

Opinnäytetyö (AMK)

Tieto- ja viestintäteknikka

2022

Anton Leinonen

Avoimen datan hyödyntäminen ja visualisointi



Opinnäytetyö (AMK) | Tiivistelmä

Turun ammattikorkeakoulu

Tieto- ja viestintäteknikka

2022 | 40 sivua

Anton Leinonen

Avoimen datan hyödyntäminen ja visualisointi

Avoim data on julkista tietoa, joka on kenen tahansa käytettävissä mihin tahansa tarkoitukseen. Avoin data on internetissä kokonaisena, ilmaisena ja koneluettavana. Data on avointa, jos sillä on jokin avoimen hyödyntämisen lisenssi, kuten Creative Commons BY 4.0 eli ”nimi mainittava” -lisenssi.

Datan visualisointi on tiedon esittämistä graafisessa muodossa. Datan visualisointi on erinomainen tapa esittää tietoa helposti ymmärrettävästi ei-tekniiselle yleisölle. Kaikesta rakenteellisesta datasta voi tehdä visualisointeja.

Tämän opinnäytetyön tarkoituksena oli selvittää, mitä on avoin data, millaisia visualisointikeinoja on olemassa ja mitkä visualisoinnit sopivat eri datan esittämiseen.

Työssä visualisoitiin Suomen kuuden suurimman kaupungin päihde- ja mielenterveyspalveluiden kustannuksia ja sitä, onko kaupunkien välillä eroja siinä, mistä kunkin kaupungin kustannukset koostuvat. THL:n latauspalvelua hyödyntäen ladattiin kaksi eri datasettiä THL:n avoimen datan tietokannasta. Visualisoinnin tarkoituksena oli tehdä esimerkkejä erilaisista kaavioista käyttäen kahta eri THL:n datasettiä ja vertailla kaavioista selviävää dataa. Esimerkeistä selvisi, millaisia kaavioita kannattaa käyttää visualisoimaan dataa.

Asiasanat:

avoin data, datan visualisointi, datan hyödyntäminen, datan avaaminen

Bachelor's Thesis | Abstract

Turku University of Applied Sciences

Information and Communications Technology

2022 | 40 pages

Anton Leinonen

Utilization and visualization of open data

Open data is public information that is available for anyone to use however they want. Open data is on the internet for free, in its entirety and in machine-readable form. Data is open if it has an open license such as Creative Commons BY 4.0 or "by attribution" license.

Data visualization is presenting data in a graphical form. Visualizing data is an excellent way to present information in an easy to understand way to non-technical people. All structured data can be visualized.

The purpose of this thesis was to explain the concepts of open data and data visualization and to find out which visualizations are suitable for presenting different kind of data.

In this thesis the costs of substance abuse and mental health services in the six largest cities in Finland were visualized and also whether there were differences between the cities in terms of what each of their costs consist of. Using the file download service of the Finnish Institute for Health and Welfare (THL), two different datasets were downloaded from THL's open data database. The purpose of the visualization of THL's data was to make examples of different charts using two different datasets and to compare the results. All visualizations were made using Microsoft Excel.

Keywords:

open data, data visualization, utilization of data, opening up data

Sisältö

Käytetyt lyhenteet	7
1 Johdanto	8
2 Avoin data	9
2.1 Avoin data ja sen hyödyt	9
2.2 Datan avaaminen	11
2.2.1 Tietovarantojen kartoitus	11
2.2.2 Hyödyntäjien tarpeet	11
2.2.3 Rajoitteet	11
2.2.4 Avattava data	12
2.2.5 Avattavan datan jakelu- ja tiedostomuodot	15
2.2.6 Datan viimeistely ja julkaisu	16
2.3 Avoimen datan löytäminen	17
3 Datan visualisointi	18
3.1 Datan visualisointi ja sen tarkoitus	18
3.2 Datan visualisoinnin historia	18
4 Avoimen datan visualisointia	23
4.1 Päihde- ja mielenterveyspalveluiden visualisointi	23
4.2 THL:n tarjoaman avoimen datan visualisointi	25
4.3 Esimerkki THL:n avoimen datan visualisoinnista: Steriloinnit	27
4.3.1 Viivakaavio	27
4.3.2 Pylväskaavio	28
4.3.3 Pinottu pylväskaavio	29
4.3.4 Rengaskaavio	30
4.4 Esimerkki THL:n avoimen datan visualisoinnista: Vastasyntyneet	32
4.4.1 Viivakaavio	32
4.4.2 Pylväskaavio	33
4.4.3 Pinottu pylväskaavio	34
4.4.4 Ympyräkaavio	35

5 Yhteenveto	36
---------------------	-----------

Lähteet	38
----------------	-----------

Kuvat

Kuva 1. Avoimen datan kriteerit. (Korhonen & Sore 2017.)	10
Kuva 2. Tim Berners-Leen viiden tähden malli. (Kim 2015a.)	13
Kuva 3. Tim Berners-Leen viiden tähden malli, avoin tiedostomuoto 3/5 tähteä. (Avoindata.fi 2020f.)	15
Kuva 4. Langrenin vuoden 1644 kaavio etäisyyden määrittämisestä pituusasteina Toledosta Roomaan. (Tuftte 1997.)	19
Kuva 5. William Playfairin vuoden 1821 aikasarjakaavio vehnän hinnoista, palkoista ja hallitsevista monarkeista 255 vuoden ajalta. (Tuftte 1983.)	19
Kuva 6. Minardin vuoden 1869 kartta Napoleonin etenemisestä Moskovaan. (Jacobs 2010.)	21
Kuva 7. THL:n latauspalvelun rajaustoiminnon eri vaihtoehtoja. (THL 2022.)	26

Kuviot

Kuvio 1. Kaupunkien väliset päihde- ja mielenterveyspalvelujen kustannukset vuonna 2021. Pylväskaavio.	24
Kuvio 2. Päihde- ja mielenterveyspalvelujen kustannusten osuus palvelujen kokonaiskustannuksista vuonna 2021. Pinottu pylväskaavio.	25
Kuvio 3. Sukupuolten väliset steriloinnit Suomessa. Viivakaavio.	28
Kuvio 4. Sukupuolten väliset steriloinnit Suomessa. Pylväskaavio.	29
Kuvio 5. Sukupuolten väliset steriloinnit Suomessa. Pinottu pylväskaavio.	30
Kuvio 6. Sukupuolten väliset steriloinnit Suomessa, vuodet 1987–2021. Rengaskaavio.	31
Kuvio 7. Sukupuolten väliset steriloinnit Suomessa, vuodet 2014–2021. Rengaskaavio.	32
Kuvio 8. Vastasyntyneet Suomessa. Viivakaavio.	33

Kuvio 9. Vastasyntyneet Suomessa. Pylväskaavio.	34
Kuvio 10. Vastasyntyneet Suomessa. Pinottu pylväskaavio.	34
Kuvio 11. Vastasyntyneet Suomessa. Ympyräkaavio.	35

