

Kalevi Rantanen

Professori Rosling näyttää tilastojen kauneuden

Hans Roslingin vauhdikkaat animaatiot tekevät numerodatasta ymmärrettävää, jopa jännittävää.

Edessä on pieni pöytä. Sen päällä on valkoinen liina, joka peittää joitakin esineitä. Olemme Aalto-yliopiston Inspire-auditoriossa, mutta asetelma tuo mieleen taikurin teltan entisajan maalaismarkkinoilla.

Monenlaista rekvisiittaa tarvitaan, kun ruotsalainen **Hans Rosling**, tilastotieteilijä, lääketieteilijä, puhuja, datavisionääri ja Karoliinisen instituutin professori, esiintyy Open Knowledge-festivaaleilla Helsingissä.

Rosling näyttää aluksi vauhdikasta animaatiota. Erikokoiset pallot ryntäävät näytön oikeasta yläkulmasta kohti vasenta alakulmaa. Pallot esittävät maailman maita. Vaaka-akselille on kirjattu lapsiluku synnyttäjää kohti. Pystyakselilla näkyy lapsikuolleisuus.

Pallojen juoksu ruudulla kertoo kehityksestä kohti kaksilapsisia perheitä, kaikkialla maailmassa.

Rosling kommentoi tapahtumia inostuneesti kuin jalkapallo-ottelun seostaja. Sitten hän menee pöydän luo. Liinan alta paljastuu kanien ja kyyhkysten asemesta vessapaperirullia. Yksi rulla esittää miljardia ihmistä. Paperirullia siirtelemällä Rosling havainnollistaa maapallon väestökehitystä.

Jättiauikko tilastotiedoissa

Hans Rosling käyttää itsestään myös titteliä *edutainer*, koulutusviihdyttäjä. Hän opiskeli 1960- ja 1970-luvulla sekä tilastotiedettä että lääketiedettä ja työskenteli sitten parikymmentä vuotta terveydenhuollon kehitystehtävissä Afrikassa. Siellä hän löysi uuden halvaannuttavan sairauden, konzon. Rosling löysi myös taudin syyt: köyhyyden ja huonosti käsitellyn kassavan.

Palattuun Ruotsiin opettamaan globaalia terveydenhuoltoa Rosling havaitsi, että tilastotietoa on vaikea välittää yleisölle.

”Parhaat ruotsalaiset opiskelijat ymmärsivät yhteiskunnallisia tilastoja huomattavasti heikommin kuin simpanssit”, hän kuvaa tilannetta.

Yksinkertaisissa valintehtävissä ruotsalaisopiskelijoiden tiedot olivat näin huonommat kuin satunnainen arvaus, mihin apinatkin kykenevät.

