



# Luvut, num3rot ja kuvat

**Kimmo Vehkalahti**

Matematiikan ja tilastotieteen laitos  
Helsingin yliopisto

## Aluksi

Tv-sarjassa Num3rot (*Numb3rs*) ratkotaan rikosmysteereitä matemaattisin ja tilastollisin keinoin. Sarja korostaa matematiikan roolia arkipäiväisissä asioissa kuten sään ennustamisessa, mutta muistuttaa sen tärkeydestä myös rikosten analysoinnissa ja käyttäytymisen mallintamisessa. Kantava teema on säännönmukaisuuksien hahmottaminen, jota todennäköisyyslaskennan ja tilastollisten menetelmien ohella tukevat erilaiset tilastolliset kuvat.

Lukujen ja numeroiden esittämisellä tilastollisina kuvina on verrattain pitkä historia, ja monet kuvatyypeistä ovat vakiintuneet osaksi tiedon esittämisen arkipäivää. Kekseliäisyydellekin on edelleen tilaa, sillä alati laajeneva informaatiotulva asettaa haasteita yhä suurempien tietomäärien esittämiselle yhä tiivistetympin.

Seuraavassa tarkastelen joitakin enemmän tai vähemmän tyypillisiä tilastollisia kuvia. Kommentoin niiden laatimista ja tulkintaa enimmäkseen tekniseltä kannalta ja jätän kuvien mahdollisen sisällön tutkimisen haasteeksi lukijalle.

## Pylväät ja piirakat

Tilastollinen kuva on parhaimmillaan tehokas ja mieleenpainuva tapa esittää lukuja ja numeroita. Yleisim-

piä lienevät erilaiset pylväät ja piirakat, jotka perustuvat pinta-alojen tulkintaan.

Kuvassa 1 on tyypillinen, lukumääriä esittävä pylväskuva. Prosenttiosuudet on lisäksi ilmaistu lukuna pylväiden päissä ja kokonaismäärä kerrottu kuvan sisään sijoitetulla tekstillä. Pylväät voitaisiin piirtää myös pystyyn, mutta tällöin nimien kanssa tulisi ongelmia. Selkeys on tärkeä kriteeri, sillä tilastollisen kuvan yleisenä päämääränä on välittää tietoa. Tieto ei mene perille, jos esitys on epäselvä. Varsinkin vinoon ladottuja nimiä on vaikea lukea.

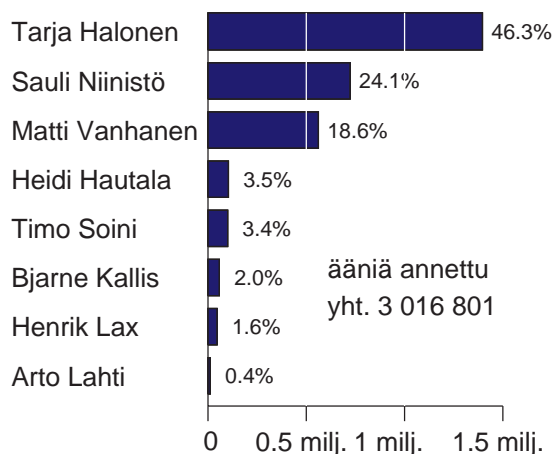
Selkeyteen liittyy paljon muutakin: ylimääräisiä raameja ja etenkin koristeita on syytä välttää, numeeristen asteikkojen kuvausten on oltava helppolukuisia ja värien käytön harkittua. Värikuvia on helppo tuottaa, mutta on hyvä muistaa että niitä saatetaan usein tulostaa paperille mustavalkoisina. Pelkästään väreihin perustuva esitys voi latistua äkkiä lukukelvottomaksi sotkuksi. Hyvin laadittu mustavalkokuva onkin monesti harkinnan arvoinen vaihtoehto värikuvalle. Tieteellisissä julkaisuissa se on usein ainoa vaihtoehto, joskin verkkojulkaisemisen myötä painopiste lienee muuttumassa.

Kuva 2 kertoo piirakkakuvan luonteen mukaisesti vain prosenttiosuudet. Kuvaan 1 verrattuna aineistoa on tiivistetty: vain kolme eniten ääniä saanutta on erikseen mukana ja loput niputettu yhteen. Kuva on sinänsä sel-

keä, mutta oleellisinta olisi pohtia, mitä kuvalla halutaan viestiä. Tässähän voisi tarkastelun tiiviistää vain kahteen eniten ääniä saaneeseen ehdokkaaseen, jotka jatkoivat toiselle kierrokselle. Kahden prosenttiluvun esittämiseen kuva tosin alkaisi olla liioittelua – tiedotahan kertoisi lyhyemmin tekstinä. Kuvan niin sanotun tietotiheyden pitäisi olla suurempi.