Taulukot

Taulukko 1. Esimerkki Tim Berners-Leen neljän tähden taulukosta. (Kim 2015b.)	14
Taulukko 2. Esimerkki Tim Berners-Leen viiden tähden taulukosta. (Kim 2015c.)	14
Taulukko 3. Rajaamaton taulukko naisten ja miesten steriloinnista. (THL 2022.)	26
Taulukko 4. Rajattu taulukko naisten ja miesten steriloinnista. (THL 2022.)	27

Käytetyt lyhenteet

API	Ohjelmointirajapinta (engl. application programming interface).
CSV	Pilkulla erotellut arvot (engl. comma-separated values) on tiedostomuoto, jolla tallennetaan taulukkomuotoista, pilkulla toisistaan erotettua tietoa tekstitiedostoon.
PDF	Portable Document Format. Adoben ohjelmistoriippumaton tiedostomuoto.
THL	Terveyden ja hyvinvoinnin laitos.
TXT	Tekstin tallentamiseen käytetty yksinkertainen tiedostomuoto.
URI	Uniform Resource Identifier on internetissä olevan resurssin yksilöivä tunnus.
XML	Laajennettava merkintäkieli (engl. Extensible Markup Language) on merkintäkieli, joka määrittää tietojen merkintämuodon loogisella rakenteella.

1 Johdanto

Tämän opinnäytetyön tarkoituksena on perehtyä avoimeen dataan ja sen visualisointiin. Työssä käydään läpi, miten dataa avataan ja mikä tekee avattavasta datasta arvokasta. Avoin data ja datan visualisointi on kasvanut viime vuosina räjähdysmäisesti. Lähes jokainen yritys, organisaatio ja jopa valtio on avannut tai haluaisi avata dataa. Avoimesta datasta on hyötyä niin sen käyttäjälle kuin sen avaajallekin.

Opinnäytetyössä tarkastellaan, mitä on avoin data ja mitä hyötyä siitä on sekä minkälaista avointa dataa on saatavilla ja mistä sitä saa. Lisäksi työssä tarkastellaan, mitä tarkoitetaan avoimella datalla ja datan visualisoinnilla. Työssä käydään läpi myös datan visualisoinnin historiaa.

Opinnäytetyössä tarkastellaan Helsingin kaupunginkanslian ylläpitämää päihde- ja mielenterveyspalvelujen tiedonkeruupohjaa, josta löytyy dataa Suomen kuuden suurimman kaupungin päihde- ja mielenterveyspalveluista. Lisäksi tutustutaan myös THL:n tarjoamaan avoimeen dataan. Näistä tehdään visualisointeja ja käydään läpi visualisointien tuloksia.

2 Avoin data

2.1 Avoin data ja sen hyödyt

Avoin data on julkista tietoa, jota kuka vaan voi käyttää, miten haluaa. Avoin data on kaikille saatavilla internetistä kokonaisena, ilmaisena ja koneluettavana. Avoin data voi olla peräisin julkishallinnolta, organisaatioilta tai yrityksiltä, ja sitä voivat käyttää sekä kansalaiset että yritykset tasavertaisesti julkishallinnon kanssa. (Helsinki Region Infoshare 2017a.)

Esimerkiksi kartat, videot, dokumentit ja äänitiedostot ovat dataa. Avointa dataa ne ovat vain, jos niillä on avoimen hyödyntämisen lisenssi, kuten Creative Commons BY 4.0 eli ”nimi mainittava” -lisenssi. Kyseisen lisenssin alle kuuluvaa dataa saa siis vapaasti käyttää, kunhan mainitsee lähteen asianmukaisesti. (Creativecommons.fi 2022a.)

Avoin data ei saa loukata yksityisyyden suojaa eikä vaarantaa yleistä turvallisuutta. Data ei saa siis sisältää henkilötunnuksia tai liikesalaisuuksia. Jos data täyttää kaikki kuvan 1 kriteerit, on kyseessä avoin data.



Kuva 1. Avoin datan kriteerit. (Korhonen & Sore 2017.)

Avoin data edistää uusien palveluiden syntymistä. Esimerkiksi linja-autoliikenteen aikataulut ja yhteyksien hakeminen sekä henkilöauton pysäköinnin maksaminen suoraan kännykkäsovelluksella on täysin avoimen datan ansiota. (Helsinki Region Infoshare 2017b.)

Avointa dataa voivat hyödyntää ketkä tahansa, esimerkiksi aktiiviset kansalaiset, joita kiinnostaa julkisen päätöksenteon seuraaminen tai hallituksen julkinen valvonta. Tämä auttaa myös vähentämään korruptiota lisäämällä hallituksen avoimuutta ja edistää näin demokratiaa.

2.2 Datan avaaminen

Datan avaaminen on pitkä prosessi, mutta siitä on hyötyä kaikille. Yksilöt, yhteisöt, yritykset ja koko yhteiskunta hyötymään avoimesta datasta. Datan avannut yritys hyötyy myös omasta avoimesta datastaan. Jos samasta datasta tehdään jatkuvasti tietopyyntöjä, yrityksen kannattaa avata data ja julkaista se avoimena datana. Näin yritys säästää omaa ja muiden aikaa sekä helpottaa kyselijöiden työtä. (Helsinki Region Infoshare 2017c.)

2.2.1 Tietovarantojen kartoitus

Datan avaaminen alkaa omien tietovarantojen kartoittamisella. Dataa ei voi avata, jos ei tiedä, mitä dataa omalla organisaatiolla on. Täytyy myös selvittää, mitä tietojärjestelmiä organisaatiolla on ja mitä dataa on missäkin tietojärjestelmässä. Onko kaikki data, joka halutaan avata, jo koneluettavassa muodossa, vai pitääkö asialle tehdä jotain? Viimeisenä asiana tulisi tarkistaa datan oikeellisuus ja ajantasaisuus. (Avoindata.fi 2020a.)

2.2.2 Hyödyntäjien tarpeet

Ennen datan avaamista täytyy kartoittaa mahdollisen avoimen datan hyödyntäjät. Ei ole mitään syytä avata dataa, jos sille ei ole käyttöä. Kohdeyleisön löydyttyä datan avaamista kannattaa pilotoida eli avata ensin rajatulle hyödyntäjäkunnalle vain pieni osa datasta. (Avoindata.fi 2020b.)

2.2.3 Rajoitteet

Kaksi kenties tärkeintä datan avaamista ohjaavaa lakia ovat viranomaisen toiminnan julkisuutta ohjaava julkisuuslaki (621/1999) ja henkilötietojen tietosuojaa käsittelevä tietosuojalaki (1050/2018). Muita tärkeitä huomioon otettavia seikkoja ovat liikesalaisuuksien paljastuminen, kolmannen osapuolen tekijänoikeudet ja kansallinen turvallisuus. (Avoindata.fi 2020c.)

Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/1024

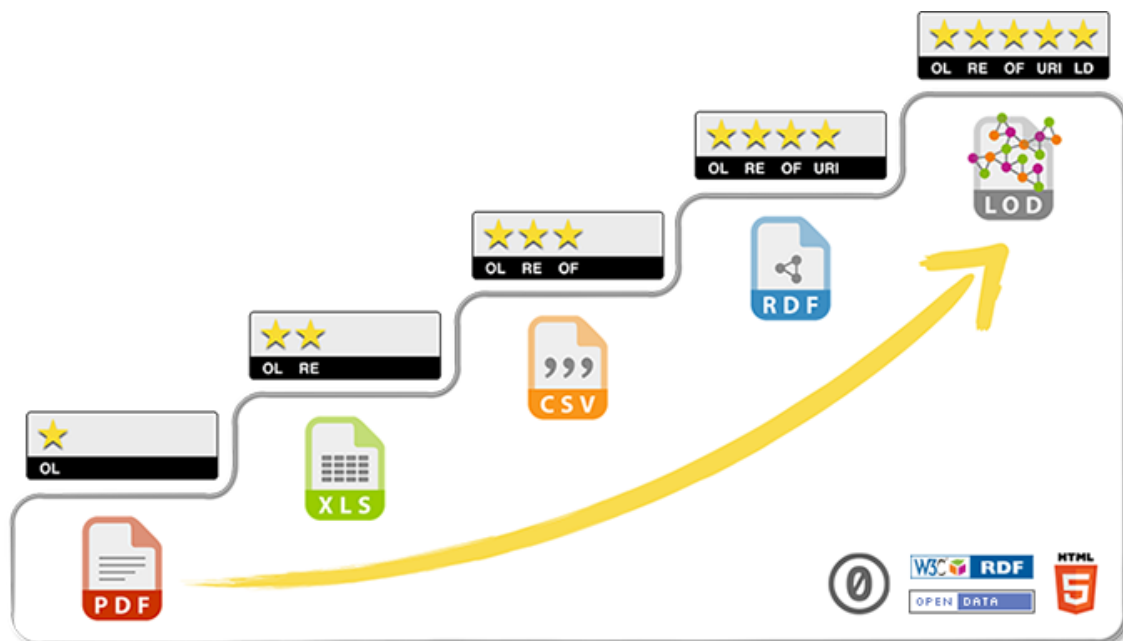
Avointa dataa koskeva direktiivi perustuu yleiseen periaatteeseen, jonka mukaan julkisia ja julkisesti rahoitettuja tietoja tulisi voida käyttää uudelleen mihin tahansa tarkoitukseen. Direktiivillä edistetään avoimen datan käyttöä. Julkisen sektorin, julkisten yritysten ja liikennealan toimijoiden on avattava julkinen data EU:n kansalaisten ja yritysten saataville. Poikkeuksina direktiiviä ei sovelleta asiakirjoihin, jotka eivät toteuta avoimen datan kriteerejä. Direktiivillä ei ole vaikutusta tietosuojalakiin. (EUR-Lex 2019.)

2.2.4 Avattava data

Avattavan datan määrästä riippumatta avattavaa dataa on priorisoitava jollakin tavalla, sillä kaiken avaaminen kerralla voi tulla hyvin kalliiksi, ja datan määrästä riippuen avaaminen voi olla jopa mahdotonta. Avattavalle datalle voidaan määrittää arvo. Datan arvoa määritettäessä sitä voidaan arvioida kolmesta eri näkökulmasta: datan uudelleenkäytettävyyden, datan arvo omistajan ja datan arvo hyödyntäjän näkökulmasta. Hyödyntäjälle datan arvo määräytyy usein uudelleenkäytettävyyden perusteella. (Avoindata.fi 2020d.)

Datan uudelleenkäytettävyys

Uudelleen käytettävä data on arvokasta, sillä sitä voi käyttää eri datasettien kanssa. Avoimen datan uudelleenkäytettävyyden arvioimisessa hyvä malli on Tim Berners-Leen 5 tähden malli (Kuva 2.).



Kuva 2. Tim Berners-Leen viiden tähden malli. (Kim 2015a.)

Laadukkaan ja arvokkaan avoimen datan tulisi saada vähintään kolme tähteä kuvassa 2 esitetyn mallin mukaisesti. Data saavuttaa ensimmäisen tähden, kun se on saatavilla verkossa, avoimella lisenssillä, missä tahansa muodossa, esimerkiksi PDF:nä. Toisen tähden data saa, kun se on, ensimmäisen tähden kriteerien lisäksi, rakenteellisessa muodossa, esimerkiksi Excel-taulukkona. Kolmannen tähden saa, kun data on rakenteellisessa, ei-kaupallisessa muodossa, kuten esimerkiksi CSV-taulukkona Excel-taulukon sijaan. (5stardata.info 2015a.)

Neljänten tähden ei riitä, että data on ladattavissa CSV-taulukkona, vaan sen täytyy olla myös verkossa esitettynä ja datan tulee sisältää yksilöllisiä URI-tunnisteita. Taulukossa 1 on esitetty XML-pohjainen, verkkosivulle upotettu taulukko, jonka jokaisella säätiedotteella on oma URI ja säätiedotteista voidaan luoda minkä tahansa kokoinen taulukko. (5stardata.info 2015b.)

Taulukko 1. Esimerkki Tim Berners-Leen neljän tähden taulukosta. (Kim 2015b.)

TEMPERATURE FORECAST FOR GALWAY, IRELAND	
Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

Viidennen tähden saamiseksi datan täytyy sisältää linkitys ulkopuoliseen dataan kontekstin tarjoamiseksi. Taulukossa 2 on esitetty muuten sama taulukko kuin taulukossa 1, mutta sana "Temperature" linkittää nyt Wikipedian lämpötila-artikkeliin. Taulukon 2 data muuttuu viiden tähden arvoiseksi avoimeksi dataksi. (5stardata.info 2015c.)

Taulukko 2. Esimerkki Tim Berners-Leen viiden tähden taulukosta. (Kim 2015c.)