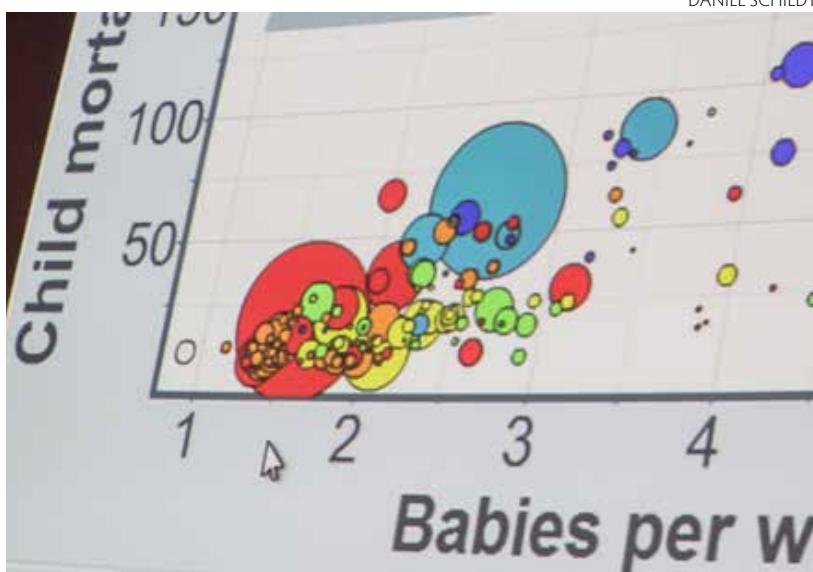
Julkista, havainnollista, liikkuvaa

Rosling keksi ratkaisun ongelmaan. Hän hyödynsi nykyaikaista tietotekniikkaa ja yhdisti tilaston, kuvan sekä liikkeen. Tilastojen vapautuminen vähitellen kaikkien käyttöön mahdollisti uuden tekniikan laajan hyödyntämisen ja tulosten julkaisemisen.

Vuonna 2005 Rosling perusti yhdessä poikansa ja miniänsä kanssa voittoa tuottamattoman Gapminder-organisaation. Sen pyrkimys on levittää tilastotietoa kaikille.

Gapminderin sivulla tietoa visualisoidaan kaksiulotteisilla piirroksilla. Pystyakselilla voi olla esimerkiksi elinajan odote ja vaaka-akselilla tulot henkeä kohti. Maailman maat puolestaan on esitetty erikokoisilla ja värisillä palloilla. Kun siirtyilee aikajanalla, saa näkymään tilanteen tietynä vuonna. Ja videonappulaa painamalla muutoksen saa esiin animaationa. Gapminder sisälsi syyskuun lopussa jo 361 tilastoindikaattoria.

Kirjoittaja on vapaa tiedetoimittaja.



Erikokoiset, maailman maita esittävät pallot ryntäävät kohti näytön alakulmaa Gapminderin lapsikuolleisuutta esittävissä animaatioissa.

"Odotetaan lapsen kuolemaa"

Gapminder-tilastot vähentävät numeroitaidottomuutta, mutta mitä tehdä muulle faktavastaisuudelle kuten auktoriteettiuskalle ja asenteille?

Hans Rosling on muuttanut tapamme katsoa tilastoja. Gapminder on jopa pelottavan hyvä työkalu. Tulee houkutus ajatella, että uusi väline on yleisratkaisu, Arkhimedeen vipu, jolla väännetään koko maailma uuteen asentoon.

Gapminder vähentää tehokkaasti numerotaidottomuutta, eli tottumattomuutta ja haluttomuutta käsitellä lukuja. Mutta entä jos vastassa on muita esteitä, kuten vahvoja tunteita ja auktoriteettiuskoa?

Helsingissä Rosling hurmasi yleisön murskaamalla myyttejä suurista perheistä kehittyvässä maailmassa. Mutta olisiko taputuksia tullut yhtä paljon, jos hän olisi käsitellyt geenimuutteleua, ydinvoimaa tai rokotuksia?

Voimme osoittaa Gapminder-tyylisellä animaatiolla esimerkiksi, että rokotukset ovat pelastaneet miljoonia ihmishenkiä ja aiheuttaneet vahinkoa vain muutamille harvoille. Silti jotkut ihmiset vastustavat rokotuskampanjoita kiivaasti.

Kantapään kautta?

Luentosalissa ei ollut mahdollista tarttua rokotusteemaan, mutta sähköpostikirjeenvaihdossa Rosling kommentoi asiaa. Mitä siis tehdä, jos ihminen torjuu rokotustiedon, vaikka selkeästi ja kauniisti esitettyä?

"Odotetaan, kun vastustajan perhe sairastuu ja hänen lapsensa kuolee", Rosling lohkaisee vastaukseksi. "Silloin vastustaja kyllä muuttaa suhtautumistaan."

Varmaan ihmiskunta oppii lopulta kantapään kautta, mutta olisiko muitakin keinoja saada tietoa perille?

Psykologi ja juristi **Dan M. Kahan** Yalen yliopistosta ja joukko muita tutkijoita selostivat viime keväänä *Nature Climate Change* -lehdessä mittaus- tuloksia tieteellisen lukutaidon vaiku-

tuksesta käsityksiin ilmastonmuutoksen riskeistä. He olivat havainneet, ettei numerotaidon ja tieteellisen lukutaidon paraneminen välttämättä lisää luottamusta ilmatieteellisiin tietoihin. Ihmiset pikemminkin pitäytyvät entistä tiukemmin mielipiteissään.

