

Kursus: Andmeanalüüs/Kvantitatiivsed ...

Seminar II: Kirjeldav statistika - tabelid ja diagrammid.

Õppejõud: Katrin Niglas
PhD, dotsent
informaatika osakond



Millest sõltub analüüsimeetodi valik?

- I. Küsimuse tüübist
e mis tüüpi on küsimus, millele tahame analüüsiga vastust saada – nt Kas kolm gruppi **erinevad**? Kas kaks nähtust on **seotud**?
- II. Andmete tüübist
Kas nimi-, järjestus, arv- või binaarsed tunnused
- III. Sihtrühmast
Kui suurt teadlikkust statistiliste meetodite osas võib eeldada?
Milline esitlusviis on selle rühma puhul kõitev ja sobilik?



Kuidas oma andmeid kokku võtta?

Struktureeritud andmete esmaseks kokkuvõtuks ning ülevaatlikuks analüüsiks saab kasutada **kirjeldava statistika** meetodeid, mis võib jagada kolme suurde rühma:

- Sagedustabelid (sh risttabelid)
- Arvnäitajad
- Arvjoonised e diagrammid

Sagedustabelid

Sagedustabelid aitavad vastust leida kahte tüüpi küsimustele:

- Kui palju on/oli? (arv, sagedus, absoluutne sagedus)
- Kui suur osa on/oli? (protsent, osakaal, suhteline sagedus)

Perekonnaseis	Arv	Osakaal
abielus	479	54,3%
vallaline	157	17,8%
lesk	92	10,4%
lahutatud	68	7,7%
vabaabielus	65	7,4%
elab lahus	21	2,4%
Vastanuid kokku	882	100,0%

Sagedustabelid

Võrdlusteks sobivad mitmemõõtmelised sagedustabelid või risttabelid:

Kas suitsetate?	mees		naine	
	Arv	Osakaal	Arv	Osakaal
ei suitseta üldse	162	39,3%	343	74,4%
vahetevahel, kuid mitte iga päev	25	6,1%	34	7,4%
iga päev	225	54,6%	84	18,2%
Kokku	412	100,0%	461	100,0%

sugu	mees	Arv	Kas suitsetate?			Kokku
			ei suitseta üldse	vahetevahel, kuid mitte iga päev	iga päev	
	Arv	162	25	225	412	
	%	39,3%	6,1%	54,6%	100,0%	
	Arv	343	34	84	461	
	%	74,4%	7,4%	18,2%	100,0%	
Kokku	Arv	505	59	309	873	
	%	57,8%	6,8%	35,4%	100,0%	



TALLINNA ÜLIKOOL

emakeel * tänuikkus kelleltki saadud toe või ilmutatud huvi eest Crosstabulation

		tänuikkus kelleltki saadud toe või ilmutatud huvi eest			Total
		pole tundnud	korra olen tundnud	palju kordi olen tundnud	
emakeel eesti	Count	85	138	178	401
	% of Total	12,6%	20,5%	26,5%	59,7%
vene	Count	58	129	84	271
	% of Total	8,6%	19,2%	12,5%	40,3%
Total	Count	143	267	262	672
	% of Total	21,3%	39,7%	39,0%	100,0%

Mõtle ja ütle, millises tabelis on sisuliselt sobiv info?!

emakeel * tänuikkus kelleltki saadud toe või ilmutatud huvi eest Crosstabulation

		tänuikkus kelleltki saadud toe või ilmutatud huvi eest			Total
		pole tundnud	korra olen tundnud	palju kordi olen tundnud	
emakeel eesti	Count	85	138	178	401
	% within tänuikkus ...	59,4%	51,7%	67,9%	59,7%
vene	Count	58	129	84	271
	% within tänuikkus ...	40,6%	48,3%	32,1%	40,3%
Total	Count	143	267	262	672
	% within tänuikkus ...	100,0%	100,0%	100,0%	100,0%

emakeel * tänuikkus kelleltki saadud toe või ilmutatud huvi eest Crosstabulation

		tänuikkus kelleltki saadud toe või ilmutatud huvi eest			Total
		pole tundnud	korra olen tundnud	palju kordi olen tundnud	
emakeel eesti	Count	85	138	178	401
	% within emakeel	21,2%	34,4%	44,4%	100,0%
vene	Count	58	129	84	271
	% within emakeel	21,4%	47,6%	31,0%	100,0%
Total	Count	143	267	262	672
	% within emakeel	21,3%	39,7%	39,0%	100,0%



TALLINNA ÜLIKOOL

Kuidas oma tulemusi esitleda?

Statistilisi andmeid ja statistilise andmeanalüüsi arvulisi tulemusi saab esitada:

- tekstina
- tabelina
- arvjoonise e diagrammina

Kuidas oma tulemusi esitleda?

Mõned käibe"tõed":

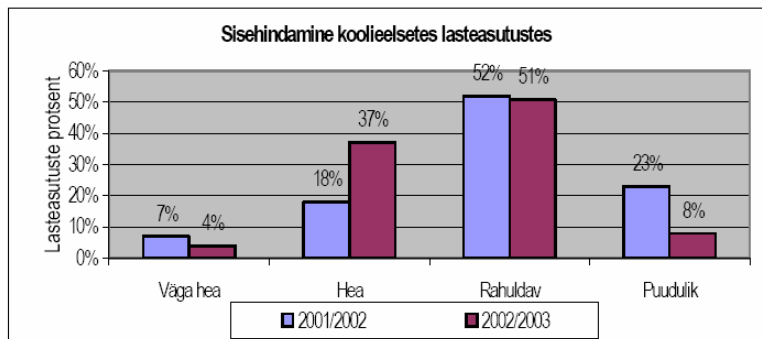
Tekst on igav ja ei jää meelde ...