TEMPERATURE FORECAST FOR GALWAY, IRELAND	
Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

Datan arvo omistajan näkökulmasta

Datan arvokkuutta tulee arvioida myös omistajan näkökulmasta. Omistajalle data on arvokasta, kun se täyttää yhden tai useamman seuraavista kriteereistä:

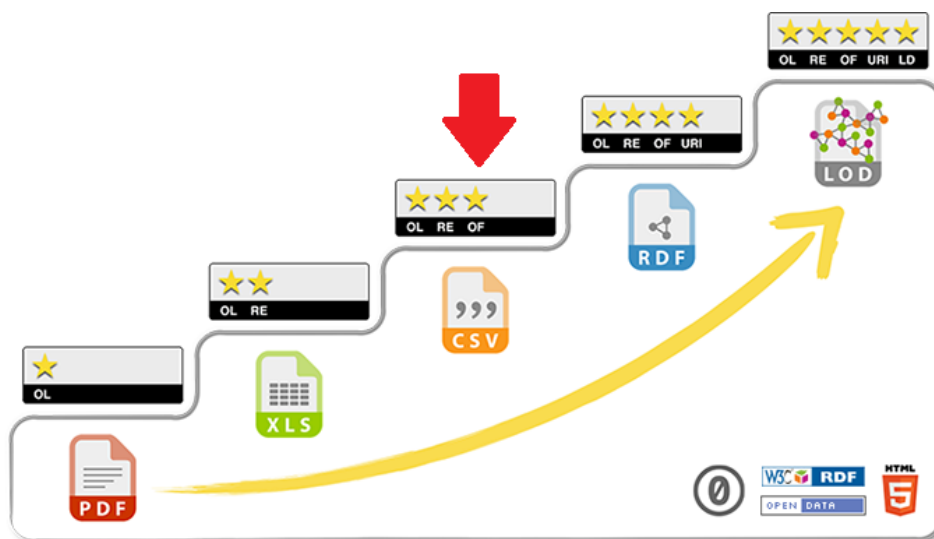
- Datan avaaminen lisää läpinäkyvyyttä.
- Datan julkaisemiseen on laillinen velvoite.
- Data liittyy omistajan julkiseen tehtävään.
- Datan avaaminen vähentää kuluja.

Kuten datan hyödyntäjälle, datan omistajalle datan todellinen arvo perustuu suurimmaksi osaksi sen uudelleenkäytettävyyteen. Jos dataa voidaan hyödyntää useaan eri käyttökohteeseen ja sillä on suuri kohdeyleisyys, data on arvokasta. (Avoindata.fi 2020d.)

2.2.5 Avattavan datan jakelu- ja tiedostomuodot

Avattavan datan valitsemisen jälkeen sille täytyy löytää sopiva jakelumuoto. Erityyppiselle datalle sopivat erityyppiset jakelumuodot. Tiedosto on yksi jakelumuodoista. Tiedosto sopii hyvin datalle, joka on pienikokoista tai ei muutu usein. (Avoindata.fi 2020e.)

Avoimen datan julkaisuun tulee käyttää avoimia tiedostomuotoja. Tekijänoikeudet, patentit, tavaramerkit ja muut rajoitukset eivät estä avoimien tiedostomuotojen käyttöä. Tim Berners-Leen viiden tähden mallin mukaisesti avoimessa tiedostomuodossa julkaistu data saa vähintään 3/5 tähteä (Kuva 3.). Avoin tiedostomuoto on kenen tahansa käytettävissä oleva, ei-kaupallinen tiedostomuoto, kuten esimerkiksi TXT ja CSV. (Avoindata.fi 2020f.)



Kuva 3. Tim Berners-Leen viiden tähden malli, avoin tiedostomuoto 3/5 tähteä. (Avoindata.fi 2020f.)

API on toinen mahdollinen datan jakelumuoto. Rajapinnan kautta on mahdollista siirtää dataa sovellusten ja käyttäjien tai sovellusten ja laitteiden välillä. Jos avattava data on suurikokoista tai usein muuttuvaa, rajapinta on sille paras vaihtoehto. API:n kautta käyttäjä saa pyydettyä juuri sen datan, mitä hän tarvitsee, eikä kaikkea mahdollista dataa. Usein voi olla myös järkevää julkaista sama data esimerkiksi tiedostona ja rajapintana. (Avoindata.fi 2020e.)

Viimeinen jakelumuoto on latauspalvelu. Latauspalvelun kautta käyttäjä saa valittua haluamansa datan ja lataa sen sitten tiedostona. Latauspalvelu sopii datan jakelumuodoksi parhaiten, jos data on suurikokoista eikä se muutu usein. (Avoindata.fi 2020e.)

2.2.6 Datan viimeistely ja julkaisu

Ennen datan julkaisua on hyvä kirjoittaa datalle metatiedot. Metatiedot ovat kuvailutietoja, hyödyllistä lisäinformaatiota, tietoaineistosta. Metatietoja ovat muun muassa tietoaineiston otsikko, avainsanat, julkaisupäivä ja kieli. Aineiston kielen määrittäminen edistää datan uudelleenkäytettävyyttä ja aineiston mahdollista kansainvälistä hyödyntämistä. Lisäksi olisi hyvä kuvailla sisältöä ja käyttötarkoitusta helposti ymmärrettävässä muodossa. Metatiedot tulisi olla sekä ihmiselle luettavassa että koneluettavassa muodossa. (Avoindata.fi 2022a.)

Avattu data kannattaa julkaista jollakin jo valmiiksi tunnetulla julkaisualustalla. Oman alustan luominen toisi huomattavia lisäkuluja ja vähentäisi datan löydettävyyttä. Avoindata.fi on suomalainen kansallinen palvelu, jonka tehtävänä on koota kaikki suomalainen avoin data yhteen paikkaan. Avoindata.fi-palvelun data löytyy Euroopan kansainvälisestä avoimen datan portaalista. Sivusto on myös menestynyt erilaisissa kansainvälisissä vertailuissa hyvin, joten Avoindata.fi on erinomainen vaihtoehto avoimen datan julkaisulle. (Avoindata.fi 2020g.)

2.3 Avoimen datan löytäminen

Avoimen datan löytäminen on nykypäivänä kohtalaisen helppoa. Helpoin ja nopein tapa löytää itseä kiinnostavaa avointa dataa on vieraila jonkin dataportaalin verkkosivuilla. Dataportaalit ovat kuin kirjastoja avoimelle datalle. Ne keräävät avointa dataa eri lähteistä, organisoivat datan kategorioihin ja tekevät kiinnostavan datan löytämisestä helppoa.

Dataportaaleja on hyvin monella tasolla. Suomen sisällä eri kunnilla ja kaupungeilla on omia dataportaaleja, jotka koskevat vain kyseisen alueen avointa dataa. Esimerkkeinä pääkaupunkiseudun dataportaali Helsinki Region Infoshare ja Tampereen kaupungin dataportaali. Myös eduskunnalla on oma dataportaali, Eduskunta Avoin Data.

Lounais-Suomen dataportaali kerää luonnollisesti avointa dataa Lounais-Suomesta. Maiden tasolle siirryttäessä Suomella on oma dataportaalinensa, Avoindata.fi. Lähes jokaisella maalla on oma avoimen datan portaali, esimerkiksi Saksalla Das Datenportal für Deutschland, Tšekillä Portál otevřených dat České republiky. Myös Euroopan mailla on yhteinen dataportaali, data.europa.eu.

Dataportaalit eivät tietenkään ole ainoita paikkoja, joista avointa dataa voi löytää. Organisaatioilla on usein myös avoimia datalähteitä omilla verkkosivuillaan. Usein nämä samat datalähteet löytyvät myös yleisistä dataportaaleista.