Suhtautuminen tietoon riippui sen sijaan voimakkaasti siitä, mitä ihmiselle tärkeä yhteisö ilmastonmuutoksesta ajatteli.

Kahan ja kumppanit kehottavat tiedeviestijöitä luomaan neuvottelevan ilmapiiriin ja osoittamaan, että tieteellinen tieto on sopusoinnussa kullekin yleisölle tärkeiden arvojen kanssa.

Ehkä rokotuskeskustelussakin olisi tilastotietojen ohella hyvä nostaa esille myös reilun pelin kaltaisia arvoja. Voisi kysyä vaikka, onko reilua nauttia laumasuojasta ja jättää riskit muille.

Kahanin tutkimusryhmä ehdottaa myös yleisökohtaisesti räätälöityjä tiedottamisen tapoja. Tilastodatan uudetkin esitystavat toimivat selvästi

parhaiten, kun ne yhdistetään muihin työkaluihin.

Lisää aiheesta:

Hans Rosling, Data Journalism and Data Visualization, Open Knowledge Festival, Helsinki, syyskuu 2012
<http://lokfestival.org/streams/inspire-foyer-1/>

Dan M. Kahan, Ellen Peters, Maggie Wittlin, Paul Slovic, Lisa Larrimore Ouellette, Donald Braman, Gregory Mandel. The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature Climate Change, 2012; doi:10.1038/nclimate1547
<http://www.nature.com/nclimate/journal/vaop/ncurrent/full/nclimate1547.html>

Open Knowledge Festival, joka tänä syksynä pidettiin Helsingissä, juhlii ja edistää nimensä mukaisesti tiedon avoimuutta. Pääjärjestäjät olivat Suomen Lontoon instituutti, Open Knowledge Foundation ja Aalto Media Factory.

DANIEL SCHILDT



Hans Rosling havainnollistaa väestönkasvua vessapaperirullien avulla.

Minna Ruckenstein

Big datan suuri lupaus – numerot kertojina

Piilaaksossa visioidaan tietokoneiden ja matkapuhelinten tuottamien valtaviin tietomassojen analyysistä, joka vie uuteen läpinäkyvään maailmaan. Kriitikot pelkäävät valtioiden ja suuryritysten tiukkenevaa otetta.

Stanfordin yliopiston kampuksella kohkataan big datasta. Quantified Self -konferenssiin kerääntyneet sadat tutkijat, hakkerit, yrittäjät ja uteliaat heittelevät viittauksia valtaviin luokittelemattomiin tietomassoihin, joita kertyy yhä kiihtyvään tahtiin. Älypuhelimet tuottavat jatkuvaa aineistoa liikkumisesta ja ihmisten tekemisistä. Harva ymmärtää mitä kaikella aineistolla pitäisi tehdä tai mitä siitä pitäisi saada irti. Silti keskusteluissa vallitsee laaja yhteisymmärrys: Big data on kiinnostavaa.

Big datalla viitataan valtaviin luo-

kittelemattomiin tietomassoihin, jotka ovat saattaneet meidät kohti uutta aikakautta. Tavoitteena on maksimoida laskennallinen voima ja algoritmitarkkuus, jotta yhä laajempia ja monimuotoisempia tietovarantoja saataisiin analysoitua ja vertailtua. Päämääränä on, että tiedonkeruu ja tietojenkäsittely olisi reaaliaikaista ja jatkuvaa. Arki ei pyörisi ilman niitä.

Analysoitava aineisto voi olla melkein mitä vain, esimerkiksi potilastietoa, paikkatietoa, sähkönen menekkiä, luottokorttitapahtumia, nettisivujen lokitiedostoja tai sosiaalisen median

sisältöä. Big datan toivotaan auttavan terveydenhuollossa, esimerkiksi syövän ennakkoinnissa ja hoitojen kohdentamisessa. Lisäksi aineistoista toivotaan tukea muun muassa liikennesuunnitteluun, kriminaalipolitiikkaan, ruoantuotantoon ja ilmastonmuutoksen hidastamiseen.

