teadvustaks. Kõige põhilisem viga selle juures on aga see, et ei teata graafilise esitluse parimaid tavasid. Eriti suureks probleemiks on see visualiseerimise valdkonnas kogenematute inimeste puhul [7].

Just pilkupüüdvus on tänapäeva kiires maailmas üks tähtsamaid küsimusi. Teades, et kõigi eluvaldkondade andmemahud kasvavad päev-päevalt, on väga kasulik osata välja tuua kõige olulisem. Parim võimalus tähtsat informatsiooni esile tuua on kogu tähelepanu sellele juhtida [1].

Laialdaselt kasutatakse visualiseerimist hariduses. Haridusliku visualiseerimise all mõeldakse mingi nähtuse graafilist kujutamist õpetuslikul eesmärgil selliselt, et kujutatud esitus toetab teoreetilist materjali. Loodud graafiline kujutis või mudel aitab eriti hästi mõista nähtusi, mida on ilma abivahendita keeruline või võimatu näha, näiteks aatomi struktuuri kujutamise mudelid keemias ja füüsikas.



Joonis 2. Rakendus Build an Atom.

Joonisel 2 kujutatud rakendusega Build an Atom on võimalik luua erinevaid perioodilisussüsteemi elemente. Rakendus aitab õpilastel saada paremat ettekujutust erinevate elementide aatomitest ning nende eripäradest. Veel aitab visualiseerimine reprodutseerida minevikus aset leidnud sündmusi – tänu visualiseerimisele osatakse ette kujutada muuhulgas ka dinosauruste ajastut [2].

Tänu ülivõimsate arvutite suurenenud kättesaadavusele on võimalik uurida väga keerulisi andmeid. Enamasti kaasneb keeruliste andmeanalüüsidega märkimisväärne kogus andmeid. Selleks, et suuri ning keerulisi andmekogumeid uurida, kasutatakse spetsiaalset tarkvara, mis suudab neid hiiglaslikke ning komplitseeritud kogumeid ilma suuremate probleemideta analüüsida ning visualiseerida. Tänu visualiseerimisele on võimalik võita

aega ning suurel hulgal ressursse, kuna mehaaniline andmete töötlus ning algoritmide rakendamine ei pruugi soovitud tulemust anda ning tihti on tarvis protsessi korduvalt teostada, muutes algoritme ning parameetreid. Seejuures visuaalsele kujule viidud andmete põhjal on võimalik lühikese ajaga tekitada seoseid ning teha järeldusi, mis algoritmiliselt või katsetades ei pruugi kunagi välja tulla [1][8].

Informatsioonist arusaamine või selgus tähendab, et informatsiooni on võimalik uurida, selgitada ning selle põhjal otsuseid vastu võtta. Mõned valdkonnad, kus visualiseerimine ning täpsus on üliolulised: finantsandmete analüüs, turu-uuringud ning tööstuse toodangu kontrollimine. Välja saab tuua nende valdkondade juures väga oluliseks osutuvad visualiseerimise eelised.

- Suurt hulka andmeid on võimalik säilitada ning neile on kerge ligi pääseda (nt kaardid).
- Andmete otsing on lihtsustatud.
- Visualiseeritud informatsiooni tuvastamine, omandamine ning selles orienteerumine on tavakasutaja jaoks lihtsamad kui samaväärse visualiseerimata informatsiooni puhul.
- Probleemid ning weakohad tulevad visualiseeritud andmete puhul selgemini esile [2].

Visualiseerimise edu ei sõltu mitte ainult tulemusest, vaid ka keskkonnast, kus visualiseerimine läbi viiakse. Visualiseerimiseks soodne keskkond sõltub järgnevatest ressurdidest: kettaruum, printerite kvaliteet, visualiseerimise tarkvara ning riistvara üldiselt. Paljud probleemid tulenevad kasutatavate masinate väiksest võimsusest või tarkvaralistest piirangutest, näiteks kasutajaliides, sisend- ning väljundandmed, mida tarkvara aktsepteerib, andmete töötlemise võimalused ning palju muud [9].

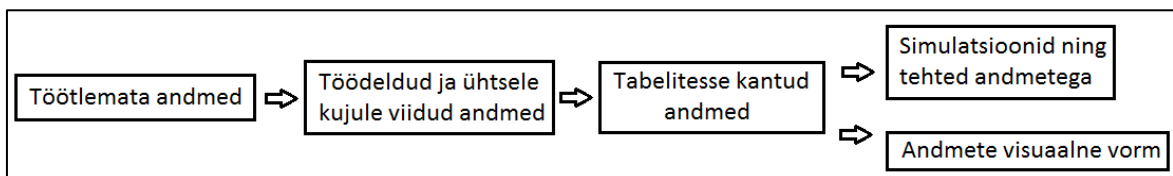
1.3 Kuidas visualiseerida?

Kõige olulisemaks visualiseerimise juures peetakse selle läbipaistvust ehk millist informatsiooni reaalselt vastavalt jooniselt lugeda saab ning kui kaua läheb aega joonise mõistmiseks ehk informatsiooni omandamiseks [10]. Andmekogumeid võib visualiseerida erinevatel meetoditel. Meetod omakorda mõjutab tehtud analüüsi ning võib kas soodustada või pärssida teatavat tõlgendust. Seepärast tuleb graafikuid, tabelleid ja ka muid visualiseeritud materjale luua kriitilise pilguga [11].

Teaduses ning erinevates projekteerimistöodes osutub visualiseerimine suureks osaks informatsiooni töötlemisest, mõistmaks detailsemalt ning selgemalt objektide ruumilist ehitust. Mõned enamlevinud visualiseerimise tehnoloogiad on järgnevad:

- a) isotermid,
- b) isobaarid,
- c) renderdamine,
- d) tabelid,
- e) maatriksid,
- f) diagrammid,
- g) graafikud,
- h) kaardid,
- i) animeeritud ilmaennustus,
- j) nomogramm,
- k) röntgenpildid [2].

Graafikute ning tabelite kasutamine on mugavam viis suuri andmemahte üheselt mõistetavalt ning kompaktselt esitada. Visualiseerimisel on vaja hoolikalt läbi mõelda, milliseid andmeid kujutada ning mis on visualiseerimise eesmärk. Sõltuvalt eesmärgist tuleks valida graafiku/joonise tüüp. Kindlasti tuleks mõelda telgedele, nende ühikutele ning vahemike peale. Telgede vahemikud valitakse sõltuvalt andmete suurimast ning väikseimast väärtusest. Tähele tuleks panna, et alati pole kõige otstarbekam alustada telgesid nullist. Seejuures peab olema kindel, et mingilt teiselt väärtuselt alustades joonis ei moonduks. Samuti on üheks oluliseks elemendiks värvid. Kohad, mida soovitakse diagrammil rõhutada, tuleks värvida erksamate värvidega ning omadused, mis peaksid jääma tagaplaanile, tuleks märkida tuhmimate toonidega [11].