### Presidentin vaalit 2006

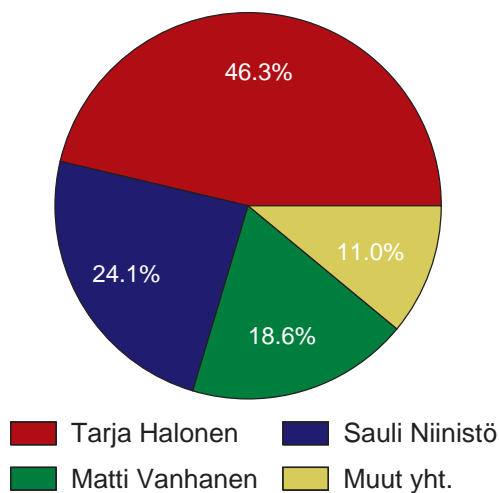
Annettujen äänien lukumäärät 1. kierroksella



Kuva 1. Pylväskuva.

### Presidentin vaalit 2006

Osuudet annetuista äänistä 1. kierroksella



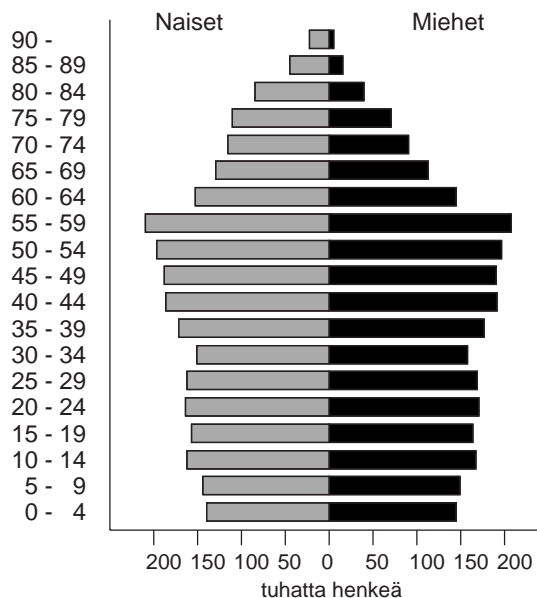
Kuva 2. Piirakkakuva.

Väestöpyramidiksi kutsutussa pylväskuvassa tietotiheys onkin jo huomattava, esittäähän se pienessä tilassa koko väestön ikä- ja sukupuolijakauman. Tottuneelle yksi vilkaisu kuvaan 3 antaa yleiskäsityksen Suomen profiilista ja suhteuttaa sen muihin maihin. Kuvan tarkempi tutkimus puolestaan valottaa yksityiskohtaisemmin miesten ja naisten ikäjakaumien eroja ja yhtäläisyyksiä sekä ikäryhmien välisiä suhteita.

Väestöpyramidissa toteutuvatkin useat tilastollisten kuvien laatukriteerit: informaatiota on monessa tapauksessa, kuva haastaa vertailemaan ja tarjoaa kiinnostuneelle tarkempaa tietoa.

### Suomen väestö iän mukaan

v.2005 lopussa (www.tilastokeskus.fi)



Kuva 3. Väestöpyramidi.

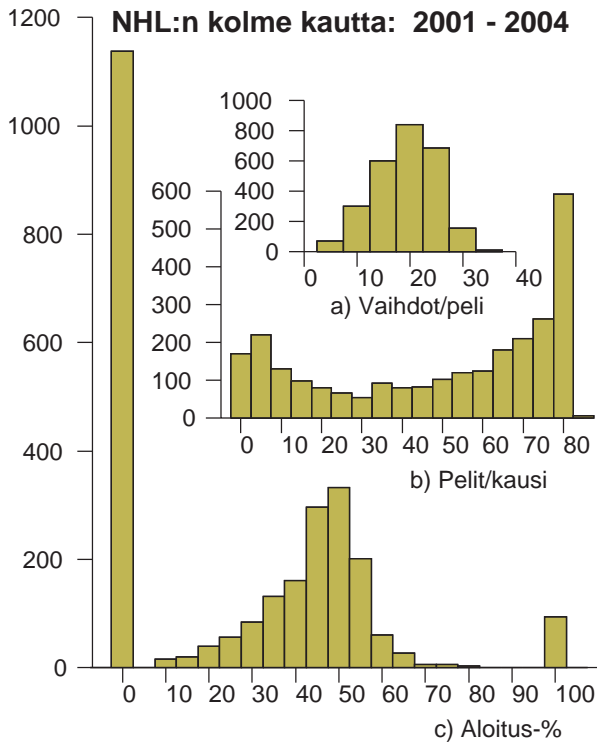
Jatkuvan muuttujan frekvenssijakaumaa esittävää pylväskuvaa kutsutaan histogrammiksi. Kuvaan 4 on tilan säästämiseksi aseteltu sisäkkäin kolme histogrammia näyttää Pohjois-Amerikan jääkiekkoliiga NHL:n lukemattomista tilastoinnin kohteista kolmelta työsulkua edeltäneeltä kaudelta. Kaikki kuvat koskevat pelaajakohtaisia tietoja. Kuva 4 a) kuvaa vaihtojen määrää, siis sitä, kuinka monesti pelaaja on päässyt kentälle pelin aikana. Tyypillisin määrä näyttää olevan 20:n paikkeilla. Pelien määrän jakauma kauden aikana näkyy kuvasta b), ja kuvan c) aloitusprosentti kertoo, kuinka usein pelaaja on onnistunut siirtämään omalle joukkueelleen erotuomarin pudottaman kiekon.

