



# Infografiikan uudet muodot

GRAAFISTA SUUNNITTELUA, DATAA JA JOURNALISMIA  
VERKKOJULKAISEMISEN AIKAKAUDELLA

Mika Minkkinen  
Lahden ammattikorkeakoulu  
Muotoilu- ja taideinstituutti  
Viestinnän koulutusohjelma  
Graafinen suunnittelu  
Kevät 2013





**Tiivistelmä**

Tämä opinnäytetyö käsittelee infografiikan verkkojulkaisemista ja tutkii perinteisen tilastografikan uusia muotoja verkkosovelluksina. Työssä keskitytään teoreettiseen perinteeseen, havaintoon ja datan visualisointitekniikoihin. Opinnäytetyön tavoitteena on koota perustietopaketti verkon visualisointeja suunnittelevalle graafikolle, toimittajalle tai ohjelmistokehittäjälle.

**Abstract**

This study deals with information graphics online publishing and examines the traditional statistical graphics in new forms of web app. The focus will be in the tradition, observation and data visualization techniques. Aim of this study is to gather basic information package on the network visualizations designed for graphic artists, editor or software developer.

**Asiasanat**

Infografiikka, datajournalismi, informaatiomuotoilu, datavisualisointi, tilastot, diagrammit, verkkojulkaiseminen

**Keywords**

Infographics, datajournalism, information design, data visualization, statistics, diagrams, web publishing

# Sisällys

<b>Johdanto</b> .....	4
<b>1 Infografiikan käsitteitä</b>	
1.1 Infografiikka .....	6
1.2 Datavisualisointi .....	8
1.3 Datajournalismi .....	8
1.4 Tilastotieteen käsitteitä .....	9
1.5 Big Data .....	9
<b>2 Visualisoinnin perusteet</b>	
2.1 Kuvan lukeminen .....	10
2.2 Kuvan tulkinta .....	11
2.3 Visualisointimuodot .....	14
2.4 Kuviomallien erityispiirteitä .....	14
<b>3 Tekniikka ja design</b>	
3.1 Tekniikan ehdot .....	20
3.2 Datan siivoaminen .....	22
3.3 Brändielementit .....	22
3.4 Vuorovaikutus .....	24
3.5 Tarinallisuus .....	25
<b>4 Visualisointiprosessi käytännössä</b>	
4.1 Lähtökohdat .....	28
4.2 Teoria ja käytäntö .....	28
4.3 Tekninen toteutus .....	29
4.4 Data-aineistot .....	31
4.5 Visualisointien suunnittelu .....	32
4.6 Lopputulokset ja jatko .....	32
<b>Lopuksi</b> .....	34
<b>Lähdeluettelo</b> .....	36
<b>Linkkejä ja esimerkkikuvia</b> .....	37
<b>Liitteet</b> .....	38

Infografiikka valtaa jatkuvasti enemmän palstatilaa tekstiltä ja valokuvilta. Ilmiö on tuttu paperilehdissä ja sama kehitys tuntuu jatkuvan verkossa. Usein näitä kuvia silmäillessä mietin, mikä onkaan infografiikkana esitetyn visualisoinnin viesti ja tehtävä? Onko visualisointi palstantäytettä vai esteettinen makupala? Oliko uutinen kiinnostava ja oliko sen pohjalla dataa? Oliko visualisointi kuvitusta vai infografiikkaa?

**USEASTI VERKOSSA JULKAISTAVAA** sisältöä, infografiikka mukaan lukien, toteutetaan paperilehden taiton ehdoilla. Erityisesti tilastografiikkaa ja datavisualisointeja julkaistaan useimmiten staattisena kuvana uutisen yhteydessä. Verkossa tällaiset visualisoinnit ovat usein kapean tekstipalstan johdosta lähes lukukelvottomia ja toteuttavat lähinnä palstantäytteen funktiota.<sup>1</sup>

Toisaalta huomaan, että myös verkkojulkaisemista varten tuotetaan kauniita julistemaisia visualisointeja, joita jaetaan ahkerasti sosiaalisessa mediassa. Usein näitä kuvia tuotetaan visuaalisuuden ehdoilla, ja usein datan määrä on tämäntyyppisissä visualisoinneissa vähäinen.<sup>2</sup>

Data-aineistojen visualisointia tehdään paljon myös tieteellisten tutkimusten yhteydessä. Näissä aineisto on laaja ja näkökulmat harvittuja, mutta lopullinen visuaalinen ulkomuoto näyttää usein kulmikkaalta ja viimeistelemättömältä.<sup>3</sup>

Verkossa kehittyy jatkuvasti uusia teknologioita visuaalisen tiedon prosessointiin ja kuvalliseen esittämiseen. Näiden teknologioiden avulla visualisointeja tuotetaan myös ohjelmoinnin näkökulmasta. Uusilla animaatiotavoilla ja käyttöliittymämalleilla esitellään ohjelmistojen suorituskykyä ja mahdollisuuksia. Näissä visualisoinneissa jään kaipaamaan käytännön sovellusta ja oikeaa data-aineistoa.

Tämän nopeasti ja monella saralla kehittyvän datavisualisointi-ilmiön ääressä myös oma kiinnostukseni on herännyt. Halusin selvittää kuinka visualisointeja tehdään ja pystyisinkö itse kehittymään ammattillisesti niin, että voisin toimia alalla journalistisesti kiinnostavien aiheiden parissa, datan käsittelyn ja visualisointien tekijänä.

Näistä ajatuksista alkoi puolivuotinen sukellukseni verkossa julkaistavan infografiikan maailmaan. Selvitin teoriaa, termejä ja käytäntöjä, sekä opettelini toteuttamaan visualisointeja itsenäisesti. Tämä opinnäytetyö on oma muistikirjani ja ikkunani verkossa julkaistavan infografiikan perusteisiin. Toivon, että tästä on hyötyä myös muille aiheesta kiinnostuneille lukijoille ja infografiikan suunnittelijoille.

## **MIKSI DATAVISUALISOINNISTA TULI TRENDI UUTISMEDIASSA?**

Parhaimmat verkossa julkaistut datavisualisoinnit tuottavat lukijalle uudenlaista lisäarvoa. Sellaista, jota ei paperilehteä lukiessa koskaan pysty kokemaan. Infografiikka ei verkossa ole ainoastaan kuvaa, vaan entistä useammin käyttöliittymä dataan, jonka avulla lukija voi suhteuttaa tietoa omaan elämäänsä, tutkia toimittajan näkökulman taustalla olevia faktoja tai luoda mielipiteitä monimutkaisista asioista.

Suunnittelijalle verkossa julkaistava infografiikka tarjoaa monenlaisia vuorovaikutuksen välineitä, loputtomasti tilaa ja mahdollisuuden tuottaa vaikka jokaiselle lukijalle omanlaisensa visualisointi.

Mediatalojen näkökulmasta verkossa julkaistava infografiikka on sisältöä, jota helposti jaetaan ja jonka vuorovaikutteisiin sovelluksiin linkitetään. Uutissivustot saavat linkityksien kautta lisää lukijoita ja mainostajat näkyvyyttä. Kun verkon uutistarjonnan rahoitusmalli vähä vähältä siirtyy maksulliseksi ja tilaajarahoitteiseksi, sijoitukset tämänkaltaiseen yksilölliseen sisältöön kuulostavat myös bisnesmielessä entistä houkuttelevimmilta.

Myös lukijat vaativat muutosta. Kritiikki perinteisen median verkkopalvelujen laatua kohtaan on jatkuvasti kasvanut. Toimitukset eivät ole enää hetken pärjänneet nopeudessa sosiaaliselle medialle ja mainosnäyttöihin perustuva rahoitusmalli on ajanut mediatalot laaduttomaan ja sensaatiohakuiseen klikkausjournalismiin. Kuukausimaksuun perustuva rahoitusmalli pakottaa mediatalojen verkkotoimitukset irti pikavoittonen kalastelusta takaisin laadukkaaseen journalismin tielle.

**Helsingin Sanomien** verkkolehden muuttuminen maksulliseksi on yksi selkeä merkki kotimaan mediamarkkinoiden muutoksesta. Uskon, että mediabisneksessä on seuraavaksi tulossa vaihe, jossa kilpailukeinoksi nousee taas laadukas sisältö. Vuodesta 2011 HS on aktiivisesti etsinyt uusia verkossa julkaistavia datajournalismisovelluksia kaikille avoimissa HS Open -tapahtumissa. Tapahtumiin on kokoontunut teknisiä osaajia, journalisteja ja graafikoita mitä erilaisimmista taustoista verkostoitumaan ja keksimään jotain uutta. Helsingin Sanomien datajournalismiin liittyvää toimintaa voi seurata HSNext-blogissa, jossa esitellään tapahtumissa syntyneitä sovelluksia.<sup>4</sup>

Myös **Yle** on perustanut oman datajournalismin osastonsa vuoden 2013 alussa. Plusdeskin nimellä kulkeva osasto tuottaa visualisointeja yhdessä eri toimitusten kanssa. Ylen Plusdesk esittelee aikaansaannoksiaan omassa, juuri perustetussa blogissaan.<sup>5</sup> Suurten mediatalojen liikehdinnän ohella visualisointeja syntyy myös ruohonjuuritasolla.

Suomessa alaa kokoa yhteen erilaiset innovaatiokilpailut, kuten *Apps for Finland*<sup>6</sup> ja Helsingin Sanomien Säätiön *Uutisraivaaja*<sup>7</sup>-kilpailu. Muita aktiivisia toimijoita ovat Helsingin kaupungin omistama *Forum Virium*<sup>8</sup> ja *Open Knowledge Foundation Suomen osasto OKF Finland*<sup>9</sup>.

Maailmalla mediatalojen kiinnostus on jo herännyt ja suuret mediatalot panostavat voimakkaasti erityisesti verkossa julkaistavaan infografiikkaan. **New York Timesin** useasti palkittu datadesk<sup>10</sup> on hyvä esimerkki siitä, mitä datajournalismiin panostamalla voi saavuttaa. Mediatalo on johdonmukaisesti panostanut datajournalismiin ja ottanut alan edelläkävijän roolin.<sup>11</sup>

On odotettavissa, että mitä enemmän viranomaiset ja yritykset avaavat omia data-aineistojaan ja ihmiset verkostoituvat, sitä enemmän saamme nauttia erityisesti verkkoa varten suunnitelluista visualisoinneista perinteisen median tuottamina tai avoimen datan käyttäjyhteisön tuottamina sovelluksina.

*Käytän tekstissä yläindeksi<sup>1</sup>-merkintää silloin, kun viittaaan verkko-osoitteeseen tai esimerkisovellukseen. Yläindekseihin liittyvät linkit löydät kirjan lopusta.*

# 1

# Infografiikan käsitteitä

Infografiikka on laaja kattoterminä valtavalle määrälle erilaisia visualisointeja. Käsitteellä viitataan monenlaisiin kuvallisiin esityksiin, opasteisiin, käyttöliittymiin tai tilastograafeihin. Tässä työssä keskityn käsittelemään journalistisista lähtökohdista suunniteltua infografiikkaa, joka perustuu numeeriseen dataan ja joka julkaistaan sähköisessä muodossa.

## 1.1 Infografiikka

Laajimmillaan infografiikaksi voidaan käsittää mikä tahansa tulkittava tai tunnistettava kuvallinen esitys, jonka lähtökohdista ei ole kuva. (Kosara 2010)

Keskeinen infografiikka-termin määrittelijä on **Edward R. Tufte**. Hänen kirjansa *Visual Display of Quantitative Information* on yksi infografiikkaan keskittyvän kirjallisuuden pääteoksista. Tufteen mukaan infografiikka on numeraalisen tiedon visualisointia (Tufte 2001, 11), joka onnistuessaan täyttää seuraavat kriteerit:

- Esittää ensisijaisesti dataa. Esityksen lähtökohdista ei voi olla mitään muuta kuin numeroaineisto.
- Johdattaa lukijaa tulkitsemaan dataa, ei visualisoinnin tekotapaa tai teknistä osaamista.
- Välttää dataan liittyvän tiedon ka-dottamista tai piilottamista.
- Esittää tietoa pienessä tilassa.
- Esittää isot tietomäärät tarkoituk-senmukaisessa muodossa.
- Johdattaa lukijan katsetta vertai-lemaan datan eri osia.
- Paljastaa datasta eri tasoja, kuten esimerkiksi yleisiä ilmiöitä ja yksityiskohtia.
- Auttaa lukijaa tekemään johtopäätöksiä visualisoinnista ja tul-kisemaan dataa oikein.
- Toimia yhdessä muun tekstin ja numeroinformaation kanssa.

(Tufte 2001, 13)

Tufteen määritelmästä on luetta-vissa hänen huolensa infografiikan ja kuvituksellisten elementtien se-koittumisesta. Tufte tutki lehdissä julkaistuja infografiikoita ja oli huolissaan siitä, kuinka visuaaliset teho-keinot vääristivät datan tulkintaa ja usein nousevat sisältöä keskeisem-pään rooliin. (Tufte 2001, 13)

Tufteen näkemyksen mukaan tiedon visualisoinnissa tärkeintä on esit-tää data mahdollisimman yksinker-taisessa ja helposti ymmärrettävässä muodossa (Tufte 2001, 13). Kuvituksellisil-le elementeille on tilaa visualisoinnin ulkopuolella (Tufte 2001, 11).

Tufte käytti visualisointien laatua arvioidessaan käsitteitä, jotka ovat edelleenkin käyttökelpoisia.

**Data-muste-suhde** (data-ink-ratio) käsitteellä Tufte viittaa siihen musteen määrään, jota käytetään datan esittämiseen. Mitä taloudellisempaa ilmaisu on, sitä tehokkaammin kuvio tietoa välittää. Voidaan myös laskea datamusteen suhdetta (data-ink-ra-tio), eli miten paljon mustetta käy-tetään datan esittämiseen suhteessa muuhun sisältöön kuten esimerkiksi koristeluun. (Tufte 2001, 91–107)

**Kuvioroinalla** (Chart junk) Tufte tarkoittaa kaikkea datan ulkopuolis-ta hälyä, jonka liittämistä visuali-sointiin kannattaa välttää. Hälyä ovat esimerkiksi: optiset illuusiot, voi-makkaat väri- tai kuviopinnat, liian

voimakkaat ja tiheät asteikkomerkin-nät sekä ankat eli ylimääräiset dataan kuulumattomat kuvitukselliset ele-mentit. (Tufte 2001, 107–123)

**Datan tiheyden** (data density) kä-si-te viittaa data-aineiston ja visuali-soinnin koon suhteeseen. Kuvion datatiheyttä voidaan mitata laske-malla kuinka monta tietoyksikköä mahtuu visualisoinnin pinta-alalle. (Tufte 2001, 161–177)

**Valekerroin** (lie factor) on yksik-kö, joka mittaa visualisoinnista tul-kittavan ja datan esittämän totuuden eroa. Jos kahden asian välinen ero on datassa esimerkiksi 50 prosent-tia, niin sen tulee olla sama myös vi-sualisoinnissa, eli viisikymmentä prosenttia suurempi. Useimmiten valheellisuutta aiheuttaa perspek-tiiviin asetetut kuvat. Esimerkik-si pylväsdiaagrammien sijoittaminen perspektiiviin hämärtää, pylväiden kokoero tulkintaa, koska edessä ole-vat pylväät ovat pidempiä kuin taa-emmaat pylväät. (Tufte, 2001, 57)

**Esteettiset kriteerit** visualisoin-nissa täyttyy Tufteen mukaan silloin, kun yksinkertainen visualisointi ja monimutkainen data kohtaavat. Esteettisesti laadukkaassa visualisoin-nissa kohtaavat:

- Oikein valittu datan esitystapa ja siihen sopiva visuaalinen tyyli.
- Sanojen, numeroiden ja kuvitus-ten laadukas yhteiskäyttö.



- Visuaalinen tasapaino, oikeat elementtien väliset suhteet ja skaala.
- Mahdollisuus tutkia datan yleiskuvaavaa ja yksityiskohtia.
- Datasta luettavat faktat ja tarina.
- Teknisesti ammattilaatuinen, huolellisesti toteutettu ja viimeistelty ulkoasu.
- Ei sisällä ylimääräistä koristelua.

(Tuftte 2001, 177–191)

Tuftten jyrkkää näkemystä on myös arvosteltu. Wheaton Collegen professori **John Grady** testasi Tuftten teorioita havaintokokeissa ja huomasi, että kaikki Tuftten oletukset eivät aina pitäneet paikkaansa. Joskus kuvitukselliset elementit auttavat datan lukemisessa, johtopäätösten tekemisessä ja muistamisessa. (Cairo, 2013, 65)

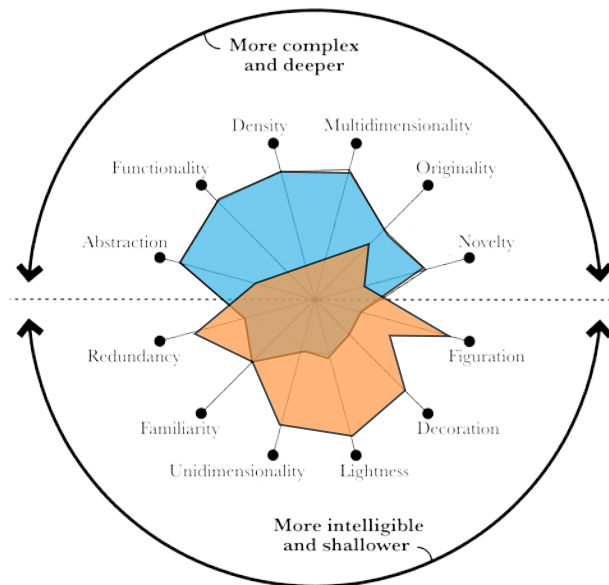
Keskustelu Tuftten määritelmistä voimistui, kun Tuftten kirjoissaan parhaana Times-lehden graafikko **Nigel Holmes** vastasi kritiikkiin.

Holmesin mukaan akateeminen infografiikkatutkimus (Tuftte) menee vikaan perusolettamuksessaan, jossa oletetaan että lukijat ovat älykkäitä, kiinnostuneita ja heillä on aikaa lehtiin sisältöjä lukiessaan. (Holmes 1991, 47)

Holmesin mielipide perustuu kokeemukseen lehtityöstä. Hän painottaa käyttäjystävällisyyttä, yksinkertaistamista ja nopeasti ymmärrettävien kuvien käyttökelpoisuutta. Hänen mukaansa lehti palvelee lukijaa parhaiten tarjoamalla helppoa ja nopeasti omaksuttavaa sisältöä, jota kuvituksellinen infografiikka tarjoaa. (Holmes 1991, 47)

**Alberto Cairo** määrittelee infografiikkaa Tuftten ja Holmesin näkemysten välimaastossa. Cairo tuntee lehtityön espanjalaisen El Mundos interaktiivisen infografiikan osaston vetäjänä, mutta myös akateemisen maailman työskenneltyään Pohjois-Carolinan ja Miamin yliopistoissa. (Holmes 1991, 47)

Kirjassaan *The Functional Art Cairo* käsittelee infografiikan käsitettä kehittämänsä *Visualization wheel* -ku-



*Visualization wheelin avulla Cairo esittää havainnollisesti mitä ominaisuuksia arvioitava visualisointi korostaa. Oranssilla värillä täytetty alue kuvaa tyyppillisiä taiteellisen, graafisen ja journalistisen visualisoinnin esittämistapoja. Sinisellä täytetty alue esittää insinöörimäiselle ja tieteelliselle ilmaisulle ominaisia visualisoinnin keinoja.*

viomallin avulla. Hänen mukaansa visualisoinneille on erilaisia tehtäviä eri julkaisuissa ja siksi monenlaisille ilmaisukeinoille tulee olla tilaa. Cairo arvioi visualisointeja kuuden ominaisuuden välillä.