Joonis 3. Visualiseerimise protsess [12] (Autor modifitseeris ja tõlkis eesti keelde).

Joonisel 3 on kujutatud protsessi, mille käigus algandmetest saab visuaalne vorm. Andmeid on väga erinevat tüüpi: kassaautomaatidest päeva jooksul läbi löödud kaupadest kuni meditsiiniliste haiguslugudeni. Enamasti on andmed alguses läbiseegi ning kasutamatud.

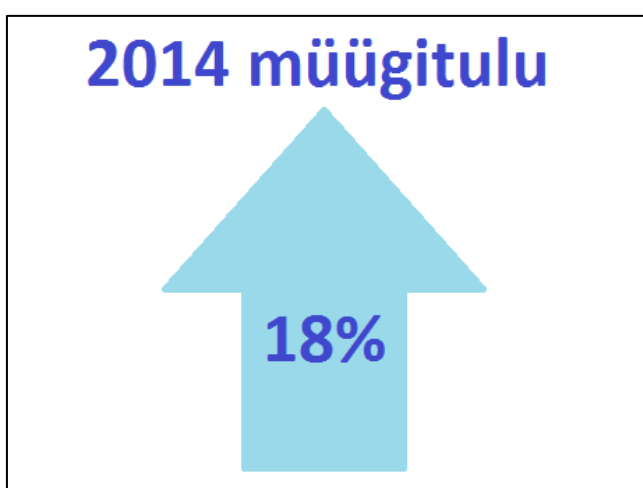
Selleks on vajalik andmed kõigepealt korrastada: eemaldada vead ning viia ühtlustatud kujule. Seejärel andmed sorteeritakse ning soovituslikult kantakse tabelisse. Pärast seda saab andmetega teostada erinevaid protseduure: arvutusi, simulatsioone, grupeerimisi ning sealhulgas ka visualiseerimist. Kusjuures visualiseerimine võib toimuda viimase etapina [12].

Andmeid saab visualiseerida ka tavaliste piltidega – näiteks ilmaennustustes märgitakse päikesepaistelist ilma päikese kujutisega ning vihmast ilma vihmapiisaga. Teise näitena võib välja tuua liiklusmärgid. Auklikku teed kujutav märk teepervel koos arvilise väärtusega annab liiklejale järgnevat informatsiooni: märgitud distantisi jooksul alates praegusest asukohast võib teekattes esineda auke või muid elemente, mis võivad autot lõhkuda või tekitada liiklusohhtliku olukorra. Sellised näited võivad tunduda iseenesestmõistetavad, kuna tarbija on taoliste metafooridega juba niivõrd harjunud [13].

1.4 Kuidas ei tohiks visualiseerida?

Visualiseerida ei tohi nii, et andmed kaotavad tähenduse. Eeltoodud olukord on võimalik, kui graafiku telgedel on märkimata pealkirjad või skaala. Sellisel juhul võib graafiline pool olla laitmatu, kuid selle mõistmine on probleemne. Tagajärjeks ongi see, et visualiseerimine on kaotanud informatsiooni tähenduse [14].

Samuti on tarvis visualiseerimisel vältida liigset lihtsustamist. Andmete mõistmiseks on nende lihtsustamine vajalik, kuid viies andmed liiga lihtsustatud kujule, võib kaduda nende mõte.

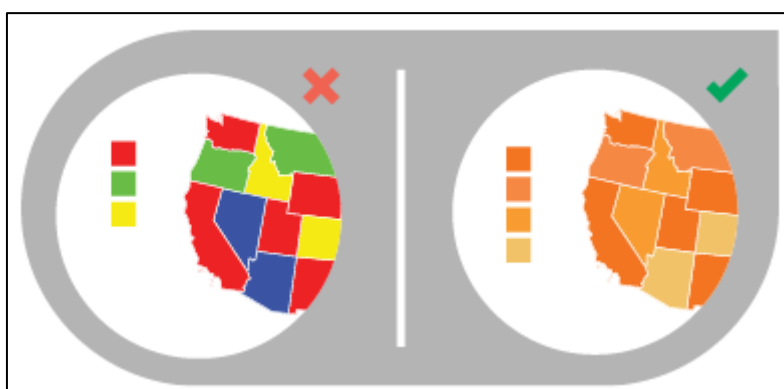


Joonis 4. Liigselt lihtsustatud visualisatsioon.

Jooniselt 4 on näha, et 2014. aasta müügitulu oli 18%. Tõenäoliselt on sihtgrupp, kellele andmeid presenteeritakse vastava ettevõttega kursis, kuid sellest hoolimata ei ole joonis üheselt mõistetav. Millega või kellega võrreldes kasvas müügitulu? Mis ajavahemikuga võrreldes kasvas müügitulu [15]?

Vastupidiselt liigsele lihtsustamisele takistab mõistmist ka liiga detailne visualiseerimine. Väga mugav on ühe joonisega edasi anda hulgaliselt olulist informatsiooni ning tahestatmata võib tekkida kiusatus kõike ühel graafikul edastada. Tihti muudab see aga kogu visualiseerimise üleliigseks, kuna nõnda detailse graafiku uurimine on samaväärne andmestiku enda uurimisega. Sellistel juhtudel kaob piir olulise ning ebaolulise vahel [15].

Viimasena võib välja tuua vale graafiku tüübi ning värviskeemi kasutamise. Vale graafiku kasutamine tuleb tihti ette andmete võrdlemise puhul. On oluline meeles pidada, et andmeid on kõige lihtsam võrrelda, kui need on üksteise kõrval. Heaks näiteks on siinkohal tulpdiagrammid. Halva tavana võib välja tuua sektordiagrammi. Ebasobiva värviskeemi näidet ilmestab soojuskaart liiga erinevate värvidega.



Joonis 5. Hea ja halb soojuskaart [16].

Jooniselt 5 on näha, et erinevate värvide kasutamisel on joonise lugemine keerulisem ning kaob võrdlusaste. Joonis sama värvi erinevate toonidega on paremini loetav [16].

2. Korrelatsioonimaatriks ja selle visualiseerimise näiteid

2.1 Korrelatsiooni definitsioon

Korrelatsiooni kasutame selleks, et saada teada, kui tugev on erinevate tunnuste vaheline seos ning missugune on seose iseloom (positiivne või negatiivne). Antud Lõputöös on kasutusel Pearson'i ehk lineaarne korrelatsioon (edaspidi „korrelatsioon“), kuna see on enimlevinud ning -kasutatud. Korrelatsiooni mõõdetakse korrelatsioonikordajaga (tähistus r). Korrelatsioon saab olla vahemikus $-1 \dots +1$. Kui korrelatsioon on suurem nullist, siis on tunnused kasvavalt seotud. Kui korrelatsioon on nullist väiksem, siis on tunnused kahanevalt seotud. Kui korrelatsioon on 0, siis mingit seost tunnuste vahel ei esine, kuid vastava korrelatsiooni graafikut uurides on võimalik sealt leida seoseid, mis pole lineaarsed ning ei oma arvulist väärtust. Kui korrelatsioon on $+1$, siis on tegemist perfektse positiivse korrelatsiooniga, mis tähendab seda, et kui üks väärtus muutub, siis teine muutub proportsionaalselt sama palju samas suunas. Kui korrelatsioon on -1 , siis on tegemist perfektse negatiivse korrelatsiooniga, mille korral ühe muutuja kasvamisel teine kahaneb proportsionaalselt. Vahemikus $|r| = 0 \dots 0,3$ on tegemist nõrga seosega, $0,3 < |r| < 0,7$ keskmise seosega ning vahemikus $|r| > 0,7$ tugeva seosega.

