

Tuomo Räsänen

**DATAVISUSALISOINNIN
HYÖDYNTÄMINEN**
Case: Etelä-Savo ennakoi

Opinnäytetyö
Tietojenkäsittelyn koulutusohjelma

2018



**Kaakkois-Suomen
ammattikorkeakoulu**

Tekijä/Tekijät	Tutkinto	Aika
Tuomo Räsänen	Tradenomi (AMK)	marraskuu 2018
Opinnäytetyön nimi		28 sivua
Datavisualisoinnin hyödyntäminen Case: Etelä-Savo Ennakoi		
Toimeksiantaja		
Etelä-Savon Maakuntaliitto		
Ohjaaja		
Tomi Numento		
Tiivistelmä		
<p>Tämä opinnäytetyö käsittelee datavisualisoinnin hyödyntämistä Etelä-Savon maakuntaliiton ”Etelä-Savon ennakointiverkko” -hankkeen toteutuksessa. Hankkeen tavoitteena on vahvistaa alueen ennakointikulttuuria ja lisätä alueen menestyksen mahdollisuuksia alati muuttuvassa toimintaympäristössä. Hankkeen yhtenä tavoitteena oli uudistaa vanha verkkosivusto ja luoda uusi verkkoalusta, jossa maakunnan tilastotiedot ovat hankkeen kohderyhmille selkeästi esillä ja maakuntaliitolle helposti lisättävissä erilaisista tietovarastoista.</p> <p>Opinnäytetyön teoriaosuudessa tarkastellaan tietokokonaisuuksien visualisointia hahmopsykologian näkökulmasta. Työssä tutustutaan yleisimpiin tilastokuvioihin, ja niiden ongelmiin. Teoriaosuudessa vertaillaan myös erilaisia visualisointimenetelmiä ja -työkaluja ja tutustutaan siihen, miten interaktiivisuus voi tukea visualisointia.</p> <p>Opinnäytetyön tuloksena syntyi verkkosivusto, joka on apuväline Etelä-Savon maakuntaliitolle tilastotietojen julkaisuun ja jakamiseen. Datavisualisoinnin työkaluksi työssä valittiin Infogram -verkkopalvelu, joka osoittautui vastaamaan hankkeen tarpeita parhaiten.</p>		
Asiasanat		
datan visualisointi, infografiikka, datavisualisoinnin työkalut, havaintopsykologia, interaktiivinen visualisointi		

Author (authors)	Degree	Time
Tuomo Räsänen	Bachelor of Business Administration	November 2018
Thesis title		28 pages
Utilizing data visualization Case: Etelä-Savo Ennakoi		
Commissioned by		
Etelä-Savon Maakuntaliitto		
Supervisor		
Tomi Numento		
Abstract		
<p>The purpose of this bachelor's thesis was to study how data visualization tools could be utilized in The Regional Council of South Savo's project "Etelä-Savon ennakointiverkko". The aim of the project was to strengthen the region's ways of observing the foreseeable future and to increase the opportunities for success in a changing environment. One of the project's priorities was to modernize the former website and create a new platform where the Regional Council's statistics would be easily available.</p> <p>The theoretical part of the thesis examined data visualization from the perspective of Gestalt psychology. The study explored the most common types of charts and their problems. The theoretical part also compared different kinds of data visualization methods and tools and explained how interactivity would support data visualization.</p> <p>The result of the empirical part of the study was a tool for the Regional Council of South Savo to publish and share statistics. The tool chosen for the project was a web service called Infogram which proved to suit the the project's needs best.</p>		
Keywords		
data visualization, information graphics, data visualization tools, Gestalt psychology, interactive visualization		

SISÄLLYS

1	JOHDANTO.....	5
2	DATAVISUALISOINTI JA INFOGRAFIikka.....	6
2.1	Tieto, data, informaatio	6
2.2	Datavisualisointi ja infografiikka.....	7
2.3	Gestaltin hahmolait (Gestalt laws)	8
3	VISUALISOINNIN TOTEUTTAMINEN.....	10
3.1	Yleisimmät tilastokuviot	10
3.2	Tilastokuvioiden ongelmia & yleisiä ohjeita.....	15
3.3	Interaktiivisuus visualisoinnin apuna.....	17
3.4	DATAVISUALISOINNIN TYÖKALUJA.....	19
3.4.1	Tableau.....	19
3.4.2	D3.js JavaScript kirjasto.....	19
3.4.3	Infogram.....	20
4	CASE: ETELÄ-SAVO ENNAKOI	21
4.1	Työn tavoitteet	21
4.2	Toteutus.....	22
5	TYÖN TULOKSET	26
6	PÄÄTÄNTÖ	27
	LÄHTEET.....	28

1 JOHDANTO

Data on uusi raaka materiaali. Nykyään lähes kaikilla on pääsy loputtomaan informaatioon muutamissa sekunneissa. Tieteellisen tiedon määrä kaksinkertaistuu joka yhdeksäs vuosi (Bormann & Mutz 2014), ja on arvioitu, että uuden digitaalisen tiedon määrä kasvaa n. 60 % vuosittain (Dahlberg 2015). Datan on kuitenkin oltava pureksittuna sellaiseen muotoon, että jokainen voi ymmärtää sen.

Datan visualisointi on ollut viime aikoina esillä mediassa yhä enemmän, ja tähän on vaikuttanut avoimen datan lisääntynyt tarjonta ja helppokäyttöiset visualisointityökalut, joita on nykyään tarjolla monia. Hyvänä esimerkkinä tästä on sanomalehtien verkkojulkaisuissa käytetyt interaktiiviset datavisualisoinnit, jotka tukevat julkaisua.

Tämä opinnäytetyö käsittelee datavisualisoinnin työkaluja, joiden avulla avoin data voidaan esittää interaktiivisessa ja helppoluikuisessa muodossa. Työn tavoitteena oli luoda Etelä-Savon maakuntaliitolle uusi verkkoalusta, jossa maakunnan tilastotiedot ovat koottuna yhteen paikkaan, ja josta ne ovat helposti saatavilla ja jaettavissa tilastotietoja tarvitseville ammattilaisille ja myös muille tilastoista kiinnostuneille.

Työn teoriaosuus keskittyy käsittelemään informaation visualisointia hahmolaerien kautta – miksi ihminen ymmärtää paremmin muotoja ja värejä, kuin pelkkää tekstiä tai numeroita. Työssä käydään läpi visualisoinnin erilaisia malleja, niin onnistuneita, kuin epäonnistuneitakin ja esitellään datavisualisoinnin työkaluja erilaisiin tarkoituksiin. Tietolähteinä olen käyttänyt alan kirjallisuutta, tiedejulkaisuja, artikkeleita ja blogeja.

Opinnäytetyö pyrkii antamaan lukijalle perusteet datavisualisoinnin hyödyntämiseen erilaisia tilastokuvioita ja visualisointiohjelmaa hyödyntäen.

2 DATAVISUALISOINTI JA INFOGRAFIikka

2.1 Tieto, data, informaatio

Useimmissa tiedon määrittelyissä korostuu sana *tosi*. *Tieto* on siis jotain, joka on totta. Tietoa on monenlaista: on kuultua, opittua, lehdistä luettua tai itse pääteltyä (Työväen Akatemia 2009). Suomen kielen sana tieto onkin moniselitteinen ja sillä saatetaan viitata tietoon, dataan, informaatioon tai tietämykseen.

Datalla tarkoitetaan yksittäisiä tiedon jyvää, joita ei ole varsinaisesti käsitelty tai jalostettu. Dataa on esimerkiksi sääaseman rekisteröimä lämpötila tietyllä hetkellä. Data muuttuu hyödylliseksi vasta, kun sitä tulkitaan ja samalla se muuttuu informaatioksi. *Informaatio* on jalostettua dataa, jolla on merkitys tarkastelijalleen. Informaatiota olisi tällöin sääaseman lämpötiloista koottu sarja, joka osoittaa keskilämpötilan nousseen merkittävästi. Dataa voidaan luokitella myös erilaisiin kategorioihin joista yleisimpiä ovat *big data*, *avoin data* ja *meta-data*. (Tiedonjohtamisen tutkimuskeskus NOVI 2014.)

Big datalla ei ole tarkkaa määritelmää, eikä se ole vielä täysin vakiintunut käsite. Siihen usein liitetään kolme määrettä: määrä (*volume*), nopeus (*velocity*) ja monimuotoisuus (*variety*).

- Määrä: Datamäärän johdosta se ei ole käytettävissä yleisesti käytössä olevilla laitteilla ja ohjelmistoilla järkevässä ajassa.
- Nopeus: Dataa kertyy monista lähteistä eri muodoissa, usein automaattisesti kerättynä, ja se kasaantuu ja/tai muuttuu nopeasti.
- Monimuotoisuus: Datalla ei välttämättä ole lainkaan rakennetta tai se on vain löyhästi määritelty, minkä johdosta sen analysointi sellaisenaan on hankalaa (Sovelto 2018).

Avoin data on dataa, joka on kenen tahansa uudelleen käytettävissä maksutta, luvallisesti ja koneluettavassa muodossa (Avoindata 2016). Avointa dataa käytettäessä tulee mainita aineiston alkuperäinen lähde.

Metadatum on dataa datasta. Se voi esimerkiksi kuvailla datan rakennetta, ja se mahdollistaa asiakirjallisten tietojen haun ja paikallistamisen (Avoindata 2016). Metadatum voi avata tiedon sisältöä, esimerkiksi GPS-ominaisuudella varustetulla kameralla tai puhelimella otetut valokuvat saattavat paljastaa sijainnin jossa kuva on otettu, ajankohdan, sekä laitteen mallin.