Histogrammin pystyakseli ilmaisee pylvään osoittamaan luokkaan kuuluvien havaintojen lukumäärän. Histogrammeihin liitetään usein normaalijakauman tiheysfunktion kuvaaja, mutta kuten tästäkin nähdään, kaikki asiat eivät ole normaalisti jakautuneita, eivät edes symmetrisiä. Vaihtojen määrän jakauma on lähimpänä normaalijakaumaa, tosin sekin melko vino.

Erikoisin jakauma on aloitusprosentilla: valtava piikki nollassa, sen jälkeen erikseen hyvin vino jakauma välillä 10–70 % ja jälleen toinen piikki kohdassa 100 %. Jääkiekkoa tuntevat keksivät heti selityksen: aloitusprosenttia ei ole mieltä tarkastella kaikkien pelaajien osalta, sillä aloitukset kuuluvat pääasiassa vain keskushyökkääjän tehtäviin. Prosenttien kanssa olisi myös

hyvä ottaa huomioon lukumäärät. Molemmat kuvan piikeistä paljastuvatkin lopulta aika turhanpäiväisiksi.

Myös kuvan 4 b) jakauma on hyvin kaukana normaali-jakaumasta. Jakauman moodiluokan muodostavat ne, jotka ovat pelanneet suurinpiirtein kaikki pelit kauden aikana.



Kuva 4. Histogrammeja.

## Histogrammeista historiaan

Tilastollisten kuvien historia on mielenkiintoinen ja pidempi kuin useimpien tilastollisten menetelmien, jotka on pääosin kehitetty vasta 1900-luvulla. Pylväät ja piirakat keksi n. 200 vuotta sitten skotlantilainen *William Playfair*. Hän oli erikoinen ja ristiriitainen hahmo, jonka ehdotuksiin suhtauduttiin tuolloin varsin kielteisesti. Playfair myös kehitti ahkerasti viivakuvia ja nykyään teemakartoiksi kutsuttuja tilastollisia esityksiä. Pylväiden ja piirakoiden osalta hänen arvellaan saaneen vaikutteita veljeltään *John Playfairiltä*, joka oli matematiikan professori ja perehtynyt *Leibnizin* ja *Eulerin* 1600–1700-luvuilla kehittämiin logiikan diagrammeihin.

Toinen mainittava henkilö 1800-luvulla oli *Florence Nightingale*, jonka saavutukset matematiikan ja tilastotieteen puolella ovat jääneet yleensä vähemmälle huomiolle kuin hänen ansionsa sairaanhoidon saralla. Nightingale keksi kuvata Krimin sodassa menehtyneiden sotilaiden kuolinsyitä napakoordinaatistoon piirretyllä

kuvalla, joka nosti dramaattisella tavalla esiin brittiarmeijan pahimman vihollisen: saniteettiongelman. Nightingalen huomiolla ja ennen kaikkea sen visualisoinnilla oli huomattavia vaikutuksia terveydenhuollon kehittämisessä.

Nightingalen ja Playfairin henkilöhistoriaan ja tilastokuvien voi perehtyä verkossa mm. Wikipedian välityksellä. *Datan Leonardo da Vinciksi* kutsutun tilastotieteen professori *Edward Tuften* teoksiin kannattaa tutustua jo yleissivistyksen vuoksi.

## Valehtelu ja lukutaito

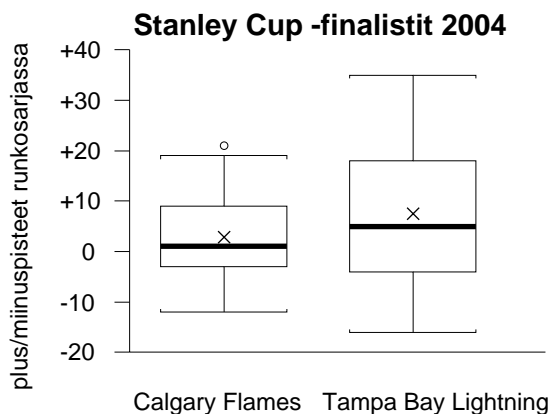
Piirakkakuvista on tullut sittemmin niin sanotun business-grafiikan symboli, ja niihin suhtaudutaan 200 vuotta Playfairin jälkeenkin nuivasti, ainakin yliopistomaailmassa. Tämä johtuu paljolti niiden holtittomasta käytöstä, johon taulukkolaskentaohjelmat suorastaan kannustavat. Kolmiulotteisen oloisilla, sopivaan perspektiiviin leivotuilla piirakoilla saadaan vaatimatonta yritystä näyttämään kilpailijoihin verrattuna markkinajohtajalta. Vääristeltyjen lukujen ja numeroiden esittämistä pidettäisiin valehteluna, mutta samaa johdopäätöstä ei välttämättä osata tehdä vastaavan kuvallisen esityksen perusteella.