3 Datan visualisointi

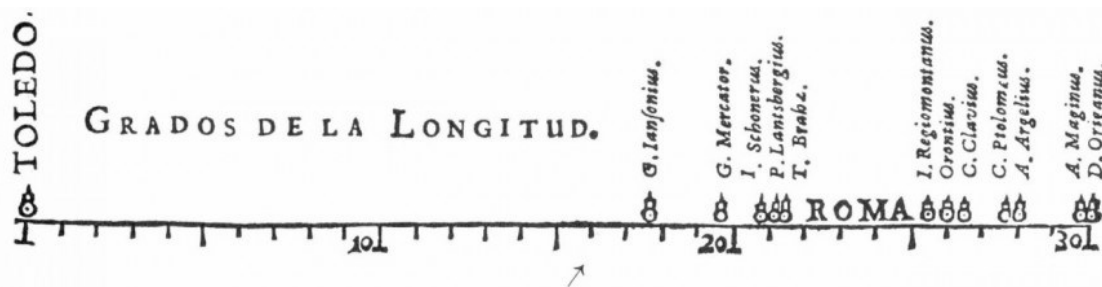
3.1 Datan visualisointi ja sen tarkoitus

Datan visualisointi on tiedon esittämistä graafisessa muodossa. Käyttämällä esimerkiksi karttoja ja kaavioita ihmisen on huomattavasti helpompi ymmärtää datassa esiintyvää informaatiota (Brush & Burns 2020). Lisäksi datan visualisointi on erinomainen tapa esittää tietoa helposti ymmärrettävästi ei-tekniiselle yleisölle. Kaikesta rakenteellisesta datasta, kuten esimerkiksi CSV-muotoisista tiedostoista, voi tehdä visualisointeja.

Suuren laskentataulukon informaation sisäistäminen on hyvin hankalaa. Visualisoinnin tarkoituksena on yksinkertaistaa monimutkaista tai vaikeasti hahmotettavaa dataa helposti ymmärrettävään visuaaliseen muotoon. Näin on helpompi sisäistää datassa esiintyviä trendejä ja poikkeavuuksia. (Tableau 2022.)

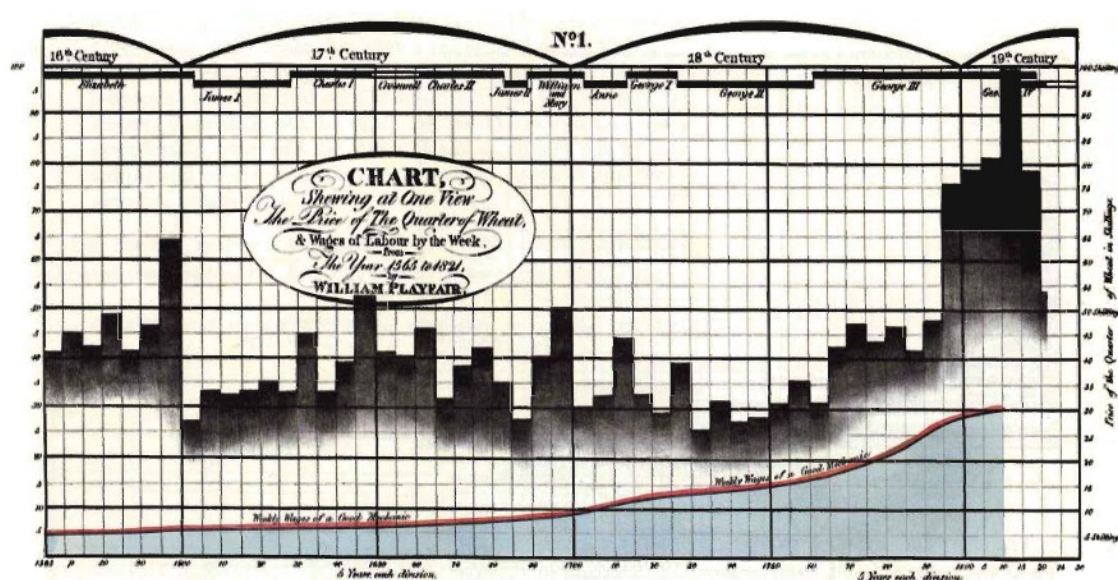
3.2 Datan visualisoinnin historia

Dataa on visualisoitu ympäri maailmaa jo satoja vuosia. Ennen 1600-lukua visualisoinnit olivat lähinnä vain karttoja tai kartan tyyppisiä esityksiä, kuten maamerkkejä, kaupunkeja tai teitä. Vuonna 1644 flaamilaisen tähtitieteilijä Michael Florent Van Langrenin uskotaan tehneen ensimmäinen tilastotieteelliseen dataan pohjautuva visualisointi. Hän loi kuvassa 4 näkyvän yksiulotteisen kaavion, josta on nähtävissä siihen aikaan ainoat 12 arviota pituusasteiden erosta Toledon ja Rooman välillä ja arvioiden tehneiden tähtitieteilijöiden nimet. (Friendly 2006.)



Kuva 4. Langrenin vuoden 1644 kaavio etäisyyden määrittämisestä pituusasteina Toledosta Roomaan. (Tuftte 1997.)

1600-luvun loppuun mennessä visualisointi oli selkeästi kiinnostuksen kohteena. Saatavilla oli merkittävää, mielenkiintoista dataa, jonninkäköisiä teorioita niiden ymmärtämiseksi ja jopa joitain ideoita niiden visualisoimiseksi. William Playfair (1759–1823) keksi ensin viiva- ja pylväsdiagrammit vuonna 1786 ja myöhemmin vuonna 1801 ympyrädiagrammin. Playfair julkaisi vuonna 1821 kaavion vehnän hinnoista, viikoittaisista palkoista ja hallitsevista monarkeista 255 vuoden ajalta. Playfair käytti samassa kaaviossa kolmea rinnakkaista diagrammia ajan kuvaamiseksi. Playfair käytti kuvassa 5 näkyvää kaaviota osoittamaan, että työläisillä oli paremmat oltavat kaavion viimeisimpinä vuosina. (Insightsoftware 2019.)