2.2 Lineaarse korrelatsiooni puudused

Lineaarne korrelatsioonikordaja on kergesti mõjutatav tunnustes esinevate erandite poolt. Kui andmed, mille vahel korrelatsiooni leitakse, sisaldavad kas või paari väga suurt erijuhetu, siis võivad need korrelatsiooni rikkuda. Sellised erandid tuleks andmetest eelnevalt eemaldada parima tulemuse saavutamiseks.

Teiseks probleemiks on erinevat tüüpi seosed. Lineaarne korrelatsioonikordaja kirjeldab vaid lineaarseid seoseid, igasuguste muude seoste korral on võimalus, et tulemus on null või nullilähedane. Taoliste olukordade puhul on soovitatav tunnuste vahelise seose leidmiseks uurida graafikut.

2.3 Kuidas Pearson'i korrelatsiooni arvutada?

Korrelatsiooni kahe tunnuse vahel arvutatakse järgmise valemiga:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Joonis 6. Korrelatsiooni arvutamise valem.

Kasutame valemi selgitamiseks järgmisi näidisandmeid:

	Tunnus X	Tunnus Y	X * Y	X * X	Y * Y
n	15	101	1515	225	10201
	42	96	4032	1764	9216
	35	123	4305	1225	15129
	83	165	13695	6889	27225
	53	142	7526	2809	20164
	42	145	6090	1764	21025
	64	117	7488	4096	13689
	25	128	3200	625	16384
	95	131	12445	9025	17161
	42	110	4620	1764	12100
	19	105	1995	361	11025
Summa	515	1363	66911	30547	173319
Summa ruudus	265225	1857769			

$(\sum x_i)^2$ $(\sum y_i)^2$ $\sum x_i y_i$ $\sum x_i^2$ $\sum y_i^2$

Joonis 7. Näidisandmed korrelatsiooni arvutamise jaoks.

Murru lugeja: Kogu tunnuste hulka tähistatakse tähega n. Joonisel 2 on tunnuste hulgak 11. $\sum x_i y_i$ on kõigi tunnuste x ja y omavaheliste korrutiste summa: $15*101 + 42*96 + \dots + 19 * 105 = 66911$. $\sum x_i \sum y_i$ on kõigi tunnuse x liikmete summa korrutis tunnuse y kõigi liikmete summaga: $(15+42+\dots+19)*(101+96+\dots+105) = 701945$. Ehk murru lugeja väärtuseks on $11*66911 - 701945 = 34076$.

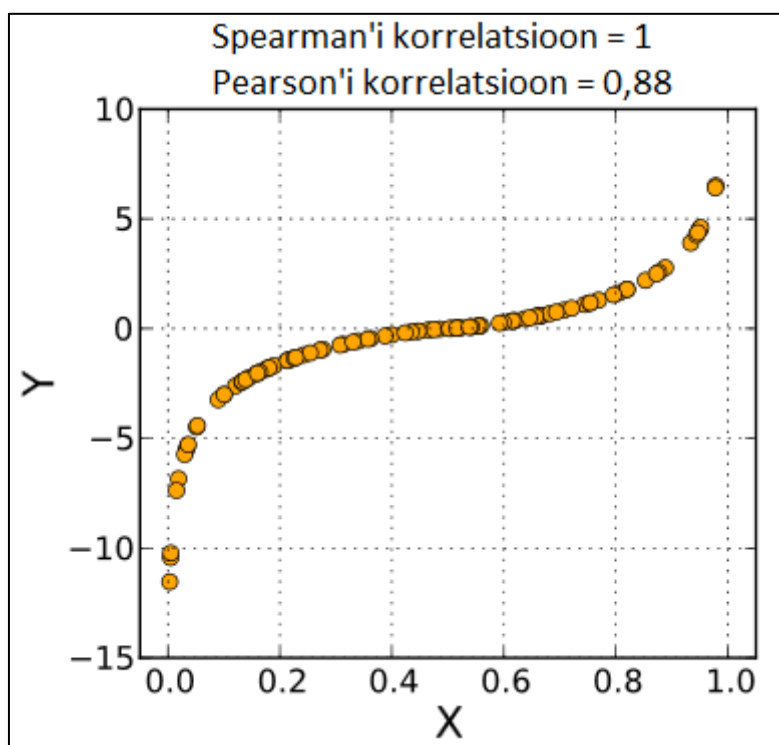
Murru nimetaja: $\sum x_i^2$ tähistab kõigi tunnuse x ruutude summat: $15*15 + 42*42 + \dots + 19*19 = 30547$. $(\sum x_i)^2$ on kõigi tunnuse x elementide summa ruudus: $(15 + 42 + \dots + 19)^2 = 265225$. Arvutame esimese ruutjuure all oleva tehte vastuse:

$\sqrt{11 * 30547 - 265225} \approx 266,0677$. Nüüd arvutame samamoodi teise juure all oleva vastuse: $\sqrt{11 * 173319 - 1857769} \approx 220,7714$. Nüüd korrutame mõlema juure all saadud vastused kokku: $266,0677 * 220,7714 \approx 58740,1386$.

Lõppvastuse saame murru lugejast jagades murru nimetaja: $34076 / 58740,1386 \approx 0,5801$.

2.4 Teisi korrelatsioone

Lisaks Pearson'i korrelatsioonile on veel teisigi korrelatsiooni tüüpe. Kuna Pearson'i korrelatsioon arvestab vaid lineaarset seost, siis ei sobi see igasse olukorda. Olukorras, kus ühe tunnuse suurenemisel suureneb või väheneb ka teine tunnus, kuid mitte proportsionaalselt, annab Pearson'i korrelatsioon tulemuseks positiivse arvu, kuid mitte 1. Sellistel juhtudel on otstarbekas kasutada Spearman'i ning Kendall'i tau astmelist korrelatsiooni.



Joonis 8. Pearson'i ning Spearman'i korrelatsiooni võrdlus [17].

Jooniselt 8 võib näha, et tegemist on kasvava graafikuga. Pearson'i korrelatsioon ei andnud tulemuseks üht, kuna kasvamine ei toimunud lineaarselt, kuid Spearman'i ning Kendall'i korrelatsioonid annavad tulemuseks ühe, kuna nende jaoks ei ole kasvamise samm oluline. Kendall'i ning Spearman'i korrelatsioonid on enamasti sama või ligilähedase tulemusega [18].