- abstrakti – esittävä (*abstraction – figuration*) asteikolla verrataan visuaalisten elementtien käyttöä. Perustuuko kuvio esittävään (esim. valokuva) vai abstraktiin (esim. ikoni) visuaaliseen kieleen.
- funktionaalinen – koristeellinen (*functionality – decoration*) asteikolla mitataan minkälaisessa suhteessa tietoa esittelevät ja koristeelliset elementit ovat. Tuftten termein ilmaisuna koristeellisen kuvioroinan (*Chartchunk*) suhde tieto-musteeseen (*Data-ink*).
- tiheys – keveys (*density – lightness*) asteikolla mitataan kuinka paljon informaatiota visualisointi välittää. Esimerkiksi ympyrädiagrammissa on mahdollista esittää vain muutamia lukuja, kun taas pistekartassa samassa tilassa voidaan esittää jopa satoja eri arvoja.

- moniulotteisuus – yksiulotteisuus (*multidimensionality – unidimensionality*) asteikolla mitataan sitä kuinka monta erilaista näkökulmaa visualisoinnista on luettavissa. Yksiulotteinen visualisointi on helposti hahmotettava ja ymmärrettävä, mutta moniulotteinen mahdollistaa useat näkökulmat aiheeseen.
  - ainutlaatuisuus – tuttuus (*originality – familiarity*) asteikolla mitataan sitä onko meillä entuudestaan tietoa kuinka visualisointia pitää lukea. Esimerkiksi väestöpyramidi on tuttu visuaalinen tapa esittää väestön ikärakennetta ja sukupuolijakaumaa.
  - olennainen – epäolennainen (*novelty – redundancy*) asteikolla mitataan sitä kuinka paljon dataa kuvioon on sisällytetty ja onko sitä tiivistetty lukijaa varten. Jotkut tietoaineistot voivat sisältää kymmeniä tuhansia rivejä tietoa, mutta visualisoinnissa näytetään vain muutama.
- (Cairo 2013, 52–59)

## 1.2 Datavisualisointi

Usein datavisualisoinnin ja infografiikan käsitteitä käytetään rinnakkain. Hienoinen merkitysero käsitteillä kuitenkin on. Infografiikan käsite sopii paremmin kuvaamaan suunnittelijan tekemää visualisointia, jota lukija tulkitsee. Datavisualisoinnin käsitteellä voidaan tällöin viitata datalähtöiseen visualisointiin, jota käyttäjä voi selailla. Datavisualisointi on ennen kaikkea käyttöliittymä data-aineistoon. (Cairo 2013, xvi)

Datavisualisoinneilla on kahdenlaisia tehtäviä, viestinnällisiä ja havainnollistavia. Havainnollistavia tehtäviä toteutetaan kun esitetään abstraktit tilastoaineistot ymmärrettävässä muodossa esimerkiksi perinteisinä tilastograafeina. (Few 2013)

Viestinnällisiä tehtäviä toteutetaan kun esitetään data-aineistoon perustuvia visualisointeja, joista on tulkittavissa tarina ja joiden avulla lukija ymmärtää ilmiön toimintaa, syitä ja seurauksia. (Few 2013)

## 1.3 Datajournalismi

Datajournalismin käsite keskittyy sisällöntuottamisprosessiin. Käsitteen alla puhutaan datalähtöisestä tuotantoprosessista, jonka lopputuotokset ovat usein visualisointeja.

Datajournalisti ja Birminghamin yliopiston professori **Paul Bradshaw** määrittelee datajournalismin olevan uusi mahdollisuus ja työkalu, jonka avulla voidaan yhdistää perinteinen tutkiva journalismi ja tarinankerronta uudenlaisiin digitaalisiin lähdeaineistoihin. (Bradshaw 2012)

Bradshaw'n mukaan datajournalismi voi olla osa mitä tahansa journalistisen prosessin vaihetta. Digitaalisista aineistoista voidaan tekniikan avulla kerätä ennen lukematonta ja näkemätöntä lähdeaineistoa, kuten esimerkiksi **Adrian Holovaty** teki luodessaan Chicagon rikollisuuskarttaa. Holovaty ohjelmoi juttuaan varten tiedonkeruua-

tomatiikan, joka haki automaattisesti rikostilastoja eri poliisipiirien verkkosivuilta. Holovaty yhdisti tulokset kartalle ja selvitti alueellisia eroja kaupungin rikostilastoissa.<sup>12</sup>

Tekniikan avulla voidaan myös analysoida suuriin data-aineistoihin perustuvia sisältöä, kuten esimerkiksi *The Telegraph* -lehti teki, kun se loi verkostanalyysin kansanedustajien rahankäytöstä. Analyysin perusteella lehti löysi rahankäytöstä uusia sidoksia ja paljasti väärinkäytöksiä. Lehti julkaisi aiheesta useita juttuja.<sup>13</sup>

Datajournalismi voi auttaa journalistia myös kertomaan tarinaa lukijalle. Esimerkiksi **Hans Roslingin** luoma *Gapminder*<sup>14</sup>-sivusto kerää miljoonia lukijoita erilaisien tilastojen äärelle. Maailman tilaa koskevat tilastot ovat olleet jo pitkään kaikille avoimena, mutta vasta Roslingin hyvän visualisoinnin avulla ne ovat saavuttaneet suuren yleisön huomion.

Datajournalismin käsitettä määrittämään jatkuvasti uudelleen myös journalistien piirissä. Toiset uskovat sen olevan vallankumouksellinen idea ja uusi käänteentekevä työtapo, joka tekee valtaapitävien toiminnan läpinäkyvämmäksi ja tuo demokraattisen päätöksenteon askeleen lähemmäs kansalaista. (Bradshaw 2012)

Esimerkiksi *Guardianin* toimittaja ja datajournalisti **Simon Rogers** ylistää datajournalismin olevan uuden journalismin alku, jossa tarvittavat työkalut ja tiedonlähteet ovat kaikkien saatavilla ja ilmaisia. Rogersin mukaan datajournalismi on samassa kehitysvaiheessa kuin punk-liike sen alkuaikoina, jolloin ihmiset halusivat vaikuttaa ja tehdä asioita itse auktoriteetteja kumartelematta. (Rogers 2012)

Datajournalismin saralla mediatalot ovat vanhoja auktoriteetteja, joista piittamatta datajournalismin harastajat kaivelevat maailman datavarantoja ja julkaisevat niistä omia datajournalistisia juttuja omilla sivuillaan. (Rogers 2012)

Rogers uskoo, että kehityksen edessä parhaat datajournalistiset käytännöt valtavirtaistuvat ja tulevat myös mediatalojen arkeen. Kuten punkin kapina aikoinaan, datajournalismi kehittyi ja kasvaa villinä harastelijoiden keskuudessa ja parhaat oivallukset kopioidaan mediatalojen työnkulkuun. Punk-liike valtavirtaistui ja musiikista tuli soittolistakelpoista. Kuitenkin suurin arvo punkin ja datajournalismin alkuaikojen vapaudelle on Rogersin mukaan se, kuinka ne muuttavat ajattelua ja näyttävät mallia kuinka asioita voi tehdä uudella tavalla. (Rogers 2012)

Datajournalisti **Jonathan Gray** hylitsee Rogersin intoa ja uskoa datajournalismin vaikutuksista ihmisen elämään. Gray muistuttaa että datan tutkiminen itsessään ei muuta maailmaa tai journalismia, eikä data ole koko totuus maailmasta. (Gray 2012)

Gray vertaa datajournalismia varhaiseen valokuvajournalismiin. Valokuvan alkuaikoina uskottiin että objektiivinen valokuva välittää totuuden maailmasta ja nostaa journalismin luotettavuutta. Nykyisin olemme kuitenkin tottuneet lukemaan kuvia, ymmärtämään rajauksia ja tunnistamaan kuvanmuokkauksia. Enää emme usko valokuvan olevan ikkuna todellisuuteen vaan tiedämme sen välittävän kuvaajan, median tai mainostajan valitsemää tulkintaa. (Gray 2012)

Gray muistuttaa, että myös data vaatii aina tulkintaa ja tulkinta ammattitaitoa. Avoimet datavarannot ei ole valtaa koko kansalle vaan ainoastaan sitä ymmärtävälle ja siitä kiinnostuneille. (Gray 2012)

Dataan perustuva journalismi ja julkishallinnon avaamat datavarannot varmasti nostavat mielenkiintoa yhteisiä asioita kohtaan, mutta data itsessään ei saa ihmisiä osallistumaan. Datan avulla voidaan vain tarkastella joitain ilmiöitä, joita on joskus keksitty mitata tai joista on jäänyt jokin sähköinen jälki. (Gray 2012)

## 1.4 Tilastotieteen käsitteitä

Tilastoaineistosta puhuttaessa on yleistä käyttää käsitteellistä erottelua tietoa, informaatio ja data -käsitteiden välillä. (Niiniluoto 1989, 27–30)

**Datan** käsitteellä viitataan laajaan ja jalostamattomaan aineistomaan (Niiniluoto 1989, 27–30). Tässä työssä puhun datasta tai data-aineistosta silloin kun viitataan alkuperäiseen, muokkaamattomaan tilastoaineistoon (Kuusela 2000, 204).

**Informaation** käsitteellä viitataan aineistosta jalostuneeseen tai valittuun sisältöön (Niiniluoto 1989, 27–30). Esimerkiksi laajoista data-aineistoista voidaan valita vain tietyt sarakkeita tai rivejä visualisoitaviksi. Tässä vaiheessa data-aineisto muuttuu informaatioksi, mutta se ei ole vielä tietoa.

**Tiedon** käsitellä viitataan data-aineistosta tuotettuun lopputulokseen, joka on tulkittavissa (Niiniluoto 1989, 27).

Tilastotieteessä datan käsitettä käytetään silloin, kun tilastoaineisto on numeerista ja se on kerätty luotettavalla mittaamistavalla. (Kuusela 2000, 204)

Mittaustuloksia voidaan kerätä kahdella tavalla. Silloin kun kerätään yhden tietyn hetken aineistoa, puhutaan poikkileikkausaineistosta. Aikasarja-aineistosta puhutaan silloin kun samaa aineistoa on kerätty useammalla mittauskerralla. (Kuusela 2000, 204)

Mittaustuloksen yksiköitä kutsutaan muuttujiksi, joita luokitellaan laadullisiin ja määrällisiin muuttujiin. (Kuusela 2000, 204)

Laadulliset muuttujat ovat sellaisia, joiden avulla järjestään tai luokitellaan tietoa. Laadullisten muuttujien yksiköt voivat olla tekstiä tai numeroita, mutta niillä ei voi koskaan suorittaa laskutoimituksia. Esimerkiksi osoitetietojen postinumerosarake on laadullinen muuttuja. (Kuusela, 2000, 204)

Määrälliset muuttujat tunnistaa siitä, että niillä voi suorittaa laskutoimituksia. Määrällisillä muuttujilla voidaan mitata esimerkiksi välimatkaa tai lukujen suhdetta. Suhdetta

mittaavien muuttujien asteikko täyttyä alottaa aina nollostasta. Välimatkaa mittaavien muuttujien yhteydessä voidaan käyttää myös muita asteikkoja. (Kuusela 2000, 204)

## 1.5 Big Data

Big Data -käsitteellä viitataan useiden yksittäisten data-aineistojen yhdistämisestä syntyvään päädata-aineistoon. Big Datan tietoyksiköt voivat syntyä vaikkapa erilaisten antureiden avulla. Esimerkiksi lämpötilaa, kosteutta ja valon määrää mittaavat anturit tuottavat jokainen erillistä ja eri yksiköissä tulkittavaa dataa. Data voi myös kasvaa loputtomasti, koska uusia tietoyksiköitä voidaan antureiden avulla vaivattomasti tuottaa vaikka joka minuutti. (Open Methodology 2013)

Vaikka käsitettä määrittää tietoa-aineiston laajuus, ei Big Dataa välttämättä tunnista tietoa-aineiston tiedostokoosta. Esimerkiksi lentokoneissa saattaa olla lennon aikana jopa 100 000 erilaista anturia käytössä. Nämä anturit tuottavat lentäjille tietoa sääolosuhteista, matkustamon ilmanlaadusta ja paljon muusta. Vaikka koneessa on antureita valtavasti niin dataa syntyy suhteellisen vähän tiedostokokoa arvioitaessa. Yksi anturi tuottaa dataa 8 bittiä, jolloin tunnin lennon aikana kerääntynyt anturitieto mahtuu noin 3 gigatavun tiedostokokoon. (Open Methodology 2013)

Toisaalta Big Data -aineisto voi olla valtavan suuri aineistomassa. Esimerkiksi Googlen toiminta perustuu Big Data analyysiin, jossa satojen erilaisten sisältöä arvottavien algoritmien avulla tiedon hakijalle tarjotaan mahdollisimman relevanttia tietoa. Googlen käsittelemät datamassat ovat bittikooltaan valtavia, mutta käytännössä kohtuullisen yhteneväisessä ja helposti vertailtavassa muodossa. (Open Methodology 2013)

Laajojen ja monimutkaisten Big Data -aineistojen käsittely vaatii usein erityisen suorituskykyisiä ja

Big Datan analysointiin erikoistuneita ohjelmia tulkitsemaan lentokone-datan kaltaisia aineistoja. Tähän tehtävään on olemassa ilmaisia palvelinsovelluksia, kuten esimerkiksi avoimen lähdekoodin Hadoop-sovellus. Ohjelmiston käyttö vaatii perehtyneisyyttä palvelinohjelmointiin.

(Open Methodology 2013)

# 2

## Visualisoinnin perusteet

Tiedon visualisoinnilla on pitkä historia. Perinteiset muodot ovat ajan saatossa hioutuneet erinomaiseksi pohjaksi data-aineiston visualisointiin. Pylväs-, viiva- ja ympyrädiagrammeja on opittu ja opeteltu tulkitsemaan. Uusia visualisointimuotoja suunniteltaessa on tärkeää tuntea nämä esitystavat ja niiden tulkintaan liittyvät ominaispiirteet.

### 2.1 Kuvan lukeminen

Infografiikkaa tai datavisualisointia suunniteltaessa on hyvä tuntea muutama perusasia ihmisen havaintomekanismeista. Erityisesti interaktiivisen ja animoidun infografiikan luomisessa on tärkeää ymmärtää ihmisen havaintokyvyn rajat. Perusopien tuntemus auttaa suunnittelijaa ohjaamaan lukijan katsetta kuvassa ja hallitsemaan visualisoinneista tehtäviä tulkintoja.

Ihmisen havaintoprosessi on monimutkainen yhdistelmä kokonaiskuvan ja yksityiskohtien havainnointia. Havaintopsykologian avulla voimme tutustua prosessin peruseräisiin ja hyödyntää niitä suunnittelutyössämme. Havaintopsykologi **Stephen Kosslyn** esittelee ihmisen havaintoprosessia kolmen mieleenpainuvan väittämän avulla:

- Ihmismieli ei ole kamera.
- Mieli arvioi kirjan sen kannen perusteella.
- Halua on mutta rahkeet eivät riitä.

(Kosslyn 1994, 3–13)

**Ihmismieli ei ole kamera** -väittämällä Kosslyn tarkoittaa sitä, kuinka ihminen havainnoi vain merkityksellisiä asioita ja sivuttaa muut. Mieli ei tallenna todellisuutta kuten kamera, vaan valikoi. Muistiimme tallentuu vain ne asiat, jotka koemme merkityksellisiksi. (Kosslyn 1994, 3–13)

**Mieli arvioi kirjan sen kannen perusteella** -väittämällä Kosslyn viittaa siihen tapaan kuinka visuaalinen tulkintamekanismi ja muistijärjestelmä ovat yhteydessä toisiinsa. Tulkitsemme esimerkiksi katkoviivaa niin, että oletamme sen muodostavan samanlaisen kokonaisuuden kuin yhtenäinen suora viiva. Todellisuudessa näemme vain sarjan samansuuntaisia lyhyitä viivoja. (Kosslyn 1994, 3–13)

**Halua on, mutta rahkeet eivät riitä** -väittämällä Kosslyn viittaa ihmismielen rajalliseen kapasiteettiin muistaa asioita. Voimme säilyttää lyhytaikaisessa muistissa keskimäärin neljä yksikköä kerralla. Visuaalisilla keinoilla yksikköjä on mahdollista yhdistää isommiksi kokonaisuuksiksi ja sen avulla voimme muistaa ja tulkita suurempia tietomääriä samanaikaisesti. (Kosslyn 1994, 3–13)

#### VERTAILU

Aivot pyrkivät vertailemaan tulkittavaa tietoa. Vertailua tehdään nähdyn havainnon ja muistikuvien välillä tai esimerkiksi aineistossa nähtävien samankaltaisten elementtien välillä. Vertailtavia tehdään pääsääntöisesti viiden ominaisuuden avulla.

**Läheisyys** (proximity). Hahmotamme toisiaan lähellä olevat elementit yhdeksi ryhmäksi. Esimerkiksi pylväskaaviossa lähekkäin olevat pylväät tulkitaan samaksi ryhmäksi.