Tietämättömiä on muutenkin helppo hämätä kikkaillemalla kuvien mittasuhteilla, mikä koskee yhtä hyvin muitakin kuin piirakoita, myös pylväskuvia. Kaikkien aikojen myydyin tilastotieteen kirja *Kuinka tilastoilla valehdellaan* esitti jo 1950-luvulla tyypilliset tavat joilla lukijaa saatetaan johtaa harhaan tilastokuvien avulla. Yhä ajankohtainen opus kuului aikoinaan tilastotieteen tutkintovaatimuksiinkin. Näin haluttiin varmistaa että ainakin opiskelijat osaisivat suhtautua tilastokuvien kriittisesti ja laatia niitä sortumatta tyypillisimpiin virheisiin.

## Viivat ja pisteet

Siitä lähtien kun tietokoneilla on voitu tehdä asiallista grafiikkaa, ovat erityisesti amerikkalaiset tilastotieteilijät *John Tukey*, *William Cleveland* ja *John Chambers* kehittäneet uusia tapoja tilastollisten aineistojen visualisointiin. Yksi on kuvan 5 laatikkokuva, jota Tukey ehdotti eksploraatiivisen data-analyysin työkaluksi 1970-luvulla.

Kuvatyyppi on vähitellen tullut tunnetummaksi, kun yhä useammat ohjelmistot ovat sisällyttäneet sen valikoimiinsa. Silti on monia jotka eivät ole tätä esitystä ennen nähneet. Uudempien kuvatyyppien vaarana on, että kuva ei tule lainkaan ymmärretyksi. Kunnolliset selitykset ovat tarpeen.



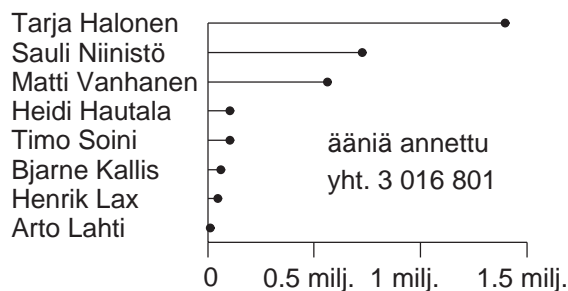
Kuva 5. Laatikkokuva.

Laatikkokuva esittää jatkuvan muuttujan jakautuman histogrammia tiivistetympin. Paksu viiva kuvaa järjestetyn aineiston keskimmäistä lukua mediaania, ja laatikko sen ympärillä piirretty vastaaviin 25 %:n ja 75 %:n kohtiin jättäen näin puolet havainnoista laatikon sisään. Suurin osa muista havainnoista kuvautuu välinä molempiin suuntiin laatikosta, ja vain poikkeavimmat havainnot piirretään yksittäin näiden ulkopuolelle. Kuvassa 5 on lisäksi havainnollistettu keskiarvot rasteina.

Laatikkokuva on tehokkaimmillaan, kun se piirretään jonkin luokittelevan muuttujan luokille rinnakkain. Kuva 5 esittää NHL:n *Stanley Cup* -pokaalista 2004 pelanneiden kahden joukkueen eroja pudotuspelejä edeltäneessä runkosarjassa. Pystyakseli kertoo joukkueen pelaajien yhteenlasketut plus/minus-pisteet. Pelaaja saa kentällä ollessaan plussan kun oma joukkue ja miinuksen kun vastustaja tekee maalin. Pelaajan positiivinen saldo osoittaa hänen pelaavan joukkueensa hyväksi, negatiivinen viestii ettei joukkuepeli oikein suju. Kuva paljastaa, että pokaalin vienyt Tampa Bay oli tässä suhteessa parempi.

## Presidentin vaalit 2006

Annettujen äänien lukumäärät 1. kierroksella

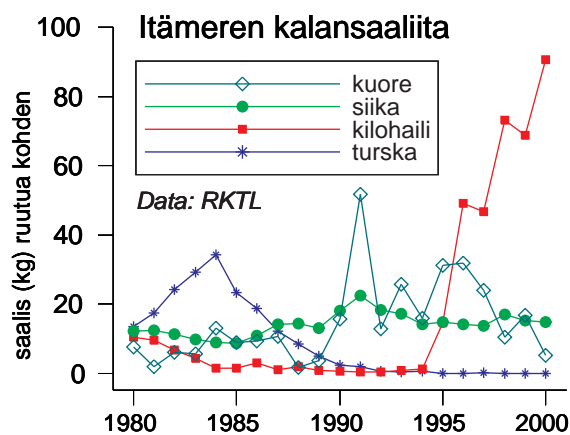


Kuva 6. Pistekuva.