2.5 Mis on korrelatsioonimaatriks?

Kui tahta leida rohkem kui kahe tunnuse vahelisi seoseid, siis tuleb appi korrelatsioonimaatriks. Korrelatsioonimaatriksis kuvatakse kõikvõimalike paaride vahelisi korrelatsioone (joonis 9).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	A	B	C	D	E								
2	0	2,3	11	1	160				A	B	C	D	E
3	2	4,5	15	2	112			A	1				
4	4	6,3	19	3	192			B	0,999748	1			
5	6	8,5	5	4	114			C	0,214413	0,207598	1		
6	8	10,3	2	5	145			D	1	0,999748	0,214413	1	
7	10	12,1	2	6	146			E	0,118927	0,118627	0,051304	0,118927	1
8	12	14,3	18	7	172								
9	14	16	23	8	101								
10	16	18,3	7	9	175								
11	18	20	21	10	164								

Joonis 9. Korrelatsioonimaatriks Excelis.

Jooniselt 9 on näha, kuidas veergudes A-E olevatest andmetest arvutab Excel nendevahelised korrelatsioonid ning loob nende põhjal maatriksi. Selle maatriksiga on andmeid hea edasi töödelda ning veel parema visuaalse ettekujutuse saamiseks on võimalik neid lahtreid automaatselt vastavalt korrelatsiooni arvulisele väärtusele värvida. Korrelatsioonimaatriksi ülemine pool on tühi, kuna see oleks alumisest poolest peegelpilt.

2.6 Olemasolevad lahendused korrelatsioonimaatriksi visualiseerimiseks

On olemas mitmeid rakendusi, mis suudavad arvutada ning joonistada korrelatsioonimaatrikseid. Aga igäüks neist on erinev ning kõigil neil on omad eelised ning puudused sõltuvalt sellest, milliseks otstarbeks vastav funktsionaalsus loodud on. Järgnevalt ongi mõned neist välja toodud.

2.6.1 Excel

Excel on üks suurimaid ning enimlevinud andmetötluse programme. Selles on mugav nii arvulisi kui tekstilisi andmeid hoida ning analüüsida. Excelil on eriline õppimis- ning äratundmisfunktsionaalsus, mis automaatselt täidab kasutaja eest lahtreid, ilma et kasutaja ise peaks sisestamata valemeid või makrosid. Excel annab väga palju alternatiive oma andmete graafiliseks kujutamiseks. Näidet korrelatsioonimaatriksist Excelis näeb jooniselt 9 [19].

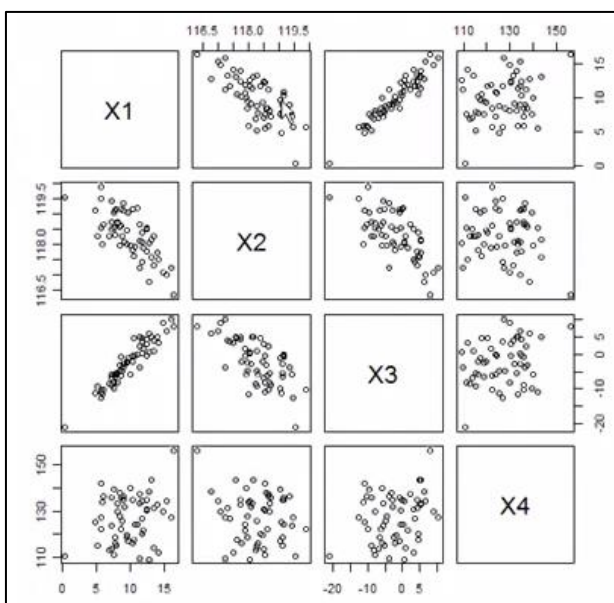
2.6.2 R

Programm R on üheaegselt nii programmeerimiskeel kui ka keskkond matemaatilise statistika arvutuste tegemiseks ning graafiliseks kujutamiseks. R pakub laialdaselt graafilisi ning statistilisi võimalusi [20].



Joonis 10. Korrelatsioonimaatriks programmis R [21].

Nagu võib näha jooniselt 10, siis on selline maatriks üsna sarnane Exceli omale, kuid sellele joonisele graafilisi elemente pärast maatriksi joonistamist lisada ei saa. Sellise maatriksi puhul tuleb eelnevalt määrata värvide piirkondade valik.

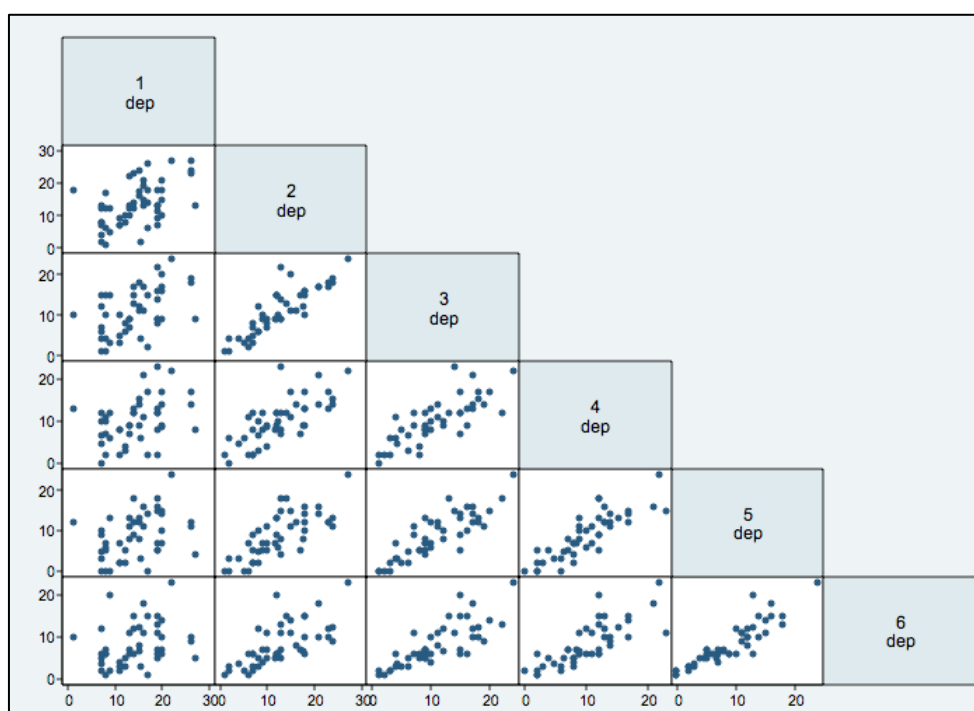


Joonis 11. Visualiseeritud korrelatsioonimaatriks programmis R.

Joonisel 11 olev korrelatsioonimaatriks ei kujuta endas korrelatsioonide arvulisi väärtusi, vaid näitab graafiliselt omadustevahelisi seoseid. Sellise maatriksi joonistamine Excelis on tülikas ning ajamahukas.