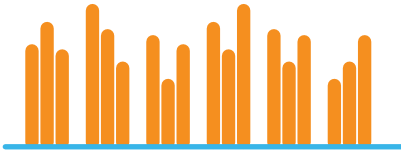
Isompi tyhjä tila pylväsryhmien välillä tulkitaan ryhmien erottajaksi. Tästä syystä pylväiden väli pitäisi pyrkiä pitämään tasavälisenä silloin kun sillä ei haluta kertoa lisä-informaatiota esitettävästä asiasta. (Cairo 2013, 114–118)

**Samankaltaisuus** (similarity). Keskenään samanlaiset elementit mielletään ryhmäksi. Jos sijoitamme kartalle esimerkiksi ympyröitä ja neliöitä, lukija tulkitsee ne kahdeksi eri kategoriaksi, joissa hän vertailee ympyröitä keskenään ja neliöitä keskenään. (Cairo 2013, 114–118)

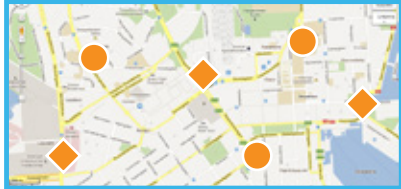
**Yhteydet** (connectedness). Toisiinsa visuaalisesti kytketyt elementit hahmotamme saman ryhmän jäseniksi. Esimerkiksi verkostovisualisoinneissa kytkemme verkoston jäseniä toisiinsa viivoilla, jotta aineistosta hahmottuu ryhmät ja niiden väliset suhteet. (Cairo 2013, 114–118)

**Jatkuvuus** (continuity). Mieleemme täydentää ja jatkaa samansuuntaisia linjoja. Suorien tai kaarevien linjojen tulkitaan helpommin jatkuvan, kuin sellaisen, joissa on teräviä kulmia. Esimerkiksi voimme kuvata pallon putoamista tai pomppaamista piirtämällä vain osittain lentorataa kuvavan viivan. (Cairo 2013, 114–118)

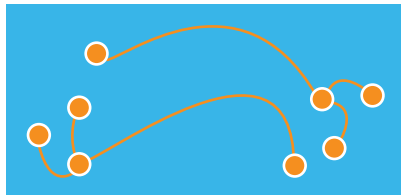
**Sulkeminen** (closure). Rajatut alueet tulkitaan ryhmiksi. Esimerkiksi pistekartan tulkintaa voidaan helpottaa rajaamalla yksittäinen piste parvi. (Cairo 2013, 114–118)



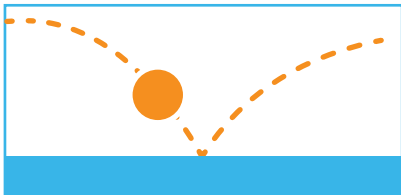
Läheisyys



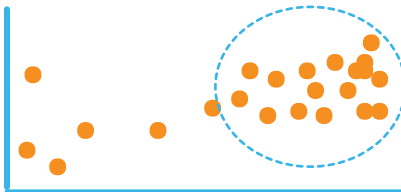
Samankaltaisuus



Yhteydet



Jatkuvuus



Sulkeminen

## KATSEHAVAINTO

Katsehavainto koostuu sarjasta yksittäisiä pistemäisiä kuvia, joista aivomme valitsevat vertailun ja muistin avulla joko tuttuja tai tavallisuudesta poikkeavia muotoja tulkittaviksi. Huomiomme kiinnittyy olennaisimpiin eroihin, joita alamme analysoida tarkemmin. Tutuista muodoista, kuten esimerkiksi kasvoista, teemme vertailuja pitkäkestoisen muistimme

kanssa ja yhdistämme kasvot muuhun muistissamme olevaan tietoon kyseisestä ihmisestä. (Cairo 2013, 110)

Tilastovisualisoinnin suunnittelussa pääsääntönä on saada lukija kohdistamaan huomio keskeisiin eroihin aineistossa. Parhaiten se onnistuu silloin kun vertailu tapahtuu vain yhdellä akselilla. Ihminen havaitsee helposti yhdellä akselilla tapahtuvan muutoksen. (Cairo 2013, 118–120)

Esimerkiksi useita pystypylväskaavioita samalle sivulle tai näytölle sijoitettaessa kannattaa asemoida graafit niin, että kaikkien nollapiste alkaa vaak-akselilla samasta kohdasta. Tällöin pylväiden välinen korkeusero saa suurimman huomion, koska muut arvot ovat muuttumattomia. (Cairo 2013, 118–120)

Seuraavaksi tehokkain tapa on kahdella akselilla tapahtuvan muutoksen vertailu. Havaintoprosessissa vertailaan aineiston yksiköiden sijaintia suhteessa muihin. (Cairo 2013, 118–120)

Kolmas tapa havaita eroja asioiden välillä on pituuden, suunnan tai kulman vertailu. Esimerkiksi ilman aikamerkintöjä olevaa kellotaulua tulkitaan viisareiden kulman avulla. (Cairo 2013, 118–120)

Selvästi edellisiä virheherkempiä vertailutapoja ovat pinta-alan, tilavuuden ja kaarevuuden vertailu. Vaikein tehtävä havaintomekanismille on tummuuden ja värikylläisyyden vertailu. Joskus koropleetti-tyyppiset karttavisualisoinnit saattavat tehokkaasti välittää yleiskuvaa tarkasteltavasta ilmiöstä, mutta tarkempien yksityiskohtien ja erojen havaitseminen niistä on vaikeaa. (Cairo 2013, 118–120)

## 2.2 Kuvan tulkinta

Hyvin toteutettu visualisointi lisää lukijan ymmärrystä tarkasteltavasta aiheesta ja houkuttelee arvioimaan ilmiön suuruutta, vertailemaan tietoa ja etsimään aineistosta syy-seuraussuhteita. (Tilastokeskus 2012)

Visualisointia suunniteltaessa on tärkeää löytää keskeisimmät asiaan

liittyvät kysymykset ja konkretisoida ilmiötä lukijalle tutuilla yksiköillä.

Esimerkiksi seuraavalla aukeamalla kuvattu jätemäärän jakautuminen voidaan tulkita monista eri näkökulmista. Visualisoinnin suunnittelijan tulee valita kiinnostava näkökulma ja päättää mihin kysymyksiin visualisoinnilla haluaa vastata. Jättemäärän kehitystä, kuten muitakin visualisointeja, voi tarkastella esimerkiksi seuraavista näkökulmista:

- Miten ilmiön koko on kehittynyt ajassa? Onko ilmiö kasvanut tasaisesti, pyrähdyksittäin vai ilmestynyt yhtäkkiä?
- Miten ilmiön koko eroaa muista vastaavista ilmiöistä? Onko esimerkiksi kaivosjätteen määrän kasvu vastaavanlaista kaikkialla maailmassa vai onko kehitys kotimainen trendi?
- Onko tilasto luotettava ja vertailtavissa muihin tilastoihin? Luokitellaanko Suomessa kaivosjätteeksi samat ainekset kuin muualla maailmassa? Onko tilaston laatija kaivosalan yhtiö tai luonnonsuojelujärjestö?
- Onko ilmiölle ilmeistä selitystä? Onko kaivostoiminta vasta alkanut Suomessa, jolloin jätteen määrän kasvu on valtavaa, vai onko kaivostoiminnassa syntyvä jättemäärä kasvanut uuden kaivamistekniikan myötä?

(Tilastokeskus, 2012)

**KORRELAATIO JA KAUSALITEETTI** Ihmisellä on taipumus tulkita samanaikaisesti tapahtuvat ilmiöt syyksi ja seuraukseksi. Tästä seuraa usein virhepäätelmiä. (Mellin 1997, 177)

Kahden ilmiön suhdetta toisiinsa kutsutaan korrelaatioksi. Tilastot voivat osoittaa, että kaksi tarkasteltavaa ilmiötä korreloivat, eli kehittyvät samantahtisesti. Tästä ei kuitenkaan suoraan voi päätellä ovatko ilmiöt sidoksissa toisiinsa, vai vaikuttaako samantahtisuuteen jokin kolmas seikka. (Mellin 1997, 177)

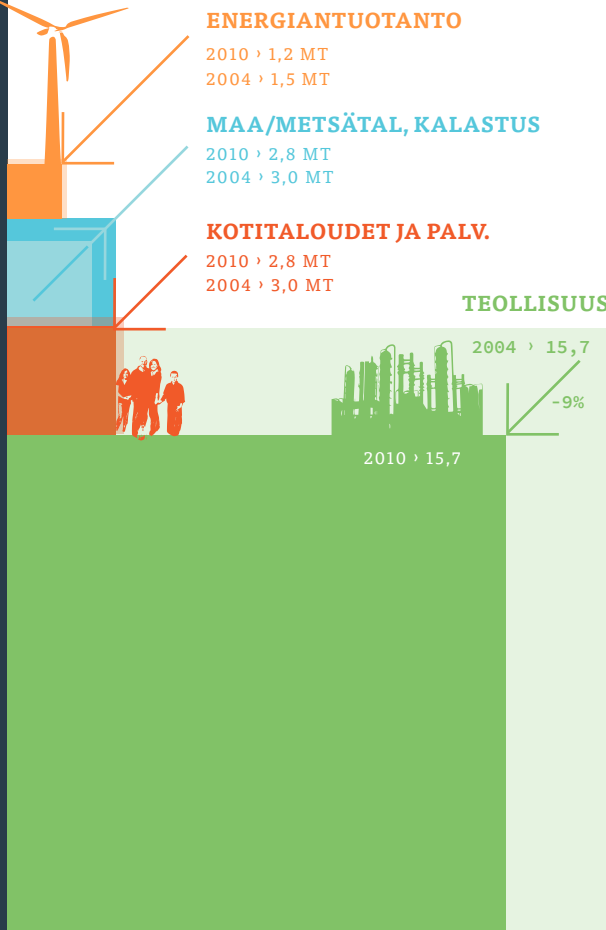


+ 38%

2004 › 23,8 MT

## KAIVUJÄTE

2010 › 49,7 MT



## Jätekertymä Suomessa vuosina 2004 ja 2010

LÄHDE: TILASTOKESKUS

## RAKENUSJÄTE

2010 › 24,6

+ 12%

2004 › 20,8



Syy ja seuraussuhteen eli kausaliteetin osoittaminen on usein monimutkaisempi prosessi kuin kahden samanaikaisen arvon vertaaminen. Korrelaation mittaamisesta on kuitenkin hyötyä, koska se mittaa muuttujien välisen yhteyden voimakkuutta, joka voi auttaa kausaliteetin selvittämisessä. (Mellin 1997, 177)

### 2.3 Visualisointimuodot

Esitystavan valinta on keskeisessä asemassa tiedon ymmärtämisessä ja sen muistamisessa. Joskus teksti tai taulukkomuotoinen esitystapa on perustellumpi valinta kuin kuvallinen ilmaisu. (Kuusela 2000, 8)

Tekstissä kirjoittajan ilmaisu on keskeisemmässä osassa kuin varsinaisen data. Tekstin avulla kirjoittaja voi tehokkaasti välittää aineistosta tekemiään havaintoja ja tuntemuksia lukijalle. (Kuusela 2000, 11–15)

Taulukkomuoto taas on erinomainen tarkkuutensa ansiosta. Taulukkomuotoinen esitystapa soveltuu tilanteisiin, joissa tietoaineistossa keskeistä ovat yksittäiset tietoalkiot, ei niinkään kokonaisuus tai asioiden väliset suhteet. (Kuusela, 2000 11–15)

Kuvallinen ilmaisu on tehokas tapa esittää data-aineistossa esiintyviä suhteita ja riippuvuuksia. Visualisoinneissa yksittäisten tietoalkioiden havainnoiminen on usein vaikeaa. (Kuusela 2000, 11–15)

Kuviotyypin valinnalla voidaan keskeisesti vaikuttaa siihen, mitä puolia tiedosta halutaan korostaa. Kaikki kuviot eivät sovellu kaikenlaiselle datalle. Kuviotyypin valintaan vaikuttaa:

- Mikä on datan mittayksikkö? Sisältääkö data esimerkiksi lukumääriä, prosenttiarvoja, keskiarvoja, aika- tai paikkatietoa, tms.
- Moniko ulotteista data on? Mikä on datasta löytyvien muuttujien määrä, sijoittuuko se nelikenttään vai koordinaatistoon, onko dataa ryhmiä tai negatiivisia arvoja.
- Mikä on muuttujien asteikko?

Tarvitaanko tasavälistä vai logaritmistä asteikkoa? Missä skaalassa datan muutokset parhaiten näyttäytyvät, kiloissa, grammoissa vai tonneissa?

- Mikä on luokiteltujen muuttujien asteikkotyyppi? Onko tieto jatkuva-arvoista vai ei?
- Kuinka monta eri elementtiä kuvioon täytyy sisällyttää ja kuinka paljon tilaa on käytettävissä lopullisessa julkaisussa?

(Kuusela 2000, 52)

#### AKSELIT, ASTEIKOT JA KOORDINAATISTOT

Koordinaatisto on avain tilastotiedon visuaaliseen hahmottamiseen. Datan sijoittamisella kuvion eri akseleille voidaan vaikuttaa merkittävästi kuvion tulkintaan. (Kuusela 2000, 69)

Ihminen mieltää koordinaatiston vaakakselin jatkumona eli jatkuva-asteikkoinen tieto, kuten esimerkiksi aika ja matka kannattaa sijoittaa vaakakselille. (Kuusela 2000, 69)

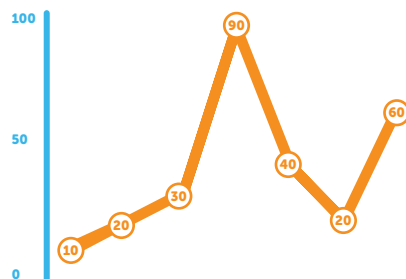
Epäjatkuva tieto kuten esimerkiksi määrä, laatu tai prosenttietä sopii koordinaatiston pystyakselille tai piirakkadiagrammiin. (Kuusela 2000, 53)

Tilastokuviossa voi olla kahdenlaisia asteikkoja, joista yleisin on tasavälinen asteikko. Siinä jokaisen asteikon luvun välimatka on täsmälleen samansuuruinen. (Kuusela 2000, 105)

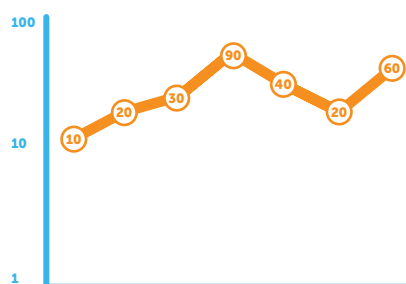
Toinen asteikkomalli on logaritminen asteikko, joka soveltuu tarkasteltavan ilmiön kasvuvauhdin esittämiseen. Logaritminen asteikko perustuu suhdelukuun, joka voidaan laskea potenssilaskusta tutulla tavalla. Esimerkiksi luvun 100 potenssi-merkintä on 10 potenssiin 2, jolloin sadan logaritmi on 2 ja tuhannen logaritmi 3 jne. Logaritmisiin lukuihin perustuvalla asteikolla on tehokasta kuvata kehitystrendien voimakkuutta ja muutosnopeutta. (Kuusela, 2000, 105)

Logaritmistä asteikkoa käytetään usein taloudellisten ilmiöiden kuvauksissa, koska se paljastaa tasavälistä asteikkoa paremmin kasvut ja taan-

tumat. Asteikon käyttö on kuitenkin ongelmallista, koska logaritmin käsite on heikosti tunnettu. (Kuusela, 2000, 105)



Tasavälinen asteikko



Logaritminen asteikko

### 2.4 Kuviomallien erityispiirteitä

Perinteiset tilastokuviot, kuten pylväs-, ympyrä- ja viivadiagrammi toimivat edelleenkin lähes samoin kuin kuvioita suunnitelleen **William Playfairin** (1759–1823) aikana 1800-luvun alussa. Tuolloin Playfair perusteli kuvallisen ilmaisun ja visualisointien käyttöä nykypäivänakin päteville kuulostavilla perusteluilla.

- Grafiikan avulla voi yksinkertaistaa monimutkaisia asioita.
- Kiireiset lukijat tarvitsevat jonkinlaisen visuaalisen apuvälineen tiedon ymmärtämiseen.
- Graafinen tieto on helpommin omaksuttavissa kuin taulukot.
- Graafisesti esitettyä tietoa on helpompi arvioida visuaalisen havainnon avulla.
- Graafisesti esitetty tieto on helpompi muistaa

(Kuusela 2000, 27)



Playfairin kehittämät visualisoinnit ovat kehittyneet yllättävän vähän verrattuna niiden pitkään historiaan.

Datavisualisointien ensimmäinen kulta-aika oli jo 1800-luvun lopussa ja toinenkin maailmansotien välissä 1918–1939. (Friendly 2009) Graafinen esitystapa kuitenkin väheni toisen maailmansodan jälkeen ja yleisty seuraavan kerran vasta 80-luvulla. (Kuusela 2000, 26–27)

### YMPYRÄDIAGRAMMI

Ympyrädiagrammi on yksi yleisimmistä ja tunnetuimmista kuvio-malleista, käytännössä se soveltuu ainoastaan prosenttijakauman esittämiseen. Kuvio on suosittu, koska sitä pidetään visuaalisesti kiinnostavana. Kuuselan mukaan sen ongelmana on kuitenkin epätarkkuus. Havaintopsykologisten tutkimusten mukaan osa ihmisistä vertaa kaaren pituutta, osa jänteen pituutta ja osa sektorin keskuskulmaa, oikean vertailukohteen eli pinta-alan sijasta. (Kuusela 2000, 26–27)

Ympyrädiagrammin lukemista voi helpottaa sijoittamalla lohkot suuruusjärjestykseen siten, että suurin arvo alkaa kello kolmesta ja seuraavat arvot kiertyvät siitä vastapäivään. Enimmillään ympyrädiagrammissa saisi olla kuusi sektoria. Kuuselan mukaan joissain tapauksissa jo vähempikin lohkojen määrä aiheuttaa sekavuutta. (Kuusela 2000, 26–27)



Ympyrädiagrammin vertailu tapahtuu jänteen tai kaaren pituuden mukaan, joskus jopa keskuskulmaa arvioimalla.

### VIIVAKUVIO

Viivakuviot sopii hyvin kuvaamaan jonkin ilmiön kehityssuuntaa tai trendiä. Viivakuvion keskeisin vaatimus on, että kummallakin akselilla on jatkuva-arvoinen ilmiö. Yleensä x-akselille sijoitetaan aika ja y-akselille määrä. (Kuusela 2000, 77–79)

Ainoa poikkeus on syy-seuraussuhteiden esittäminen viivakuviossa, jolloin x-akselille voidaan sijoittaa syy ja y-akselille seuraus. (Kuusela 2000, 78)

Viivakuvion visualisoinnissa keskeisintä on valita oikea tarkkuusaste. Viivakuviot koostuu koordinaatistoon merkityistä pisteistä ja niiden väliin piirretyistä viivoista.



Viivakuviot syntyvät piirtämällä tietoyksiköiden väliin viiva. Jos yksiköitä on liian harvassa saattaa viiva aiheuttaa tulkintavirheitä.

Kuvion tulkinnaissa on syytä muistaa, että dataa sisältävät ainoastaan pisteet. Pisteiden piirretty viiva kuvaa kuviteltua kehitystä. Tämä voi olla kuvioista tehtävien johtopäätösten kannalta ongelmallista. Jos aineiston tietoalkiot ovat ajallisesti harvassa niistä ei ole mahdollista lukea lyhytaikaisia poikkeamia. (Kuusela 2000, 77–79)

Tietoja yhdisteltäessä yhteen taulukkoon kannattaa varmistua siitä että kaikki taulukkoon sijoitettava data käyttää samaa asteikkoa. Jos tiedot käyttävät kahta eri asteikkoa on selkeää tehdä kaksi eri taulukkoa.

Kuvioon ei myöskään kannata mahdollistaa liian montaa viivaa. Kuvio voi olla täynnä ja vaikeaselkoinen jo kahdestakin viivasta, jos ne risteävät usein tai menevät päällekkäin.

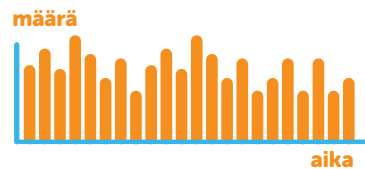
Riittävän suurella visuaalisella kontrastilla eli esimerkiksi värin tai koon muutoksilla viivat voidaan erottaa toisistaan kuvion kaikissa kohdissa. (Kuusela 2000, 78)

Viivakuviosta käytetään myös erilaisia muunnelmia, joista tunnetuimmat ovat alue- ja summaviivakuviot. Alueviivakuviossa on yksi viivakuvio, jonka alla oleva tila on täytetty värialueella. Summaviivakuvaajassa alueita on useampia. Muunnelmien käyttöä perustellaan usein sillä, että viivojen alapuolen täytöllä voidaan saavuttaa kuviolle visuaalista lisähuomiota. (Kuusela, 2000, 97)

Muita muunnelmia ovat erotus- ja nettoerotustyyppiset viivakuviot. Ne koostuvat kahdesta viivasta, joiden väliin jäävä alue korostetaan. Kuviolla voidaan verrata ilmiöiden vaihteluvälejä. (Kuusela 2000, 99)

### PYLVÄSKUVIOT

Pylväskuvioita on kahta pääalajia, pysty- ja vaakapylväikköjä. Pystypylväskuvio sopii parhaiten jatkuva-arvoisen ilmiön, esimerkiksi aikasarjan esittämiseen. Vaakapylväikkö soveltuu epäjatkuvaa arvoisten muuttujien esittämiseen. (Kuusela 2000, 124)



Pysty- ja vaakapylväikkö eivät ole vaihtoehtoja toisilleen. Pystypylväikkö soveltuu hyvin määrä- ja aikatiedon esittämiseen.



Vaakapylväikkö soveltuu usean eri kategorian ja määrätiedon esittämiseen.

Esimerkiksi työttömyyden kehitystä voidaan tutkia eri tavoin pysty- ja vaakapylväiden avulla. Jos pystypylväskuvioon sijoitetaan x-akselille kuukaudet ja y-akselille työttömien määrä, niin voidaan tarkastella työttömyyden muutoksia ajassa.

Vaakapylväikkö soveltuu paremmin esimerkiksi ikäluokkien välisen työttömyysmäärän vertailuun. Tällöin y-akselille asetetaan ikäluokka ja x-akselille työttömien määrä.