2.2 Datavisualisointi ja infografiikka

Datavisualisoinnin tavoitteena on auttaa ihmisiä ymmärtämään dataa sijoittamalla se visuaaliseen muotoon. Visualisointi pohjautuu siis ei-visuaaliseen dataan, josta on vaikea lähteä tekemään havaintoja. Sijoittamalla data visualisointiohjelmiin, voidaan niistä havaita korrelaatioita ja trendejä helpommin.

Datavisualisoinnin ekspertti Edward Tufte (1983) määritteli kirjassaan *The Visual Display of Quantitative Information* listan laatukriteerejä datavisualisoinnille. Hänen mukaansa hyvän visualisoinnin tulee:

- näyttää itse data
- saada katselija keskittymään sisältöön esitystavan sijaan
- välttää dataa vääristymästä
- pystyä esittämään suuria tietomassoja tiiviissä paketissa
- esittää datajoukot yhtenäisessä muodossa
- rohkaista katsojaa vertailemaan datajoukkojen eri osia
- esittää data yleisellä sekä yksityiskohtaisella tasolla
- olla selkeästi joko selittävä, tutkaileva tai järjestelvä
- sanallisten kuvausten tulee tukea visualisointia.

Myös infografiikka perustuu datan käyttöön, mutta sitä ei ole usein luotu koneellisesti, vaan käsin. Datavisualisoinnissa tietoa voidaan lisätä ja poistaa myös jälkikäteen, mutta infograafi on valmis esitys, joka osoittaa jonkin tietyn asian tai havainnon. Infografiikka auttaa ihmistä ymmärtämään informaatiota, joka on vaikea selittää sanoin. Toimiva infograafi syntyy hyvästä sisällöstä ja tarkoituksenmukaisesta kuvituksesta. Kartta on mainio klassinen esimerkki infografiikasta.



Kuva 1. Harry Beckin ensimmäinen kartta Lontoon metrotunneleista (Rendgren 2012, 82)

Harry Beck oli ensimmäinen henkilö, joka piirsi vuonna 1933 Lontoon metroreitit kaavioimaisesti. Siinä täyttyy moni Tuften luettelema kriteeri hyvästä visualisoinnista. Kartasta ei voi havaita selkeitä välimatkoja, ja Beckin mukaan tämä olikin toissijainen seikka – paljon tärkeämpää oli nähdä nopealla vilkaisulla, miten asemalta pääsee helpoiten toiselle, ja missä kohti täytyy vaihtaa metroa. Se saa katelijan keskittymään sisältöön esitystavan sijaan, on selkeästi selittävä ja näyttää itse datan. Beckin kartta onkin yksi tunnistettavimmista julkisen liikenteen kartoista, ja samaa mallia on käytetty monissa suurkaupungeissa ympäri maailmaa.

2.3 Gestaltin hahmolait (Gestalt laws)

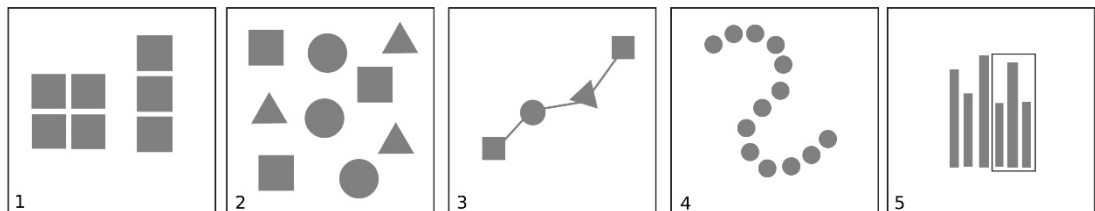
Visualisointi on hyvä aloittaa tutustumalla hahmolakeihin. Hahmopsykologiaan perustuvilla Gestalt-hahmolaeilla pyritään osoittamaan sitä, miten yksittäiset kohteet havaitaan laajempina kokonaisuuksia. Havaintojärjestelmämme pyrkii ryhmittelemään kuviot yhteenkuuluviksi.

Hahmolakien ymmärtäminen voi auttaa esimerkiksi ohjelman käyttöliittymän suunnittelussa luomaan loogisia kokonaisuuksia käyttöliittymän elementeistä.

Käyttöliittymän elementtien on hyvä myös pohjautua yleisimpiin käytäntöihin ja tapoihin. Esimerkiksi nuolet toimivat eteen/taakse navigoinnin merkkeinä.

Alberto Cairo (2013, 114-117) korostaa infografiikassa erityisesti seuraavia Gestalt-hahmolakeja:

1. Läheisyyden laki (law of proximity): toisiaan lähellä olevat kuviot mielletään kuuluvan yhteen.
2. Samankaltaisuus (law of similarity): keskenään samankaltaiset kuviot (kooltaan, muodoltaan, väriltään) mielletään yhteenkuuluviksi.
3. Yhteenliittyminen (law of connectedness): toisiinsa esimerkiksi viivalla yhdistetyt kuviot mielletään samaan ryhmään kuuluviksi
4. Jatkuvuus (law of continuity): yhtenäinen viiva tai viivan muotoa noudattava kuviojoukko mielletään yhtenäiseksi
5. Sulkeutuvuus (closure): jos kuviot sijaitsevat esimerkiksi viivalla suljetulla alueella, katsoja mieltää alueeseen kuuluvat objektit samaan ryhmään



Kuva 2. Hahmolait havainnollistettuna (Cairoa 2013 mukailen Tuomo Räsänen 2018).

Pelkkiin hahmolakeihin suunnittelutyötä ei kannata kuitenkaan perustaa. Hahmolait ovat vain apuväline, ja jokainen ihminen käsittelee havaintojaan eri tavalla. Pessimisti näkee lasin puoliksi tyhjänä, optimisti puoliksi täynnä ja realististi puolillaan. (Mustonen 2001, 21.)

3 VISUALISOINNIN TOTEUTTAMINEN

3.1 Yleisimmät tilastokuviot

Tilastotietoa esitetään perinteisesti tekstin, taulukoiden tai kuvioiden avulla. Näistä kolmesta taulukko on yleisin esitystapa, mutta tilastokuvioita on alettu käyttää yhä enemmän, koska hyvin tehtynä ne voivat välittää asian tehokkaammin ja helpommin kuin perinteinen taulukko. (Kuusela, 2015.)

Tilastokuvion tärkein tehtävä on välittää tieto visuaalisesti. Tilastokuvioita voidaan käyttää mm. esittämään aineiston rakenteellisia ominaisuuksia, tiivistämään suuria tietomääriä ja välittämään ajatuksia ja johtopäätöksiä. Hyvä tilastokuvio kertoo tarinan. (Kuusela, 2015.)

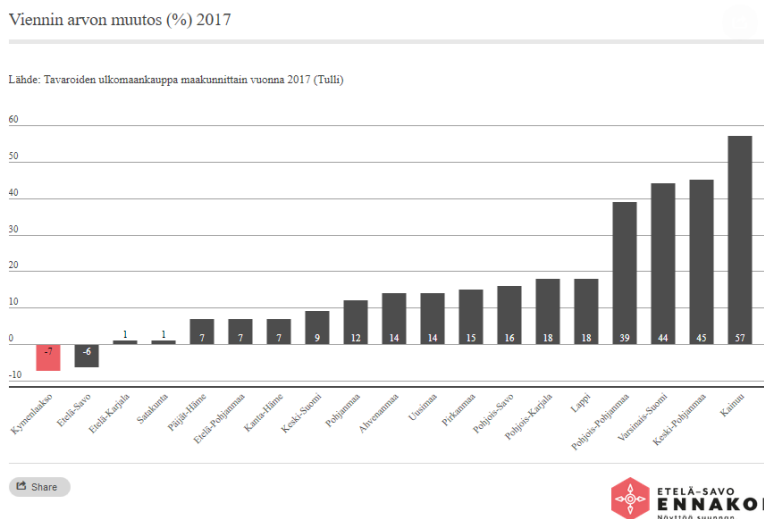
Suurin osa käytetyistä tilastokuvioista lukeutuu **viiva-**, **pylväs-**, ja **piirakkakuvioidiin**. Monilla näistä kuvioista on myös lukuisia alatyyppejä. Kuviotyypin valinnassa täytyy huomioida:

- esitettävän tiedon luonne (lukumäärät, prosentit, keskiarvo)
- kuvattavien ominaisuuksien määrä
- muuttujien mitta-asteikko
- luokiteltujen muuttujien asteikkotyyppi
- luokkien määrä (Kuusela, 2001, 52).

Tilastokuvio valitaan yleensä muuttujien asteikkotyyppien perusteella. Nämä muuttujat voivat olla laadullisia (kvalitatiivisia) tai määrällisiä (kvantitatiivisia). Määrällisten arvot ovat lukuja, joiden välillä voi suorittaa laskutoimituksia. Esimerkkejä kvantitatiivisista muuttujista ovat esimerkiksi lämpötila, ikä tai väkiluku. (Kuusela, 2001, 204-205.)