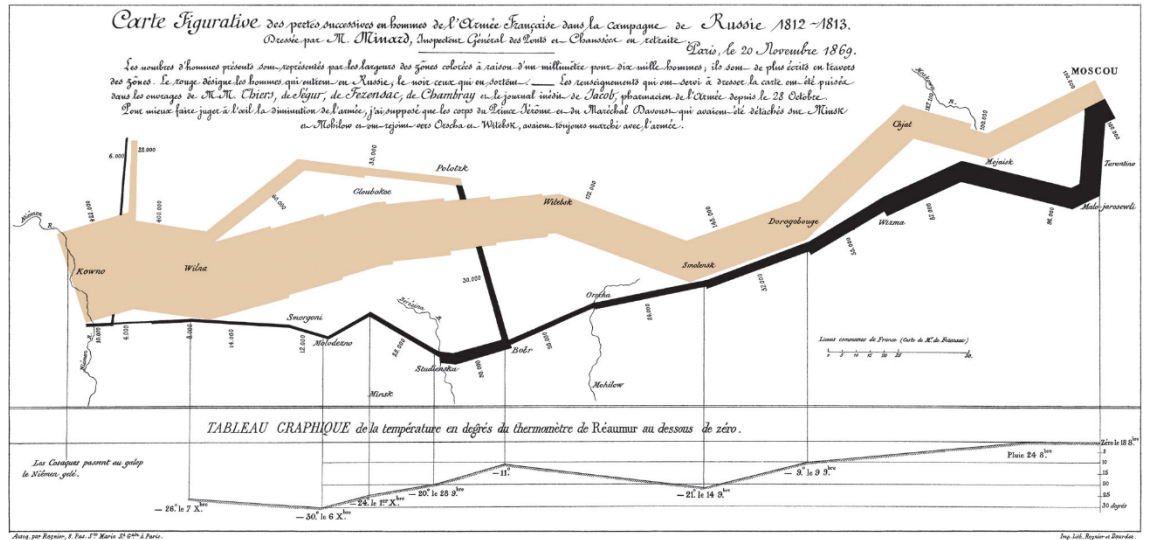


Kuva 5. William Playfairin vuoden 1821 aikasarjakaavio vehnän hinnoista, palkoista ja hallitsevista monarkeista 255 vuoden ajalta. (Tuftte 1983.)

1800-luvun loppupuolella visualisointi oli nopeassa kasvussa ja valtiot ympäri Euroopan perustivat virallisia tilastotoimistoja. Monimutkaisten tietojen ymmärtämisen kasvamisen myötä monia uusia graafisia muotoja keksittiin ja niitä laajennettiin uusille tutkimusalueille. Charles Joseph Minardin vuonna 1869 laatima kaavio Napoleonin armeijan etenemisestä Moskovaan on tilastotieteilijä Edward Tuften mukaan paras koskaan piirretty tilastollinen visualisointi. Kuvassa 6 esitetty Minardin kartta yhdistää kuusi eri datasettiä.

- Karttaan on nimetty joet ja kaupungit, kuten tavallisissakin kartoissa.
- Napoleonin armeijan kulku Moskovaan ja takaisin on merkitty reittinä.
- Armeijan kulun suunta on osoitettu väreillä: hiekan värinen viiva Moskovaan ja musta viiva takaisin.
- Sotilaiden määrä on ilmoitettu tekstinä ja viivan kapenemisena matkan edetessä. Kartalla jokainen millimetri vastaa 10 000 miestä.
- Lämpötila on osoitettu kartan alareunassa Réaumur-asteikolla. Tällä asteikolla vesi jäätyy 0 °Ré ja kiehuu 80 °Ré.
- Aika on osoitettu lämpötilan ohella kartan alareunassa päivämäärinä oikealta vasemmalle kuvaamaan Napoleonin armeijan poistumista Venäjältä.

Armeijan suuret menetykset on kuvattu viivojen suuruudella. Moskovaan lähti 442 000 miestä ja heistä kotiin palasi 10 000. Minardia pidetään yhtenä modernin infografiikan ensimmäisinä tekijöinä. (Jacobs 2010.)



Kuva 6. Minardin vuoden 1869 kartta Napoleonin etenemisestä Moskovaan. (Jacobs 2010.)

1900-luvun alku oli datan visualisaation kehittämisen kannalta synkkää aikaa. Uusia innovaatioita ei juurikaan ollut, ja 1940-lukuun mennessä 1800-luvun visualisointivimma oli loppunut. Tänä aikana kuitenkin tilastolliset grafiikat yleistyivät, ja niitä alkoi löytyä oppikirjoista, opetussuunnitelmista ja hallitusten tavallisesta käytöstä.

1960-luvulta alkaen kiinnostus datan visualisointiin lähti jälleen kasvuun. Vuonna 1962 Yhdysvalloissa John W. Tukey julkaisi pyynnön data-analyysin tunnustamiseksi omaksi tilastotieteen alaksi eikä vain matemaattisen tilastotieteen haaraksi. Vuonna 1957 John Backusin luoma FORTRAN, ensimmäinen korkean tason ohjelmointikieli, mahdollisti tietokoneiden tilastollisen datan prosessoinnin. Myöhemmin 1960-luvun lopulla joidenkin yliopistojen tietokoneet pystyivät luomaan graafisia visualisointeja. Kesti kuitenkin vielä kauan, ennen kuin tämä oli mahdollista kenelle tahansa. (Friendly 2006.)

1900-luvun lopulla, tietokoneiden yleistyttyä ja datan graafiseen visualisointiin käytettävien sovellusten kehittyttyä, datan visualisointi kasvoi räjähdysmäisesti. Nykypäivänä on olemassa dashboardeja, tiedonhakuun suunnattuja työkaluja, analytiikkaohjelmistoja ja lukuisia muita ohjelmistotyökaluja, jotka antavat

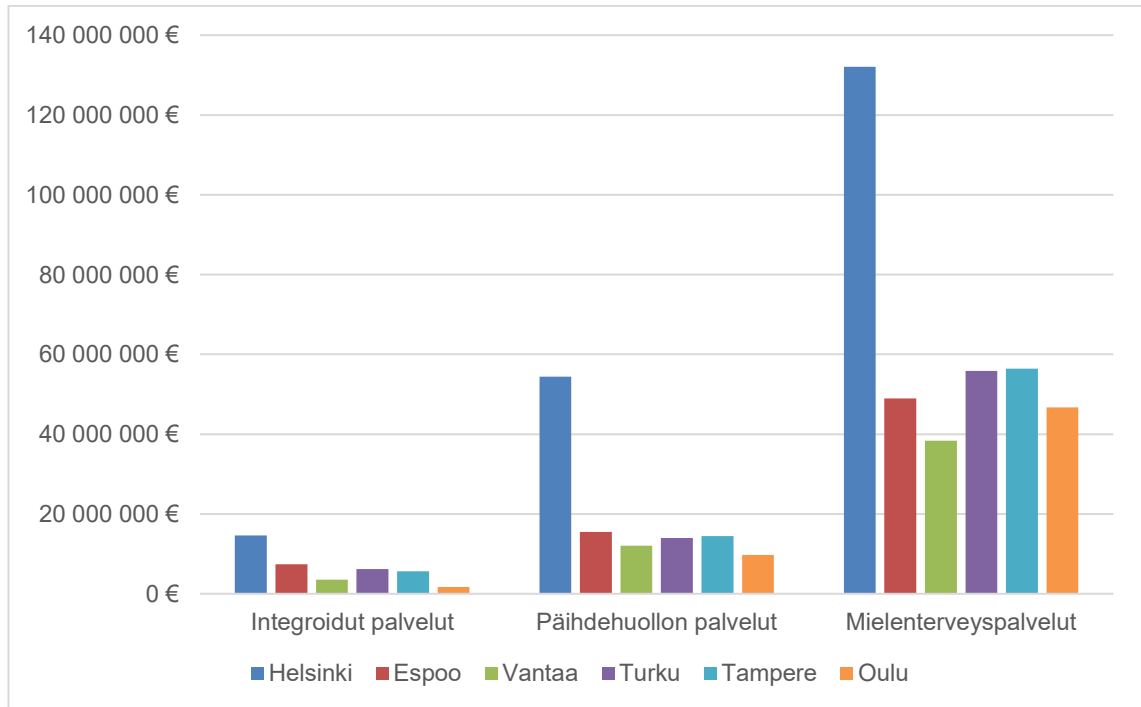
jokaiselle mahdollisuuden tutkia, käsitellä ja visualisoida dataa uusilla ja entistä paremmilla tavoilla. Varjopuolena kaikille näille uusille tavoille kerätä, analysoida ja visualisoida dataa on myös yhteiskunnallisia haasteita. Yksityisyyden loukkaus ja tietojen väärinkäyttö, tahallista tai ei, ovat yksiä suurimmista haasteista avoimen datan visualisoinnissa.