2.6.3 Stata

Stata on statistiline tarkvara komplekt, mis pakub kõike, mida on vaja andmete analüüsiks, töötlemiseks ning graafiliseks esitluseks. Stata võimaldab kasutada nii interaktiivset liidest kui ka kasutajale intuiitivset käsurit. Programmis on nii lihtsaid kui ka keerulisi funktsiooni ning tehnikaid sõltuvalt kasutaja vajadustest ning oskustest. Lisaks pakub Stata andmete haldamise funktsionaalsust [22].



Joonis 12. Visualiseeritud korrelatsioonimaatriks programmis Stata.

Kõigi nende erinevate rakenduste juures hakkasid silma erinevad aspektid – mõnel ei saanud mugaval kujul andmeid sisse lugeda, mõnel jällegi polnud graafiline esitus piisavalt ilus. Just need põhjused eristavadki lõputöös valminud rakendust eelnevatest: näiterakendusega saab hõlpsalt andmed failist sisse lugeda ning nii arvuline korrelatsioon kui ka graafiline tunnuste võrdlus on esitatud ühes maatriksis.

3. Korrelatsioonide visualiseerimise näiterakendus

Rakenduse idee ning valmimise põhjus on tingitud sellest, et lõputöö kirjutamise hetkel ei leidnud autor ühtegi rakendust, mis täidaks soovitud funktsionaalsust: üks pool maatriksist on korrelatsioonide arvuliste väärtustega ning teine pool annab graafilise ettekujutuse seostest. Niisiis valminud rakenduses on maatriksilt võimalik lugeda tunnuste vahelisi korrelatsioone nii arvuliselt kui ka graafilise kujutusena.

3.1 Tehnoloogiad

Rakendus on koostatud objekt-orienteeritud programmeerimiskeeles Java, mis tähendab seda, et kogu programmi sisu koosneb objektidest ning kõik klassid pärinevad juurklassist „*Object*“. Esimene Java versioon anti välja aastal 1995 Sun Microsystems'i poolt. 2015. aasta aprillikuu seisuga on Java maailma enimlevinud programmeerimiskeel [23]. Java on platvormist sõltumatu, mis tähendab seda, et lähtekoodist kompileeritud baitkood töötab kõigil platvormidel, mis toetavad Javat. Sellel keelel on väga rikkalik baasteekide kogumik, näiteks sisend- ja väljund-, võrgutegevuse, kasutajaliidese, lõimede, erindite ning paljude muude funktsionaalsuste teegid. Lisaks sellele, et Javas on võimalik teha eraldi-seisvaid rakendusi, on võimalik teha programme, mis töötavad veebibrauseris. Tänu sellele, et Java süntaks sarnaneb populaarse C++ keele omale, on seda keelt kerge omandada. Javas programmeerides ei pea mõtlema mälu kasutusele ja vabastamisele, mis teeb programmeerimise lihtsamaks ning programmid pole mäluprobleemide suhtes nii haavatavad [24][25].

Rakenduse loomiseks otsustas autor kasutada programmeerimiskeelt Java, kuna sellega on enim kokkupuudet olnud. Lisaks soodustas valikut fakt, et Java on platvormist sõltumatu ning et Java abil on võimalik luua tänu suurele teegipagasile mitmekülgeid rakendusi. Märkimisväärne on asjaolu, et järgnevad suured organisatsioonid on kasutanud enda projektides peamiselt Javat: Google, NATO, Blu-Ray, erinevad pangad ning paljud teised. Näiteks Google'i projektid Gmail ning Android on programmeeritud suures osas Javas [26][27].

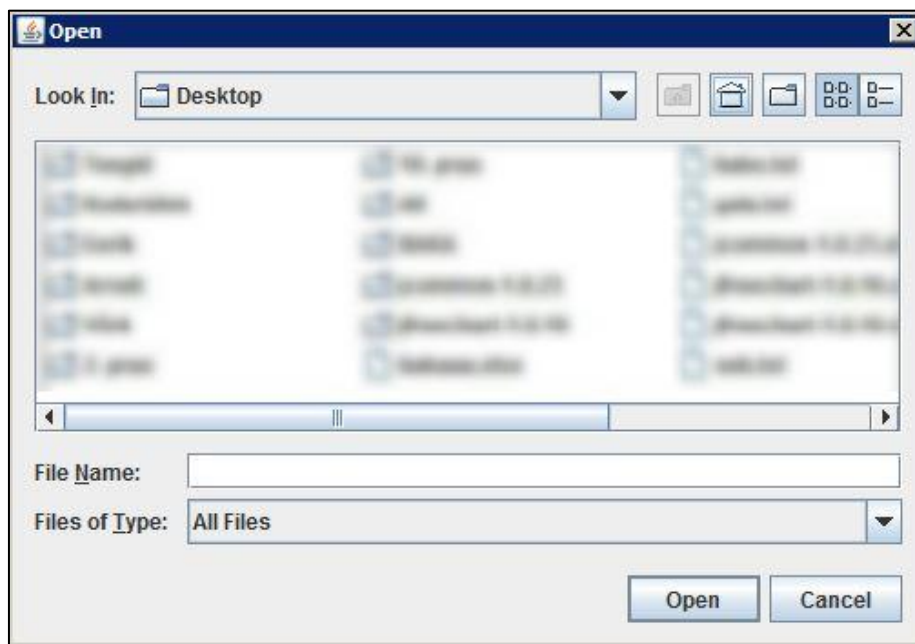
Visuaalne pool rakendusest on programmeeritud kasutades teek Swing ning JFreeChart. Sarnaselt Java keelele on Swing platvormist sõltumatu ning kaasab endaga nuppe, paneele, tabeleid, tekstialasid ja palju muud. Tänu Swingile on võimalik kogu rakendus vastavalt soovile ning vajadusele kujundada [28][29]. JFreeChart on tasuta teek, mis on loodud

programmeerimiskeelele Java. JFreeChart on mõeldud spetsiaalselt nii interaktiivsete kui ka mitteinteraktiivsete diagrammide loomiseks ning just seetõttu sobib see ka näiterakendusse. See teek loodi aastal 2000 ning on hetkel maailmas enimkasutatud diagrammide loomise teek programmeerimiskeelele Java. JFreeChart toetab Swingi ning pilditüüpi PNG, mida mõlemat ka lõputöös on kasutatud [30].