... tabel on arvude surnuaed ...

... esita oma tulemused alati diagrammina!

mille (paikapidavuse) üle tasub enne tegutsemist pisut pikemalt mõtiskleda!

Kuidas oma tulemusi esitleda? – näide 1

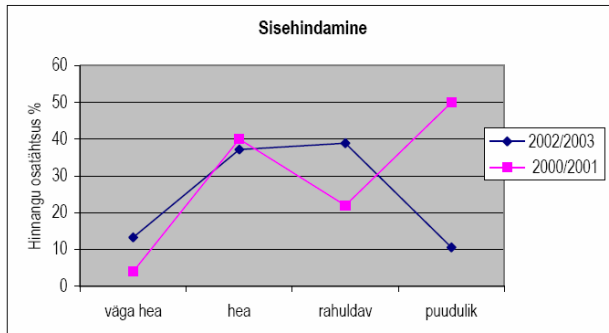


Joonis 1. Sisehindamine koolieelsetes lasteasutustes

Kuidas oma tulemusi esitleda? – näide 1 Kommentaariid

- + Valitud on vaatajale harjumuspärane ning sobiv diagrammi tüüp (lähtuvalt analüüsi eesmärgist ja andmete tüübist), mistõttu on diagrammi üldiselt lihtne jälgida.
- Väärtuste järjekord teljel võiks olla vastupidine, sest harjumuspäraselt on teljel paremal pool kõrgemad väärtused, mis seotud parema hinnanguga, mistõttu praegune visuaalne lahendus jätab esialgu tagurpidise mulje võrreldes tegeliku tendentsiga
- Diagrammil toodud numbrid on halvasti loetavad, kuna nad lõikuvad joonisele kantud abijoontega (ühest tuleks loobuda!)
- Diagrammil kasutatud hall taustavärv on pisut segav ja esineb dubleerimist (protsent telje pealkirjas ning ka arvu kõrval)

Kuidas oma tulemusi esitleda? – näide 2



Joonis 2. Hinnangu muutumine võtmealale *Sisehindamine*.

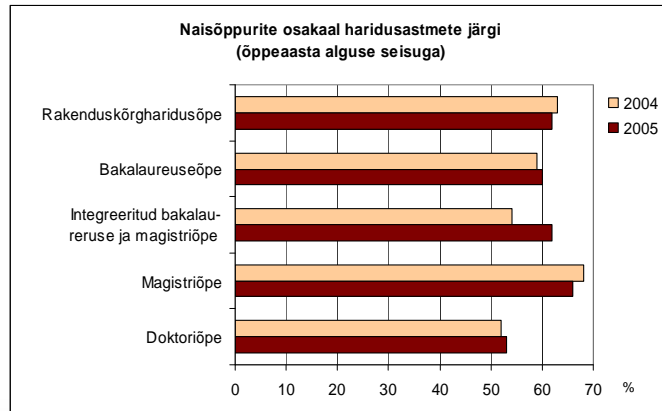
Riikliku järelevalve tulemustest selgub, et enamikes koolides tegeldakse sisehindamisega, kuid vajakajäämisi on süsteemsuses (www.hm.ee > järelevalve > järelevalve tulemused).

86

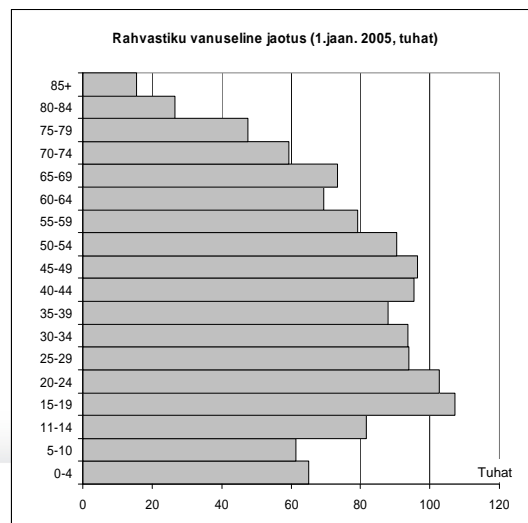
Kuidas oma tulemusi esitleda? – näide 2 Kommentaariid

- Joondiagramm ei ole parim võimalik diagrammi tüüp antud olukorras, kuna harjumuspäraselt peaks sellisel puhul olema ajaühikud alumisel teljel, ning joone tõusmine või langemine näitama hinnangu muutumist ajas
- Olukorra teeb lugejale veelgi keeruka-maks asjaolu, et samas materjalis eelmisel lehel oli täpselt sama loogikaga andmete esitamiseks kasutatud tulppiagrammi (vt. näide 1)
- Tõusev joon tähendab harjumuspäraselt muutust paremuse poole, mistõttu tekib antud jooniselt eelmise näitega võrreldes veelgi rõhutatumalt vale visuaalne mulje nagu oleks 2002/2003 aastal hinnangud võrreldes eelmise aastaga halvemaks läinud (probleemiks väärtuste järjekord teljel)
- + Vertikaalsel teljel on välditud dubleerimist (telje pealkiri viitab protsendile, millest lähtuvalt arvude juures enam ühikut toodud ei ole)