Uudenlaista yrittämistä ja täsmällisyyttä

Keskusteluissa kierrätetään usein samoja esimerkkejä, tulvien ja flunssa-aalto-

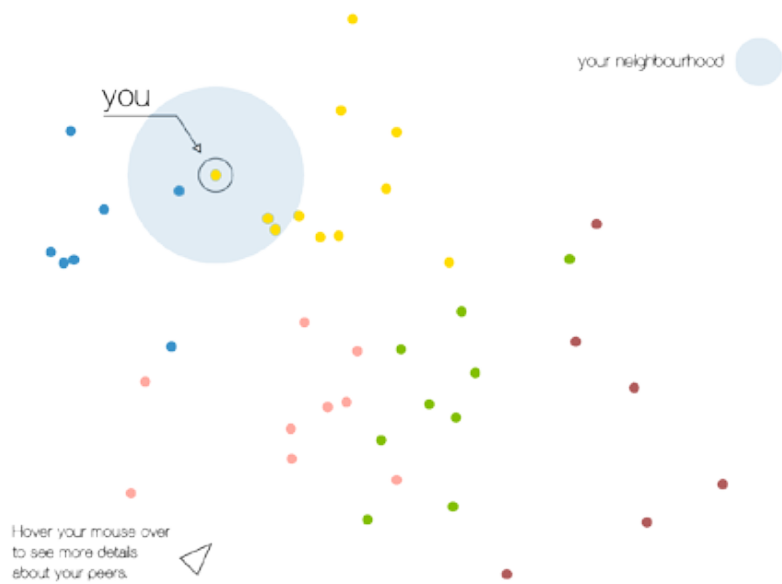
Kiehtovaa big datassa on sen lupaus tuottaa uusia näkökulmia.

Suurten aineistojen avulla voi numeraalisesti kertoa mistä tulemme, mihin pyrimme ja miten tekemisiimme voi vaikuttaa.

STRESS MAP

NəRV

Your peers positioned according to their similarity with you and each other. Additional grouping by colours helps you to distinguish between different peer groups.

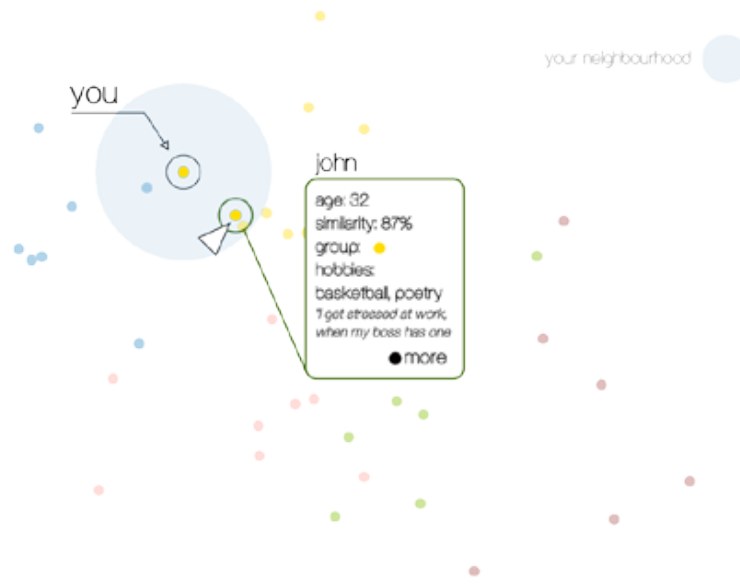


Kuvat ovat Antti Heikkilän Aalto-yliopistoon tekemästä diplomityöstä Information Visualisation in a Peer Support Application. Aineisto koostuu ihmisen vastauksista nettissä julkaistuun stressikyselyyn. Tutkimus on osa Krista Laguksen johtamaa VirtualCoach-projektia.