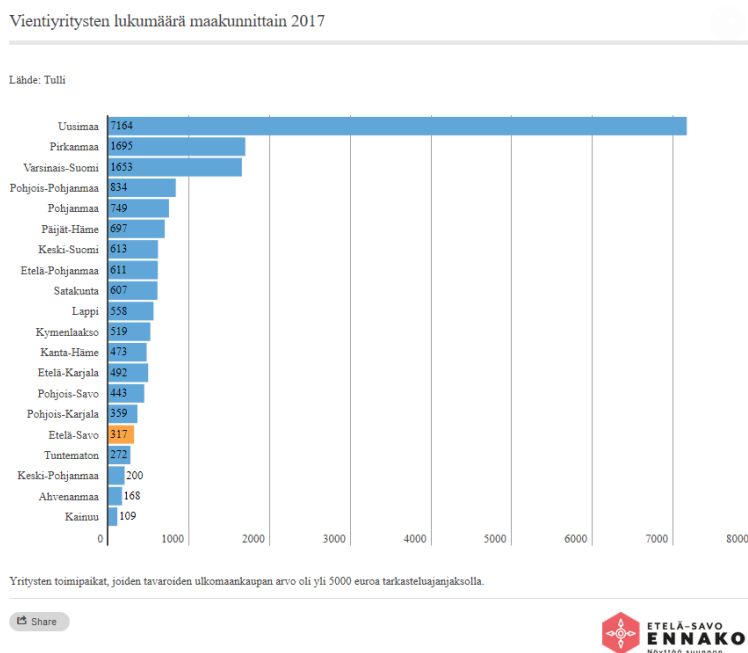
Laadullisia muuttujia voidaan kuvata myös numeroilla, mutta numeroilla ei ole lukujen ominaisuuksia. Laadullisten arvojen välillä ei voi suorittaa laskutoimituksia ja niiden muuttujat voivat esiintyä missä järjestyksessä tahansa. (Kuusela 2001, 54, 204-205.)

Pystyylväskuvio korostaa määrää ja määrän muutoksia, ja sen x-akselilla on oltava ominaisuus, jolla on jatkuva-arvoinen mitta-asteikko, esim. lämpötila. Pystyylväskuvion määräästeikko alkaa aina nolasta ja sekä y- että x-akselin asteikkojen on oltava tasavälisiä. Pylväsryhmissä tulisi olla korkeintaan kolme pylvästä rinnakkain. (Kuusela, 2015.)



Kuva 3. Pystyylväskuvio (Ruutukaappaus, Etelä-Savo Ennakoi).

Vaakapylväskuvio soveltuu luokkien tai ryhmien määrien kuvaamiseen. Y-akselilla olevan ominaisuuden ei tarvitse olla jatkuva-arvoinen. Vaakapylväskuvio siis ei voi olla vaihtoehto pystyylväskuviolle, kuten yleensä virheellisesti luullaan. (Kuusela 2015.)

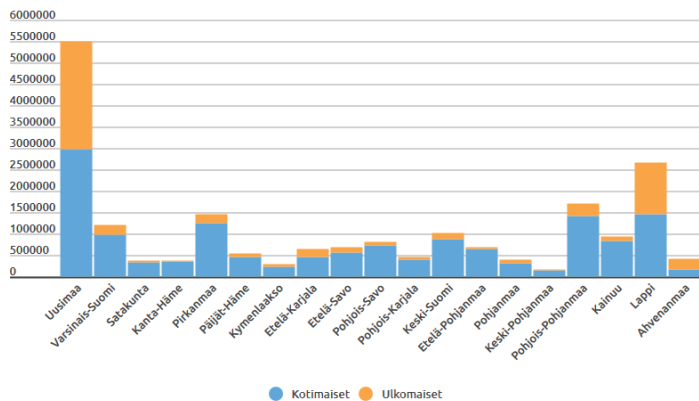


Kuva 4. Vaakapylväskuvio (Ruutukaappaus, Etelä-Savo Ennakoi).

Summapylväskuviossa tarkastellaan kokonaismäärän ja sen osien kehitystä ja sen osilla täytyy olla samat mitta-asteikot. Sitä voidaan myös käyttää prosentiosien vertailuun ja piirakkakuvion sijaan käytettäväksi. (Kuusela 2001, 119-120.)

Kotimaiset ja ulkomaiset yöpyjät majoitusliikkeissä v. 2016 maakunnittain

Lähde: Tilastokeskus / Majoitustilastot

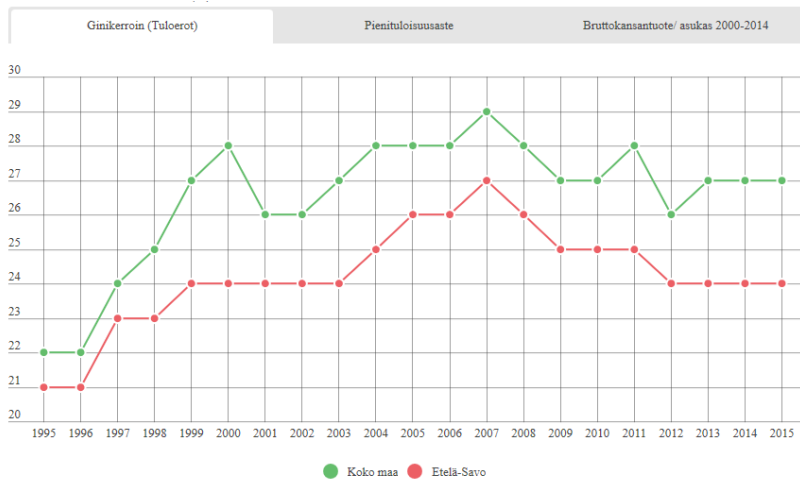


Share



Kuva 5. Summapylväskuvio (Ruutukaappaus, Etelä-Savo Ennakoi).

Viivakuviot kuvaavat kehitystä tai muutosta, joka muodostaa jatkumon. Tästäkin tilastokuviossa kuvattavan ilmiön tulee olla jatkuva-arvoinen. Viivan muodostaa havaintopisteet, jotka yhdistyvät toisiinsa. Viivakuviot poikkeaa muista tilastokuvioista, koska sen välittämä tieto ei perustu pinta-alaan. (Kuusela 2001, 52.)



- Ginikerroin on yksi yleisimmistä käytetyistä tuloerojen mittareista. Ginikerroinassa tuloerojen taso on kiteytetty yhteen lukuarvoon, mikä voi vaihdella nolian ja yhden välillä. Mitä suuremman arvon Ginikerroin saa, sitä epätasaisempi tulotilanne on. Työllisyys ja työttömyys ovat keskeisimmät tuloeroihin vaikuttavat tekijät. Hyvin suuren tuloerojen kannon on heikentävien sosiaalisten kokeistojen, hoitamattomuus ja liikkuvuus, joiden heikentämisellä on negatiivisia vaikutuksia väestön hyvinvointiin ja talouteen.
- Pienituloisuusaste -tilasto kuvaa, kuinka suuri osa alueen väestöstä kuuluu kottaloukseen, josden tulot jäävät suhteellisen pienituloisuusrajan alapuolelle. Kyse on mittarista, joka kuvaa väestön asemaa suhteellisesti tulotilakannalla.
- Bruttokansantuote mittaa tietyn valtion kokonaistuotantoa, johon lasketaan mukaan kaikki tietynä aikana tuotetut tavarat ja palvelut. Kun lasketaan BKT asukaista kohden, BKT jaetaan valtion asukkaisten määrällä.

Share

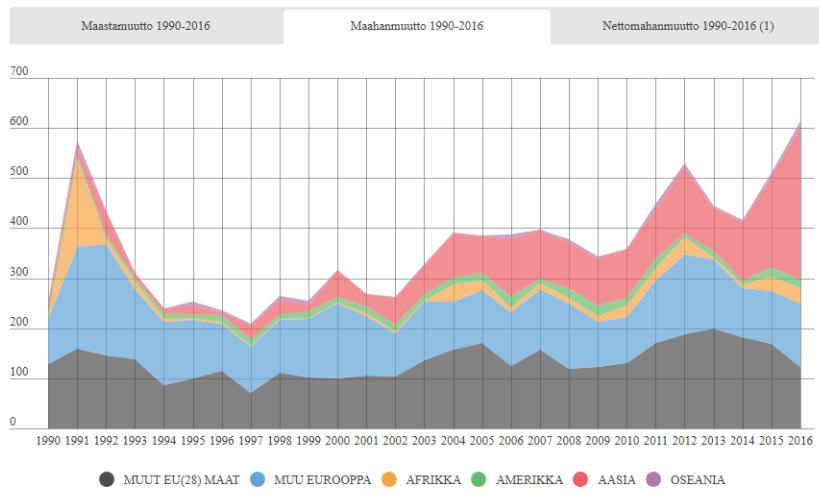


Kuva 6. Viivakuviot (Ruutukaappaus, Etelä-Savo Ennakoi).

Aluekuvio korostaa viivan sijaan sen alle jäävää aluetta. Se on yksi viivakuvion muunnoksista. Tässäkin tilastokuviossa määräästeikon on alettava nolasta.

Summaviivakuvio on aluekuvion alatyyppejä, jolla pystyy esittämään monta asiaa samanaikaisesti. Tässä kuviossa on tärkeää sijoittaa ne kuviot alimmaisiksi, missä tapahtuu vähintään vaihtelua, ja ylimmäksi ne missä vaihtelu on suurin (Kuusela 2001, 96-98).