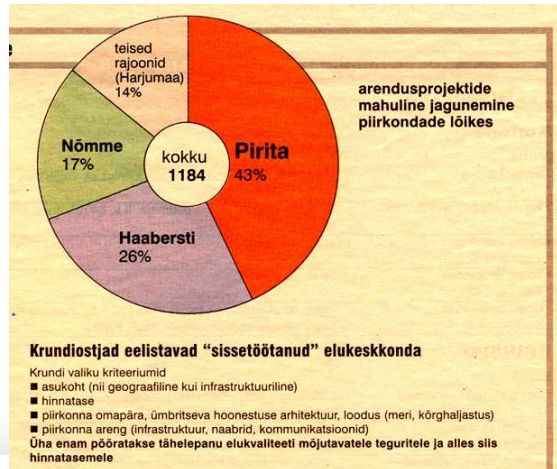
Enamkasutatavad diagrammid - tulpdiaagramm



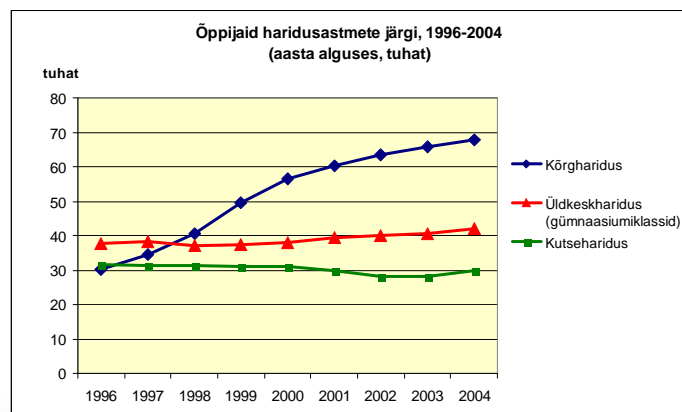
Enamkasutatavad diagrammid - histogramm



Enamkasutatavad diagrammid - sektordiagramm



Enamkasutatavad diagrammid - joondiagramm



Õpime teiste vigadest!

Tõepoolest, diagrammist on materjalist arusaamisel tihti abi, aga kas alati ja kas igasugusest?

Järgmised näited on vaid väike kolleksioon tegelikest esitlustest pärit diagrammidest, kus miskit on mitte päris nii nagu peaks.

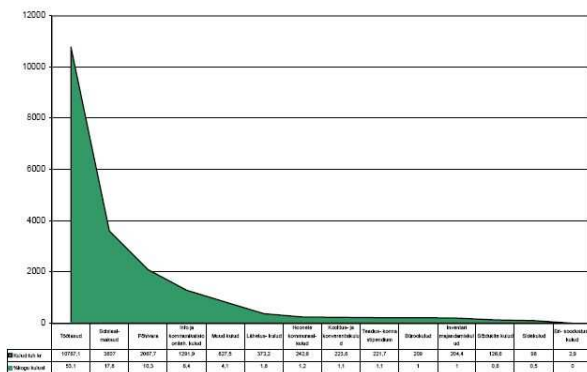
Mõttele ja otsustele, mida tuleks teha teisiti, et esitletava sisu tõepoolest võimalikult hästi vaatajani jõuaks!



TALLINNA ÜLIKOOL

☺ Nagu te kõik kõrvalolevalt jooniselt **näha** võite ...

MATEMAATIKA-LOOUSTEADUSKONNA KULUD 2004
(20 283,2 tuh kr)

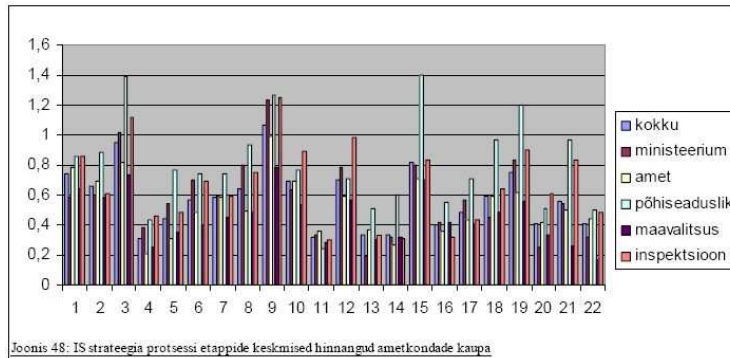


☹ Kas ma peaksin prille kandma?



TALLINNA ÜLIKOOL

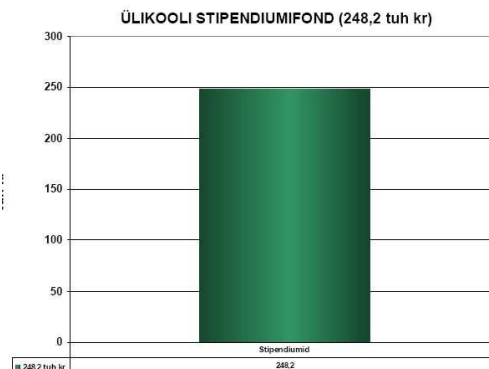
☺ Analüüsi põhjal selgusid **mitmed** huvitavad erinevused ...