*Naomi Robbins* on tuoreessa kirjassaan ansiokkaasti nostanut monia Tukeyn ja erityisesti Clevelandin

jo aiemmin ehdottamia kuvatyyppejä uudelleen esille. Yksi näistä on pylväskuvan vaihtoehdoksi tarjottu, visuaalisesti keveämpi pistekuva, jossa pylväät on korvattu pisteillä ja viivoilla. Kuva 6 esittää tällä tekniikalla oleellisesti samat tiedot kuin kuva 1, hieman pienemmässä koossa ja kokonaan mustavalkoisena. Säännönmukaisuudet hahmottuvat helpommin pistekuvasta kuin pylväskuvasta, erityisesti suuremmilla aineistoilla.

Perinteisemmin viivoja käytetään aikasarjojen kuvaamisessa. Kuva 7 on tyypillinen viivakuva, joka kertoo Itämeren kalakannoissa ajassa tapahtuneista muutoksista. Aineiston keruu ja tulosten käyttötarkoitus ratkaisevat, miltä aikaväliltä ja miten tiheästi tietoja kuvassa esitetään. Tässä tiedot vuosilta 1980–2000 on esitetty vuoden tarkkuudella, vaikka ne onkin alunperin mitattu tiheämmin.



Kuva 7. Viivakuva.

Kuvassa 7 on neljä aikasarjaa, jotka on piirretty eri väreillä. Värit eivät ole välttämättömiä, sillä joka sarjalla on myös erilainen piste havainnon kohdalla. Pisteet ja viivat on selostettu omassa laatikossaan, joka on sijoitettu kuva-alueen tyhjään tilaan ja näin saatu kuva selityksineen mahtumaan verrattain pieneen kokoon.

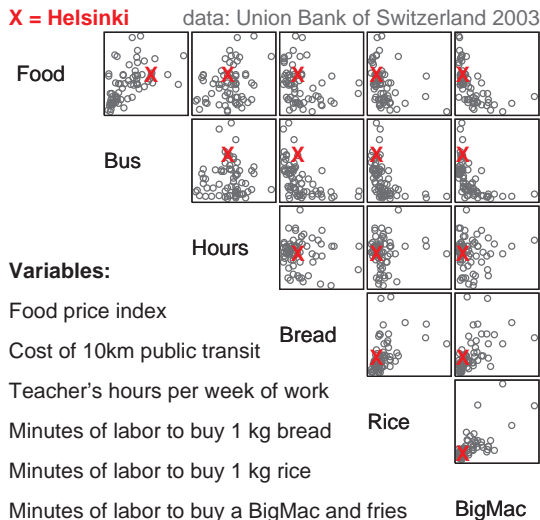
## Hajontakuvat

Vielä pienempään kokoon tietoa on tiivistetty kuvan 8 hajontakuvamatriisissa. Tässä esillä on vain muuttujien parittaisten hajontakuvien muodostaman matriisin yläkolmio. Lävistäjällä ovat muuttujien nimet, ja alakolmion tila on käytetty selityksiin. Aineiston kuvaamista kaupungeista erottuu punaisella rastilla piirretty Helsinki.

Hajontakuvamatriisi tarjoaa hyvän yleisnäkymän aineistoon: se paljastaa sekä säännönmukaisuudet että poikkeamat säännönmukaisuuksista. Muuttujien välisten riippuvuuksien luonne, mahdolliset

epälinearisuudet ja poikkeavat havainnot tulevat äkkiä esille. Tarkemmin niihin on parasta syventyä piirtämällä parittaisia hajontakuvia erikseen kiinnostavista muuttujista. On syytä muistaa, että hajontakuvamatriisi ei ole moniulotteinen kuva vaan kokoelma tavallisia kaksikulotteisia kuvia.