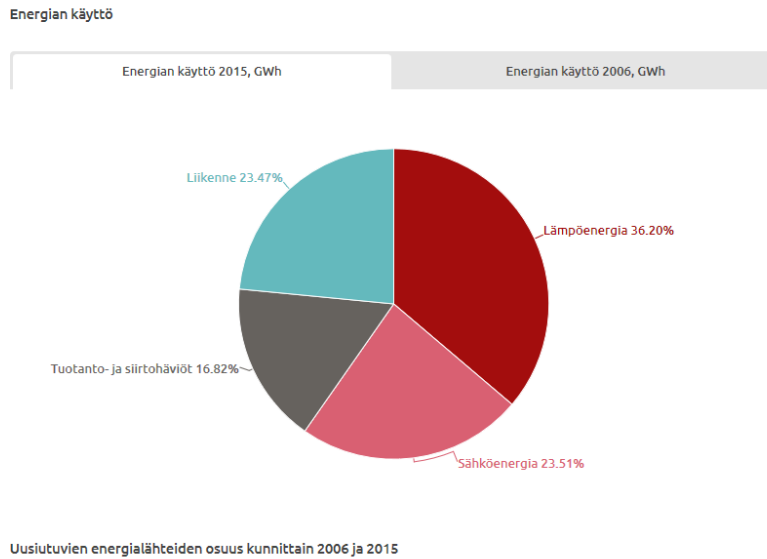
Etelä-Savon siirtolaisuuden kehitys 1990-2016



Kuva 7. Summaviivakuvio (Ruutukaappaus, Etelä-Savo Ennakoi)

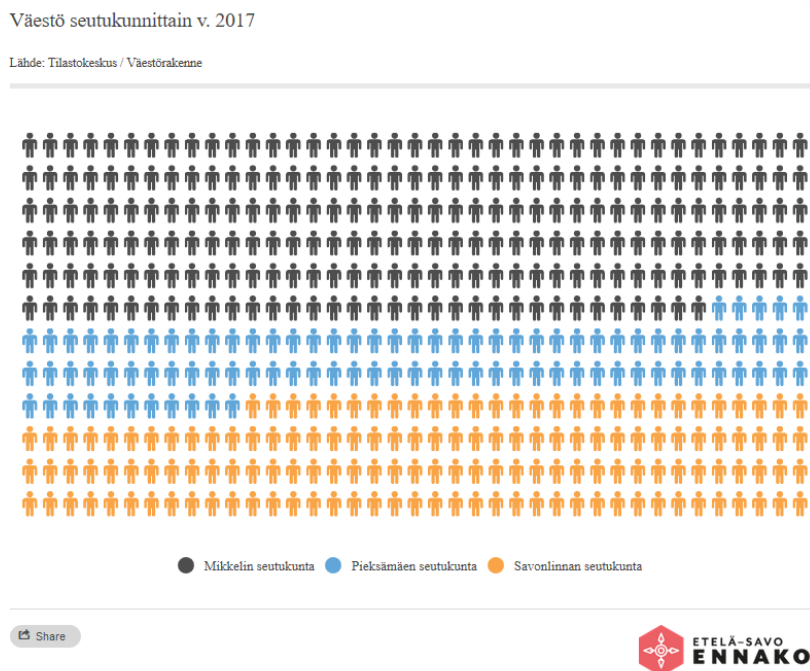
Piirakkakuvio eli ympyrädiagrammi on hyvin suosittu ja se soveltuu ainoastaan kuvaamaan sitä, miten jokin kokonaisuus jakautuu eri osiin. Kokonaisuus on jaettu sektoreihin, jotka muodostavat yhdessä sata prosenttia.

Piirakkakuvion sektorit tulisi järjestää suurusjärjestykseen ja sektoreita saisi olla maksimissaan kuusi kappaletta (Kuusela 2001, 148-149.)



Kuva 8. Piirakkakuviio (Ruutukaappaus, Etelä-Savo Ennakoi).

Yksikkösymbolikuviossa mitta-asteikkona toimii samankokoiset symbolit. Esimerkiksi väestöä kuvatessa yksi ihmisen kuva voi tarkoittaa 100 000 ihmistä. Yksikkösymbolikuvio noudattaa pysty- tai vaakapylväskuvion muotoa. (Kuusela 2001, 186.)

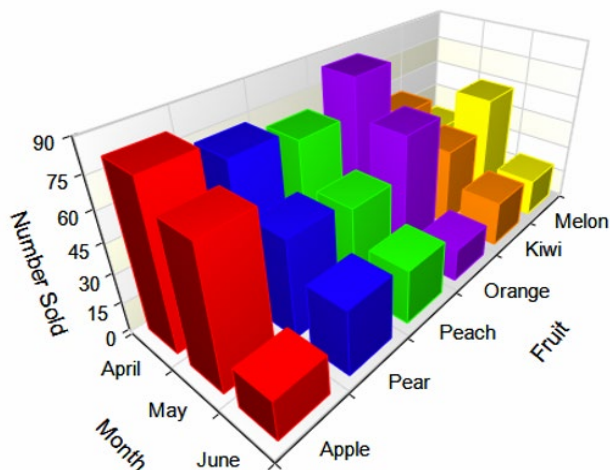


Kuva 9. Yksikkösymbolikuvio (Ruutukaappaus, Etelä-Savo Ennakoi)

3.2 Tilastokuvioden ongelmia & yleisiä ohjeita

Visualisointia suunniteltaessa on tärkeintä löytää oikea esitystapa tiedolle. Tietoa ei saa yrittää pakata samaan kuvaan liikaa. Erityistä huomiota tulee kiinnittää skaalattavien symboleihin käyttöön ja kolmiulotteisiin kuviin.

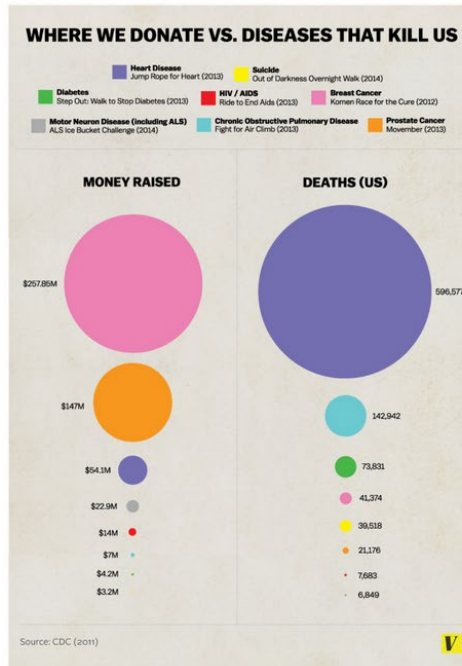
Kolmiulotteisuuden käyttö pitäisi rajoittaa vain tapauksiin, jossa kolmannesta ulottuvuudesta on todella hyötyä (Kuusela, 2015). Kuvassa on rakennettu esimerkki huonosta kolmiulotteisesta tilastokuviosta, ja niitä käytetään harmillisen usein. Kuvio voi näyttää nopealla vilkaisulla hienolta, mutta tarkemmalla tarkastelulla jotkin pylväät jäävät toisten piiloon.



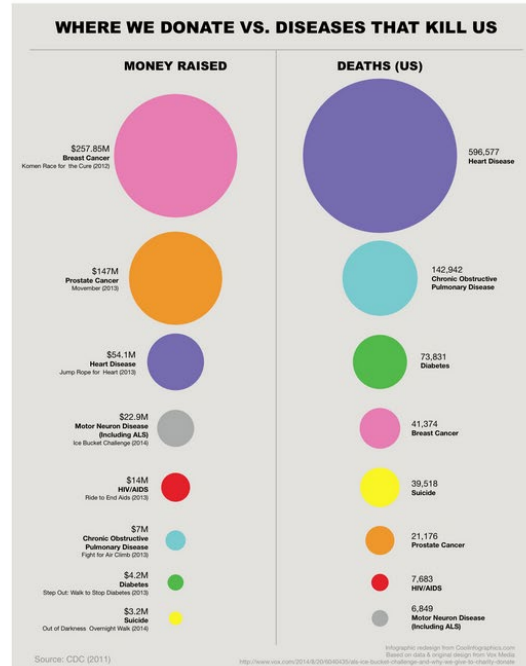
Kuva 10. Epäonnistunut visualisointi, kolmannen ulottuvuuden käyttö (KevinJ-Z, 2016)

Skaalattavien symbolien kohdalla tulee muistaa, että kuvioden vertailu tapahtuu pinta-alan mukaan. Esimerkkikuvassa pallojen koot eivät vastaa niitä kuvaavia rahasummia suhteessa toisiinsa. Kuvassa on myös samasta visualisoinnista korjattu versio.

Original Design



Corrected Design



Kuva 11. Esimerkki skaalattavista symboleista (Ruutukaappaus, Visual learning center)

Kuusela (2015) antaa oppaassaan ohjeita hyvään visualisointiin:

Tunne yleisösi ja mediasii. Visualisointia tehdessä täytyy pitää koko ajan mielessä sen kohdeyleisö. Suosittuun sanomalehteen ei kannata tehdä samaa visualisointia kuin vaikkapa asiantuntijapiirille. Myös esitystapa on otettava huomioon, sillä jokainen media esittää visualisoinnin eri tavoin.

Tunne rajoituksesi. Ota selvää millainen on hyvä visualisointi äläkä lähde tekemään kuviota tuntematta esitettävää asiaa tarpeeksi hyvin.

Älä tyydy grafiikkaohjelmien esitykseen. Visualisointiohjelmia on markkinoilla monia, ja Kuuselan (2015) mukaan yksikään niistä ei ole täydellinen ja ”on suuri kiusaus tyytyä siihen, mitä grafiikkaohjelmat antavat myöten, vaikka tulos ei tuntuisikaan kovin hyvältä.”