STRESS MAP

NeRV

Your peers positioned according to their similarity with you and each other.
Additional grouping by colours helps you to distinguish between different peer groups.



Toisaalta suuret aineistot ovat alttiita inhimillisille vääristymille. Sosiaalisessa mediassa kierrätetään tutkimusartikkeliä, joka todistaa korrelaatiota pörssin S&P indeksin ja bangladeshilaisen voituotannon välillä.

jen ennakoinnista (google.org) ruotsalaisen professori **Hans Roslingin** tilastoihin pohjaviin visualisointeihin (gapminder.org). MITin Media Lab taas keräsi kännyköiden paikkatietoa, erityisesti parkkipaikoilla, ja ennusti sen perusteella tulevaa joulukauppaa.

Tiedonanalyysin ympärille on nopeasti syntynyt uudenlaisia yrityskäytäntöjä. Lentokentän logistiikka on saatu aiempaa sujuvammaksi, ja kulluttajaa voidaan puhutella entistä kohdistetummin. Verkon kauppapaikoilla ja palveluissa ihmisten käyttäytymistä tutkitaan lokitiedostojen perusteella. Näin saadaan tietoa ihmisten tiedontarpeista tai esimerkiksi siitä, milloin he kirjautuvat sisään verkkopalveluun, kuinka kauan he siellä viihtyvät ja mitä he ostavat.

Julkinen sektori on erityisen kiinnostunut big datan mahdollisuuksista sysätä liikkeelle uudenlaista kasvutaloutta. Yhteisiä tietovarantoja jakamalla pyritään edistämään kansalaisaktiivisuutta ja yritystoimintaa. Suomessa Maanmittauslaitos on avannut maastotietoa-aineistonsa kansalaisten ja yritysten vapaaseen käyttöön. Avatussa aineistossa on esimerkiksi karttoja, ilmakuvia ja korkeusmalleja.

Taloudellisen hyvinvoinnin ohella suurten aineistojen ajatellaan tuottavan yhteiskunnallisesti aiempaa osuvampia

valintoja ja poliittista ohjausta. Big datan uskotaan lisäävän läpinäkyvyyttä päätöksenteossa. Tietomassoja analysoimalla on mahdollista nähdä asioita, joita ei muuten nähtäisi. Näin big data tarjoaa totuuksia ja täsmällisyyttä.

Microsoftilla tutkimusta tekevä, teknologiatutkimuksen vaikuttaja **danah boyd** (hän kirjoittaa nimensä tarkoituksellisesti pienillä alkukirjaimilla) on listannut kriittisiä näkökulmia big dataan. Kriitikot näkevät yksityisyyden ja tietoturvan loukkauksia, valtioiden ja suuryritysten yhä tiukempaa otetta kansalaisista ja kuluttajista. Totalitarismin uhkaakin on jo välätelty.

Numeroiden laatu

Big datan ehkä kiehtovin piirre on sen lupaus tuottaa uusia näkökulmia. Suurten aineistojen avulla voi piirtää esiin ajallisia ja paikallisia yhteyksiä ja riippuvuuksia. Aineisto voi tehdä ehdotuksia ihmisten toiminnasta, heidän välisistään verkostoista ja yhteistyöstä. Näin big datalla on valtaa numeraalisesti kertoa meille mistä tulemme, mihin pyrimme ja miten tekemisiimme voisi vaikuttaa. Terveystieteissä ja vakuutusyhtiöissä tätä pidetään valtavana mahdollisuutena. Aineistot voivat auttaa kroonisten sairauksien ennalta-

ehkäisevässä hoidossa tai erotella onnettomuuksiin johtavat kehityskulut jo ennen kuin ne tapahtuvat.

Big data -analyysien yleistyessä yhä polttavampi kysymys on, miten ja mihin mitattavaa ja luokiteltavaa tietoa käytetään. Kun yritykset etsivät säännönmukaisuuksia ihmisten käyttäytymisestä, ne saavat myös uudenlaisen otteen arjesta. Sosiaalinen media, älypuhelimien sovellukset ja arkiympäristöön istutetut sensorit ja mittalaitteet toimivat aineistonkerääjinä ja välittäjinä, jotka osallistuvat uusien alueiden ja ilmiöiden näkyväksi tekemiseen ja tunnistamiseen.