Pylväiden tasaleveys on tärkeää. Lukija vertailee pylväitä pylväiden pinta-alaa arvioimalla ja jos pinta-ala eroilla ei haluta kertoa jotain aineistoon perustuvaa tietoa, niin pylväiden ja niiden välien tulee olla keskenään yhtä leveitä. (Kuusela 2000, 115)

Pylväikön määräästeikko on aloitettava aina nolasta. Pystypylväikössä tämä on ainut mahdollinen asteikko, mutta myös muissa pylväsmalleissa tästä poikkeamiseen tulee olla vahvat perustelut. Jos asteikko alkaa jostain muusta arvosta kuin nolasta, se hankaloittaa ja voi vääristää tulkintaa. (Kuusela 2000, 121–136)

Erilaisia pylväskuvion variaatioita on paljon, tunnetuin on luultavasti pylväsryhmäkuvio, jossa pylväitä ryhmitellään samaan kuvioon. Ryhmitely on hyödyllistä pienissä määrissä, mutta usein yli kolmen pylvään ryhmät ovat jo vaikeasti tulkittavia ja vertailtavia. (Kuusela 2000, 121–136)



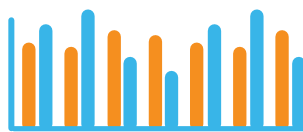
*Jos pieni osa pylväistä on huomattavasti pidempiä tai lyhyempiä kuin muut, ne voi katkaista ja erotella ne omaksi kuvioiksi.*

Summapylväskuviossa pylväaseen sijoitetaan useampia arvoja. Netto- ja bruttopoikkeamakuviossa esitetään myös negatiivisia arvoja niin, että pylväät piirretään nollakohdan ylä- ja alapuolelle. (Kuusela 2000, 121–136)

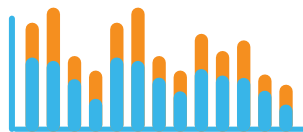
Kelluvassa pylväikössä jokainen pylväs piirretään suurimman ja pienimmän arvon välille, jolloin pylväät

korostavat kuinka ilmiön positiiviset ja negatiiviset arvot prosentuaalisesti jakautuvat. (Kuusela 2000, 121–136)

Vaihteluvälipylväikössä pylväiden huippuarvojen välille on merkitään keskiarvo, jolloin ilmiön jakauman esittäminen korostuu. (Kuusela, 2000, 121–136)



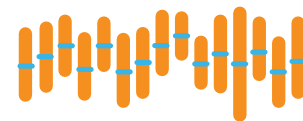
**Pylväsryhmäkuvio helpottaa kategoriatiedon esittämässä.**



**Summapylväskuvio sopii prosentuaalisen jakauman esittämiseen.**



**Nettopoikkeamakuvion avulla voidaan esittää myös negatiivisia arvoja.**

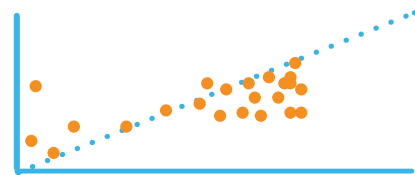


**Vaihteluvälipylväikkö esittää vaihteluvälin lisäksi keskiarvon.**

Pylväskuvio voidaan korvata myös helminauhakuviolla (Cleveland dot plot). Kuviota käyttää runsaasti muun muassa The Economist-lehti.

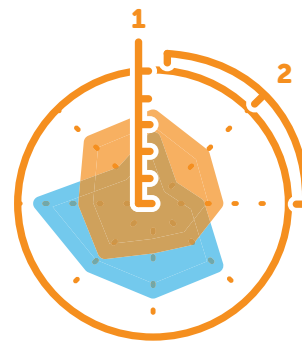
## PARVIKUVIO

Parvikuvio on yksinkertainen, tehokas ja harmillisen vähän nähty visualisointi. Parvikuviossa jokaisella tietoyksiköllä on kaksi arvoa jonka mukaan ne sijoitetaan koordinaatiston x- ja y-akseleille. Kuvio on tehokas, koska ihmissilmä on hyvä huomaamaan pisteiden muodostamat kuviot, "parvet" sekä poikkeamat aineistossa. (Kuusela 2000, 70)



**Mitä tarkemmin pisteet sijoittuvat kuvion lävistäjälle, sitä voimakampi on ilmiöiden välinen riippuvuus.**

Polaarikoordinaatisto on parvikuvion ympyränmallinen sovellus, jossa yksi arvoasteikko lähtee kuvion keskeltä (nollapistestä) kohti kuvion reunaa ja toinen kuvion ylimmästä kohdasta asteittain alaspäin. Polaarikoordinaatisto soveltuu hyvin toistensa poissulkevien asioiden kuvaamiseen. Esimerkiksi jonkin tapahtuman yleisyys eri vuorokaudenaikoina sopii polaarikoordinaatistoon. (Kuusela 2000, 70)



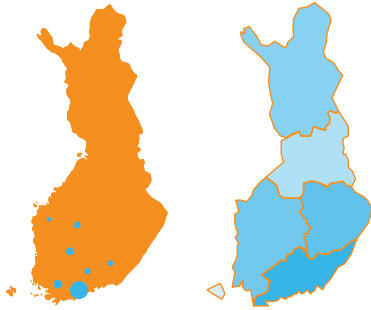
**Polaarikoordinaatiston ensimmäiselle asteikolle voidaan sijoittaa määrätietoa ja toiselle ajanjakso, esimerkiksi tunnit.**

## KARTAT

Jos tilastollisella tiedolla on arvo myös maantieteellisesti, karttavisualisointi voi olla hyvä tapa esittää tieto. Kartat voivat perustua alueluokitukseen (esim. kunnat) tai yhteen maantieteelliseen sijaintiin (esim. ka-tuosoite).

Pistesymbolikartalla symbolit sijoittuvat yhteen maantieteelliseen sijaintiin ja ne voivat olla keskenään samankokoisia tai niiden koolla kerrotaan kyseiseen paikkaan liittyvän arvon suuruutta. (Kuusela 2000, 161)

Alueluokitukseen perustuvia karttoja kutsutaan koropleettikartoiksi. Niissä aineiston erot esitellään kartalla alueittain tummuusastevaihtelun avulla. Visualisointi toteutetaan niin, että esitettävä tieto luokitellaan ja luokittelun perusteella kullekin alueelle määritellään tummuusaste.



*Pistekartan (vas.) pisteet välittävät koolaan määrätietoa. Koropleettikartta (oik.) perustuu aluejakoon ja sen tieto välittyy väri- tai tummuusastevaihtelun avulla.*

Koropleettikartalle ominaista on se että sillä voidaan esittää vain yhtä asiaa kerrallaan, ja sen avulla voidaan tarkastella vain yhden arvon suhteellista vaihtelua. Ongelmia koropleettikartan käytössä on:

- Pinta-alaraha. Alueellisesti suuret, mutta väkimäärältään pienet alueet korostuvat kartassa.
- Alueiden sisäinen vaihtelu ei tule esille. Alueen sisällä voi olla suuriakin vaihteluita, jotka eivät pääse esille.
- Rajaamisharha. Vaikka muutokset ovat usein asteittaisia, voi alueiden väliset tarkat rajaukset esittää erot dramaattisimpina kuin ovatkaan.

*(Kuusela, 2000, 161)*

Ruutukarttaa käytetään kuten koropleettikarttaa, mutta siinä kartan alueellisuus perustuu samansuuruisiin yksikköihin, esimerkiksi neliökilometreihin, jotka sävy- tai tummuusvaihtelun avulla välittävät tietoa aineistosta. *(Kuusela, 2000, 161)*

Pistetrieheyskartta on samankoisista pisteistä muodostuva kartta, jossa pisteiden tiheys indikoi jonkun tietyn ilmiön yleisyyttä jossain maantieteellisessä sijainnissa. *(Kuusela, 2000, 161)*

Virtauskartoilla esitetään eri maantieteellisten alueiden välillä tapahtuvaa liikettä. Liikettä voidaan esittää nuolilla viivoilla tai virtausnauhoilla. Symbolit voivat esittää tietoa virtojen suuruudesta, laadusta tai suunnasta. *(Kuusela 2000, 161)*

### VERKOSTOKUVIOT

Verkostoanalyysi tutkii ilmiöiden välisiä suhteita ja niiden riippuvuutta toisistaan. Usein verkostoanalyysin lopputuotoksena on kuvallinen esitys tutkittavien yksiköiden suhteista.

*(Jyväskylän Yliopisto 2013)*

Tutkimuskohteina voi olla esimerkiksi ihmisten tai organisaation välisiä suhteita, tapahtumien välisiä ajallisia suhteita tai ilmiöiden välisiä suhteita *(Jyväskylän yliopisto 2013).*



*Tämä verkostoanalyysikuva esittää Facebook-kavereideni keskinäisiä sidoksia. Jokainen piste edustaa yhtä kaveria ja välille piirretyt viivan kaverein sidoksia muihin. Isossa tiiviissä pallossa alhaalla löytyvät työkaverini, jotka ovat keskenään tiiviisti Facebook-yhteydessä toisiinsa. Ylhäällä oikealla koulukaverit ja keskeillä harrastuskavereita, joilla on vähemmän keskinäisiä yhteyksiä.*

Verkostoanalyysien visuaalista mallintamista tehdään aiheeseen erikoistuneilla ohjelmilla, jotka osaavat yksinkertaisella tasolla jopa tulkita ja luokitella lähdeaineistoa. Esimerkiksi Facebook-verkostokuvio on tuotettu suosituilla avoimen lähdekoodin verkostoanalyysiohjelmalla.

### TEKSTIN VISUALISOINTI

Tekstin visuaalinen olemus eli koko, muoto, sijainti ja värit voivat viestittää sanallisen ilmaisun lisäksi myös numeraalisia arvoja.

Sanapilvi-tyyppisissä visualisoinneissa esitetään usein sanan yleisyyttä tekstikokoja kasvattamalla. Sanoja voidaan kategorisoida eri teemojen alle esimerkiksi tekstin väriä muuttamalla. *(Halvey&Keane 2007)*



*Tässä esimerkkikuvassa on Wordle-palvelussa luotu sanapilvi kotisivuni sisällöstä. Eniten käytetyt sanat esiintyvät kuviossa suurimmalla pistekoolalla.*

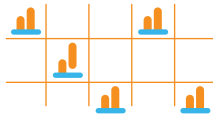
Infografiikkaan erikoistunut graafinen suunnittelija **Juuso Koponen** kritisoi yllä kuvatun tyyppistä visualisointia harhaanjohtavaksi. Hän erittelee kuvasta viisi tietoa välittävää ominaisuutta. Sanojen sijainti, pituus, lukusuunta, pinta-ala ja väri. Kuitenkin yllä kuvattu sanapilvi käyttää vain pinta-alatietoa. *(Koponen 2011)*

Sanapilviä-tyyppisiä visualisointeja on yleisemminkin pidetty epäluotettavina. Parempia keinoja tekstin visualisointiin löytyy esimerkiksi verkosto-analyysityökaluista, joissa sanojen välisiä suhteita visualisoidaan kuten oheisessa Facebook-kaveri esimerkkikuvassa.

MUUTTUVALEVEYKSI-  
NEN  
PYLVÄSDIAGRAMMI



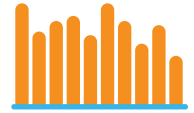
KUVIORYHMÄ



VAAKAPYLVÄS  
DIAGRAMMI



PYSTYPYLVÄS  
DIAGRAMMI



PALJON TIETOYKSIKÖITÄ

VÄHÄN TIETOYKSIKÖITÄ

MONTA KATEGORIAA

VÄHÄN KATEGORIOITA

KAKSI MUUTTUJAA  
PER TIETOYKSIKKÖ

YKSI MUUTTUJA  
PER TIETOYKSIKKÖ

KATEGORIAMUUTTUJAT

VERTA



PARVIKUVIO

YKSI  
MUUTTUJA

SUHTEITA



YMPYRÄKUVIO

KAKSI  
MUUTTUJAA

Mitä h  
visuali

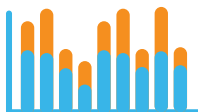
MUUTOKSIA AJASSA

KOKONAI

VÄHÄN  
AJANJAKSOJA

USEITA  
AJANJAKSOJA

SUHTELLISIA  
EROJA



SUMMAPYLVÄS-  
DIAGRAMMI

SUHTEELLISIA JA  
ABSOLUUTTISIA EROJA



TÄYTETTY Pylväs-  
DIAGRAMMI

SUHTELLISIA  
EROJA



100% TÄYTETTY  
VIIVADIAGRAMMI

SUHTEELLISIA JA  
ABSOLUUTTISIA EROJA



TÄYTETTY  
VIIVADIAGRAMMI

YMPYRÄDIAGRAMMI



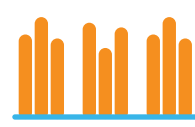
SYKLINEN AIKA

VIIVADIAGRAMMI



EI SYKLINEN AIKA

PYLVÄSRYHMÄ  
DIAGRAMMI



VÄHÄN KATEGORIOITA

VIIVADIAGRAMMI



PALJON KATEGORIOITA



ILUA

aluaat  
soida?

EROJA

YKSI  
MUUTTUJA

VÄHÄN  
TIETO-  
YKSIKÖITÄ



PYLVÄSDIAGRAMMI

PALJON  
TIETO-  
YKSIKÖITÄ



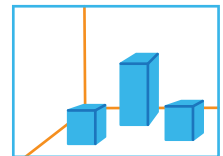
VIIVADIAGRAMMI

KAKSI  
MUUTTUJAA



PARVIKUVIO

KOLME  
MUUTTUJAA



3D PYLVÄS-  
DIAGRAMMI

SUUTTA

YHDEN HETKEN TILANNE

KOKONAISUUDEN  
JAKAUTUMINEN



YMPYRÄDIAGRAMMI

KOKONAISUUDEN  
KEHITTYMINEN



VESIPUTOUS-  
DIAGRAMMI

USEAN KOKONAISUUDEN  
KEHITTYMINEN



100% PYLVÄSDIA-  
GRAMMI

# 3

## Tekniikka ja design

Verkkototeutuksia joudutaan harmillisen usein suunnittelemaan tekniikan – tai vielä useammin, teknisen osaamisen ehdoilla. Verkkovisualisointien tekemiseen on tarjolla valtavasti erilaisiin visualisointitehtäviin erikoistuneita työkaluja. Esittelen kokoamillani esimerkkisivuilla erilaisia visualisointityökaluja ja niiden käyttöä.

### 3.1 Tekniikan ehdot

Verkossa julkaistavaa infografiikkaa selataan valtavalla määrällä erilaisia laitteita. Laitteiden käyttöliittymät voivat toimia hiirellä, kosketuksella tai kynällä. Lisäksi näyttökokoja ja kuvasuhteita on valtava variaatio. Pöytäkoneet (desktop), tablet-laitteet (tablet) ja älypuhelimet (smartphones) ovat yläkategorioita, joiden sisään mahtuu suuri skaala eri kokoja, ominaisuuksia ja käyttötapoja. Näitä eri käyttöliittymävariaatioita yrittään hallita muutamalla ajankohtaisella suunnitteluperiaatteella, joista viime vuosina puhutuimmat ovat responsive design ja mobile first -suunnitteluperiaatteet.

**Responsive designilla** tarkoitetaan sisällön skaalautumista näyttökoon ja käyttöliittymän mukaan (Marcotte 2011). Teknisesti on mahdollista tunnistaa minkälaisella laitteella käyttäjä lukee sivuja ja tarjota sisältöä kyseisen laitteen ominaisuuksiin sopivassa muodossa.

Koska laite ja käyttöliittymäyhdistelmiä on satoja, ei käytännössä ole mahdollista suunnitella kaikille omaa näkymää vaan usein suunnitellaan näkymät ja toiminnallisuudet niin, että ne toimivat pöytäkoneiden isoilla näyttöillä ja hiiren avulla, tablettien keskikokoisissa kosketus- ja kynäohjatuissa näytöissä sekä puhelinten pienillä kosketus tai tikkuohjatuilla näyttöillä.

Jos jokin laite on erityisen suuressa osassa tilastoja edustettuna (usein esimerkiksi iPhone ja iPad) voidaan tarjota jopa laitekohtaista käyttöliittymää. (Marcotte 2011)

**Mobile First** -suunnitteluperiaate ohjaa suunnittelijaa aloittamaan suunnittelun pienimmästä näyttökoosta, jolloin visualisoinnissa valitaan tärkein sisältö ensin ja suuremmille näyttökoille voidaan tarjota täydentävää sisältöä tai kuvituksellisia elementtejä. (Wroblewski 2011)

Verkkovisualisointien suunnittelijalle on tärkeä tiedostaa edellä esitelty suunnittelumetodit, sillä yhä useammin mediatilat haluavat tuottaa sisältöä kaikille laitteille.

Erityisesti isoja data-aineistoja visualisoitaessa älypuhelimien ruudun näyttöpinta-ala, hitaat yhteysnopeudet ja rajoittunut prosessointiteho on haastava julkaisualusta visualisoinneille. Joissain tapauksissa voi olla järkevää keskittyä teknisistä-, aika- ja kustannussyistä vain pöytäkoneiden näyttöille suunniteltavaan visualisointiin.

### JULKAISUJÄRJESTELMÄT

Käytännössä uutisten julkaiseminen verkkosivuilla toimii mediatalon käyttämän julkaisujärjestelmän avulla. Järjestelmä on yhteydessä toimitusjärjestelmään, johon toimittajat ja kuvaajat tuottavat sisältöä. Toimittajilla voi olla tämän lisäksi oma sisäl-

lönsyöttönäkymä suoraa verkkojulkaisua varten. Julkaisujärjestelmän avulla sisältö voidaan esittää median verkkosivuilla halutussa muodossa, ilman että jokaista uutista tarvitsee erikseen muokata internetselaimen ymmärtämään muotoon.

Koska sisällönhallintajärjestelmät toimivat tekstin ja kuvien julkaisemisen ehdoilla, kaikki siitä muodostaa poikkeava sisältö on usein lisättävä sivuille toisella tavalla.

Ulkopuolista sisältöä, tässä tapauksessa visualisointeja lisätään yksittäisen verkkosivun sisään pääsääntöisesti Iframe tai JavaScript upotuksina. Iframe-upotus toimii kuin ikkuna verkkosivun sisällä. Iframele määritellään paikka ja koko, jonka sisällä näytetään toisen verkko-osoitteen sisältöä. JavaScript-upotuksessa voidaan monipuolisesti valita mitä osia tai sisältöjä ulkopuolisesta lähteestä halutaan sivuilla esittää.

Näiden lisäksi on monenlaisia muita upotustapoja, joiden käyttö riippuu pitkälti toimitusjärjestelmästä, julkaisujärjestelmästä tai visualisoinnin käyttämästä tekniikasta. Paras upotustapa valitaan tapauskohtaisesti ja yleensä lehden julkaisutekniikkaa tuntevien tahojen kanssa.

### FORMAATIT

Visualisoinnit voivat olla teknisesti näkökulmasta katsottuna monessa eri formaatissa. Ne voivat olla staatti-

sia kuvia, Svg- tai Canvas-elementtejä tai Flash-muotoisia mediaesityksiä.

**Staattiset kuvat** ovat yleisin ja tuuin visualisointien julkaisuformaatti. Niitä voidaan liittää sivustoille vaivatta julkaisujärjestelmän perustoimintojen kautta. Staattisten kuvien hyviä puolia on niiden helppo skaalaaminen eri näyttökoille.

Usein julkaisujärjestelmien kuva- paikka on kuitenkin liian pieni infografiikka esittäville kuville ja siksi kuvagrafiikkakin voi toimia linkkinä koko sivun levvyiseen visualisointiin joka sijaitsee julkaisujärjestelmän ulkopuolella.