4 Avoimen datan visualisointia

4.1 Päihde- ja mielenterveyspalveluiden visualisointi

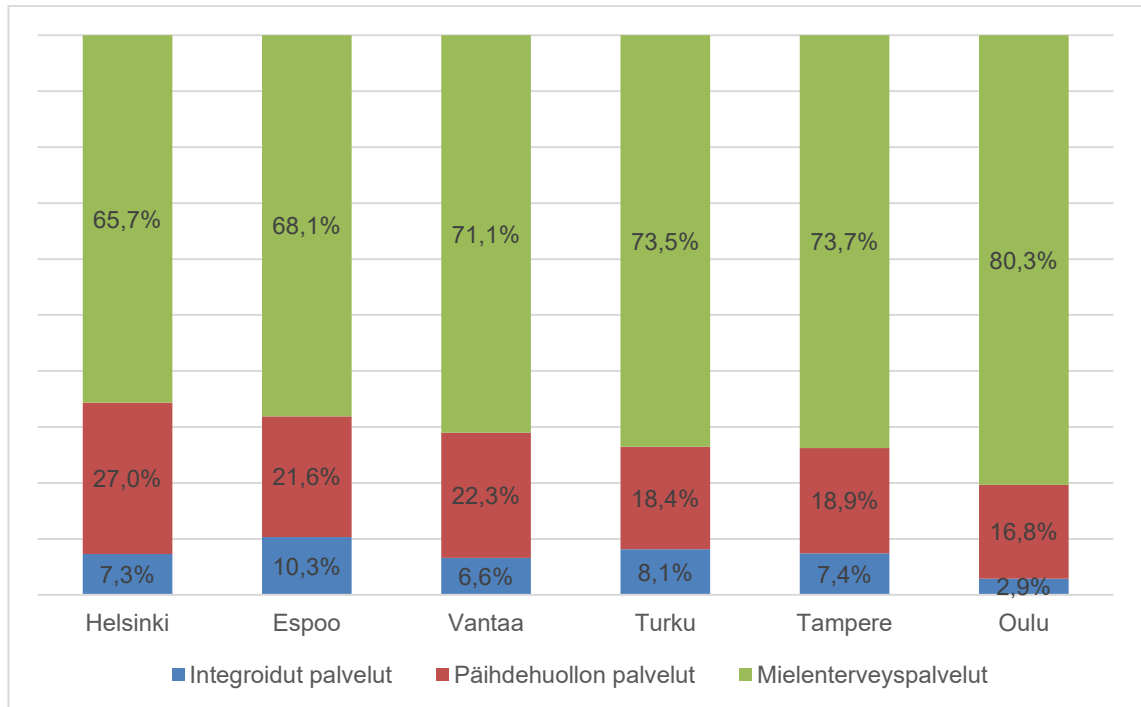
Tässä visualisoinnissa tarkastellaan Helsingin kaupunginkanslian tarjoamaa ja ylläpitämää dataa päihde- ja mielenterveyspalvelujen asiakasmääristä, suoritteista ja kustannuksista Suomen kuudessa suurimmassa kaupungissa. Nämä kaupungit ovat Helsinki, Espoo, Tampere, Vantaa, Oulu ja Turku. (Avoindata.fi 2022b.) Visualisoinnin tarkoituksena on vertailla kaupunkien välisiä kustannuksia ja selvittää, onko kaupunkien välillä rahan käyttöön liittyviä painotuksia.

Kuviossa 1 on esitetty edellä mainittujen kuuden kaupungin päihde- ja mielenterveyspalvelujen kustannukset vuodelta 2021. Tästä kuvioista näemme, että Helsingin osuus on kaikissa palveluissa selvästi suurin. Toiseksi suurimpina kustannuksiltaan ovat Turku ja Tampere. Turun ja Tampereen kustannukset ovat joka luokassa lähes yhtä suuret. Kolmantena on Espoo ja vähiten kustannuksia on Vantaalla ja Oululla. Mielenkiintoista on se, että Turku on jaetulla toisella sijalla kustannuksissa Tampereen kanssa, vaikka Turun asukasluku on kyseisistä kuudesta kaupungista pienin.



Kuvio 1. Kaupunkien väliset päihde- ja mielenterveyspalvelujen kustannukset vuonna 2021. Pylväskaavio.

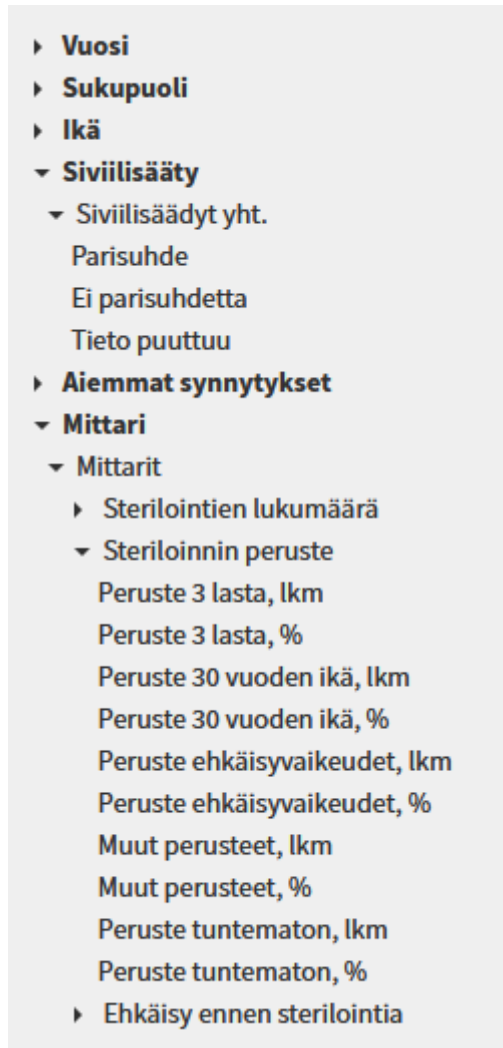
Kuviossa 2 vertaillaan kustannusten osuutta kyseisten kaupunkien kokonaiskustannuksiin. Jokaisessa kaupungissa suurin osuus kustannuksista tulee mielenterveyspalveluista, toiseksi suurin osuus päihdehuollon palveluista ja pienin osuus integroiduista palveluista. Helsingissä käytetään huomattavasti suurempi osuus päihdehuollon palveluihin kuin muissa kaupungeissa. Oulussa puolestaan mielenterveyspalvelut ovat kustannusten kärjessä. Kuvion 2 pinotusta pylväskaaviosta on huomattavasti helpommin luettavissa, kuinka suuri osuus kustannuksista menee millekin palvelulle, kuin kuvion 1 pylväskaaviosta.