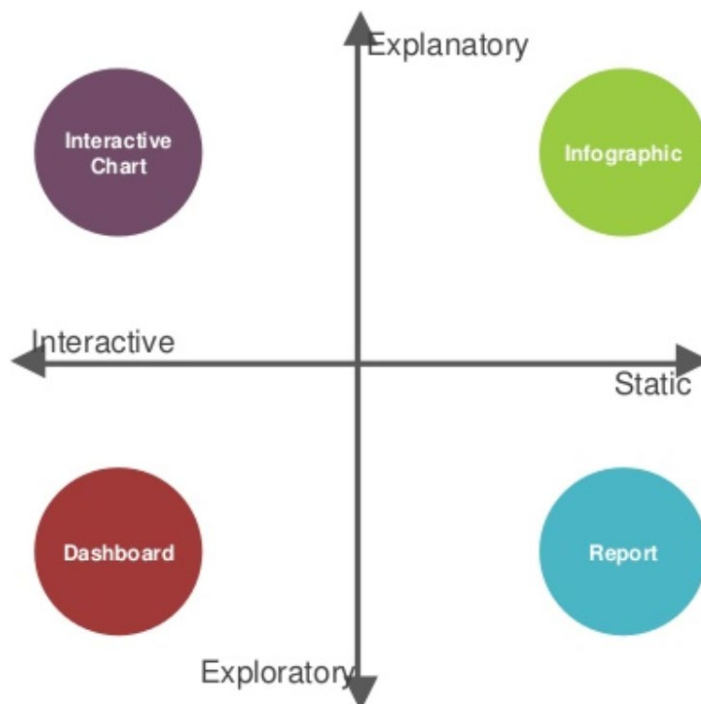
Älä ainakaan tyydy grafiikkaohjelmien oletusarvoihin. Grafiikkaohjelmien oletusasetukset on tehty lähinnä myyntitarkoituksiin ja siksi kannattaakin tutustua ohjelman ominaisuuksiin huolella ja kokeilla erilaisia vaihtoehtoja.

Vältä ”ankkoja” ja kuvioroinan käyttöä. Ankoiksi kutsutaan kuvioita, jotka eivät millään tapaa liity esitettävään asiaan eivätkä tuo kuvioon mitään lisäarvoa. Ne ovat vain kiinnittämässä huomiota. Tätä kutsutaan myös kuvioroinaksi. Esimerkiksi kolmas ulottuvuus tilastokuviossa voidaan laskea kuvioroinaksi.

3.3 Interaktiivisuus visualisoinnin apuna

Interaktiivisia visualisointiohjelmia ja kirjastoja on viime vuosina tullut markkinoille monia erilaisia. Interaktiiviset visualisoinnit vaativat ennen paljon ohjelmointiosaamista, mutta nykyään niitä voidaan tehdä graafisten käyttöliittymien kautta suhteellisen vaivattomasti.

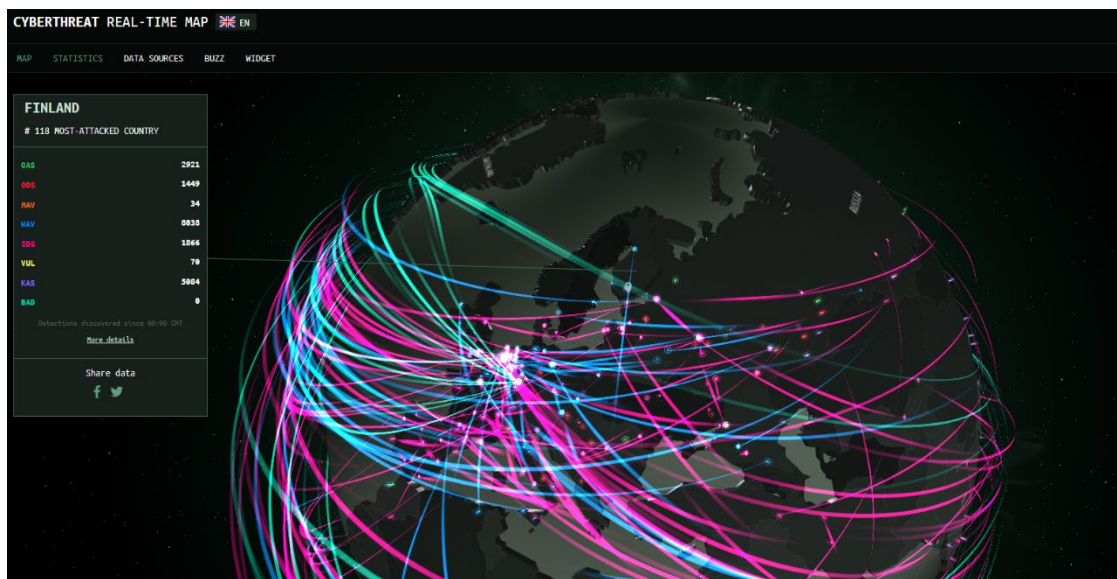
Interaktiivinen visualisointi antaa käyttäjälle mahdollisuuden tutkia dataa itsenäisesti. Se voidaankin määritellä eksploraatiiviseksi (exploratory) visualisoinniksi. Käyttäjälle annetaan vastuu ja mahdollisuus omien johtopäätösten tekemiseen. Interaktiivisen visualisoinnin avulla voidaan esittää dataa useammasta eri näkökulmasta verrattuna staattiseen kuvaajaan. Interaktiivisuus lisää datan saavutettavuutta ja parantaa käyttökokemusta (Väisänen 2017). Interaktiivisuudesta on erityisesti hyötyä, kun tarkastellaan suuria tietomassoja.



Kuva 12. Visualisointimallit eroteltuina (Ruutukaappaus AlmPartners)

Eksploratiivisen visualisoinnin vastakohta on selittävä (explanatory) visualisointi. Staattiset visualisoinnit pyrkivät olemaan selittäviä, ja niissä kuvan laattija on määritellyt millaisen viestin hän pyrkii välittämään. (Väisänen 2017.)

Interaktiivista visualisointia tulisi käyttää staattisen sijaan, kun käyttäjän halutaan itse tutkivan dataa. Interaktiivisuudesta on varsinkin silloin hyötyä, kun data visualisoinnin sisällä päivittyy (Väisänen 2017). Esimerkiksi Kaspersky Labsin reaaliaikaisessa virushyökkäysten karttavisualisoinnissa käyttäjä pystyy itse määrittelemään minkä maan hyökkäystilastoja haluaa tarkastella.



Kuva 13. Reaaliaikainen interaktiivinen visualisointi (Ruutukaappaus, Kaspersky Cybermap)

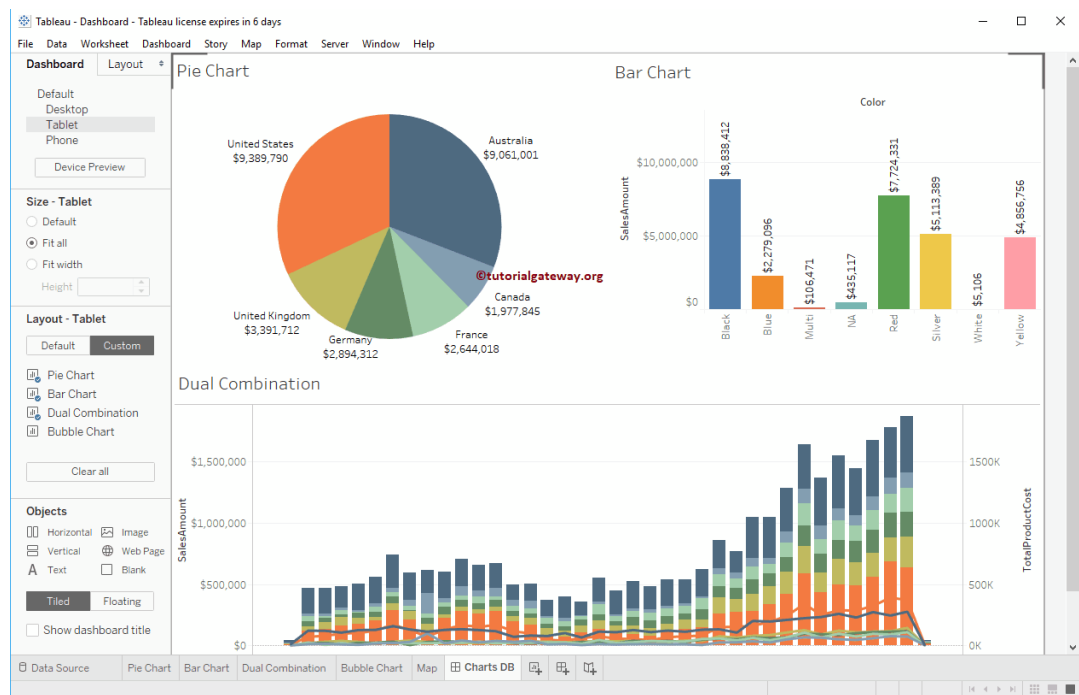
Internetissä näkee nykyään monia saman kaltaisia visualisointeja, missä käyttäjä pääsee ikään kuin ”sukeltamaan” datan sisälle. Yleisimmät toiminnot interaktiivisessa datan visualisoinnissa ovat **suodattaminen**, **porautuminen** ja **siirtyminen**.