3.2 Peamised vaated

3.2.1 Faili avamise vaade

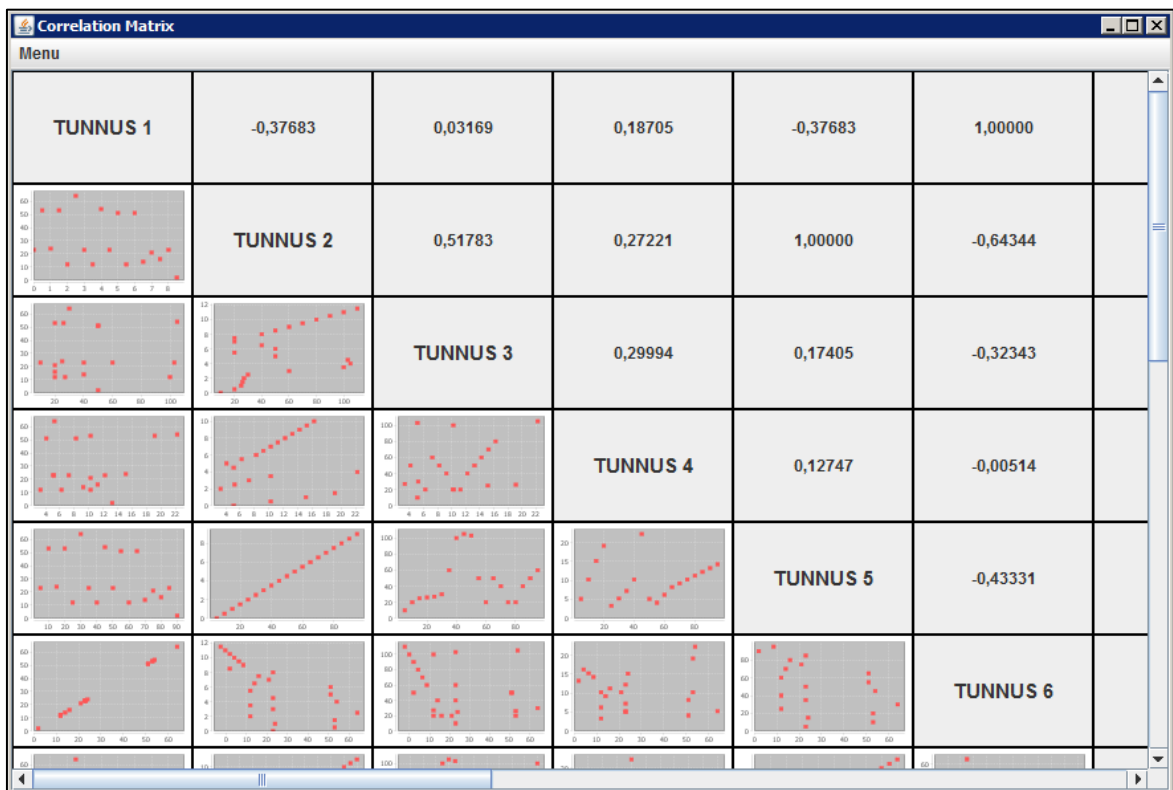
Selles vaates toimub rakendusele sisendfaili andmine. Kasutaja saab lihtsasti navigeerida arvutis ning valida sobiva faili. Sellest failist loobki rakendus korrelatsioonimaatriksi. Vajutades „Cancel“ või „X“ nupule sulgub faili avamise vaade ning sisse loetakse näidisfail „Test.csv“.



Joonis 13 „Faili avamise vaade“

3.2.2 Maatriksi vaade

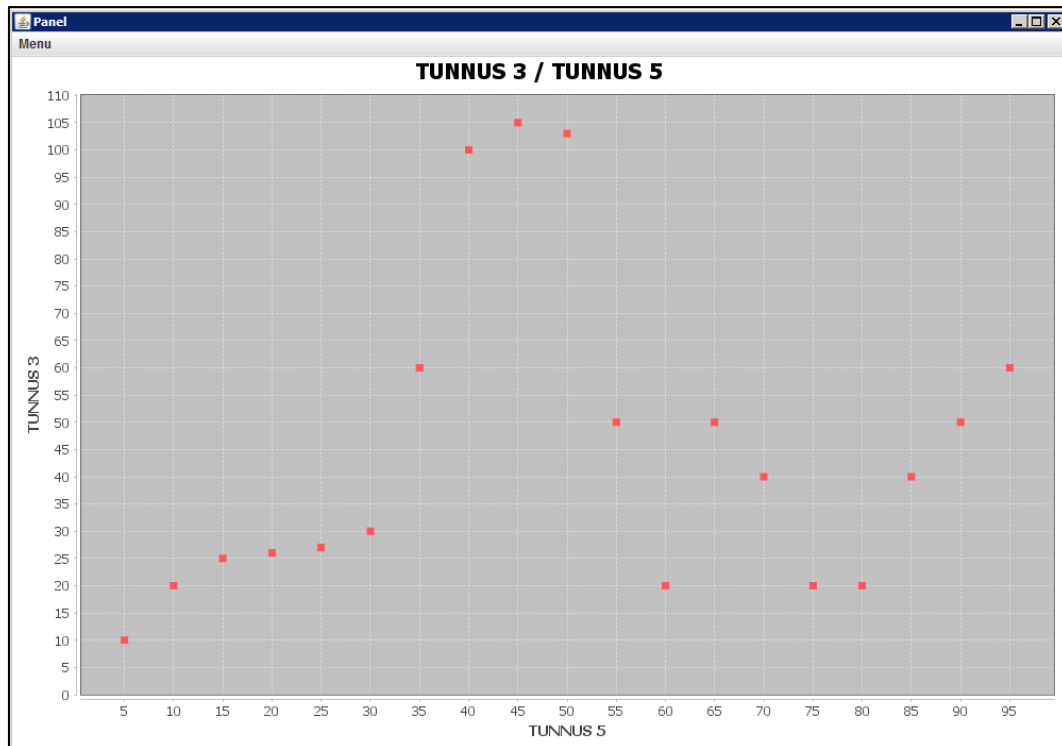
Selles vaates on näha sisendfaili põhjal koostatud korrelatsioonimaatriksit. Peadiagonaalil on näha veergude nimed. Peadiagonaalist üleval pool on tunnustevahelised korrelatsioonid arvuliste väärtustena. Peadiagonaalist all on samad seosed graafikuna. Kui maatriks on suurem, kui ekraanile mahub, on võimalik seda nähtavale tuua selleks ettenähtud kerimisribadega, mis asuvad rakenduse all ning paremal osas. Vajutades diagrammile, avaneb diagrammi vaade (Joonis 15).



Joonis 14 „Maatriksi vaade“

3.2.3 Diagrammi vaade

Selles vaates on sama diagramm, millele vajutati, kujutatud suuremana ning rohkema informatsiooniga. Diagrammil on pealkiri, milleks on tunnuste nimed. Telgedel on nimed ning skaaladel on rohkem vahejaotusi, mis tagab täpsema ülevaate. Diagrammi saab sulgeda vajutades üleval paremal asuvale „X“ nupule.



Joonis 15 „Diagrammi vaade“

3.3 Rakenduse funktsionaalsus

1. Rakendus peab suutma sisse lugeda CSV-formaadis faile.
2. Rakendus peab kasutama „Test.CSV“ faili sisendfailina, kui kasutaja poolt sisestatud fail ei vasta nõuetele.
3. Rakendus peab suutma sisseloetud failist luua korrelatsioonimaatriksi (sealhulgas korrelatsioonide arvutamine ning diagrammide joonistamine).
4. Rakendus peab võimaldama kasutajal nii maatriksist kui diagrammist pilte salvestada.
5. Rakendus peab võimaldama maatriksil oleval diagrammil klikkides avada uues aknas sama diagrammi detailsemas ning suuremas vaates.