Kuvien puute teknisessä mielessä on se, että niihin ei pysty lisäämään interaktiivisia toiminnallisuksia ja liikettäkin vain hyvin rajatussa animated-gif muodossa.

Esimerkiksi toteuttamani verkostonalyysikaavio on niin laaja, että yhdenkään lehden verkkosivujen kuvapaikka ei tee oikeutta sille, vaan parhaimmillaan se on isona, ruudunkokoisena toteutusena.

**Flash-teknologia** on ollut verkkojulkaisujen historiassa perinteisesti se tapa, jolla selaimien ominaisuuksiin liittyvät rajoitukset on kierretty ja näyttävät interaktiiviset toteutukset on lisätty verkkosivuille.

Nykyisin laitteistokirjon kasvettua Flash-tekniikalla toteutetut sovellukset eivät kaikilla laitteilla aukea katsottaviksi tai käytettäviksi. Erityisesti koska Applen mobiililaitteissa (iPad, Iphone, iPod) Flash-teknologiaa ei tueta, ovat Flash-teknologialla toteutetut infografiikat kadonneet jo lähes täysin. Poikkeuksena ovat kuitenkin verkossa julkaistavat videot, joista valtaosa perustuu Flash-tekniikkaan.

Applen laitteiden Flash-tekniikkaan liittyvät rajoitukset kierretään julkaisemalla videomateriaali YouTube-palvelun kautta, jolle Apple on rakentanut erityisen tuen. Omis- sa esimerkeissäni en käsittele Flash-muotoista infografiikkaa enkä videoinfografiikkaa.

**Svg- ja Canvas-elementit** ovat kiinnostavin ja nykyaikaisin tapa toteuttaa verkkovisualisointeja. Svg-muoto on suhteellisen vanha tiedostomuoto, mutta vasta viimeisien vuosien aikana selainvalmistajat ovat alkaneet tukea sitä laajemmin. Formaatin etuna on, että sitä voi suurentaa ja pienentää laadun kärsimättä. Lisäksi Svg-elementin sisältöä voi animoida ja sen sisään voi sijoittaa interaktiivisia toimintoja.

Canvas-elementti on yksi Html5-merkintäkielen uusista ominaisuuksista. Html-merkintäkieltä kehitetään jatkuvasti ja selainvalmistajat pyrkivät ottamaan kielen uusia ominaisuuksia mahdollisimman varhaisessa vaiheessa käyttöön julkaisemalla uusia selainversioita.

Canvas-elementti on Svg:tä vastaava, vektoreihin perustuva visualisointien toteutustapa. Sitä voi skaalata, animoida ja sillä voi toteuttaa interaktiivisia visualisointeja.

Käytännössä sekä Svg- että Canvas-elementit ovat numeromuotoisia tiedostoja, jotka sisältävät kuvaelementin koordinaatteja. Näitä koordinaatteja käyttäjän selain tulkitsee ja piirtää ruudulle kuvaksi.

Molempien muotojen ongelmana on se, että kaikki selaimet eivät osaa lukea näitä formaatteja. Osa selaimista tukee toista muotoa ja osa toista.

Käytännössä visualisoinnin toteuttamisen alkuvaiheessa täytyy kuitenkin valita kumpaa tekniikka käytetään, koska eri toteutustavat vaativat erilaisen rakenteen ohjelmakoodiin. Lopputulosta ei voi tallentaa kuten kuvaa useammassa eri muodossa, vaan visualisoinnin toteuttajan tulee päättää kuinka hallitaan ne tilanteet, jolloin käyttäjän selain ei osaa lukea visualisointia. Molemmissa tekniikossa on omat hyvät puolensa ja paljon valmiita ratkaisuja. Lopullinen visualisointi ja valitut jakelualustat, sekä oman osaamisen painottuminen kuitenkin määrittävät toteutuksessa käytettävän tekniikan.

## TOTEUTUSVAIHTOEHTOJA

Perusgraafien toteuttamisessa on paljon valmiita vaihtoehtoja. Itse olen keskittynyt opettelemaan JavaScript-kirjastojen käyttöä, joista laajin ja monipuolisin D3-kirjasto on vienyt valtaosan ajastani. Muita vaihtoehtoja on lukemattomasti, ja jokaiselle löytyy varmasti luontevimmat käyttötapsansa ja tilanteensa. Alla lyhyt lista erilaisia teknisiä toteutusvaihtoehtoja.

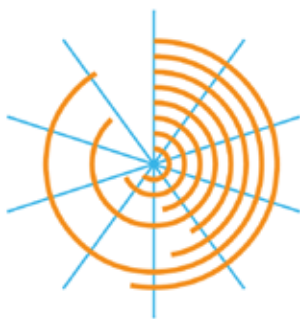
**Google Charts** on verkkovisuaalisointien tekijöille varmaankin tunnetuin helpokäyttöinen vaihtoehto, josta löytyy paljon erilaisia kuviomalleja. Oletusvisualisoinnit eivät ole erityisen viehättäviä, mutta ulkoasua ja toiminnallisuksia on mahdollista kustomoida. Käytännöllisimmillään Google Charts toteutukset ovat kertajulkaistaviin toteutuksiin. Kirjaston visualisointien ja kustomointien tekeminen vaatii syvempää perehtymistä, johon Google tarjoaa testiympäristön ja kattavan ohjeistuksen.



GoogleCharts-kirjastolla toteutettu esimerkkivisualisointi verkossa osoitteessa: [demo.newpie.fi/googlecharts.php](http://demo.newpie.fi/googlecharts.php).

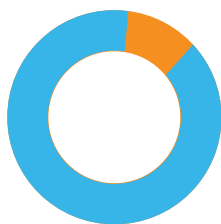
**Chart.js**-JavaScript-kirjasto tuottaa Canvas-elementtiin perustuvia visualisointeja. Sen vahvuutena on tyylikäs oletusulkoasu ja yksinkertaisuus. Kirjastossa on seitsemän oletusmallista kuviomuotoa, joissa kaikissa on tyylikäs ja kevyt animointi. Kuvion ulkoasua pystyy muokkaamaan, mutta interaktiivisuutta kuvioon ei tue.

**Chartwell-webfontti** on muista poikkeava ratkaisu visualisointien toteuttamiseen. Verkkosivuille syötetään data-aineisto normaalisti numeroina ja määritellään numeroille ulkoasuksi kyseinen fontti. Älykäs fontti luo numeroaineistosta visualisoinnin. Tällä tekniikalla yksinkertaisesta numeroaineistosta saa nopeasti toteutettua visualisoinnin. Animointi tai interaktiiviset toiminnot eivät ole tässä ratkaisussa mahdollisia.



Esimerkkitoteutus Chartwell-kirjasintyyppin verkkokäytöstä löytyy osoitteesta [demo.newpie.fi/chartwell.php](http://demo.newpie.fi/chartwell.php).

**D3.js** on kustomointimahdollisuuksiltaan erinomainen, laaja ja suorituskykyinen JavaScript-kirjasto. D3:n käyttö vaatii opettelua, mutta sen hallinta avaa valtavasti mahdollisuuksia yksilöllisten visualisointien toteuttamiseen. Kirjaston toiminta-ajatus poikkeaa muista kirjastoista siinä, että sen käytössä jokainen visualisoinnin osa kirjoitetaan erikseen visualisointia tehdessä. Kirjasto tarjoaa hyvät työkalut visualisointien tekemiseen, mutta samalla vaatii enemmän javascript osaamista kuin muut kirjastot.



D3-kirjaston käyttöesimerkki verkossa: [demo.newpie.fi/d3.php](http://demo.newpie.fi/d3.php).

### 3.2 Datan siivoaminen

Käytännössä visualisointityö alkaa ideoinnin ja data-aineiston hankinnan jälkeen datan siivouksella. Alkuperäinen data on tuskin koskaan visualisoinnille sopivassa muodossa. Usein data-aineistosta on valittava visualisoinnissa tarvittavat rivit ja sarakkeet, poistettava tyhjiä rivejä ja muokattava soluja oikeisiin ja yhtäläisiin tietomutoihin.

Datan muotoilua ja siivousta esityskuntoon tehdään pääosin taulukkolaskentaohjelmissa, joista monipuolisoin on **Microsoft Excel**. Ohjelmalle on ilmaisia vaihtoehtoja kuten Libre office tai verkkopohjaisia toteutuksia, kuten **Google Spreadsheets**. Data-aineistojen siivoamiseen erikoistuneista ohjelmista erityisesti Googlen ilmainen sovellus **Google Refine** on erinomainen työkalu.

Datan siivoamisen jälkeen on päätettävä missä formaatissa data-aineisto tallennetaan. Yksinkertaisimmillaan data voidaan tallentaa pilkuilla erotettuna numerojonona. Data voi olla myös rakenteellisessa muodossa, jolloin data-aineistossa on useita rivejä ja sarakkeita. Tällöin tallennettavan aineiston tiedostomuoto on yleensä .tsv tai .csv.

Monipuolisimmille aineistoille paras tiedostomuoto on Json tai xml-formaatit. Niiden rakenne on monikerroksellinen. Json- tai xml-muotoista dataa luettaessa JavaScript-ohjelmassa voidaan tarkasti määritellä mitä data-aineiston osaa halutaan visualisointiin milloinkin lukea ja määrittely voidaan tehdä vaikka yksi solu kerrallaan.

Kun data-aineisto on siivottu ja valmis visualisointia varten, se sijoitetaan palvelimelle. Data-aineisto voitaisiin sijoittaa suoraan visualisointia tuottavan JavaScript-koodin sekaan, tai visualisointia esittävän html-dokumentin rakenteeseen. Usein kuitenkin kannattaa sijoittaa data-aineistot omaan tiedostoon selvyden vuoksi. Erilliseen tiedostoon

voidaan viitata visualisointia tuottavassa JavaScript-koodissa.

Erillinen data-aineiston tallennuspaikka on käytännöllistä visualisoinnin jatkokäytön kannalta. Aikaisemmin luotua visualisointia voidaan käyttää sellaisenaan uusilla data-aineistoilla, koskematta alkuperäiseen ohjelmakoodiin. Kehittyneimmissä ratkaisuissa data-aineistoa luetaan jopa reaaliaikaisesti datan tuottajan palvelimelta, jolloin julkaistu visualisointi päivittyy kun datalähteeseen lisätään uusia lukuja.

### 3.3 Brändielementit

Vaikka perusmallisen tilastokaavion visualisointi muodostuu valtaosaltaan data-aineistosta, jää visualisoinnin ulkoasuun paljon lehden identiteettiä korostavia elementtejä. Värien, typografian ja viivatyörien sääntönmukaisella käytöllä voidaan tukea lehden visuaalista maailmaa.

Värien käyttö visualisoinnissa perustuu kontrasteihin. On tärkeää esittää kuvion erot riittävän suurina kontrasteina, jotta kuvion elementit eivät sekoitu toisiinsa. Toisaalta valituilla väreillä on helppo tehostaa lehden muussa visualisoinnissa käytettävää värimaailmaa ja tukea siten lehden visuaalista linjaa. Esimerkiksi demovisualisoinneissa olen rakentanut kolmesta eri kuviotyypistä kolme visuaalista variaatiota.

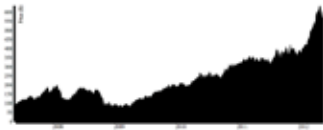
#### VÄRIT

Ensimmäinen variaatio esittää visualisoinnin ilman ulkopuolisia tyyllitelyjä. Kuvio toimii vertailukohtana sille miten erilaiset visualisointielementit vaikuttavat kuvion oleukseen. Kuviossa on vain harmaasävykontrasteja, joiden avulla eri tietoelementtien eroja voi verrata väriin vaihtoehtoihin.

Toinen variaatio on monivärinen ja voimakkaita brändielementtejä esittävä malli. Toisen variaation kuviot esittävät kuinka tietoelementtien kontrastierot toimivat värien avulla.



Kolmas variaatio esittää hillitymän ja yleiskäyttöisen tyylin. Kolmannessa variaatioissa kuvioiden värierot toteutetaan yhden värin sävyvaihteluina.



Variaatio 1 esittää visualisoinnin raaka-muodossa eli ilman tyyliasetuksia.



Variaatio 2 esittää visualisoinnin voimakkailla tyyliasetuksilla.



Variaatio 3 esittää visualisoinnin perinteisillä tyyliasetuksilla.

Esimerkkivisualisoitinta löytyvät verkko-osoitteista:

[demo.newpie.fi/brandi-viiva.php](http://demo.newpie.fi/brandi-viiva.php)

[demo.newpie.fi/brandi-pylvasryhma.php](http://demo.newpie.fi/brandi-pylvasryhma.php)

[demo.newpie.fi/brandi-kalenteri.php](http://demo.newpie.fi/brandi-kalenteri.php)

## TYPOGRAFIA

Visualisointien typografista viimeistelyä ja kirjaintyyliä on verkossa hie-man monimutkaisempaa toteuttaa kuin paperijulkaisuissa.

Ensimmäisessä ulkoasuvariaatiossa kuviossa käytettävää typografiaa ei ole erikseen määritelty. Tästä syystä jokaisen käyttäjän selain käyttää oletusasetuksia näyttäessään ensimmäisen visualisoinnin typografiaa. Esimerkiksi Mac-koneelle asennetun Chrome-selaimen oletuskirjasintyyli on Helvetica.

Toisessa variaatioissa visualisoinnin typografia on voimakkaan ilmai-

sullista ja luo voimakasta visuaalista identiteettiä. Otsikon sript-tyylinen kirjasin (Lobster) ja asteikoissa epätyypillinen päätteellinen kirjaintyyppi (Tisa pro) pyrkivät erottautumaan yleisesti käytetyistä ratkaisuksista ja vahvistamaan yksilöllistä visuaalista identiteettiä.

Kirjasimet on upotettu sivulle Typekit-kirjastosta. Tällä yleisesti käytetyllä tekniikalla Typekitin vuosimaksua maksavat suunnittelijat voivat käyttää sivuillaan laadukkaita kirjaintyyppiä niiden lisenssiehtojen mukaisesti.

Kirjasintyyppiä voi lisätä sivulle lataamalla ne itse verkkoon. Ongelmana tällaisessa ratkaisussa on se, että kirjasinperheet voi kuka tahansa osaava lukija kopioida itselleen. Kaikki kirjasinlisenssit eivät hyväksy tällaista kirjainten verkkokäyttöä. Itse olen tästä syystä tottunut käyttämään edullista ja lisenssin suhteen huolellista Typekit-palvelua.

Kolmannessa variaatioissa typografia on hillittyä ja perinteistä. Typografiassa käytetään käyttäjän koneella olevia kirjaimia, mutta niiden kirjasinperheet ja koot on määritelty tarkasti. Tätä niin sanottua web-sa-fe-kirjasintyyppien käyttöä suositetaan, koska kirjainten upottaminen tai tuominen sivuille kasvattaa hie-man sivujen latausaikaa. Ongelma korostuu erityisesti silloin, jos tarvitaan latinalaisen aakkoston lisäksi muita, esimerkiksi kyrillisiä kirjaimia. Tällöin ladattavan kirjaintiedoston koko helposti jopa kolminkertaisuus.

Joskus ongelmana on myös se, että kirjainten rajat eivät toistu tuoduissa kirjaintyypeissä yhtä terävästi, kuin käyttäjän koneen omia kirjaimia käytettäessä.

## VIIVATYYLIT

Viivatyyliä visualisoinnissa ovat voimakas tyylillinen keino luoda yksilöllistä visuaalista ilmettä. Viivatyylien toistuminen näytöllä vaihtelee suuresti käyttäjän laitteistosta. Suu-

riresoluutioiset näytöt pystyvät toistamaan tarkasti ohuitakin viivoja, toisaalta taas osa näytöistä on niin heikkolaatuisia että hienojakoisimpia vaihteluita viivojen tyyliä ei niissä erota.

Selaimissa on myös oletuksena päällä reunapehmenys (antialiasing)-ominaisuus, joka pehmentää kaarevien pikseliviivojen reunoja niin, että pikselien rakentama pisteiviiva ei ole rosainen. Suorissa viivoissa toiminto joskus tarpeettomasti pehmentää viivan reunaa.

Toisessa ja kolmannessa visualisointivariaatioissa antialiasing toiminnallisuus on otettu pois päältä svg-muotoilumäärittelyn (shape-rendering: crisp edges;) avulla.

Ensimmäisessä variaatioissa esityksen viivat ovat muotoilemattomia ja niissä on selaimen oma reunapehmenys päällä.

Toisessa variaatioissa viivatyyliksi on valittu tavallista paksumpi viiva, jossa on pyöristetyt päät. Tälläkin ratkaisulla pyritään erottautumiseen yleisesti käytetyistä ohuista ja yksivärisistä viivoista.

Kolmannessa variaatioissa käytetään mahdollisimman ohuita ja vähäkontrastisia viivoja, jotta itse kuvion data korostuu. Tässä variaatioissa on otettu selaimen reunapehmenysominaisuus pois käytöstä x- ja y-asteikkojen osalta.

## 3.4 Vuorovaikutus

Käyttäjällä on monia tapoja olla vuorovaikutuksessa visualisoinnin kanssa. Professorit **Yvonne Rogers**, **Heleen Sharp** ja **Jennifer Preece** ovat luokitelleet vuorovaikutustapoja:

- Vaikuttaminen (instruction).

Käyttäjä vaikuttaa infografiikan sisältöön ja laittaa sen "liikkeelle". Grafiikka voi perustua käyttäjän itse syöttämään dataan, kuten esimerkiksi erilaiset painoindexsimittarit. Niissä käyttäjä syöttää pituus- ja painotietonsa, jolloin mittari laskee painoindexsin ja

mahdollisesti vertaa sitä ikäluokan keskiarvoon tai muiden käyttäjien syöttämiin tuloksiin.

- Keskustelu (conversation). Käyttäjä voi olla myös dialogissa visualisoinnin kanssa. Esimerkiksi vaalikoneiden kaltaiset infografiikat kulkevat erilaisia polkuja aina sen mukaan miten käyttäjä niihin vastaa.
- Muokkaaminen (manipulation). Käyttäjä voi myös muokata esitystä. Jotkut visualisoinnit antavat käyttäjälle mahdollisuuden valita taustakuvia tai värejä visualisoinneilleen.
- Selailu (exploration). Käyttäjä voi selata esitystä ja löytää itseään kiinnostavia näkökulmia. Esimerkiksi sanapilvet esittävät sanoja esiintymistiheyden mukaan. Käyttäjä voi selata sanoihin liittyviä artikkeleja ja tutkia aineistoa aiheen yleisyyden mukaan.

(Cairo 2013, 200–204)

## KÄYTTÖLIITTYMÄT

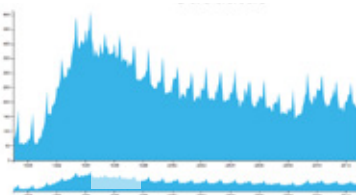
Perinteisen käyttöliittymäsuunnittelun periaatteissa löytyy paljon sovellettavaa tietoa myös verkossa julkaistavan infografiikan suunnitteluun. Vuonna 1988 julkaistu **Donald A.**

**Normanin** kirja *The Design of Everyday Things* nostaa neljä keskeistä elementtiä käyttöliittymäsuunnittelussa:

- Ulkoasu (visibility). Kaikki kohdat, joissa käyttäjällä on mahdollista vaikuttaa tulisi visuaalisesti merkitä niin, että niin että ne jo muodollaan kertovat mitä niillä täytyy tehdä. Esimerkiksi napin muoto ja varjot kertovat että sitä pitää painaa. Liukukytkimessä tyhjä tila näyttää mihin suuntaan sitä on mahdollista liikuttaa.
- Palaute (feedback). Käyttäjän tulisi saada jokin palautekaikesta toiminnastaan. Yleisimmillään tämä toteutetaan niin, että painonapin väri muuttuu silloin kun sitä painetaan. Joskus käytetään myös ääni- tai värinäefektejä.