Suodattamisessa käyttäjä pystyy suodattamaan dataa yhden tai useamman kategorian avulla, jolloin kuvaaja päivittyy näyttämään valittuja kategorioita. Porautuminen on visualisoinnin sisällä siirtymistä datan hierarkiassa alemmalle tasolle. Siirtyminen taas tarkoittaa tietojen siirtämistä toiselle sivulle, koska yhdelle sivulle ei täydy yrittää saada mahtumaan kaikkea tietoa. (Tolvanen 2018.)

3.4 DATAVISUALISOINNIN TYÖKALUJA

3.4.1 Tableau

Tableau on maailman johtava BI-työkalu (liiketoimintatiedon visuaalinen hallintaohjelmisto). Sen takana on Yhdysvaltalainen yhtiö, joka perustettiin vuonna 2003 ja jonka toiminta keskittyy liiketoiminnan hallintaan ja datavisualisointeihin.



Kuva 14. Tableau Desktop ohjelmalla tehty "dashboard" (Googlen kuvahaku)

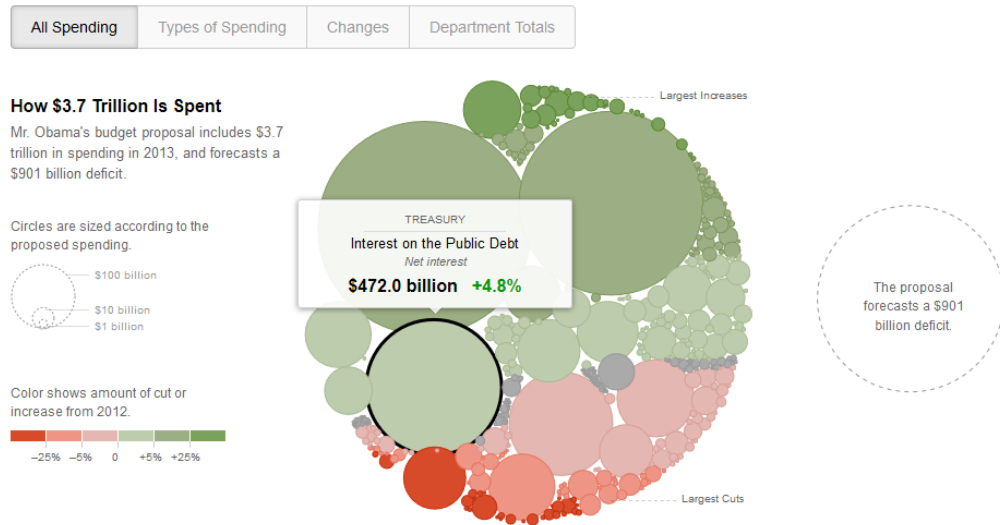
Tableau Desktop on interaktiivinen datavisualisointityökalu yrityksille. Sitä markkinoidaan helpoksi käyttää, eikä sitä varten tarvitse opetella ohjelmointikieliä. Siihen voidaan tuoda dataa monesta eri lähteestä, Excelistä Big Dataan. Kun data on tuotu ohjelmaan, käyttäjä pystyy rakentamaan itselleen halutun näköisen dashboardin ohjelman oletusvisualisoinneista.

3.4.2 D3.js JavaScript kirjasto

D3.js on JavaScript -kirjasto, joka on tehty datan visualisoinnista varten. Kirjasto käyttää HTML, CSS ja SVG tekniikoita, ja siksi se vaatiikin ohjelmointiosaamista. D3-kirjastolla pystyy tekemään juuri sellaisen visualisoinnin kuin itse haluaa, mutta se vaatii myös opettelua.

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.



Kuva 15. D3-kirjastolla toteutettu vuoden 2013 Obaman budjettiehdotus. (Ruutukaappaus, The New York Times Archives, 2013)

Kun osaamista löytyy, pystyy kirjastolla toteuttamaan mitä erilaisempia interaktiivisia visualisointeja. Esimerkiksi The New York Times käyttää verkkoartikleissaan usein D3-kirjastolla tehtyjä kuvaajia. Alla olevassa kuvassa on visualisointi Obaman 2013 kauden budjetointiehdotuksesta. Kuvaajassa on neljä eri välilehteä, joista jokainen avaa hieman erilaisen näkymän samasta datasta.

3.4.3 Infogram

Infogram on helppokäyttöinen ja monipuolinen verkkoselaimessa toimiva interaktiivinen visualisointiohjelma. Näistä syistä se valittiinkin käyttöön toimeksiantossa. Ohjelma sisältää erilaisia valmiita suunnittelumalleja infografiikalle, raporteille, dashboardeille, kartoille ja somejulkaisuille. Infogramista otettiin käyttöön Business-versio, ja seuraavaksi luetellut ominaisuudet ovat osa sitä.

Dataa Infogramiin pystyy tuomaan monista eri lähteistä Excel -tiedostoista JSON syötteisiin. Infogram sisältää yli 37 interaktiivista kuvaajaa ja yli 550 karttaa. Valmiit visualisoinnit voidaan tuoda ohjelmasta eri tiedostomuodoissa (PNG, JPG, GIF, PDF, HTML) ja ne voidaan upottaa verkkosivuille responsiiv-

visesti. Ohjelma sisältää myös kuva- ja ikonipankin mitä voi hyödyntää visuaalisoinneissa huolehtimatta tekijänoikeuksista. Infogramilla on myös oma ohjelmointirajapinta.

Siinä missä Tableau on enemmän yrityksille ja D3-kirjasto ohjelmoijille, Infogram on palvelu, joka sopii hyvin datavisualisoinnin vasta-alkajille, ja joka pystyy myös skaalautumaan vaativiin tarkoituksiin. Kaikki tämän työn tilastokuvioesimerkit on toteutettu Infogram palvelulla.

4 CASE: ETELÄ-SAVO ENNAKOI

Tämän opinnäytetyön case tuli Etelä-Savon maakuntaliitolta toimeksiantona Mainostoimisto Kixit Oy:lle, ja se on osa Etelä-Savon ennakointiverkko hankkeen toteutusta. Projekti on tehty yhteistyössä mainostoimiston ja maakuntaliiton kanssa.

Hankkeen kuvaus tiivistettynä:

”Ennakointiverkko -hankkeen tavoitteena on vahvistaa alueen ennakointikulttuuria ja lisätä alueen menestyksen mahdollisuuksia alati muuttuvassa toimintaympäristössä. Hankkeessa luodaan jatkuvasti päivittyvä, tulevaisuuteen suuntaava tilannekuva, jossa huomioidaan aluetta kuvaavat indikaattorit sekä ulkoisen toimintaympäristön muutossuunnat. Tilannekuvaa varten rakennetaan uudenlainen osallistava ennakointialusta. Tilannekuva toteutetaan osin ostopalveluna ja sen konkreettinen muoto tulee olemaan verkkosivu/ verkkoalusta, jossa hyödynnetään monipuolisesti erilaisia tilastolähteitä ja aineistoja sekä esitetään tiedot helposti ja nopeasti omaksuttavassa, visuaalisesti tiiviissä muodossa.” (Eura Järjestelmä).

4.1 Työn tavoitteet

Työn tavoitteena oli rakentaa uusi verkkosivusto vanhan tilalle. Uuden verkkosivuston toivottiin olevan helppo päivittää, sen toivottiin sisältävän jokin yhteisöllinen elementti, ja sieltä tulisi löytää tilastotietoja maakunnasta interaktiivisessa, helposti ja nopeasti omaksuttavassa muodossa. Sivustolle haluttiin myös raikas uusi ilme, ja logo. Sivuston haluttiin olevan tarkoitettu kaikkien Etelä-Savon tilannekuvasta ja tulevaisuudennäkymistä kiinnostuneiden käyttöön ja hyödynnettäväksi.

Vanhan sivuston tilastotiedot olivat suurin osa staattisia kuvaajia, jotka oli upotettu sivustolle kuvina, tai ne olivat linkin takana PDF-muodossa. Uudelle sivustolle toivottiin interaktiivisuutta tilastografiikkaan, ja tätä varten tuli selvittää mikä työkalu olisi tähän tarkoitukseen paras mahdollinen.

4.2 Toteutus

Sivuston toteutus lähti liikkeelle julkaisualustan valitsemisesta. Suurin osa Kixitin tekemistä verkkosivuista on toteutettu joko WordPress-, tai Bonsait -julkaisualustalle. Verkkosivuille toivottiin helppoa päivitettävyyttä, ja tästä syystä julkaisualustaksi valikoitui Bonsait. Bonsait tarjoaa helppokäyttöiset työkalut sisällön tuottamiseen, ja kaikki näkymät ovat itse muokattavissa käyttäen html- ja css- kieliä. Bonsait on kuopiolaisen Entecore Ky:n tavaramerkki.