3.4 Kasutusjuhend

3.4.1 Rakenduse käivitamine

Rakenduse käivitamiseks tuleb käivitada fail „CorrelationVisualizer.jar“.

3.4.2 Vajaminevad tehnoloogiad

Selleks, et rakendus tööle hakkaks, peab arvutisse olema installeeritud Java.

3.4.3 Sisendfaili tingimused

1. Fail peab olema CSV formaadis.
2. Faili esimene rida peab koosnema veergude pealkirjadest.
3. Kõik ülejäänud read, v.a esimene rida, peavad olema arvuliste väärtustega, sealjuures pole vahet, kas komana kasutatakse punkti või koma.
4. Failis peab olema vähemalt kaks rida ning kaks veergu.
5. Andmed peavad failis veergudes järjestikku paiknema. See tähendab seda, et veerus ei tohi andmete vahele tühju lahtreid jätta.
6. Kui sisendfail ei ole CSV formaadis, siis kasutatakse näidisandmeid failist „Test.csv“.
7. Kui sisendfailis on kahel veerul erinev arv andmeid, siis kasutatakse nii palju andmeid, nagu sisaldub lühemas veerus.

3.4.4 Rakenduse sulgemine

Maatriksi vaates on võimalik rakendust sulgeda kas vajutades üleval paremal asuvale nupule „X“ või klikkides üleval paremal asuvale „Menu“ nupule ning sealt avanevale valikule „Exit“.

3.4.5 Piltide salvestamine

Maatriksi vaates on maatriksist võimalik pilti teha. Selleks vajutada üleval vasakul asuvale „Menu“ nupule ning sealt omakorda „Screenshot“ nupule. Pilt salvestatakse PNG formaadis rakenduse endaga samasse kausta.

Diagrammist on võimalik samamoodi nagu maatriksist pilti teha. Selleks tuleb esmalt vajutada diagrammi üleval vasakul asuvale „Menu“ nupule ning seejärel rippmenüüst tekki- vale „Screenshot“ nupule. Pilt salvestatakse PNG formaadis ning samasse kausta, kus asub rakendus.

3.5 Edasised arendused

Rakendus pole autori arvates veel valmis ning palju funktsionaalsust on veel võimalik lisada. Esimeseks asjaks, mida lisada, oleks korrelatsioonide arvutamismeetodite valiku- võimalused. Lisaks Pearson'i korrelatsioonile võiks kindlasti valikus olla ka Spearman'i korrelatsioon, kuna see on populaarsuselt ning kasutatavuselt teisel kohal. Tüüpi oleks võimalik valida nii sisendfaili sisselugemisel kui ka pärast maatriksi joonistamist. Sellisel

juhul genereeritakse samadest andmetest lihtsalt uus maatriks kasutades teistsugust meetodit.

Teiseks muudatuseks võiksid olla täiendused sisendfailis. Kindlasti erinevate failitüüpide aktsepteerimine. Praegu toetab rakendus vaid CSV faile, kuid minimaalselt TXT, XLS ning XLSX failide tugi oleks vajalik. Lisaks on plaanis sisendfaili andmete sorteerimine ning töötlemine. Töötlemine on vajalik seetõttu, et elimineerida vigased sisendid, kuna praegu peab sisendfail olema kindlal kujul ning vigaseid andmeid eksisteerida ei tohi. Kui sisendfail oleks siiski sisselugematul kujul, siis rakendus annaks veateate ning paluks sisse lugeda uut faili.

Kolmanda suure arendusena võib välja tuua rakenduse interaktiivsuse. See tähendab, et pärast faili sisselugemist ning maatriksi joonistamist oleks võimalik interaktiivselt seal erinevaid toiminguid sooritada. Kui praegu on võimalik diagramme vaid suuremana kujutada ning salvestada, siis arendusena on plaanis näiteks maatriksi skaala muutmine ehk lahtrid pole fikseeritud suurusega. Lisaks erinevate värvilahenduste kasutamine, mis tähendab seda, et vastavalt korrelatsiooni tugevusele antakse maatriksi lahtrile värv, näiteks korrelatsioon -1 oleks tumesinine, 0 valge ning 1 tumeroheline. Veel oleks võimalik soovitud tunnuseid maatriksist eemaldada. Seda nimelt soovitud korrelatsioone üksikhaaval eemaldades või eemaldades kõik tunnused, mis vastavad teatud kriteeriumitele. Näiteks eemaldada kõik tunnused, millel on alla viiekümne elemendi, kuna valimi hulk ei ole kasutaja jaoks piisavalt suur.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärgiks oli luua rakendus, mis kasutaja poolt ette antud sisendfailist saadud andmete põhjal loob korrelatsioonimaatriksi. Maatriksi sisu pidi välja nägema järgmine: peadiagonaalil ühel pool on korrelatsioonid arvuliste väärtustena ning teisel pool tunnustevaheliste diagrammidena. Lõputöös seatud eesmärk saavutati valminud rakendusega.

Rakendus loodi programmeerimiskeeles Java ning peamiste teekidena kasutati Java'sse sisseehitatud Swing'i ning vabavaralist teeki JFreeChart. Loodud rakendus küsib käivitusel kasutajalt CSV formaadis sisendfaili, mis koosneb veergudest, mille esimene rida on tunnuse nimi ning ülejäänud read veerus koosnevad tunnuse elementidest. Peadiagonaal moodustub veergude esimestest ridadest. Korrelatsioonimaatriksi peadiagonaalil üleval pool asuvad lahtrid on täidetud arvuliste väärtustega. Nendeks väärtusteks on erinevate tunnuste vahelised korrelatsioonid, mis on arvatud, kasutades Pearson'i korrelatsiooni. Peadiagonaalil allpool asuvad tunnustevahelisi seoseid kujutavad diagrammid. Diagrammil klikkides avaneb sama diagramm suurema ning detailsemana uues aknas. Nii maatriksit tervikuna kui ka detailvaates avatud diagrammi on võimalik salvestada.

Tulevikus on võimalik rakendust edasi arendada, lisades sellele mitmesugust funktsionaalsust: korrelatsiooni arvutamine teistsuguseid meetodeid kasutades, erinevate sisendfailitüüpide võimalus, vigase sisendfaili aktsepteerimine ning parandus, rakenduse skaala muutmine, maatriksile värvilahenduste lisamise võimalus, tunnuste eemaldamine maatriksist nii üksikhaaval kui ka parameetrite alusel ning palju muud.

Lõputöö kirjutamise käigus sai autor juurde teadmisi nii Java kui ka Java teekide kohta. Kindlasti sai autor väärtusliku kogemuse visuaalse kasutajaliidese arendamise osas. Palju uut informatsiooni omandas autor ka visualiseerimise vallas üldiselt. Lisaks tutvus autor mitmete programmidega, millega varem kokkupuude puudus. Autor ei olnud varem teadlik, et visualiseerimisega on seotud terve teaduse liik nimega visuaalne andmeanalüüs.