Toisaalta suuret aineistot ovat yhtä lailla alttiita inhimillisille vääristymille ja painotuksille kuin muukin tiedontuotanto. Aineistomassojen tulkinnassa virheiden mahdollisuus on todennäköisempi, koska niissä voidaan nähdä järjestäviä periaatteita, joita ei tosiasiassa ole olemassa. Sosiaalisessa mediassa on kierrätetty tutkimusartikkeliä, joka todistaa tiedonlouhinnan tekniikoiden esittävän vahvoja korrelaatioita pörssin S&P indeksin ja bangladeshilaisen voituotannon välillä.

Suuret aineistot voivat epäilemättä myös latistaa käsitystä ihmisestä ja hänen sosiaalisista aikeistaan. Kontekstit ja kiinnostava variaatio katoaa, kun esitetään yhteisiä nimitäjiä.

Harvard Business Review julistaa lokakuun numerossaan aineistoanalyysin vuosisadan seksikkäimmäksi työksi.



Numerot eivät ole neutraaleja vaan niillä on oma laatunsa riippuen siitä, mitä niillä halutaan tehdä ja saavuttaa. Jotta big datan laskennallista voimaa ei käytettäisi väärin, se täytyy alistaa myös eettiselle tarkastelulle. Kun tietotekniikan osaajat siirtyvät ihmistutkimuksen kentille, heille on tärkeää opettaa tiedontuotantoon liittyviä valta- ja vastuukysymyksiä.

Käytännöt ratkaisevat

Big dataa ei voi hallita, jos ei ymmärrä määrällisiä aineistoja. Näin se tuottaa uudenlaisia asiantuntija-asemia, kun osa tiedontuottajista osaa järjestää ja lukea numeroita ja toiset eivät. *Harvard Business Review* julistaa lokakuun numerossa aineistoanalyysin vuosisadan seksikkäimmäksi työksi. Suuriin aineistojen liittyvien kehityskulkujen seuraaminen ja dokumentoiminen on tärkeää, jotta opimme tunnistamaan mitä se todella sysää liikkeelle. Julkisten aineistojen avaamista kannattaa myös seurata erityisen tarkkaan.

Piilaakson big data -keskustelut ovat ajoittain jopa hurmoshenkisiä. Suhteellisuudentajun puute vie ilmiöltä laajempaa uskottavuutta. Suurten aineistojen ajatellaan vievän uuteen maailmaan, jossa aineistot tuottavat

täydellisen läpinäkyvyyden. Suuryritykset tekevät hyvää. Ihminen muutuu järkeväksi ja vastuulliseksi. Hän tekee vain tietoon perustuvia valintoja.

Piilaaksoilaisesta näkökulmasta muutoksen voi kuitenkin aistia. Yrittäjät ja tutkijat ovat valjastaneet kaiken tarmonsä siihen, että jäljittämisteknologiat ja niiden avulla tuotetut aineistot muuttavat maailmaa paremmaksi paikaksi elää.

Viime kädessä big dataan pätevät samat säännöt kuin kaikkkeen muuhunkin teknologiseen edistykseen. Toimivat uudet käytännöt syntyvät tekemisestä ja dialogista, eivät suurista lupauksista.

Kirjoittaja tekee Kuluttajatutkimuskeskuksessa tutkimusta itsen mittaamisen teknologioista ja arjen analytiikasta. Hän vieraili syyskuussa Stanfordin yliopistossa ja osallistui Quantified Self-konferenssiin.

Lisää luettavaa:

danah boyd & Kate Crawford (2012) Critical questions for Big data. Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society 15(5), 662-679.

Yhä polttavampi kysymys on, miten ja mihin big dataa käytetään. Kun yritykset etsivät säännönmukaisuuksia ihmisten käyttäytymisestä, ne saavat myös uudenlaisen otteen arjesta.