☺ Mida rohkem seda ... Kas rohkem ikka on alati parem!?



☺ Nagu juba enne mainitud, on meil stipendiumifondis **284 tuh kr**



☺ Kordamine on ... Huvitav, kas tulp oleks lühem ja peenem tulnud, kui stipendiumiteks oleks vähem raha olnud!?

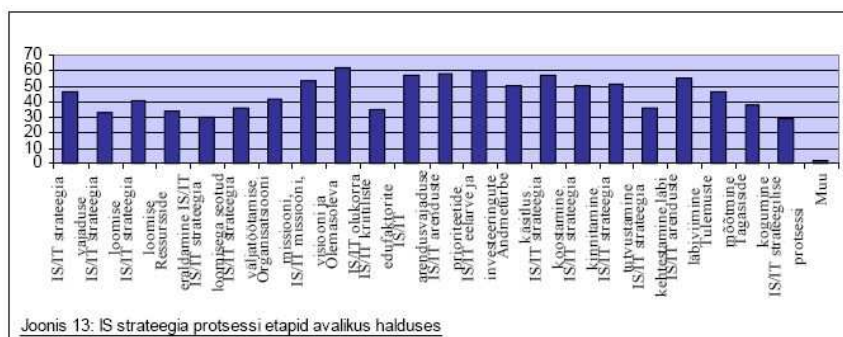


- ☹️ Kaasaegse **tarkvara piiramatuid võimalusi** kasutades saab mõjuvaid esitlusi teha ...



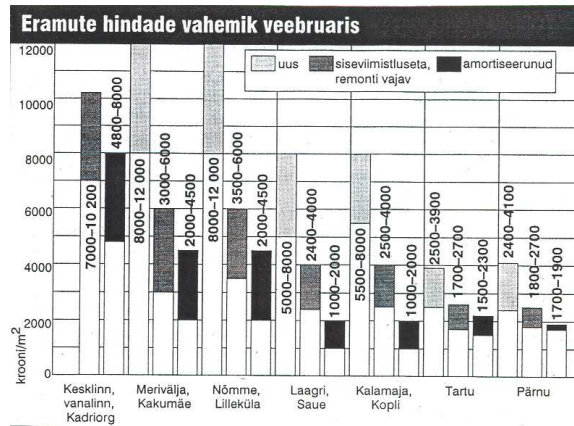
- ☹️ Seda küll, aga kas selline vorm ei „ahista“ sisu!?

- ☹️ Kaasaegse tarkvara abil saab **kiiresti ja vaevata** mõjuvaid esitlusi teha ...



- ☹️ Seda küll, aga kas maksab ikka lõpuni tarkvara peale loota!?

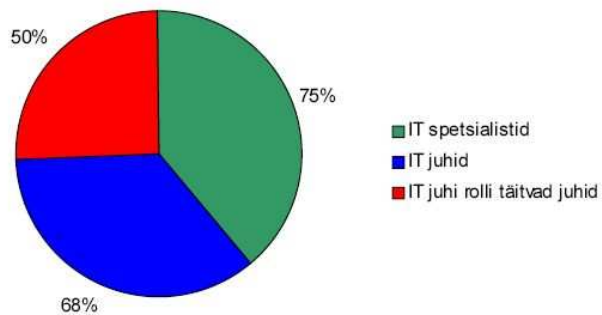
☺ Andmetest **parema ülevaate** saamiseks ...



☹ Kas see ikka on kõige lihtsam ja selgem viis esitluseks!?



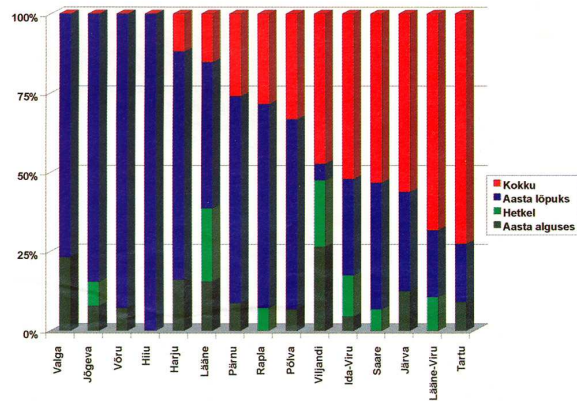
☺ Andmetest **parema ülevaate** saamiseks ...



☹ Oot, nende protsentidega on alati üks jama, aga 50% on ikka üsna kindlasti pool, mitte veerand!?



☺ "Jooniselt on näha püsiühendusega omavalitsuste arv ..."

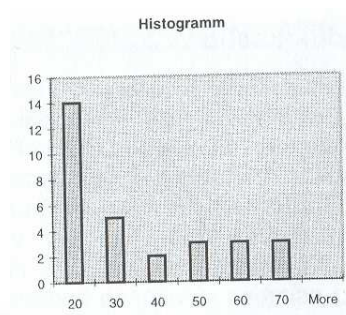
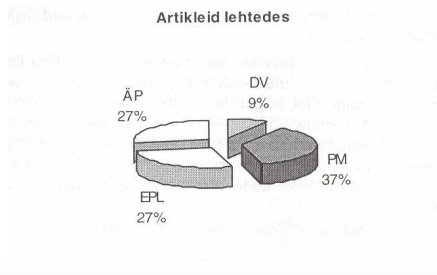


☹ Arvu oleks tõesti võinud kuskil ära tuua ... aga kuidas "kokku" saab olla suvaline osa tervikust!?