Kasutatud kirjandus

- [1] Suda M, Hampton-Smith S. The 37 best tools for data visualization. *Creative Bloq*. <http://www.creativebloq.com/design-tools/data-visualization-712402> (vaadatud 10.05.15)
- [2] Wikipedia. Visualiseerimine. <http://et.wikipedia.org/wiki/Visualiseerimine> (vaadatud 10.05.15)
- [3] Mikko E, Hinsberg H, Kärp T. Andmete visualiseerimise oskuste arendamine. <http://www.praxis.ee/tood/andmete-visualiseerimise-oskuste-arendamine/> (vaadatud 10.05.15)
- [4] Infovara. Andmete visualiseerimine. <http://www.infovara.ee/andmete-visualiseerimine/> (vaadatud 10.05.15)
- [5] Georgia College of Tech Computing. Scientific Visualization Tutorial. <http://www.cc.gatech.edu/scivis/tutorial/linked/whatissscivis.html> (vaadatud 10.05.15)
- [6] Keim D, Kohlhammer J, Ellis G, Mansmann G. Mastering the Information Age: Solving Problems with Visual Analytics. Goslar, Germany: Eurographics Association; 2010. <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf> (vaadatud 10.05.15)
- [7] Lundblad B. Dissecting How to Choose the Right Chart. *Qlik Blog*. <http://www.qlik.com/us/blog/posts/patrik-lundblad/dissecting-how-to-choose-the-right-chart> (vaadatud 10.05.15)
- [8] Georgia College of Tech Computing. Scientific Visualization Tutorials: Motivations. <http://www.cc.gatech.edu/scivis/tutorial/linked/motivations.htm> (vaadatud 10.05.15)
- [9] Georgia College of Tech Computing. Scientific Visualization Tutorials: Common Questions and Concerns. <http://www.cc.gatech.edu/scivis/tutorial/linked/concerns.htm> (vaadatud 10.05.15)
- [10] Quigley A. Introduction to Information and Data Visualisation. <http://www.csi.ucd.ie/staff/hcarr/home/department/infovisintro.pdf> (vaadatud 10.05.15)
- [11] Tudengiveeb. Kuidas esitada andmeid? <https://tudengiveeb.ee/et/esileht/toeoturule-sisenemine/16-oppimine/kirjalike-toeode-kirjutamine/126-kuidas-esitada-andmeid> (vaadatud 10.05.15)
- [12] Myatt GJ, Johnson WP. *Making Sense of Data III: A Practical Guide to Designing Interactive Data Visualizations*. Hoboken, NJ, USA: John Wiley & Sons; 2011.
- [13] Lewis A. 10 Best Or Worst Ways To Visualise Web Analytics Data. *Online Behaviour*. <http://online-behavior.com/analytics/data-visualization> (vaadatud 10.05.15)
- [14] Perez R. Four Easy Visualization Mistakes to Avoid. *Visually*. <http://blog.visual.ly/data-visualization-mistakes-to-avoid/> (vaadatud 10.05.15)

- [15] Chang K, Eyler-Werve K, Cairo A. Common Visualization Mistakes. <https://infoactive.co/data-design/ch18.html> (vaadatud 10.05.15)
- [16] Crooks R. Why Most People's Charts & Graphs Look Like Crap. *HubSpot Blogs: Marketing*. <http://blog.hubspot.com/marketing/data-visualization-mistakes> (vaadatud 10.05.15)
- [17] Wikipedia. Spearman's rank correlation coefficient. http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient (vaadatud: 10.05.15)
- [18] Minitab. A comparison of the Pearson and Spearman correlation methods. <http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearman-correlation-methods/> (vaadatud 10.05.15)
- [19] Microsoft. Excel. <https://products.office.com/en-us/Excel> (vaadatud 10.05.15)
- [20] The R Foundation. What is R? <http://www.r-project.org/about.html> (vaadatud 10.05.15)
- [21] Stackoverflow. <http://stackoverflow.com/questions/11581879/plot-of-a-correlation-matrix-in-r-like-in-excel-example> (vaadatud 10.05.15)
- [22] Stata. Why use Stata? <http://www.stata.com/why-use-stata/> (vaadatud 10.05.15)
- [23] TIOBE Software. TIOBE Index for April 2015.
- [24] The Java Programming Language. <http://groups.engin.umd.umich.edu/CIS/course.des/cis400/java/java.html> (vaadatud 10.05.15)
- [25] Wikipedia. Java (programming language). [http://en.wikipedia.org/wiki/Java_\(programming_language\)](http://en.wikipedia.org/wiki/Java_(programming_language)) (vaadatud: 10.05.15)
- [26] Black Duck Software. Android: Project Summary. <https://www.openhub.net/p/android> (vaadatud 10.05.15)
- [27] Siegler MG. The Key To Gmail: Sh*t Umbrellas. *TechCrunch*. <http://techcrunch.com/2010/03/14/key-to-gmail/> (vaadatud 10.05.15)
- [28] Techopedia. Java Swing. <http://www.techopedia.com/definition/26102/java-swing> (vaadatud 10.05.15)
<http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html> (vaadatud 10.05.15)
- [29] Wikipedia. Swing (Java). http://en.wikipedia.org/wiki/Swing_%28Java%29 (vaadatud 10.05.15)
- [30] JFreeChart. <http://www.jfree.org/jfreechart/> (vaadatud 10.05.15)

Lisad

I. Terminid

CSV Komaeraldusega väärtused. Failivorming, kus kirjed on üksteisest eraldatud komaga või semikooloni-ga.	CSV Comma-separated values. File, where items are separated with commas.
Ekraanitõmmis Pilt, mis on salvestatud visuaalsetest elementidest, mis asuvad ekraanil.	Screenshot Image taken to record the visible items displayed on the screen.
PNG Graafilise faili failinime laiend.	PNG Portable Network Graphics. A graphical file filename extension.
XLS/XLSX Programmi Microsoft Excel failiformaat.	XLS/XLSX File format for Microsoft Excel.
Maatriks Ristkülikukujuline tabel, mis koosneb arvudest või mingitest muudest etteantud hulga elementidest.	Matrix Rectangular table, which consists of number or some other elements from given array.

II. Rakendus, selle lähtekood ning testandmestik

- Lõputöös valminud rakendus on lisana kättesaadav failina „CorrelationVisualizer.jar“.
- Lõputöös valminud rakenduse lähtekood on lisana kättesaadav järgmises failis: „CorrelationVisualizer.java“.
- Testandmestik on lisana kättesaadav failis „Test.csv“.

III. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina **Eerik Muuli** (sünnikuupäev: 21.02.1993)

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Korrelatsioonimaatriksi visualiseerimine,

(lõputöö pealkiri)

mille juhendaja on Vambola Leping,

(juhendaja nimi)

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **14.05.2015**