- Rajaus (constraints). Jos visualisoinnin suunnittelija tekee rajoja käyttäjän puolesta, se tulisi kertoa käyttäjälle visuaalisesti. Esimerkiksi kaikki napit eivät välttämättä ole kaikissa näkymissä tai kaikilla laitteilla käytössä. Usein näkee esimerkiksi harmaita, ei-aktiivisia nappeja.
- Johdonmukaisuus (consistency). Käyttöliittymän tulisi toimia johdonmukaisesti. Nappien ja navigaation sijainti samassa paikassa joka näkymässä on ymmärrettävä ja helppokäyttöinen ratkaisu.

(Cairo 2013, 189–195)



Esimerkkitoteutus datan käyttöliittymästä löytyy verkosta osoitteesta: [demo.newpie.fi/suodatus.php](http://demo.newpie.fi/suodatus.php)

## NAVIGOINTI

Mitä vuorovaikutuksen ja navigoinnin tekniikoita suunnittelijalla on käytössään? Verkossa julkaistavaa visualisoinnin sanomaa voi korostaa merkittävästi animoinneilla ja interaktiivisuutta hyödyntävillä keinoilla.

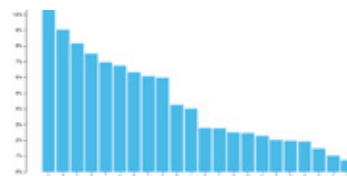
Kokenut käyttöliittymäsuunnittelija ja designer **Jenifer Tidwell** listaa viisi keskeisintä näkemystä erilaisista navigointi- ja selailutekniikoista:

- Pysty- ja sivuttaisvieritys. Käyttäjä voi vierittää kuva-alaa vaaka- ja pystysuunnassa, joka mahdollistaa kuva-alaa isompien kokonaisuuksien esittämistä. Tätä hyödynnetään usein esimerkiksi aikajanoissa.
- Zoomaus. Zoomauksen avulla voidaan näyttää aluksi laaja kokonaiskuva visualisoitavasta aiheesta, ja myöhemmin tarkat yksityiskohdat. Esimerkiksi laajoissa

verkostoaanalyysissä tästä on hyötyä, jotta saadaan suuret kartat mahtumaan yhdelle näytölle.

- Auki/kiinni. Valinnasta aukeava lisätietoikkuna auttaa suurten tietomassojen esittämisen esimerkiksi karttavisualisoinneissa. Kuva-alan päälle aukeavat ikkunat ovat käytännöllisiä, koska niiden avulla lukijan sijainti sivustolla ei muutu ja näin voidaan välttää ylimääräiset käyttöliittymäelementit
- Järjestys ja uudelleenjärjestys. Tämän toiminnallisuuden avulla voidaan esimerkiksi animoida miten jonkun tietyn asian järjestys muuttuu jos näkökulmaa vaihdetaan. Perus pylvädiagrammin muutokset tulevat pienellä animoinnilla hyvinkin havainnollisiksi.
- Haku ja suodattaminen. Tämän toiminnallisuuden avulla voidaan julkaista suuria data-aineistoja ilman että näytetään tuhansia rivejä tietoa kerralla. Käyttäjä voi hakea tai filttaroida itseään lähellä olevia näkökulmia aineistoon.

(Cairo 2013, 195–200)



Järjestämiseen perustuva käyttöliittymä: [demo.newpie.fi/jarjestys.php](http://demo.newpie.fi/jarjestys.php)

## KOHDENTAMINEN JA PERSONOINTI

Median verkossa julkaisemissa infografiikoissa hyödynnetään vielä marginaalisen vähän automaattisesti lukijasta käytössä olevaa tietoa. Mainostajat ovat jo pitkällä tämä tiedon hyödyntämisessä, esimerkiksi Googlen mainonnat kohdentuvat käyttäjän mielenkiinnon mukaan. Samoja tekniikoita käyttämällä olisi mahdollista kohdentaa myös infografiikkaa lukijalle. Vielä toistaiseksi yleisim-

mät lukijan tietoja hyödyntävät sovelukset ovat karttoja, joissa kysytään lupa käyttää lukijan sijaintitietoa ja sijoitetaan karttanäkymä suhteessa käyttäjän sen hetkiseen sijaintiin.<sup>15</sup>

Tekniikan avulla tiedämme paljon muutakin lukijasta. Tiedämme hänen lukulaitteensa tekniset tiedot ja ruudun koon. Tiedämme mistä linkistä hän on saapunut sivuillemme, mikä hänen selaushistoriansa on ja paljon muuta. Tämän tiedon hyödyntäminen mahdollistaisi monen toimenpiteen automatisoinnin ja sisällön personoinnin. Parhaassa tapauksessa saamme lukijasta riittävästi tietoa jotta voisimme tarjota hänelle kohdistettua sisältöä, jopa ilman käyttöliittymävalintoja.<sup>16</sup>

Kehitys näyttäisi tälläkin kertaa edistyvän nopeimmin verkon ulkopuolella, muun muassa älykkäissä käyttöesineissä. Esimerkiksi auton lukot eivät enää välttämättä tarvitse avainta sisäänsä, vaan ne voivat tunnistaa tutun käyttäjän lähietäisyydeltä ja avata ovet automaattisesti. Infografiikan saralla tämä voisi tarkoittaa suoraan sisältöjen ja näkökulmien, jopa esimerkkitapausten automaattista suodatusta käyttäjän mukaan. (Norman 2012)

### 3.5 Tarinallisuus

Yksi visuaalisten esityksien suurimmista voimavaroista on niiden mahdollisuus kertoa tarinoita. On kiinnostavaa kuinka datavisualisoinnit kertovat tarinan numeroiden takaa.

Journalistisesta näkökulmasta tarinallisuus laajentaa ja jatkaa uutisprosessia askeleen pidemmälle. Jos uutistuotannossa on perinteisesti vastattu kysymyksiin kuka, mitä, missä, milloin ja miksi, niin listaan voidaan tarinallisia elementtejä tuottaessa lisätä kysymys kuinka.

(Segel&Heer 2010, 1)

Stanfordin yliopiston tutkijat **Edward Segel** ja **Jeffrey Heer** ovat tutkineet datavisualisointien tarinankerrontatapoja ja sitä kuinka ne

eroavat perinteisestä kerronnasta. Tutkimuksissaan he käsittelevät eroja tyyllilajeissa, kerrontatavoissa ja tarinan rakenteissa. (Segel&Heer 2010, 1)

#### **TYYLILAJIT** (GENRE)

Tyyllilajin valinta voidaan tehdä tarinan mukaan. Eri tyyllilajit toimivat erilaisille tarinatyypeille, ja niiden valintaan vaikuttaa datan tai tarinan monimutkaisuus, kohdeyleisö tai julkaisumedia. (Segel&Heer 2010, 8)

Usein yksi visualisointi voi edustaa samanaikaisesti useaa eri tyyllilajia. Datavisualisoinneissa käytetyt tyyllilajit voidaan jaotella seuraavasti:

- Aikakauslehtityyli (magazine style)
- Selitetty taulukko (annotated chart)
- Moniosainen juliste -tyyli (partitioned poster)
- Virtauskartta (flowchart)
- Sarjakuva (comic strip)
- Diasarja (slideshow)
- Video (video)

(Segel&Heer 2010, 6)

#### **KERRONTATAVAT** (NARRATIVE TACTICS)

Visuaalisten tarinankerrontatapojen avulla lukijaa ohjataan liikkumaan visualisoinnissa. Kolme keskeistä tapaa käyttää visuaalista kerrontaa ovat korostaminen, visuaalinen kerroksellisuus ja liikkeellä ohjaaminen.

(Segel&Heer 2010, 7)

**Korostaminen** (highlighting) on toinen tapa johdattaa lukijaa kuvaan tai visualisoinnin esillepanossa eteenpäin. Yksinkertaisimmillaan liike voi tarkoittaa yksittäisen kuvan katsomistapahtumaa, jossa katse lukee kuvaa ensin kokonaisuutena ja sitten lähtee kiertämään yksityiskohdasta toiseen. Taitava suunnittelija osaa rakentaa katsomisreitit niin, että sitä kulkiessa datasta syntyy tarina. Käytännön visualisointikeinoja voivat olla esimerkiksi linjat, värikontrastit tai liike. Sarjakuvat ja animaatiot perustuvat tämän malliseen tarinankerrontaan. (Segel&Heer 2010, 7)

#### **Visuaalinen kerroksellisuus**

(visual structuring) on visuaalisen kerronnan tapa, jossa suunnittelija tarjoaa käyttäjälle mahdollisuuden tarkastella ja tutkia visualisointia omien valintojen kautta. Visualisoinnissa on alkutila, jota käyttäjä voi valinnoillaan muuttaa ja etsiä erilaisia yksityiskohtia ja ja toisia näkökulmia. Tällaisessa visualisointimallissa käytetään usein valintaruutuja tai prosessipalkkeja, joilla havainnollistetaan lukijalle mitä osaa aineistosta milloinkin tarkastellaan. Tarinallisuus voi syntyä esimerkiksi lukijan toiminnasta eri näkökulmien välillä. (Segel&Heer 2010, 7)

#### **Liikkeellä ohjaaminen** (transition

guidance) asettaa lukijan katsojan rooliin ja kertoo tarinan liikuttamalla dataa. Visualisointi voi vertailla tai esittää samasta aineistosta useita valmiiksi valittuja puolia. Yksinkertaisimmillaan tämänkaltaista tarinankerrontaa voi toteuttaa esittämällä animoidun aikasarjan, jossa graafin kuvaaja liikkuu vuosiluvun vaihtuessa. Myös videoinfografiikka perustuu tämänkaltaiseen kerrontatapaan. (Segel&Heer 2010, 7)

#### **RAKENNE** (NARRATIVE STRUCTURE)

Tarinallinen visualisointi voidaan toteuttaa erilaisten rakenteiden avulla.

**Tarinan kulku** (ordering) voi toimia ennalta suunnitellussa järjestyksessä (linear) sattumanvaraisessa järjestyksessä (random access) tai käyttäjän valitsemassa (user-directed) järjestyksessä. Suunnittelijan toimesta tehty tarinajärjestys etenee lineaarisesti alusta loppuun.

(Segel&Heer 2010, 8)

#### **Vuorovaikutusmahdollisuudet**

(interactive) voivat muokata visualisoinnista luettavaa tarinaa. Käyttäjälle voidaan tarjota suodatus-, valinta-, haku- tai navigointimahdollisuuksia, joiden avulla käyttäjä voi säätää visualisoinnista henkilökohtaisen näkökulman tarkasteltavaan aiheeseen.

(Segel&Heer 2010, 8)

**Kommunikointi** (messaging) visualisoinnin kanssa voi olla myös visualisointilähtöistä. Järjestelmä voi ehdottaa käyttäjälle lisävalintoja, kommentoida visualisoinnissa esiintyvää kehitystä tai tarjota yksityiskohtaisempaa tietoa eri vaiheissa.

*(Segel&Heer 2010, 8)*

### **TEKIJÄ- TAI LUKIJALÄHTÖISYYS**

Olenainen osa tarinallisuuden rakentamista on interaktiivisten elementtien osuuden pohtiminen. Osa visualisoinneista, kuten esimerkiksi videot ovat voimakkaasti tekijälähtöisiä, joissa tarina etenee täsmälleen siinä tahdissa kuin se on suunniteltu. Osa visualisoinneista on täysin lukijälähtöisiä, joissa ei tapahdu tai ole luettavissa mitään ennen käyttäjän tekemiä valintoja. Näiden välille jää vuorovaikutteiset sovellukset, joissa sekä tekijän suunnittelemalla järjestyksellä, että käyttäjän tekemillä päätöksillä ohjataan tarinan kulkua. Näitä kolmea erilaista tarinankerrontamallia eritellään johdattelevan rakenteen, interaktiivisen kuvasarjan ja selailtavan rakenteen mukaan. *(Segel&Heer 2010, 8)*

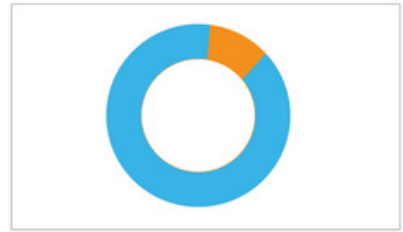
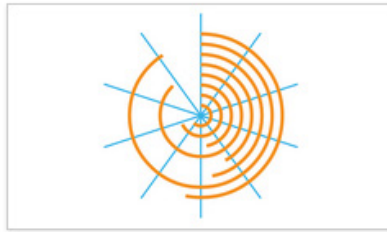
**Johdatteleva rakenne** (Martini glass structure) on tekijälähtöinen tarinankerrontamalli. Mallissa lukija johdatellaan aiheeseen kuvilla, tekstillä, kysymyksillä tai pienillä huomioilla, jota lukija seuraa tekijän suunnittelemassa järjestyksessä. Kun tekijän johdattelutarina on luettu, visualisointi avautuu lukijalle laajemmin tarkasteltavaksi. Käytännössä tämän voi toteuttaa esimerkiksi niin että, visualisoinnissa on johdattelevia sivuja, joiden lukemisen jälkeen lukijalle avautuu interaktiivisia toimintoja aineiston selailuun. *(Segel&Heer 2010, 8)*

**Interaktiivinen kuvasarja** (interactive slideshow) on rakenteeltaan tuttu esimerkiksi sarjakuvista. Interaktiivinen kuvasarja toimii samalla rakenteella, mutta kun kuva-ruutuihin lisätään interaktiivisuutta,

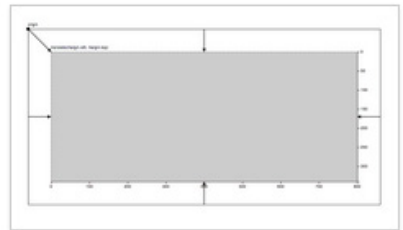
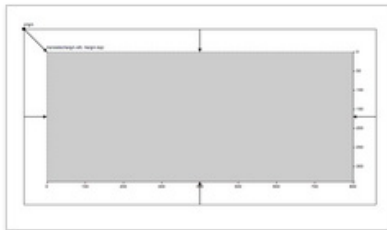
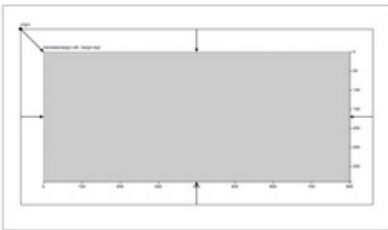
tarinan suunnittelija ja lukija muuttuvat tasavertaisemmiksi. Suunnittelija on valinnut tarinan elementit ja rakenteen, lukija kirjoittaa valinnoillaan oman tarinan. Tämänkaltaisen tarinankerrontamalli sopii hyvin moniuloitteisille data-aineistoille ja polveileville tarinoille. *(Segel&Heer 2010, 8)*

**Selailtava rakenne** (drill-down story) perustuu yhteen näkymään jota voi tarkastella monipuolisesti. Käytännössä tämä voi olla toteutettu zoomaustoiminnoilla, erilaisilla sivulle aukeavilla ikkunoilla tai muilla tietosisäältä tarkemmin esittelevillä yksityiskohdilla. Tässä mallissa pääosin lukijan toiminta määrittelee tarinan synnyn. Suunnitellun näkymän tulisi kannustaa lukijaa selaamaan ja tutkimaan aineistoa. Parhaimmillaan esimerkiksi Wikileaks-sivuston vuotamat valtioiden salaiset raportit voitaisiin visualisoinnin avulla nostaa luettaviksi niin, että niistä syntyy tarinoita. *(Segel&Heer 2010, 8)*

## Erilaisia visualisointitekniikoita



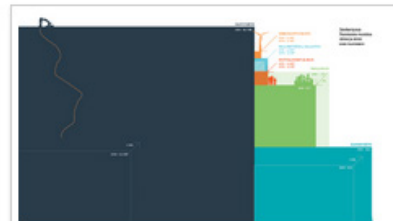
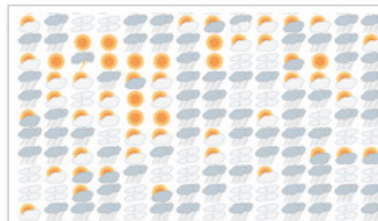
## Eri näyttökokojen hallinta



## Karttavisualisointeja



## Kuvavisualisointeja



Esimerkkitoteutukset löytyvät osoitteesta: [demo.newpie.fi](http://demo.newpie.fi).

# 4

## Visualisointiprosessi käytännössä

Tämän opinnäytetyön tavoitteena on tutustua verkossa julkaistavan infografiikan ja datavisualisointien tuottamiseen. Aiheesta kirjoitetun teorian lisäksi työhön liittyy olennaisesti erilaisten visualisointitekniikoiden opiskelu. Kuvaan tässä huomioita, jotka esimerkkitoteutusten tekemisen yhteydessä ovat nousseet keskeisimmiksi.

### 4.1 Lähtökohdat

Tekemiseni työelämässä suuntautuu vahvasti verkkoon, jonne myös jatkossa haluan keskittää suunnitteluosaamistani. Infografiikkaa käsittelevä opinnäytetyö oli hyvä tapa tutustua verkossa tehtäviin visualisointitapoihin ja hankkia käytännön kokemusta tekemisestä.

Aiheeseen perehtyminen sai myös epäilyksiä heräämään. Pystyisinkö opettelemaan tarvittavat tekniset taidot opinnäytetyhön varaamani puolen vuoden opiskeluaikana? Voisiko oppimillani taidoilla tuottaa sellaista visuaalista jälkeä kun haluan vai onko tekniikassa rajoitteita? Osaamista täytyy kartuttaa graafisen osaamisen lisäksi myös tilastotieteellisellä osaamisella. Taulukkolaskentaohjelmiston käyttötaito on yksi uusi haaste. Hieman epäilyttää, että saanko kiinni näin isosta kokonaisuudesta puolesta vuodesta? Suuria käytännön kysymyksiä työssäni on paljon.

- Teknisen osaamisen puute. En tunne riittävän hyvin nykyaikaisia visualisointitekniikoita, enkä osaa koodata. Ehdinkö oppimaan riittävät taidon prosessin aikana?
- Miten osaan luoda jotain uutta ja omaperäistä infografiikkaa tähän olevaan verkkoon?
- Onko minulla riittävästi matemaattista tai tilastollista osaamista visualisointien tekemiseen?

- Tuleeko työstäni riittävän kiinnostava visuaalisesti vai jääkö se koodausharjoitukseksi, onko tämä tosiaan graafikon työtä vai kuuluisiko tätä tehdä jollain toisella ammattialalla?
- Kuinka voin väittää työtä omakseeni, jos käytän avointa lähdekoodia ja valmiita visualisointielementtejä harjoituksissani?

### 4.2 Teoria ja käytäntö

Aloitin aiheeseen tutustumisen teorian ja esimerkkeihin tutustumalla. Selvitin lähdekirjallisuutta ja selasin verkossa visualisointiesimerkkejä.

Teoriakirjallisuuden pääteoksiksi valikoituivat Cairon, Kuuselan ja Tuften kirjat. Aluksi yritin vielä mahdollistaa mukaan Otto Neurathin Isotype-aiheista ideologiaa, mutta sen anti ei tuntunut istuvan aiheeseeni yhtä hyvin kuin muut valitsemani teokset.

Teoriaan tutustumisen ohella etsin erilaisia visualisointitekniikoita visualisointien tuottamiseen. Selasin kymmeniä erilaisia JavaScript -kirjastoja. Kiinnostavimmaksi tekniikaksi valikoitui D3-JavaScript-kirjasto, jonka käytössä oli paljon mahdollisuuksia ja siihen löytyi paljon käyttöohjeita ja esimerkkitoteutuksia.