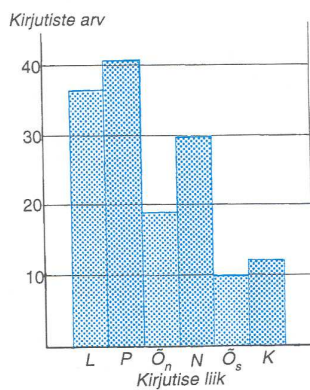


Ehk on õpikust abi?!

Kahjuks ei ole ka seal, kus diagrammide tegemist õpetatakse, alati asjad päris korras ...



Nii soovitati õpikus:



Kuid kas nii pole pisut selgem?



Vaata ja mõtle, kas kujundus
aitab või segab
sisust arusaamist!?

Kas nii või naa ... ?!

Nii oli originaalis!

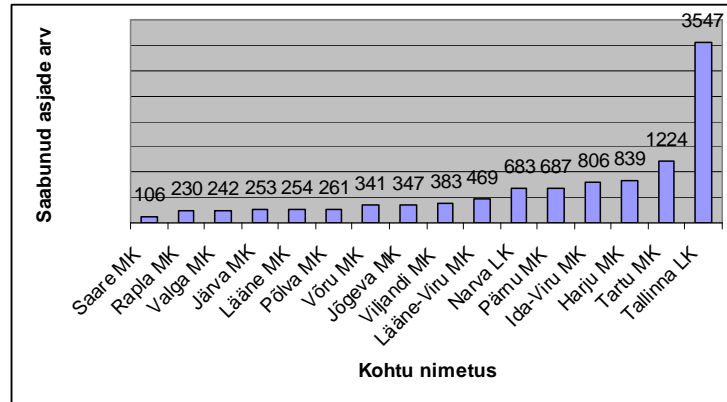
1.1. Kuritegevuse dünaamika

	1993.a	1994.a	1995.a	1996.a	1997.a.
maa- ja linnakohtud	6192	6199	5835	6301	6413
muut %-des		+0,1%	-5,9%	+7,9%	+1,8%
	1998.a	1999. a	2000.a		
maa- ja linnakohtud	6290	7279	9224		
muut %-des	-1,9%	+15,7%	+26,7%		

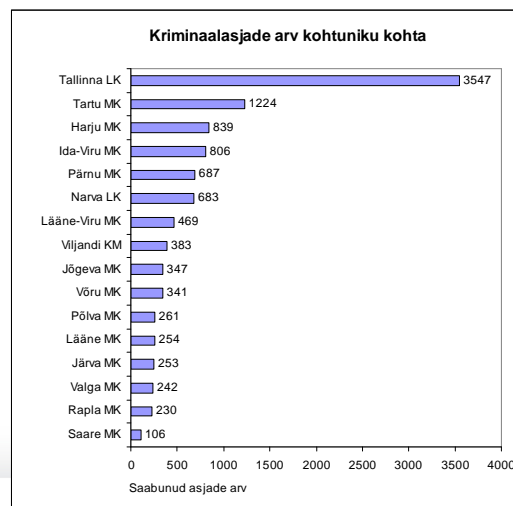
Kas pole pisut selgem?

Kuritegevuse dünaamika (maa- ja linnakohtud)		
	Kriminaal- asjade arv	Muut võrreldes eelmise aastaga
2000.a	9224	26,7 %
1999.a	7279	15,7 %
1998.a	6290	-1,9 %
1997.a	6413	1,8 %
1996.a	6301	7,9 %
1995.a	5835	-5,9 %
1994.a	6199	0,1 %
1993.a	6192	

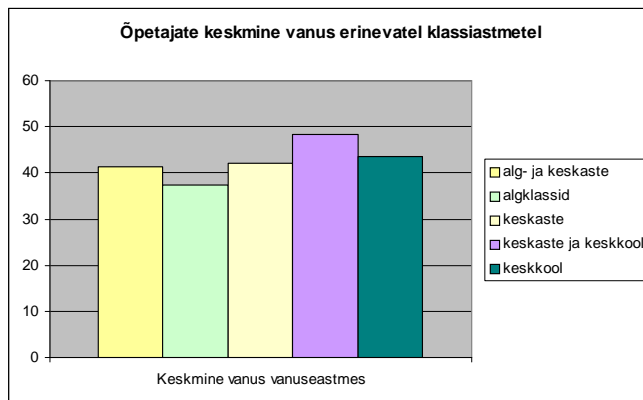
Nii oli originaalis!



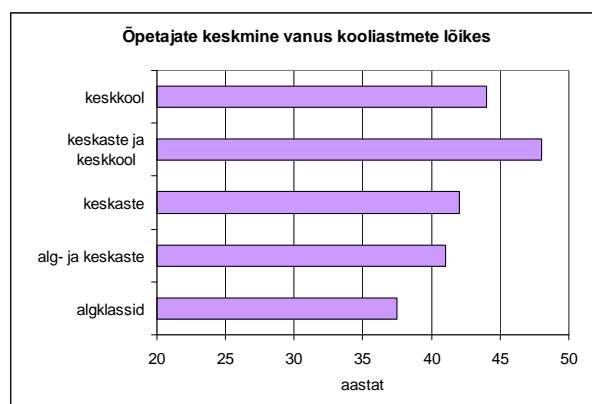
Kas pole pisut selgem?