**Economic indicators for 69 world cities**

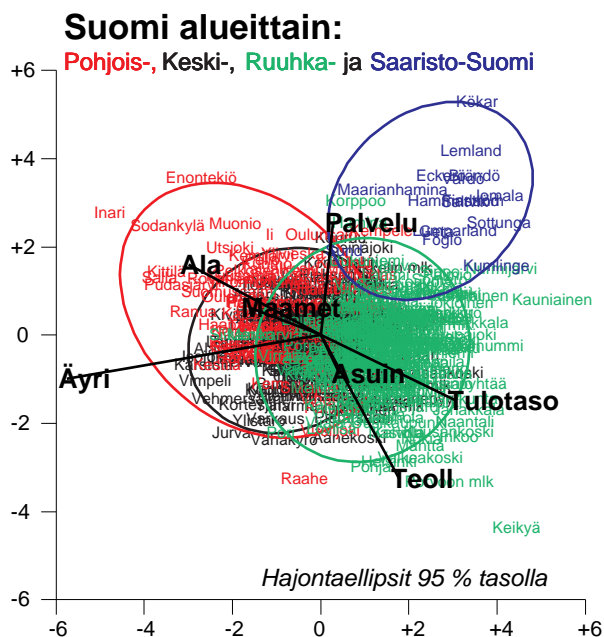


Kuva 8. Hajontakuvamatriisi.

Tilastollisessa tutkimuksessa hajontakuvat ja muut pistediagrammit ovat yleisempiä kuin pylväät ja piirakat. Etenkin monimuuttujamenetelmillä on tyypillistä pyrkiä tiivistämään useampiulotteisten ilmiöiden välisiä suhteita ja kuvailemaan niitä helpommin tulkittavina kaksikulotteisina hajontakuvan muunnelmoina. Kolmiulotteiset esitykset eivät käytännössä auta paljoakaan, sillä reaali maailman ilmiöt ovat joka tapauksessa moniulotteisempia.

Kuva 9 visualisoi erotteluanalyysia, jossa tutkitaan mikä erottaa tunnetut ryhmät toisistaan. Suomen kunnista muodostetut maantieteelliset alueet eroavat toisistaan mm. elinkeinoprofiileiltaan, vauraudeltaan ja pinta-alaltaan. Kuva on samalla esimerkki kaksoiskuvasta: samaan koordinaatistoon on piirretty sekä havainnot että muuttujat. Erotteluavaruudeksi kutsuttu kuva avaa useita näkymiä tutkittavaan aineistoon.

Hajontaellipsit auttavat hahmottamaan ryhmien muotoa ja sijoittumista toisiinsa nähden. Ryhmien keskivaiheilla ja siten myös koko kuvan keskellä on eniten ruuhkaa, eikä sieltä ole tarkoituskaan erottaa yksittäisiä kuntia. Kiintoisampia ovat ryhmien reunoilla esiintyvät ääritapaukset, kuten teollisuusvaltaiset Keikyä ja Raahen kooltaan valtavat Inari ja Sodankylä, poikkeuksellisen vauras Kauniainen ja palveluun keskittynyt Kökar. Korppoo kuuluu Ruuhka-Suomeen, mutta sen elinkeinoprofiili on samankaltainen kuin Saaristo-Suomen kuntien.



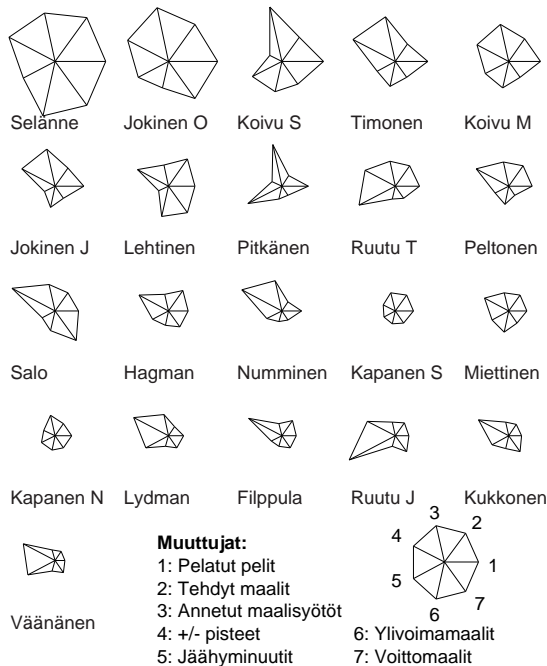
Kuva 9. Erotteluavaruus

**Profilikuvat**

Edellä on viitattu profileihin niin erotteluavaruuden kuin väestöpyramidinkin yhteydessä. Profiilien tunnistaminen ja vertailu onkin hyvä esimerkiksi säännönmukaisuuksien hahmottamisesta. Palataan hetkeksi NHL-teemaan ja tarkastellaan kuluneen kauden parhaiden suomalaispelaajien profileja tilastomerkintöjen valossa.