Alkuperäinen ajatus opinnäytetyöprosessin kulusta oli tuottaa erilaisia yksinkertaisia demoja ja kirjoittaa samalla teoriakirjallisuudesta blogi-

tekstejä. Teoriablogin kirjoittaminen ei kuitenkaan oikein tuntunut luonnistuvan, joten keskityin käytännön taitojen opiskeluun. Myöhemmin hylkäsin blogi-idean ja kirjoitin teoriaosuuden perinteisin metodein.

Käytännön tekeminen alkoi html- ja css- kuvauskielten opiskelulla. Työssäni olin tehnyt verkkosivukonsepteja ja visualisointeja ja tekniikat olivat pintapuolisesti tuttuja, mutta työssäni varsinaisen verkkopalveluiden teknisen toteutuksen hoitivat toiset, joten minulla ei ollut rutiinia sivujen koostamiseen.

Vuoden 2012 aikana huomasin verkkopalveluiden suunnitteluvaatimusten muuttuvan. Yhä useammin vaatimuksena oli responsiivinen ulkoasu, eli palvelu, jonka tuli toimia useilla eri näyttökoolla.

Responsiivinen suunnittelu vaatii suunnittelijalta uudenlaista ajattelua ja uudenlaisten työkalujen hallintaa. Perinteinen ulkoasusuunnitelma ei ole enää riittävä, koska ulkoasusuunnitelman toiminta tulee varmistaa jokaisella näyttökoolla.

Tästä johtuen ajattelin, että suoraan koodilla toteutettava ulkoasusuunnittelu tulee olemaan omassa työssäni välttämätöntä ja siksi tämäkin opinnäytetyö sisältää niin paljon teknistä harjoittelua. Datan visualisointisuunnitelmia ei käytännön työelämässä voida piirtää

ulkoasuunnielmiksi vaan ulkoasu on kiinteä osa ohjelman toimintaa. Suunnittelijan täytyy pystyä tuottamaan visualisointeja myös koodin avulla. Kaikkea ei kuitenkaan tarvitse avoimen lähdekoodin maailmassa tehdä itse, vaan teknisiä ratkaisuja muun muassa näyttökokojen hallintaan on runsasti saatavilla.

Yksi hyväksi koettu sivupohja on Twitter Bootstrap, jonka valitsin opinäytetyön pohjaksi. Sen avulla rakensin eri näyttökokoihin skaalautuvaa sivustorunkoa.

Harjoitusten edetessä opin ymmärtämään Twitter Bootstrapin toimintaa ja sen ominaisuuksien avulla opin responsiivisen designin periaatteita.

Samalla opin hyödyntämään työkalussani uusia työkaluja ja työkaluja. Kehitin koodausosaamistani Codecademyn verkkokursseilla.

Esimerkkivisualisointien tekoa harjoittelin D3-kirjaston harjoitusten avulla. Esimerkiksi Scott Murrayn kirjoittamat harjoitustehtävät osoittautuivat erittäin hyödyllisiksi.

Harjoituksia tehdessä ymmärsin miten vaikeaa joidenkin visualisointien tekeminen on, ja epäily teknistä valintaa ja osaamistani kohtaan kasvoi. Muissa tutkimissani javascript-kirjastoissa visualisointeja oli huomattavasti helpompi tuottaa. Niiden ero D3-visualisointien tekoon oli se, että muissa kirjastoissa visualisoinnit oli jo oletuksena animoituja ja tyylieltyjä. Muiden kirjastojen käyttö oli helppoa, mutta niillä ei saanut tuotettua kovin yksilöllisiä toteutuksia. D3-kirjasto toimii päinvastoin, sen avulla jokaiseen visualisoinnin ominaisuuteen ja toiminnallisuuteen voidaan vaikuttaa helposti, mutta kaikki pitää tehdä alusta asti itse.

### 4.3 Tekninen toteutus

JavaScript-kirjastoilla piirtämisessä kohtasin yllätyksiä ja ongelmia monissa kohdissa. Ehkä suurimmat niistä liittyivät datan tai visualisoinnin skaalautumiseen.

Verkkovisualisointien hyvä puoli on se, että kun ne kerran toteuttaa, niitä voi jatkossa käyttää uudelleen tuoreella data-aineistolla. Tästä syystä visualisointien skaalautuminen on tärkeää. Alla muutamia datan skaalautumiseen liittyviä ongelmia ja niiden ratkaisuja.

#### VISUALISOINTI D3:N AVULLA

Ensimmäisten pylväskaavioiden toteutusten yhteydessä aloin ymmärtää kuinka javascript visualisointi toimii ja mitä siinä täytyy huomioida. Alla on esimerkkikoodia yksinkertaisen pylväsdiagrammin toteuttamisesta D3-javascript-kirjastolla.

```
1 //data-aineisto
2 var dataset = [5, 10, 13, 21];
3
4 // määritellään svg-ikkunan
5 koko
6 var w = 85;
7 var h = 25;
8
9 //luodaan svg-ikkuna
10 var svg = d3.select("body")
11   .append("svg")
12   .attr("width", w)
13   .attr("height", h);
14
15 //luodaan pylväät
16 svg.selectAll("rect")
17   .data(dataset)
18   .enter()
19   .append("rect")
20   .attr("x", function(d, i)
21     {return i * 21;})
22   .attr("y", 0)
23   .attr("width", 20)
24
25 //määritellään pylväiden
26 korkeus data-aineiston
27 perusteella
28   .attr("height",
29     function(d)
30       {return d});
```

Rivillä kaksi annetaan koodille data-aineisto. Riveillä 6–7 määritellään visualisoinnin koko. Riveillä 10–13 luodaan svg-ikkuna ja kerrotaan että ikkuna sijoitetaan html-dokumentin body-elementtiin.

Rivillä 16–18 luetaan data-aineisto ja rivillä 19 luodaan yksi pylväs. Riveillä 15–16 yhdistetään data-aineisto ja pylväs. Rivillä 20 kerrotaan

koodille, että pitää piirtää yhtä monta pylvästä kuin data-aineistossa on tieto-yksiköitä. Jokaiselle pylväälle varataan tilaa 21 pikseliä. Rivillä 21 pylväiden alareuna asetetaan svg-ikkunan alareunaan. Rivillä 23 määritellään piirrettävän pylvään leveydeksi 20 pikseliä, jolloin jokaisen pylvään väliin jää yhden pikselin marginaali. Riveillä 28–30 määritellään pylväiden korkeus yhtä korkeaksi kuin data-aineiston arvo on, tällöin ensimmäinen pylväs on 5 pikseliä korkea ja seuraava 10 jne.

#### PYLVÄÄ KOON SKAALAUTUMINEN

Edellä kuvatussa esimerkkikoodissa svg-ikkuna on mitoitettu 85 pikseliä leveäksi, jolloin siihen mahtuu neljä 20 pikseliä leveää pylvästä 1 pikselin marginaalilla. Jos visualisointia halutaan käyttää laajemmalla data-aineistolla, ylimääräiset pylväät jäävät näkymättömiin ikkunan ulkopuolelle.

```
1//luodaan margin-niminen
2//muuttuja
3 var margin = 1;
4
5//korvataan staattinen leveys
6//.attr("width", 20);
7
8//dynaamisella leveydellä
9.attr("width", w /dataset.length
10 - margin);
```

Tämä ongelma voidaan ratkaista muokkaamalla javascript-koodia yllä kuvatulla tavalla. Pylväiden leveys voidaan määritellä ensimmäisen esimerkin staattisen määrittelyn sijaan dynaamisesti. Rivillä 8 pylvään leveys lasketaan jakamalla svg-ikkunan leveys pylväiden määrällä. Lisäksi rivillä 2 luodaan muuttuja, joka määrittää pylväiden välisen marginaalin yhdeksi pikseliksi.

Tämäntyyppisten skaalausratkaisujen avulla visualisointeja voidaan rakentaa niin, että visualisointien jatkokäyttö uudella data-aineistolla on helppoa, eikä samoja ratkaisuja tarvitse tehdä joka kerta uudelleen.

## VAAKAPYLVÄIKÖN SKAALAUUS

Vastaava skaalausongelma syntyy myös vaakapylväikön yhteydessä. Tässä visualisointimallissa skaalaus voidaan toteuttaa toisella tavalla, koska pystysuunnassa tilaa on loputtomasti.

```
1 //data-aineisto
2 var dataset = [5, 10, 13, 21];
3
4 //määritellään svg-ikkunan
5 //koko
6 var w = 25;
7 var h = dataset.length * 21;
8
9 //luodaan svg-ikkuna
10 var svg = d3.select("body")
11     .append("svg")
12     .attr("width", w)
13     .attr("height", h);
14
15 //luodaan pylväät
16 svg.selectAll("rect")
17     .data(dataset)
18     .enter()
19     .append("rect")
20     .attr("y", function(d, i)
21         {return i * 21;})
22     .attr("x", 0)
23     .attr("width", 20)
24
25 //määritellään pylväiden
26 //korkeus data-aineiston
27 //perusteella
28     .attr("height",
29         function(d)
30             {return d});
```

Yllä kuvatussa esimerkikoodissa luodaan aikaisemman kaltainen pylväikkö, mutta vaakamallina. Koodi on muuten samanlainen mutta x ja y-arvot riveillä 20 ja 22 ovat vaihtaneet paikkoja. Lisäksi riveillä 28-30 määritellään data-aineiston pohjalta pylväiden korkeuden sijasta niiden pituus. Visualisoinnin skaalaus toteutetaan rivillä 7, jossa svg-ikkunan korkeus syntyy kertomalla tietoyksiköiden määrä 21:llä. Tällöin ikkunan korkeudeksi määrittyy 4 x 21 pikseliä.

Yllä esitelty skaalaustapa on kätevä tapa muuttaa svg-ikkunan kokoa. Samalla periaatteella toimivaa skaalausta käytetään myös eri näyttökoille suunniteltaessa. Tekniikan etuna on sen suorituskyky. Näyttökokoja tarkasteleva koodi suoritetaan vain keran - lukijan saapessa sivulle.

## ERI NÄYTTÖKOOT

Suuri haaste kaikissa javascript-visualisoinneissa on svg-ikkunan skaalautuminen lukijan näytön koon mukaan. Skaalautuminen voidaan tehdä adaptiivisesti, jolloin lukijan näytön koko tarkistetaan lukijan saapessa sivustolle, jonka jälkeen visualisointi piirretään lukijan näytölle. Tämä on yksinkertainen javascriptissä oleva ominaisuus, jonka käyttö on helppoa.

```
1//määritellään svg-ikkunan
2//koko lukijan näytön koon
3//mukaan
4 var padding = window.innerWidth
5 * 0.1;
6 var w = window.innerWidth -
7 padding;
8 var h = data.length * 25 +
9 padding;
10
11//luodaan svg-ikkuna
12 var svg = d3.select("body")
13     .append("svg")
14     .attr("width", w)
15     .attr("height", h);
16
```

Yllä olevassa koodissa rivillä neljä ja viisi määritellään svg-ikkunan ympärille tyhjää tilaa 10 prosenttia lukijan näytöstä. Rivillä kuusi svg-ikkunan leveydeksi määritellään lukijan näytön leveys vähennettynä yllä määritellyllä tyhjällä tilalla. Riveillä 12-15 piirretään aikaisemmin määriteltujen mittojen mukaan svg-ikkuna, johon visualisointia voidaan piirtää.

## DATAN SKAALAUTUMINEN

Yllä kuvatuissa esimerkeissä visualisointi toteutetaan suoraviivaisesti lukemalla data-aineisto visualisointiin. Tässä ratkaisussa ongelmia kuitenkin syntyy heti, kun data-aineisto sisältää esimerkkitietoa huomattavasti suurempia tai pienempiä lukuja. Siksi data-aineistosta luettava data tulisi suhteuttaa visualisoinnissa käytettäviin lukuihin. D3-kirjastossa tätä voidaan hallita domain ja range-funktioilla. Seuraava koodi on yksinkertainen esimerkki siitä, kuinka data-aineistoja ja visualisoinnin lukuja voidaan suhteuttaa toisiinsa.

```
1 //data-aineisto
2 var dataset = [5, 10, 13, 21];
3
4 var scale = d3.scale.linear()
5     .domain([5, 21])
6     .range([10, 100]);
```

Rivillä 2 sijaitsee edellisistä esimerkeistä tuttu data-aineisto. Rivillä neljä luodaan muuttuja, joka skaalaa data-aineiston haluamimme arvojen välille. Rivillä viisi määritellään mitkä ovat aineistomme pienimmät ja suurimmat arvot. Rivillä kuusi määrittelemme mille arvovälille haluamme muuttaa aineiston arvot. Tämän koodin avulla data-aineiston pienin luku eli numero viisi muunnetaan skaalan pienimmäksi luvuksi eli numero kymmeneksi. Samalla periaatteella suurin luku, eli 21 muunnetaan skaalan suurimmaksi eli 100:ksi.

## VISUALISOINNIN MUOTOILU

Svg-elementtiä muotoillaan hieman eri tavalla kuin css-kielessä on tututtu. Esimerkiksi svg:n tekstielementtiä ei väritetä css:stä tutulla color:black-komennolla, vaan tekstielementille määritellään erikseen viivan väri ja täyte.

```
1//tekstin väri html-elementille
2//tyylitiedostossa
3 .axis text {
4     color: "red";
5 }
6
7 //tekstin väri SVG-elementille
8 .axis text {
9     fill: "red";
10 }
```

## MUIDEN KIRJASTOJEN KÄYTTÖ

JavaScript-kirjastoissa visualisointeja luodaan monella tavalla. On yleistä, että visualisointikirjasto tuottaa jo oletusmuodossaan tyyliä ja animaatioita esityksiä, joihin kirjaston käyttäjä voi vain syöttää data-aineiston. JavaScript-kirjastojen käyttöä oppii parhaiten niiden omien ohjeistusten ja rajapintakuvausten avulla. Ohjeissa yleensä kuvataan ohjelmiston toiminta ja opastetaan käyttä-



mään kirjastoon liittyviä toiminnallisuuksia. Kirjaston rajapinnan avulla voidaan esimerkiksi korvata joitain visualisoinnin oletusasetuksia ja tyytlejä. Esimerkiksi Chart.js-kirjasto tarjoaa alla kuvatun rajapinnan.

```
Pie.defaults = {
  //Segmentin viivan
  //näkyvyyden säätäminen
  segmentShowStroke : true,

  //Segmentin
  //viivanvärin säätäminen
  segmentStrokeColor : "#fff",

  //Segmentin täyteen
  //värin säätäminen.
  segmentStrokeWidth : 2,

  //Animointi päällä
  //tai pois päältä.
  animation : true,

  //Liikkeen askelmien
  //määrän säätäminen
  animationSteps : 100,

  //Animaation nopeuden säätäminen
  animationEasing : "easeOutBounce",

  //Kallistusanimointien säätäminen
  animateRotate : true,

  //Skaalautumisanimaation
  //säätäminen
  animateScale : false,

  //Functio joka käynnistyy kun
  //animaatio on ladattu
  onAnimationComplete : null
}
```

Rajapinnan ominaisuuksia saa toimimaan omassa visualisoinnissa kun lisää yllä kuvattuja määrityksiä visualisoinnin tuottavan JavaScript-koodin joukkoon.

#### KARTTAVISUALISOINTEJA

Ensimmäiset karttavisualisoinnit toteutin Googlen ilmaisipalvelulla Fusion Tablesilla. Verkossa toimivan palvelun käyttöön löytyy hyviä ohjeistuksia Fusion Tablesin omilta sivuilta. Suomenkielisiä ohjeita on tuotettu datajournalismi.fi-sivustolle.

Tutkin myös muita karttasovelluksia ja nopeiten näyttäviä verkkototeutuksia sai aikaan Map-

Box-palvelulla. Palvelussa pystyy monipuolisesti tyyliin kartoja värejä ja toteuttamaan yksinkertaisia karttavisualisointeja. Monipuolisempia toteutuksia varten palvelussa on rajapinta.



MapBoxilla toteutettu kartta Irtisanomisaalto-visualisointiin.

Kiinnostava tuttavuus karttasovellusten tekemiseen on kartograph.js-kirjasto. Kirjastossa on laajat visualisointimahdollisuudet, mutta sen käyttö on vaikeaa puutteellisen ohjeistuksen takia.

#### 4.4 Data-aineistot

Pylväsdemo harjoitusten jälkeen lähdin etsimään kiinnostavaa dataa visualisointeja varten. Löysin hienon toteutuksen dataa suodattavasta pylväsdiaagrammista, johon päätin etsiä siihen sopivaa dataa ja kokeilla onnistuisinko tekemään oman version.

Data-aineistoksi etsin työttömyyden kehitystä ikäryhmittäin ja alueittain kuvaavaa aineistoa. Halusin selvittää miten esimerkiksi nuorisotyöttömyys on kehittynyt viime vuosina ja onko kehityksessä nähtävissä alueellisia eroja.

Visualisoinnissa tarvittavaa dataa etsin aluksi Tilastokeskuksen avoimista tietokannoista. Tietoa ei Tilastokeskuksen kannoista haluumassani laajuudessa löytynyt. Uskoin kuitenkin, että data on jossain olemassa, ja seuraava etsintäpaikkani oli Kuntaliitto. Heidän sivuiltaan ei löytynyt tarpeeksi tarkkaa tietoa data-aineistoista ja uskaltauduin otamaan yhteyttä Kuntaliiton tilastoasiantuntijaan Elisa Rantaan. Ranta tutki asiaa ja osasi opastaa minut

takaisin Tilastokeskukseen oikean tilaston äärelle. Harmillisesti haluamani tilastoaineisto oli maksullinen. Ranta vinkkasi, että kyseistä tietokantaa ollaan lähiaikoina avaamassa maksuttomaksi, ja kehotti ottaman yhteyttä Tilastokeskukseen.

Tilastokeskus suostuikin avaamaan minulle tunnukset maksutta toistaiseksi suljettuun tietokantaan josta sain haluamani datan.

Tilastokeskuksen data-aineiston ja demototeutuksen data olivat huomattavan eri muodossa ja huomasi, että pelkkä datan siivoaminen ei riitä, vaan minun täytyy tehdä merkittäviä muutoksia koodiin. Jatkoin toteutuksen muokkaamista, mutta en löytänyt ratkaisua edes luotokoodaajani avulla. Huomasin, että hakemani datan rakenne oli turhan monimutkainen enkä taidoillani pystynyt toteuttamaan visualisointia.

#### DATAN SIIVOAMINEN

Datan siivoaminen on yksi datavisualisoinnin työläimpiä vaiheita. Tämän työn aikana harjoittelin eri tapoja siivota data-aineistoja visualisointeja varten. Löysin kolme keskeistä metodia, joiden avulla data-aineistosta saadaan käyttökelpoista sisältöä visualisoitavaksi.

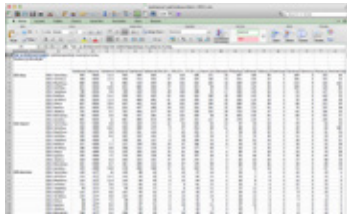
Ensimmäinen ja tärkein metodi on **journalistinen valinta**. Aineistoa tutkailtaessa ja journalistista harkintaa käyttämällä aineistosta valitaan olennaisin sisältö. Tätä metodia käytimme kun työstimme datajournalismikurssin toimittajaystävieni kanssa Irtisanomisaalto-visualisointia.

Alun perin päätimme toteuttaa visualisoinnin SAK:n keräämän irtisanomisdatan perusteella. Aineisto oli tarjolla keskusjärjestön sivuilla suoraan Excel-muodossa.