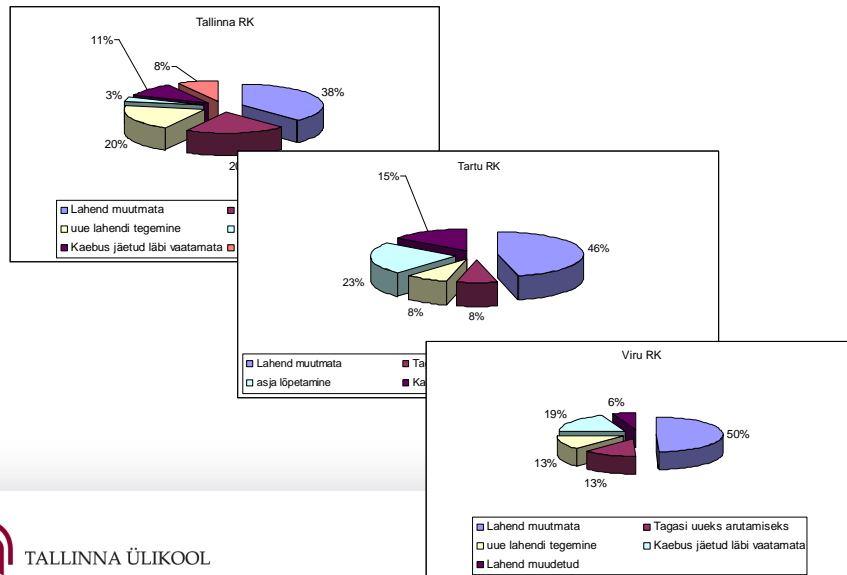
Nii oli originaalis!



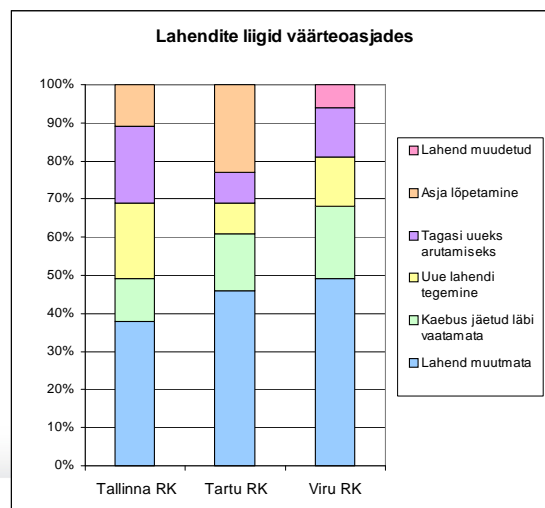
Kas pole pisut selgem?



Nii oli originaalis!



Kas pole pisut selgem?



Mõned lihtsad reeglid!

Üldjuhul:

- Vali esitluseks diagramm, kui soovid eelkõige anda kiiret ülevaadet üldtendentsi(de)st ja suundumus(te)st
- Vali esitluseks tabel, kui on vajalik anda edasi täpset arvulist infot või kui võrreldavate arvnäitajate suurusjärgud on väga erinevad
- Vali arvulise info edastusviisiks tekst, kui korraga on vaja esitada vaid üks-kaks arvulist näitajat



Mõned lihtsad reeglid!

Enne arvutilt "abi palumist" mõtle täpselt välja, millist tüüpi tabel või diagramm aitab parimal viisil edasi anda sinu andmetes peituvat sisulist mõtet ning sobib kõige enam silmas pidades valitud sihtrühma.

Kui oled arvutilt tulemi tellinud, siis esimesena veendu, et tabeli või diagrammi sisu oleks õige ja alles siis asu kujundusliku külje kallale.

Jäta meelde, et tabel ja diagramm peavad olema sisuliselt mõistetavad ka ilma toetava tekstita (sisu mõistmiseks vajalikku infot saab anda peal- või allkirjas, legendis, allmärkustes, jne)

Enne "suurele avalikkusele" esitamist kontrolli, kas sinu asjassepühendamatud sõbrad suudavad sinu koostatud tabelist või diagrammilt kiiresti ja muretult õige info välja lugeda.



Kujundamise juures pea meeles:

Kogu kujundus peab teenima ainult üht eesmärki -
äratama huvi ja usaldust esitatava info vastu ning aitama kaasa selle sisulisele mõistmisele!

Kõik info, mis on tabelis või joonisel, peab olema ilma pingutuseta loetav normaalse nägemisega vaatajale (PS! vajalik tekstide suurus sõltub suuresti esitluse viisist: lähedalt loetaval paberil, plakatil või stendil, avalik esitus konverentsil või loengus, jne).

Tekstid ja toodud arvuline info ei tohi lõikuda diagrammil olevate kujutiste ja joontega. Diagonaalis ja vertikaalis esitatud tekste tuleks võimalusel vältida, sest need on raskesti loetavad.

Kui diagrammil või tabelis võrreldavad alamgrupid ei ole sisulises mõttes üheselt järjestatavad (st tegemist on nimiskaalaga), siis on mõistlik alamgrupe esitada read/tulbad/sektoore esitada kasvavasse või kahanevasse järjekorda sorteeritult.