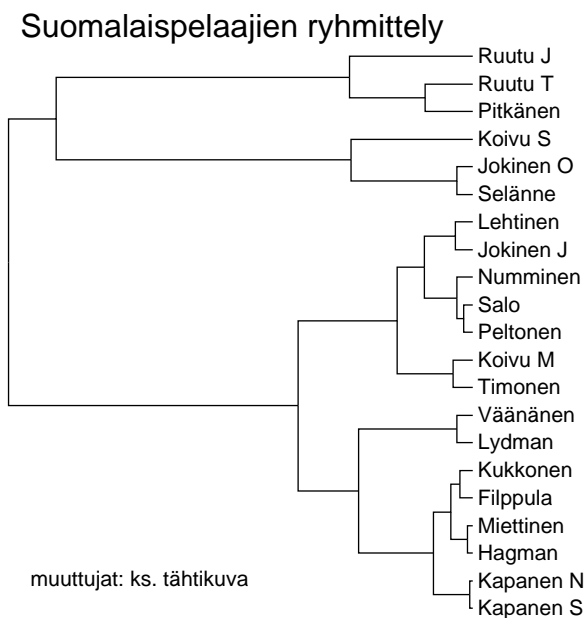
**Suomalaistähnten pelaajaprofiilit**

NHL:n runkosarja 2006-2007, yli 50 ottelua pelanneet



Kuva 10. Tähtikuva.

Kuvan 10 tähtikartta esittää 21 pelaajan profiilit seitsemän muuttujan suhteen. Kuvaa tarkastelemalla on helppo huomata samankaltaisia pelaajaprofileja, esimerkiksi Ruudun veljekset kunnostautuvat samoilla pelin osa-alueilla ja Kapasilla on muutakin yhteistä kuin nimi.



Kuva 11. Dendrogrammi.

Tähtikuva on yksi lukuisista edellä mainittujen tilastotieteilijöiden kehittämistä kuvatyypeistä, jotka eivät ole hyödyllisyydestään huolimatta yleistyneet. Sen sijaan biologian sovellusten puolelta periytyvä puukuva, jota kutsutaan dendrogrammiksi, on vakiintunut hierarkisen ryhmittelyanalyysin tulosten esitystavaksi.

Kuvassa 11 pelaajatähdet on ryhmitelty tähtikuvan muuttujien perusteella. Kuvien avulla selviää, että Ruutujen ohella myös Pitkänen on istunut runsaasti jäähdyllä. Kolme tehokkainta pelaajaa (*joista kukaan ei valitettavasti ole mukana kevään 2007 MM-kisoissa*) erottuu omana ryhmänään, ja lopuista muodostuu pari isompaa ryhmää. Kapaset ovat todellakin pelaajaprofileiltaan lähimpänä toisiaan.

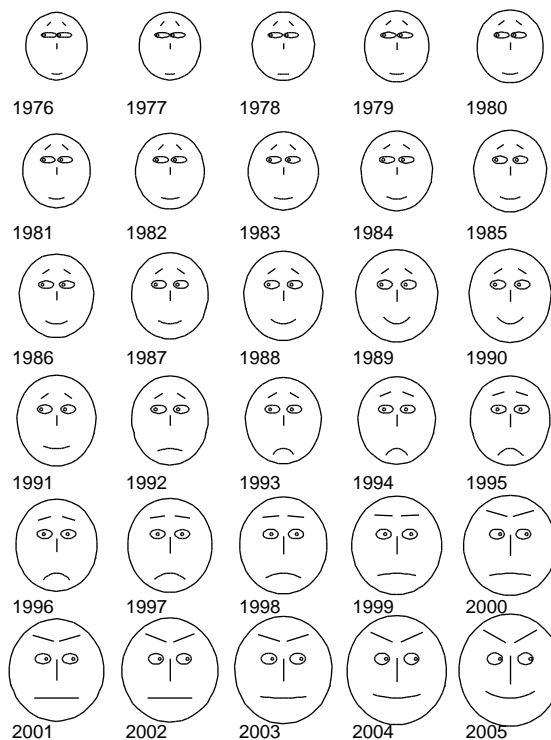
Profiilien hahmottamiseen on lukuisia muitakin keinoja. Ehdottomasti ilmeikkäimmän tavan lukujen ja numeroiden kuvaamiseen on esittänyt *Herman Chernoff*, niin ikään tilastotieteen professori. Chernoffin naamoina tunnettu tekniikka perustuu ihmisen ilmiömäiseen kykyyn tunnistaa muita ihmisiä kasvopiirteiden perusteella. Chernoffin alkuperäiseen ehdotukseen sisältyy kaikkiaan 18 kasvopiirrettä, joita voi varioida aineiston muuttujien perusteella.

Kuvan 12 naamat heijastelevat viiden keskeisen taloudellisen aikasarjan kehittymistä 30 vuoden aikana. Aikasarjat on kytketty kasvopiirteisiin, joista osa on

kokonaisilmeen kannalta näkyvämpiä kuin toiset. Esimerkiksi työttömyysaste vaikuttaa suun kaarevuuteen, bruttokansantuote (BKT) naaman kokoon ja suun leveyteen, tuonti nenän pituuteen ja vienti katseen suuntaan.

## Suomen kansantalous 1976 - 2005

BKT, tuonti, vienti, työttömyysaste ja kulutusmenot



Kuva 12. Chernoffin naamat.