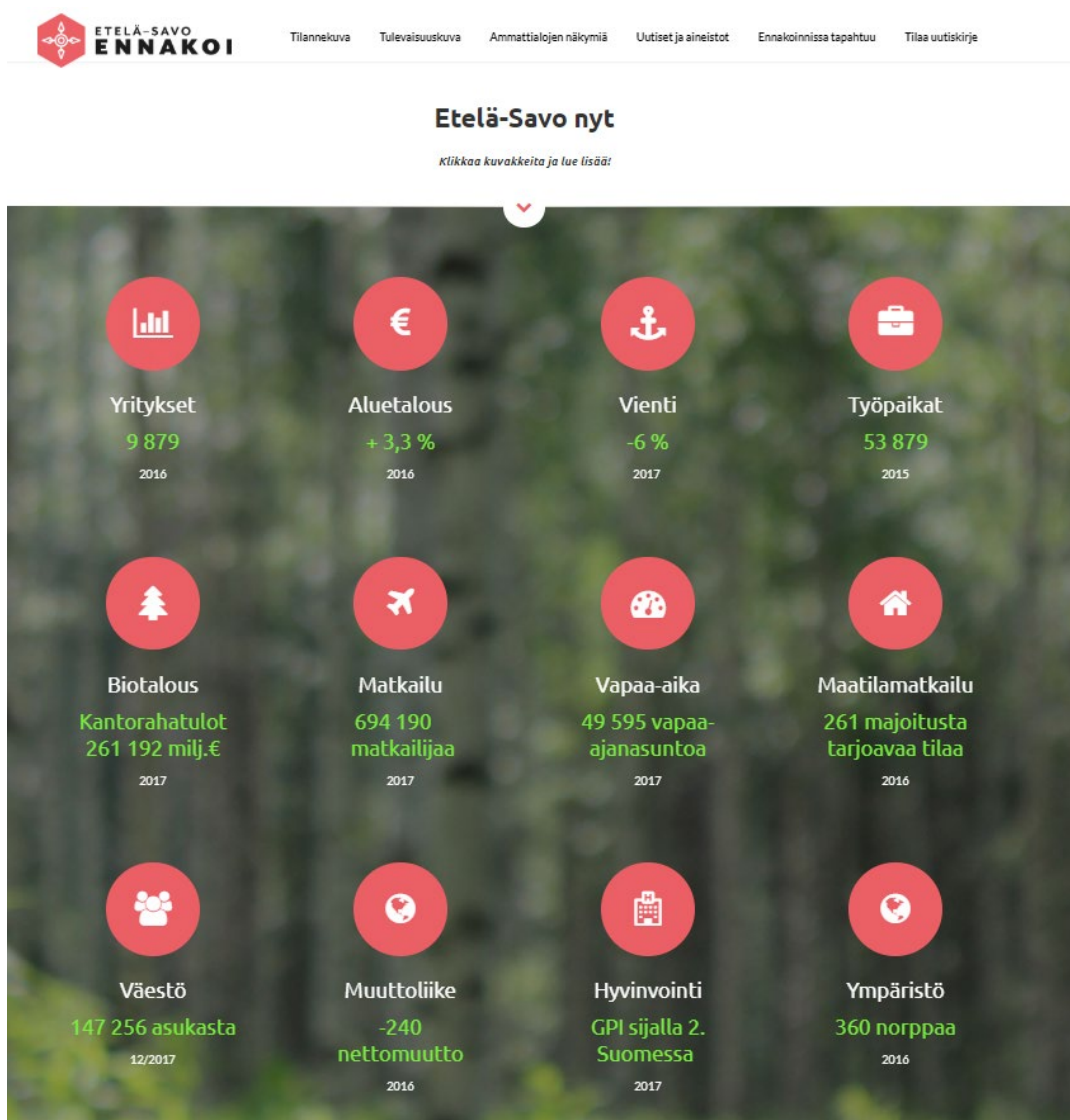
Kuva 16. Vanha esavo.fi etusivu (Ruutukaappaus)

Ensimmäiseksi sivustolle suunniteltiin uusi ilme. Vanha sivusto oli jo hieman ajan kuluttama, ja asiakkaan toiveena oli kokonaan uusi värimaailma. Vanhan sivuston vihreä ja oranssi värimaailma vaihtui marjapuuron punaiseen, ja typografiaa muutettiin selkeämmäksi.



Kuva 17. Uusi Etelä-Savo Ennakoi logo (Ruutukaappaus)

Asiakastapaamisessa maakuntaliiton kanssa sovimme mitkä sisällöt vanhoilta sivuilta jätetään, ja millainen on uuden sivuston sivurakenne. Ensimmäisiä kokonaisuuksia mitä lähdettiin miettimään, oli etusivulta löytyvä tilannekuva. Etusivulle haluttiin yksi selkeä näkymä, missä on Etelä-Savon tämän hetkisiä indikaattoreita eri osa-alueilta. Kuvakkeita klikkaamalla aukeaa kunkin osion oma alasivu.



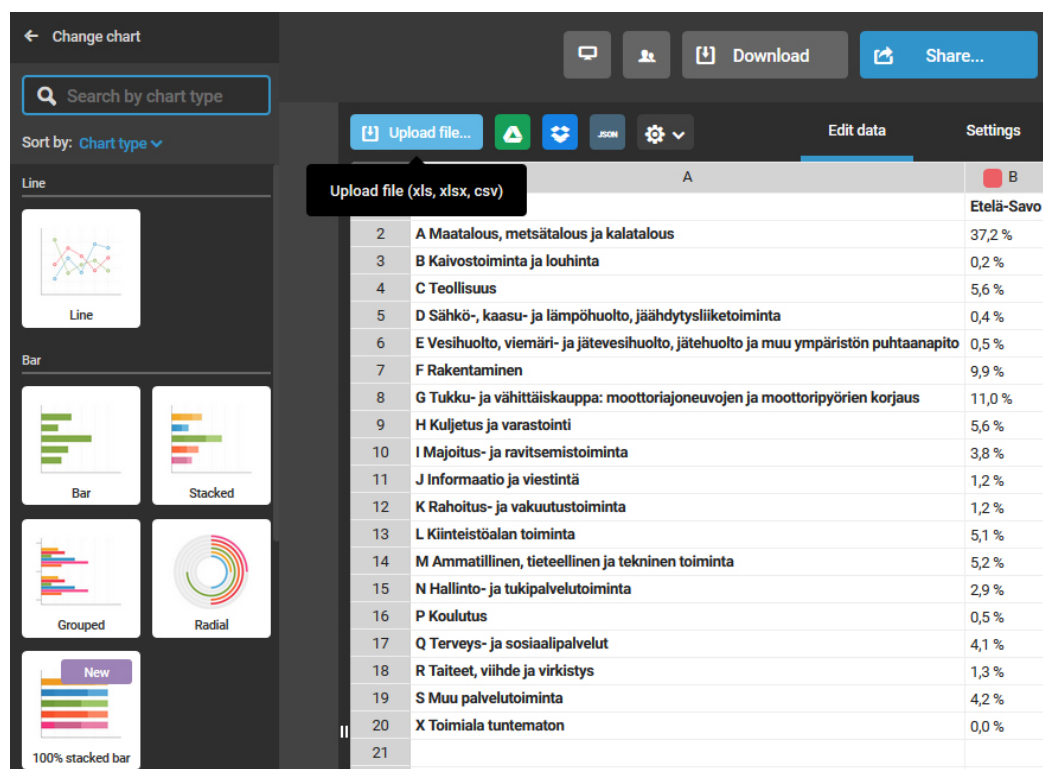
Kuva 18. Esavoennakoi.fi etusivun tilannekuva -osio (Ruutukaappaus)

Tämän jälkeen täytyi tehdä päätös mitä datavisualisointiohjelmaa verkkosivustolla lähdetäisiin käyttämään tilannekuvan alasuivilla. Heti alussa oli selvää, ettei tähän tarkoitukseen kannata valita mitään liian monimutkaista BI-työkalua tai ohjelmointitaitoja vaativaa JavaScript -kirjastoa.

Maakuntaliiton ennakointiasiantuntijoilla oli jo kokemusta muutamista datavisualisointityökaluista, ja heidän toiveena oli löytää ohjelma, joka olisi helppo ottaa käyttöön, ja jonka ominaisuudet eivät loppuisi ihan heti kesken.

Työkalulla täytyi pystyä tekemään tilastokuvioita erilaisista datalähteistä, siitä piti löytyä karttaominaisuus sekä mahdollisuus tehdä interaktiivista infografiikkaa. Valmiit visualisoinnit piti myös pystyä upottamaan helposti verkkosivulle, ja jakamaan niin somessa, kuin tallentaa tiedostoinakin.

Vertailin markkinoilla olevia visualisointiohjelmaa, ja tämän projektin tarkoituksiin ei ollut montaa muuta työkalua, joka sisälsi kaikki toivotut ominaisuudet. **Infogram** nousi esille monissa eri arvosteluissa, ja otimmekin tästä kokeiluversion käyttöön. Lähimmäksi samoja ominaisuuksia ylsivät **Plotly** ja **Datawrapper**. Plotlyn käyttöliittymä olisi ollut liian monimutkainen opetella lyhyessä ajassa, ja Datawrapperin ominaisuudet eivät aivan ylittäneet Infogramin tasolle.