Kujundamise juures pea meeles:

Diagrammil ei ole mõistlik esitada korraga liiga palju detailset infot (st liiga kirjud diagrammid ei ole sisuliselt loetavad) – olukorra lahendamiseks koonda infot või vali esitlusviisiks tabel.

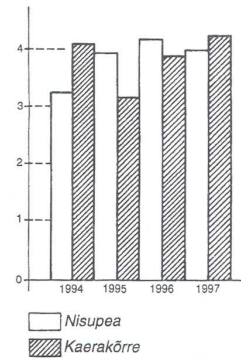
Mida rohkem ja keerulisem on tabelis/diagrammil toodud sisuline info, seda tagasihoidlikum ja rahulikum peab olema kujundus.

Efektide ja värvide valikul arvesta ka sihtrühma ja avaldamise konteksti: akadeemiline, populaar-teaduslik, avalik meedia, jne

Ühe materjali/esitluse piires püüa kasutada võimalikult ühtset kujundusstiili (st ära kasuta igal järgmisel joonisel uusi värve ja uusi efekte!). Samas diagrammi tüüp tuleb valida igal konkreetsel juhul lähtuvalt andmete iseloomust!

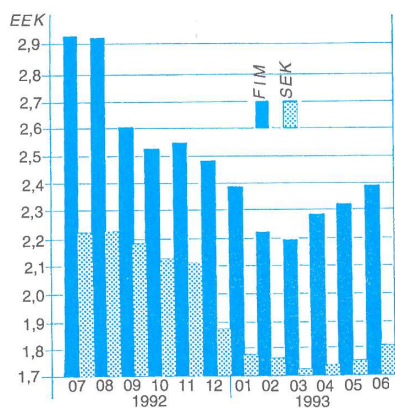
Mõtle ja ütle, mis tuleks teisiti teha?!

1072. Joonisel 4.19 on kujutatud kahe naabri – Nisupea ja Kaerakõrre talu odrasaagid tonnides hektari kohta ($\frac{t}{ha}$) neljal järjestikusel aastal. Kirjelda seda diagrammi ja tee võimalikke arvutusi.



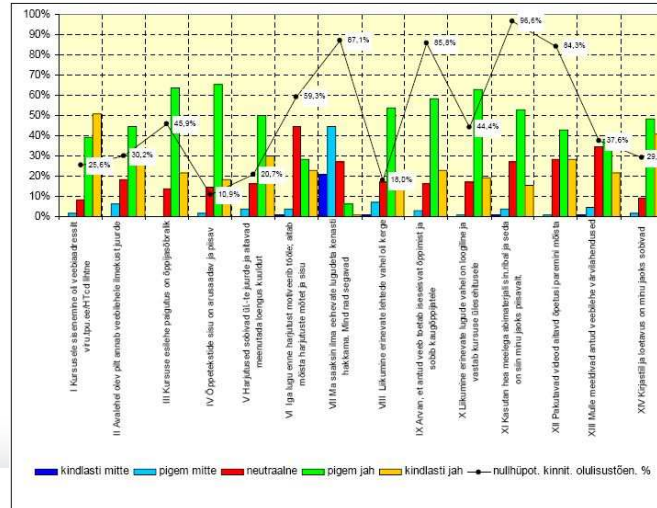
JOON. 4.19

Mõtle ja ütle, mis tuleks teisiti teha?!



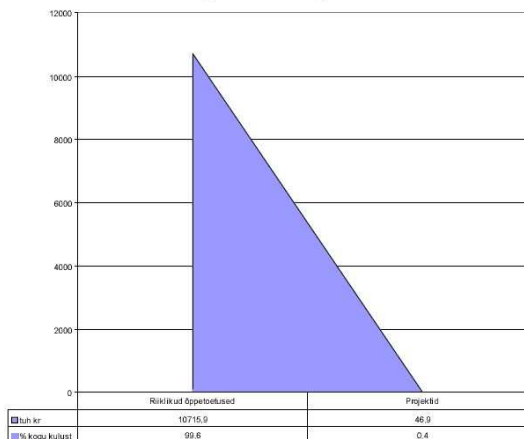
JOON. 6.11

Mõtle ja ütle, mis tuleks teisiti teha?!



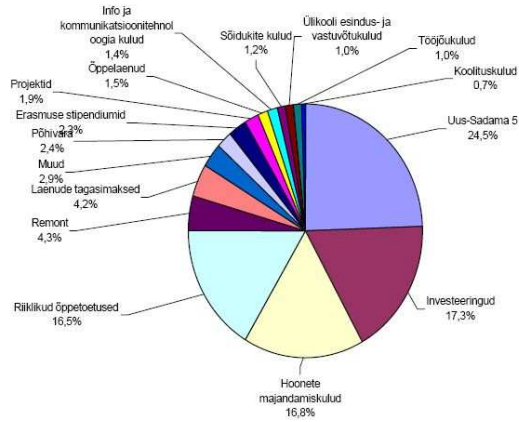
Mõtle ja ütle, mis tuleks teisiti teha?!