Poimimme aineistosta merkittävimmät tapahtumat eli yli kymmenen henkilön irtisanomiset. Lisäksi luokittelimme aineistoa kolmeen kategoriaan irtisanottujen määrän mukaan. Luokassa 1 oli yli 100 hengen

irtisanomiset, luokassa kaksi yli 50 ja luokassa kolme yli 10 hengen irtisanomiset. Näillä karsinta ja luokittelu-toimenpiteillä olimme suodattaneet datasta olennaisimman tiedon visualisointia varten.

Toinen metodi datan siivoamiseen on **aineiston järjestäminen ja yhdenmukaistaminen**. Toinen irtisanomisaallossa visualisoitava data-aineisto löydettiin Tilastokeskukselta. Siinä esitetään työttömyys-tilastoja vuosien 1990–2012 välisenä aikana. Alkuperäinen aineisto oli valtava, usean tuhannen rivin taulukko. Siivoamalla ja karsimalla siitä saatiin tallennettua yksinkertainen-csv-tiedosto D3:n ymmärtämässä muodossa.



*Tilastokekuksen data-aineistossa oli alunperin 4000 riviä ja kymmeniä sarakkeita dataa. Lopullisessa visualisoinnissa on käytössä 4000 tietoyksikköä.*

Kolmas metodi tiedon siivouksessa on **tiedon muuttaminen**. Toiteutin erään ohjelmointikurssin visualisointia Helsingin säädatasta. Data-aineistoa ei ollut saatavilla tilastomuodossa, joten jouduin kopioimaan sen ulkomaisen sääpalvelun sivulta. Aineistossa oli aurinkoinen, pilvinen tyypinen sää tietojen kategorialuokittelu. Yhdistin aineistoon toiselta sivulta kaivamani valoisan ajan tiedot. Tämän data-aineiston siivous oli hyvä taulukkolaskentaohjelman käyttöharjoitus. Aineiston sääkategorian muutin numeroluokitteluksi ja kellonaikoihin perustuvat valoisan ajan tiedot laskin minuuteiksi. Näin sain aineiston käännettyä dataksi visualisointia varten.

## 4.5 Visualisointien suunnittelu

Suunnittelijan näkökulmasta on vaikea erottaa mikä vaihe on suunnittelutyötä ja mikä toteutusta. Teknisen toteutuksen ja visuaalisen suunnittelun raja on hämärä ja visualisoinnin onnistumiseen vaikuttaa myös data-aineiston muoto. Toteutus tämän kaltaisessa työssä on yhtä olennainen ja erottamaton osa prosessia kuin piirtäjälle kuvan suunnittelu ja varsinainen piirtäminen.

Omassa suunnittelutyössäni olen huomannut hyväksi käytännöksi tehdä koevisualisointeja taulukkolaskentaohjelmassa sen tarjoamalla perusmuodoilla. Jos niitä muokkamalla data-aineisto alkaa näyttää järkevälle ja kiinnostavalle, visualisointia kannattaa lähteä suunnittelemaan pidemmälle.

Itselleni tutut graafisen suunnittelun ohjelmistot ovat myös hyvä apu visualisoinnin suunnittelussa. Niiden avulla voin suunnitella visualisoinnin värimaailmaa, typografiaa ja viivatyylejä. Tosin kaikkea ei aina voi ihan sellaisenaan toteuttaa, mutta luonnosteluun ohjelmistot soveltuvat mainiosti.

Esimerkiksi irtisanomisaallon visualisointia ja sivun ulkoasua suunnittelin Illustratorissa ja toteutin visualisoinnin myöhemmin koodilla. Lopullinen toteutus kuitenkin muotoutui toteutusta ja visualisointeja koodatessa. Visualisointi tulee toimimaan itselleni hyvänä suunnittelutyön ja työnkulun harjoituksena.

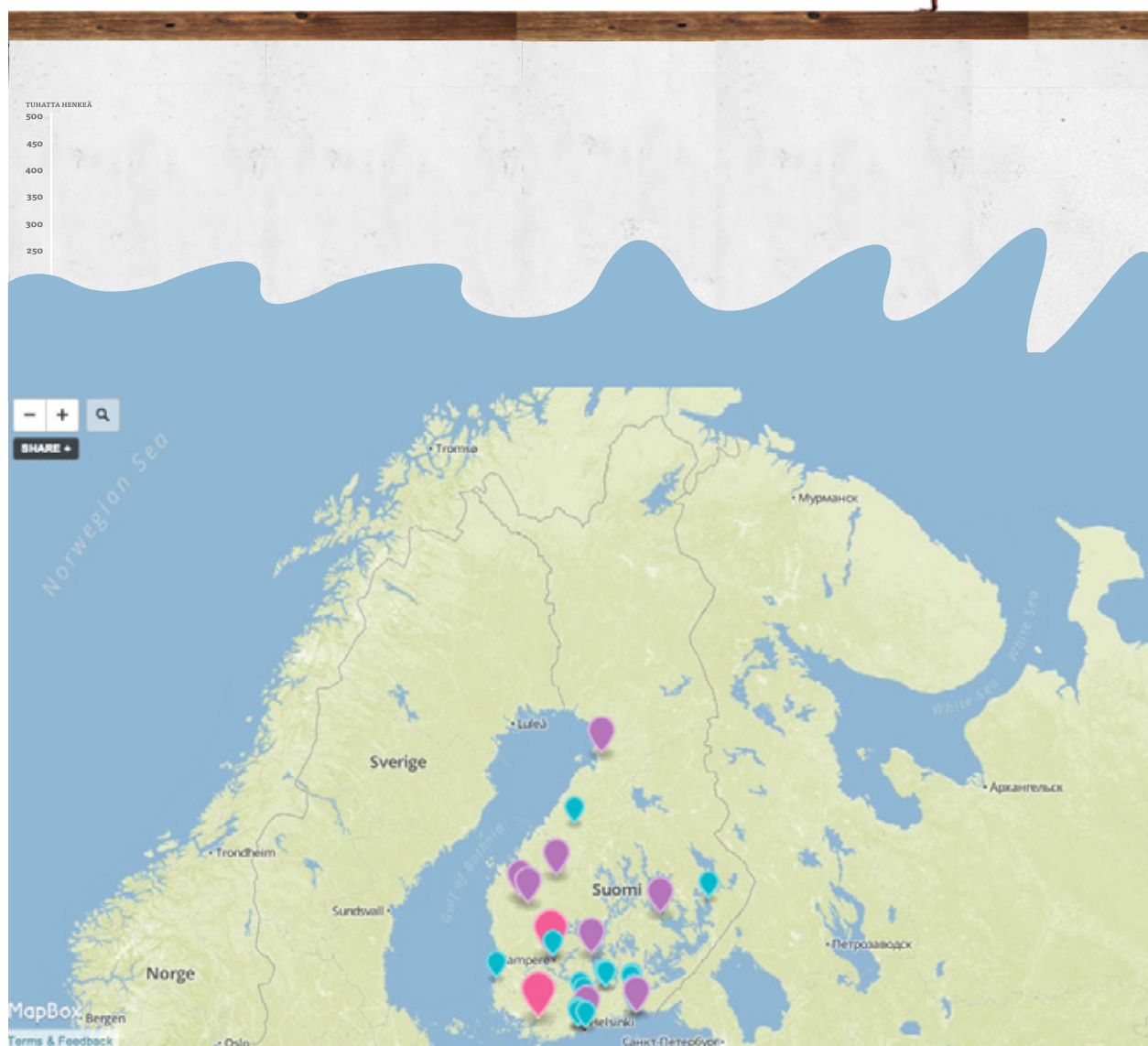
## 4.6 Lopputulokset ja jatko

Tämä projekti on ollut työläs ja opettavainen. Erinomaisen tyytyväinen olen löytämiini lähteisiin ja linkkeihin, joiden avulla tiedän pääseväni eteenpäin Datajournalismin parissa.

Vaikka yksikään tekemäni visualisointi ei ole niin vaikuttava kuin työn alussa kuvittelin tekeväni, olen tyytyväinen läpikäymääni prosessiin. Tämä työ on tuonut minulle kokemusta

ja opettanut datavisualisointiprosessin työnkulkuja. Projektin myötä olen saanut tutkia erilaisia visualisointitapoja käytännössä.

# Irtisanomis- aalto



Irtisanomisaalto visualisointi löytyy verkko-osoitteesta: [demo.newpie.fi/irtisanomisaalto](http://demo.newpie.fi/irtisanomisaalto)

Tämä lopputyö on ollut pitkä ja perusteellinen matka verkossa julkaistavan infografiikan maailmaan. Työtä aloittaessani ajattelin käyväni pikaisesti läpi perusgraafit ja visualisoinnit ja päätyväni kehittelemään uusia ja harvinaisemmin nähtyjä visualisointeja.

**TAVOITTEENA OLI TUOTTAA** työelämässä jatkokäyttöä saavia sovelluksia uutismedialle. Tutkimus- ja analyysityö kuitenkin osoitti, että hyvintehdyt perinteiset visualisoinnit ovat tehokkaimpia tiedonvälittäjiä ja siten käyttökelpoisimpia pohjia tulevalle työlleni.

Tästä huomiosta johtuen päädyin syventymään perusmuotoihin ja niihin liittyviin lainalaisuuksiin graafisen suunnittelun, datajournalismin ja tilastotieteiden näkökulmista.

Graafisen suunnittelun tärkein oppi on ollut datalähtöinen suunnittelu. Alunperin suunnittelemani esimerkkidatalla tapahtuvien graafien suunnittelu ei enää syvällisemmän opiskelun jälkeen tuntunut mielekkäältä. Valtaosan esimerkkivisualisoinneista olenkin tehnyt pääosin oikean datan pohjalta. Visualisointien suunnittelussa keskeiseksi on noussut esitysmuodon valinta sisällön ehdoilla.

Esimerkkivisualisointeja varten tekemäni datan valinta, etsiminen ja siivoaminen on kehittänyt minua journalistisessa roolissa ja tiedon ymmärtämiseen liittyvissä kysymyksissä. Erityisesti datalähteiden etsiminen ja tilastokeskuksen ulkopuoliset lähteet ovat entistä tutumpia.

Työhön liittyvä tilastotieteellinen osuus on myös ollut kasvattava ja herättänyt mielenkiintoa aiheen syvempään opiskeluun. Ilman visualisoidavan datan ymmärrystä visualisoinnin tuottajan tehtävä on yhtä vaikea kuin vieraskielisen tekstin taittaminen. Käytännön tasolla Excelin käyttötaidon karttuminen ja tilastograafien ominaispiirteiden oppiminen ovat kullannarvoisia reittejä parempiin visualisointeihin.

Verkossa julkaistavien datavisualisointien osalta myös tekniikka ja sen hallinta ovat olleet valitettavan merkittävässä asemassa. Omassa työssäni teknisten ratkaisujen etsimiseen ja osaamisen kartuttamiseen kului tuhattomasti aikaa. Muistutankin nyt itseäni ja muita samanlaisiin projekteihin lähteviä, että työ, jonka tekemiseen täytyy opetella uusi kieli, on pitkä ja kivikkoinen prosessi.

Käytännössä tämä työ ja siitä saatu oppi kuljettaa minua eteenpäin työelämässä ja toimii hyvänä alkupisteenä tuleville projekteille.

## **KIITOKSET**

Suurenmoiset kiitokset kaikille mukanaoppijoille, yhteistyö- ja keskustelukumppaneille. Ilman teitä olisin ollut pulassa. Suurena apuna oppimisprosessissa on ollut Lahden Muotoiluinstituutin opettajien lisäksi opiskelijatoverini graafisen suunnittelun opinnoista sekä datajournalismiopinnoista. Kiitos avusta ja kannustuksesta.

Erityiskiitoksen ansaitsevat kannustava ja asiantunteva ohjaajani Juuso Koponen, tarkkaakin tarkempi opponetti Taru Tarvainen ja kärsivällinen kodin hengetär vaimoni Maija Minkkinen.



# Lähdeluettelo

**Abela, Andrew** (2010) The Presentation. A Story About Communicating Successfully With Very Few Slides.

<http://www.extremepresentation.com/uploads/documents/The-Presentation-Ebook-Abela.pdf> (luettu 15.4.2013)

**Bradshaw, Paul** (2012) Data Journalism Handbook.

[http://datajournalismhandbook.org/1.0/en/introduction\\_0.html](http://datajournalismhandbook.org/1.0/en/introduction_0.html) (luettu 15.4.2013)

**Cairo, Alberto** (2013) The Functional Art. Berkley, CA: New Riders.

**Few, Stephen** (2013) Data Visualization for Human Perception. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction". Aarhus, Denmark: The Interaction Design Foundation.

[http://www.interaction-design.org/encyclopedia/data\\_visualization\\_for\\_human\\_perception.html](http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html) (luettu 15.4.2013)

**Friendly, Michael** (2009) Milestones In The History of Thematic Cartography, Statistical Graphics, And Data Visualization.

<http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf> (luettu 15.4.2013)

**Gray, Jonathan** (2012) What Data Can And Cannot Do.

<http://www.guardian.co.uk/news/datablog/2012/may/31/data-journalism-focused-critical> (luettu 15.4.2013)

**Halvey, Martin & Keane, Mark T.** (2007) An Assessment of Tag Presentation Techniques.

<http://www2007.org/htmlposters/poster988/> (luettu 15.4.2013)

**Holmes, Nigel** (1991) Pictorial Maps. Watson-Guption.

**Jyväskylän Yliopisto** (2013) Verkostoanalyysi.

<https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/aineiston-analyysimenetelmat/verkostoanalyysi> (luettu 15.4.2013)

**Kosara, Robert** (2010) Definition of visualization.

<http://eagereyes.org/criticism/definition-of-visualization> <http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf> (luettu 15.4.2013)

**Kosslyn, S.M** (1994) Elements of Graph Design. W.H. Freeman & Company, New York.

**Koponen, Juuso** (2011) Sanapilvet ovat huonoa informaatiomuotoilua.

<http://informaatiomuotoilu.fi/2011/10/sanapilvet-ovat-huonoa-informaatiomuotoilua/> (luettu 15.4.2013)

**Kuusela, Vesa** (2000) Tilastografiikan perusteet, Helsinki.

Tilastokeskus ja Oy Edita Ab.

**Mellin, Ilkka** (1997) Johdatus tilastotieteeseen. Helsingin Yliopisto.

**Niiniluoto, Ilkka** (1989) Informaatio, tieto ja yhteiskunta.

Filosofinen käsiteanalyysi. Helsinki. Valtion painatuskeskus.

**Norman, Donald** (2002) Design Every Day Things.

<http://sse.tongji.edu.cn/liangshuang/hci2013spring/readings/the-design-of-everyday-things.pdf> (luettu 15.4.2013)

**Open Methodology** (2013) Big Data definition.

[http://mike2.openmethodology.org/wiki/Big\\_Data\\_Definition](http://mike2.openmethodology.org/wiki/Big_Data_Definition) (luettu 15.4.2013)

**Rogers, Simon** (2012) Anyone Can Do It. Data Journalism Is The New Punk.

<http://www.guardian.co.uk/news/datablog/2012/may/24/data-journalism-punk>  
(luettu 15.4.2013)

**Segel, Edward & Heer, Jeffrey** (2010) Narrative Visualization.

Telling Stories with Data. Stanford vis group.

<http://vis.stanford.edu/papers/narrative> (luettu 15.4.2013)

**Tilastokeskus** (2012) Tilastokoulu.

[http://tilastokoulu.stat.fi/?page\\_type=sisalto&course\\_id=tkoulu\\_tlkt&lesson\\_id=9&subject\\_id=1](http://tilastokoulu.stat.fi/?page_type=sisalto&course_id=tkoulu_tlkt&lesson_id=9&subject_id=1) (luettu 15.4.2013)

**Tufte, Edward R.** (1983) The Visual Display of Quantitative Information.

Cheshire, CT. Graphics Press. 2nd Edition.

**Wroblewski, Luke** (2011) Mobile First.

<http://alistapart.com/article/organizing-mobile> (luettu 15.4.2013)

**Marcotte, Ethan** (2011) Responsive Web Design.

<http://alistapart.com/article/fluid-images> (luettu 15.4.2013)

## *Esimerkkitoteutukset*

Hakemistosivu

<http://demo.newpie.fi>

Eri teknologioiden käyttöesimerkkejä

<http://demo.newpie.fi/googlecharts.php>

<http://demo.newpie.fi/chartjs.php>

<http://demo.newpie.fi/chartwell.php>

<http://demo.newpie.fi/d3.php>

Esimerkejä brändielementeistä

<http://demo.newpie.fi/brandi-viiva.php>

<http://demo.newpie.fi/brandi-pylvasryhma.php>

<http://demo.newpie.fi/brandi-kalenteri.php>

Käyttöliittymäesimerkkejä

<http://demo.newpie.fi/suodatus.php>

<http://demo.newpie.fi/jarjestys.php>

<http://demo.newpie.fi/zoom.php>

Tarinallinen esimerkkitoteutus

<http://demo.newpie.fi/irtisanomisaalto>



## Linkkejä ja esimerkkikuvia

- <sup>1</sup> Esimerkkikuva: Paperilehden taiton ehdoilla suunniteltu.  
<http://xkcd.com/980/huge/#x=-1504&y=-2024&z=4> (luettu 15.4.2013)
- <sup>2</sup> Esimerkkikuva: Vähän dataa.  
<http://maclab.edisonchargers.com/blogs/ya392033/files/2013/01/lemons-info-graphic.jpg> (luettu 15.4.2013)
- <sup>3</sup> Esimerkkikuva: Kulmikkaita graafeja.  
<http://aijaa.com/O015711845457.jpg> (luettu 15.4.2013)
- <sup>4</sup> HS Next blogi  
<http://blogit.hs.fi/hsnext/> (luettu 15.4.2013)
- <sup>5</sup> Yle Plusdesk visualisointeja  
<http://yle.fi/plus> (luettu 15.4.2013)
- <sup>6</sup> Apps For Finland kilpailu.  
<http://apps4finland.fi/> (luettu 15.4.2013)
- <sup>7</sup> Uutisraivaaja. Kilpailu.  
<http://uutisraivaaja.fi/> (luettu 15.4.2013)
- <sup>8</sup> Forum Virium  
<http://forumvirium.fi/> (luettu 15.4.2013)
- <sup>9</sup> OKF Finland  
<http://fi.okfn.org/> (luettu 15.4.2013)
- <sup>10</sup> New York Timesin Datavisualisointeja  
<http://www.nytimes.com/interactive/2012/12/30/multimedia/2012-the-year-in-graphics.html?hp> (luettu 15.4.2013)
- <sup>11</sup> Malofiej infographic awards  
<http://www.malofiej21.com/> (luettu 15.4.2013)
- <sup>12</sup> Adrian Holovaty: Chicago Crime map.  
[http://mappinghacks.com/talks/bmug\\_20060824/img/gmaps\\_chicago\\_crime.png](http://mappinghacks.com/talks/bmug_20060824/img/gmaps_chicago_crime.png) (luettu 15.4.2013)
- <sup>13</sup> The Telegraph: MA`'s Expencies.  
<http://www.telegraph.co.uk/news/newstopics/mps-expenses/>
- <sup>14</sup> Hans Roslingin Gapminder-verkkopalvelu.  
<http://www.gapminder.org/>
- <sup>15</sup> Lista kiinnostavista sijaintiin perustuvista infografiikoista:  
<http://socialwayne.com/2010/08/22/location-based-infographics-showing-how-and-why-the-mobile-world-checking/> (luettu 15.4.2013)
- <sup>16</sup> Käyttäjätietoa hyödyntäviä palveluita.  
<http://kaiserthesage.com/content-marketing-ux/> (luettu 15.4.2013)





JOS EMME YMMÄRRÄ KODIA, MEIDÄN TÄYTYY KERTO A TARINOITA.