Ilmeistä on helppo havaita 1990-luvun alun laman draamaattinen vaikutus. Piirsin kuvan ensimmäisen keran 1990-luvun lopulla, ja päivitin sen vasta hiljattain muuttamatta määriksi. Hymy oli vihdoin palannut, mutta ilme ei mielestäni ole aivan terve. Ken haluaa, vetäköön tästä mielenkiintoisia johtopäätöksiä hyvinvointiyhteiskunnan tilasta.

Chernoffin naamakuva on joka tapauksessa erikoislaatuinen, voidaanhan sillä aidosti kuvata useampia asioita yhtäaikaan. Toisaalta kasvokuvat jo sinällään ja sopivien kasvopiirteiden valikointi herättävät herkästi kysymyksen subjektiivisuudesta. Hienosta ideasta ja eräistä näyttäivistä sovelluksista huolimatta naamakuva on jäänyt enemmänkin kuriositeetiksi. Kuvan 10 kuviot ovat kieltämättä neutraalimpi tapa kuvata havaintojen tai havaintoryhmien profileja.

## Lopuksi

Ei riitä että tilastotiedettä ja matematiikkaa opiskelleet olisivat perillä tässä jutussa kuvatuista asioista. Ti-

lastollisten kuvien avulla esitetään niin paljon yhteiskunnan tilaa kuvaavia tietoja ja tutkimustuloksia, että kuvien perusteiden ja tulkinnan ymmärtämisen pitäisi kuulua tietoyhteiskunnassa kansalaistaitoihin.

Usein pelkkä kuva ei riitä vaan informaation tiivistämiseksi tarvitaan myös erilaisia tilastollisia menetelmiä. Tilastolliset kuvat olisikin hyvä piirtää mahdollisimman julkaisuvalmiiksi samassa ympäristössä, jossa aineistoa muutenkin käsitellään. Tehokas työskentely edellyttää, että aineistojen siirtelyt analyysi- ja kuvanpiirto-ohjelmien välillä minimoidaan. Toinen hyvä periaate on, että kuvat piirretään ohjelmallisesti, ei siis hiirellä klikkaillen. Tällöin työvaiheiden toistaminen on yksinkertaisempaa, virheet on helppo korjata ja työ voidaan tarvittaessa automatisoida.

Tämän jutun kuvat olen piirtänyt Survo-ohjelmiston PLOT-toimintoja käyttäen PostScript-muotoon, jolloin ne saa suoraan mukaan mm.  $\text{\LaTeX}$ -dokumentteihin. Olenkin kirjoittanut jutun valmiiksi Survossa hyödyntäen sen  $\text{\LaTeX}$ -liittymää. Kuvissa ei ole yhtään käsin tehtyä kohtaa vaan kaikkia yksityiskohtia säädetään kuvanpiirtokaavioissa annetuilla täsmennystiedoilla. Voin milloin tahansa piirtää kaikki kuvat uudelleen yhdellä napinpainalluksella. Aineistojen vaihtuessa uudet kuvat on kätevä ja nopea laatia valmiita pohjia hyödyntäen. Tällaiset ominaisuudet ovat tarpeen muuallakin kuin jutun alussa mainitun *Num3rot*-sarjan rikostutkinnassa, sillä ylimääräistä aikaa ei tunnu enää olevan lainkaan.

## Viitteet

1. Num3rot: [www.mtv3.fi/num3rot/](http://www.mtv3.fi/num3rot/)

2. Numb3rs: [www.cbs.com/primetime/numb3rs/](http://www.cbs.com/primetime/numb3rs/)
3. Wikipedia: a multilingual, free content encyclopedia, <http://en.wikipedia.org/>
4. Herman Chernoff, The Use of Faces to Represent Points in K-Dimensional Space Graphically, *Journal of the American Statistical Association* 68:342, 1973, 361–368.
5. William S. Cleveland, *Visualizing Data*, Hobart Press, Summit, New Jersey, 1993.
6. J. C. Gower and D. J. Hand, *Biplots*, Chapman & Hall, London, 1996.
7. Darrell Huff, *Kuinka tilastoilla valehdellaan*, alkup. *How to Lie with Statistics* (1954), Otava, 1974.
8. Vesa Kuusela, *Tilastografiikan perusteet*, Edita, Helsinki, 2000.
9. Seppo Mustonen, SURVO MM: käyttöympäristö tekstin ja numeerisen tiedon luovaan käsittelyyn, [www.survo.fi](http://www.survo.fi).
10. Seppo Mustonen, *Survo ja minä*, Survo Systems, Helsinki, 1996.
11. Naomi B. Robbins, *Creating More Effective Graphs*, Wiley, New York, 2005.
12. Ian Spence, No Humble Pie: The Origins and Usage of a Statistical Chart, *Journal of Educational and Behavioral Statistics* 30, 2005, 353–368.
13. Edward R. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut, 1983.