ÕPPEOSAKONNA ÜLEÜLKOOLILISED KULUD 2004
(10 762,8 tuh kr)

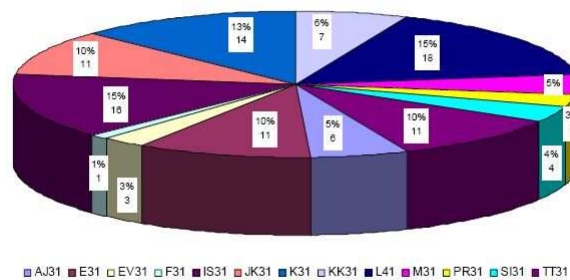


Mõtle ja ütle, mis tuleks teisiti teha?!

TUGISTRUKTUURIOSAKONDADE ÜLEÜLIKOOILISED KULUD 2004
(64 901,8 tuh kr)

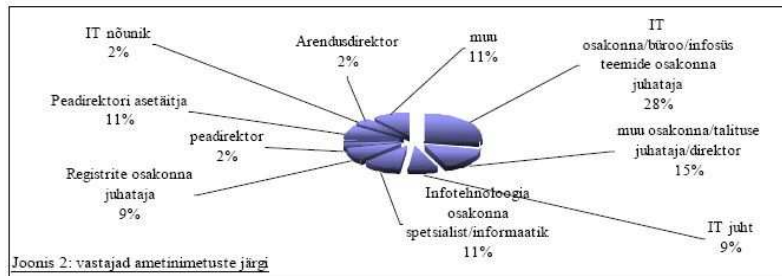


Mõtle ja ütle, mis tuleks teisiti teha?!



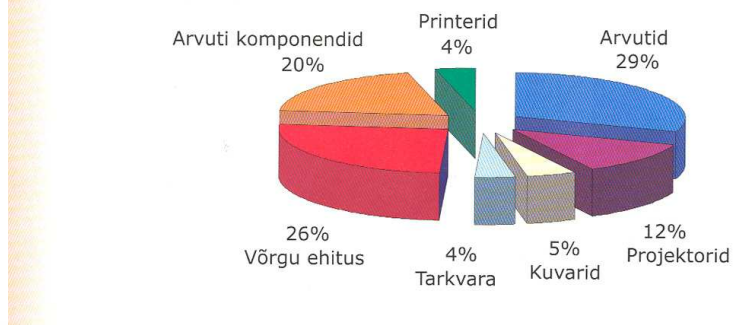
Joonis 23. Respondentide erialaline jaotuvus

Mõtle ja ütle, mis tuleks teisiti teha?!

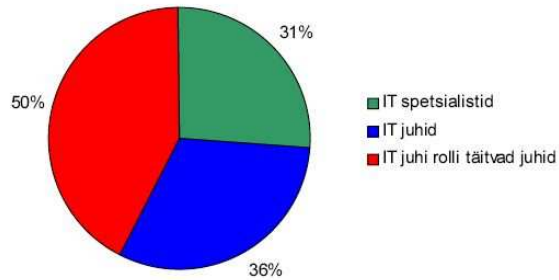


Mõtle ja ütle, mis tuleks teisiti teha?!

IKT rahade kasutamine Tallinna viie kooli näitel.



Mõtle ja ütle, mis tuleks teisiti teha?!



Mõtle ja ütle, mis tuleks teisiti teha?!

Küsimused

1. Mis on oksüdatsiooniaste?
2. Kuidas märgitakse elemendi oksüdatsiooniastet ühendis?
3. Missugune on järgmiste elementide oksüdatsiooniaste ühendites:
O, H, Na, Ca, Al?
4. Määra elementide oksüdatsiooniastmed järgmistes ühendites:
NO₂, Fe₂O₃, H₂S, SO₂, H₂SO₄.
5. Selgita ühendi valemi koostamist elementide oksüdatsiooniastmete alusel.
6. Koosta oksiidide valemid nende koostises oleva elemendi oksüdatsiooniastme järgi:
Cu^{I, IV}, N^V, N^{III}, S^{VI}.
7. Missugust nimetamisviisi kasutatakse tavaliselt:
a) metallioksiidide.

Õiguserikkumise asja liik	1995.a	1996.a	1997.a	1998.a	1999.a	2000.a
Alkohoolsete jookide pruukimine avalikus kohas või avalikku kohta joobnud olekus ilmumine	Andmed puuduvad	1396 isikut	1244 isikut	1233 isikut	1759 isikut	1877 isikut
Pisihuligaansus	1942 isikut	1313 isikut	1132 isikut	936 isikut	838 isikut	835 isikut
Avalikku korda ja ühiskondlikku julgeolekut ründavad haldusõiguserikkumised	3270 isikut	1277 isikut	1063 isikut	1272 isikut	1529 isikut	1651 isikut
Pisivargus	Andmed puuduvad	663 isikut	745 isikut	821 isikut	865 isikut	732 isikut
Politseiametniku või ühiskondliku politseitöötaja seadusliku korralduse eiramine	1168 isikut	813 isikut	495 isikut	459 isikut	432 isikut	420 isikut
Välismaalaste seaduse rikkumine	518 isikut	460 isikut	279 isikut	254 isikut	435 isikut	267 isikut
Võimuesindaja või avalikku korda kaitsva muu isiku solvamine	96 isikut	323 isikut	244 isikut	233 isikut	308 isikut	414 isikut
Ebaseaduslik ettevõtlus	Andmed puuduvad	756 isikut	265 isikut	319 isikut	513 isikut	445 isikut
Haldusõiguserikkumisi kokku	9139 isikut	8459 isikut	7392 isikut	7790 isikut	9743 isikut	8641 isikut