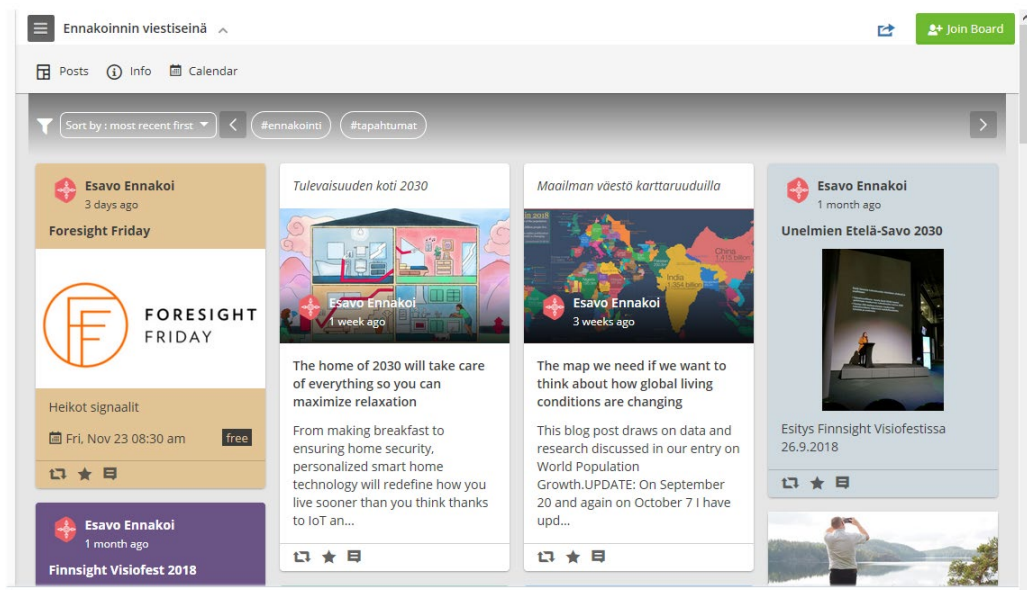


Kuva 19. Infogramin käyttöliittymä (Ruutukaappaus)

Sain maakuntaliitolta muutamia esimerkkitaulukkoja dataa, minkä pohjalta tein erilaisia kokeiluversioita tilastokuvioista, ja opettelin samalla ohjelman käyttöä. Datan syöttämisen jälkeen ohjelma osasi muodostaa valmiita tilastokuvioita datasta. Dataa piti hieman muokata käsin riippuen mitä kuviotyyppejä aikoi käyttää.

Testiversio annettiin käyttöön maakuntaliiton ennakointiasiantuntijoille, ja tämän jälkeen tehtiin päätös ottaa käyttöön ohjelman Business -versio, jonka mukana tuli mm. karttaominaisuudet, linkitykset kuvioihin ja oman logon lisäys mahdollisuus. Tilastokuvioille valittiin kaikille sama teema, ja typografia, ja niihin lisättiin uusi Etelä-Savo Ennakoi logo.

Tämän jälkeen pidin maakuntaliiton ennakointityöntekijöille koulutuksen ohjelman käytöstä, ja siitä miten valmiit visualisoinnit saadaan upotettua uusille verkkosivuille.



Kuva 20. Ennakointiseinä (Ruutukaappaus, Etelä-Savo ennakoi)

Yksi hankkeen tavoitteista oli jokin yhteisöllisyyttä lisäävä ominaisuus sivustolle. Sivustolle haluttiin osio, missä käyttäjät voivat halutessaan lisätä omia havaintoja liittyen alueen ennakointiin. Sivustolle lisättiin kommentointimahdollisuus jokaiselle tilannekuvan alasivulle ja etusivulle nostettiin ”Ennakoinnin viestiseinä”. Tämä viestiseinä toteutettiin **Hootboard** -palvelun avulla. Seinällä

käyttäjät pystyvät rekisteröitymisen jälkeen jättämään omia kirjoituksia, tai linkittämään uutisia muilta sivustoilta.

5 TYÖN TULOKSET

Työn tuloksena maakuntaliitto sai uuden ja raikkaan työkalun ennakkointitiedon jakamiseen. Hankkeen tärkeimmät kokonaisuudet kuten tilannekuva onnistuttiin mielestäni toteuttamaan järkevällä tavalla ja toimeksiantajan tavoitteisiin päästiin hyvin.

Maakuntaliiton ennakkointiasiantuntijat ottivat työkalun erinomaisesti haltuun, ja nyt kun sivusto on ollut noin vuoden verran käytössä, on työkalusta löytynyt ominaisuuksia lukuisiin erilaisiin datavisualisointeihin.



Klikkaa megatrendejä ja lue lisää!



Kuva 21. Megatrendit (Ruutukaappaus, Etelä-Savo ennakoi)

Sivustolle on toimeksiannon jälkeen tehty uusia kokonaisuuksia, joita ovat esimerkiksi "Ammattialojen näkymät" ja "Megatrendit". Viimeisen 90 päivän aikana Esavoennakoi.fi sivuston datavisualisoinneilla on ollut yli 7000 näyttökertaa. Maakuntaliitto kertoi myös vertailleensa sivustoa muiden kuntien vastaviiin, ja Etelä-Savo on selvästi parhaiden joukossa.

6 PÄÄTÄNTÖ

Tätä opinnäytetyötä tehdessäni opin valtavasti uusia asioita datan visualisoinnista sekä projektityöskentelystä. Toimeksiantoa lähdettiin kuitenkin toteuttamaan ennen kuin olin kunnolla perehtynyt tilastografiikan hyviin käytäntöihin, ja nyt jälkeenpäin katsottuna olisin muutaman esimerkkivisualisoinnin ehkä tehnyt nyt toisin. Esimerkiksi joissain piirakkakuvioiden oli liikaa dataa yritetty laittaa samaan kuvioon.

Työskentelyn tahti opinnäytetyötä kirjoitettaessa olisi voinut olla paljon nopeampi. Koska olin samalla töissä työn tekemisen ajan, pitkittyi myös kirjoitusprosessi.

Datan visualisointi onnistuu nykyään jokaiselta. Sitä varten ei tarvitse opetella käyttämään monimutkaisia ohjelmia, vaan ennemminkin perustietämystä tilastokuvien hyvistä käytännöistä. Ennen visualisointiohjelman valintaa kannattaa tutustua erilaisiin markkinoilla oleviin työkaluihin ja etsiä itselleen sopivat ominaisuudet sisältävä ohjelma. Työkalua valittaessa kannattaa ottaa huomioon missä muodossa data tullaan syöttämään ohjelmaan, ja kuinka paljon on valmis käyttämään aikaa ohjelman/koodikirjaston opettelemiseen.

LÄHTEET

Bornmann, L. & Mutz, R. 2014. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. WWW-dokumentti. Saatavissa: <https://arxiv.org/ftp/arxiv/papers/1402/1402.4578.pdf> [viitattu 13.9.2018]

Dahlberg, T. 2015. Miten hallitsemme digitaalista tietoa vuonna 2040 jos sitä on 33 miljoonaa kertaa nykyistä enemmän? WWW-dokumentti. Saatavissa: <https://www.sfs.fi/files/8009/Dahlberg-vuosiseminaari-2015.pdf> [viitattu 14.9.2018]

Työväen Akatemia. Tieto ja tiedonhankinnan prosessi. 2009. WWW-dokumentti. Päivitetty: 2.9.2009. Saatavissa: <http://www.akatemia.org/projektit/tiedonhaku/1.htm> [viitattu 15.9.2018]

Tietojohtamisen tutkimuskeskus NOVI. 2014. Tietojohtamisen perusteet. WWW-dokumentti. Saatavissa: <https://www.slideshare.net/Noviresearch/tjop-osa3-tiedontasot> [viitattu 15.9.2018]

Avoin Data. 2016. Sanasto. WWW-dokumentti. Päivitetty 15.12.2016. Saatavissa: <https://www.avoindata.fi/fi/content/sanasto> [viitattu 14.10.2018]

Tufte, T. 1989. The Visual Display of Quantitative Information. WWW-dokumentti. Saatavissa: <https://cyber.rms.moe/books/03%20-%20General%20Science/The%20Visual%20Display%20of%20Quantitative%20Information%20-%20Edward%20Tufte.pdf> [viitattu 17.10.2018]

Rendgen, S., Wiedemann, J., Ciuccarelli, P., Wurman, R. S. & Rogers, S. 2012. Information Graphics. Köln: Taschen.

Sovelto. Avain syvempään asiakasymmärrykseen. WWW-dokumentti. Saatavissa: <https://www.sovelto.fi/ratkaisut/ict-ja-uudet-teknologiat/big-data/> [viitattu 17.10.2018]

Cairo, A. 2013. The Functional Art: An introduction to information graphics and visualization. New Riders. Berkeley.

Mustonen, A. 2001. Mediapsykologia. Porvoo: Ws Bookwell Oy.

Kuusela, V. 2015. Hyvin laadittu tilastokuvio. WWW-dokumentti. Saatavissa: <http://www.uta.fi/tutkijakoulu/metodifestivaali/2015/ohjelma/torstai20082015/viisuaalinen/index/MEFE2015Kuusela20.8.2015moniste.pdf> [viitattu 6.11.2018]

Kuusela, V. 2000. Tilastografiikan perusteet. Helsinki. Edita.

KevinJZ. Examples of Good and Bad Visualization. WWW-dokumentti. Päivitetty 27.1.2016. Saatavissa: <https://kaijiezhou.wordpress.com/2016/01/27/examples-of-good-and-bad-visualization/> [viitattu 8.11.2018]

Chibana, N. Visual Learning Center. Bad Infographics: 11 Mistakes You Never Want to Make. Blogi. Saatavissa: <https://blog.visme.co/bad-infographics/> [viitattu 8.11.2018]

Väisänen, A. 2017. Lisäarvoa datalle visualisoinnin kautta. WWW-dokumentti. Saatavissa: <https://www.almpartners.fi/tag/interaktiivinen-visualisointi/> [viitattu 10.11.2018]

Tolvanen, T. 2018. Tietotekniikan koulutusohjelma. Lappeenrannan teknillinen yliopisto. Interaktiivinen datan visualisointi Power BI -työkalulla. Opinnäytetyö. PDF-dokumentti. Saatavissa: http://lutpub.lut.fi/bitstream/handle/10024/158567/kandidaatintyo_tolvanen_arttu.pdf?sequence=1&isAlloved=y [viitattu 22.11.2018]