Kuvio 2. Päihde- ja mielenterveyspalvelujen kustannusten osuus palvelujen kokonaiskustannuksista vuonna 2021. Pinottu pylväskaavio.

4.2 THL:n tarjoaman avoimen datan visualisointi

Tässä luvussa tutkitaan THL:n tarjoaman avoimen datan visualisoinnin mahdollisuuksia. THL on tehnyt tarjoamansa avoimen datan hyödyntämisestä helppoa. Sopivan aihealueen löydyttyä voi joko käyttää ohjelmointirajapintaa tai latauspalvelua. Tässä työssä käytettiin latauspalvelua. Sivuston vasemmassa reunassa on kuvassa 7 näkyviä luokkia, joita voi käyttää rajatakseen latauspalvelun luomaa taulukkoa. Taulukon rivejä ja sarakkeita voi vaihtaa raahaamalla luokkia vasemmasta reunasta taulukkoon. Taulukoissa 3 ja 4 on kaksi esimerkkiä samasta latauspalvelusta, jonka taulukkoon on rajattu eri tietoja. Taulukossa 3 on esitetty rajaamaton taulukko, ja taulukossa 4 ovat rajattuna ne, jotka ovat parisuhteessa ja joilla on kolme lasta. THL:n taulukot saa ladattua joko CSV-tiedostoon tai Excel-tiedostoon. Jotkin taulukot saa myös ladattua PDF-tiedostoon.







Kuva 7. THL:n latauspalvelun rajaustoiminnon eri vaihtoehtoja. (THL 2022.)

Taulukko 3. Rajaamaton taulukko naisten ja miesten steriloinnista. (THL 2022.)

Steriloinnit, lkm ⓘ	⚙️	Naiset ▼	Miehet ▼	Sukupuolet yht. ▼
⚙️				
1987 ▼		12 083	695	12 778
1988 ▼		11 954	589	12 543
1989 ▼		12 392	611	13 003
1990 ▼		12 864	593	13 457

Taulukko 4. Rajattu taulukko naisten ja miesten steriloinnista. (THL 2022.)

	 Naiset ▾	Miehet ▾	Sukupuolet yht. ▾
	 Parisuhde ▾	Parisuhde ▾	Parisuhde ▾
	 Peruste 3 lasta, lkm ▾	Peruste 3 lasta, lkm ▾	Peruste 3 lasta, lkm ▾
			
1987 ▾	4 826	39	4 865
1988 ▾	4 871	21	4 892
1989 ▾	4 905	32	4 937
1990 ▾	5 176	21	5 197

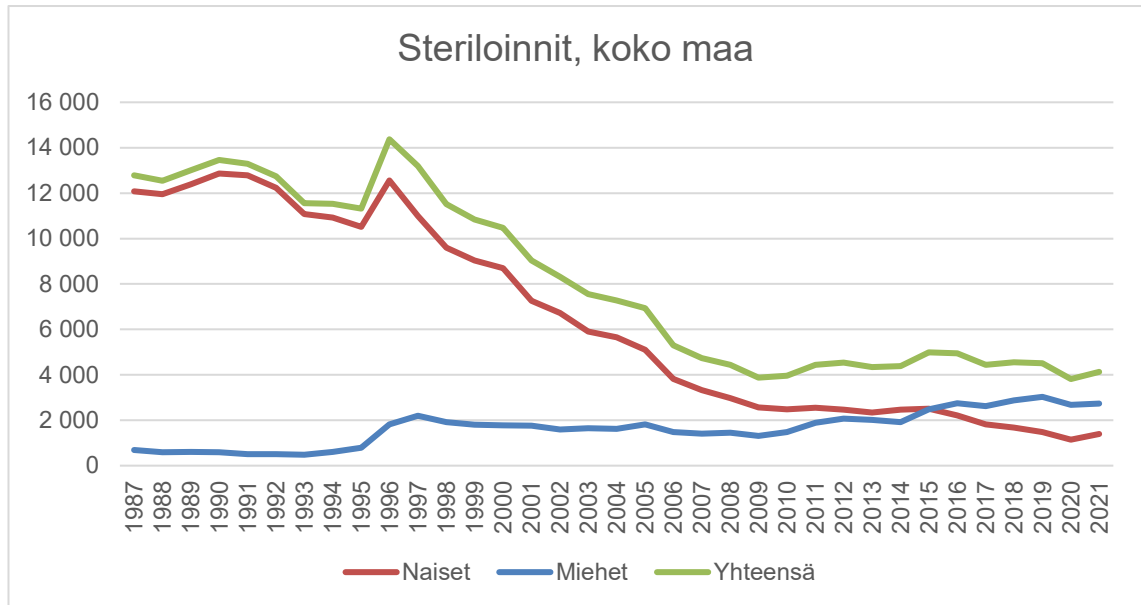
4.3 Esimerkki THL:n avoimen datan visualisoinnista: Steriloinnit

Näiden visualisointien luomiseen käytetty data löytyi THL:n avoimen datan tilastotietokannasta (THL 2022). THL:n taulukko ladattiin CSV-tiedostoon ja Excelin Tiedot-välilehdeltä data noudettiin suoraan Exceliin. Saman olisi voinut tehdä lataamalla suoraan Excel-tiedoston THL:n latauspalvelusta, mutta se ei onnistu jokaisessa latauspalvelussa. Data osoittaa naisten ja miesten steriloidut määrät Suomessa vuodesta 1987 alkaen vuoteen 2021 asti. Tässä luvussa käsitellään erilaisia mahdollisia kaavioita, joilla dataa voidaan visualisoida. Kaikissa luvun kaavioissa käytetty data on samasta tietokannasta.

4.3.1 Viivakaavio

Kuviossa 3 esitetyssä kaaviossa näkyy selkeästi sekä naisten että miesten sterilointien määrä vuosien kuluessa. Kaavion alussa yhteenlaskettujen sterilointien määrä mukaillee steriloitujen naisten määrää, sillä tilastoinnin alkuvuosina miesten sterilointi oli huomattavasti harvinaisempaa kuin naisten. Vuosien edetessä naisten sterilointi käy harvinaisemmaksi ja miesten sterilointi yleistyy. Näin yhteenlaskettu määräkin tasoittuu eikä mukaille oikeastaan kumpaakaan naisten tai miesten viivoista. Viivakaavio on hyvä kaavio

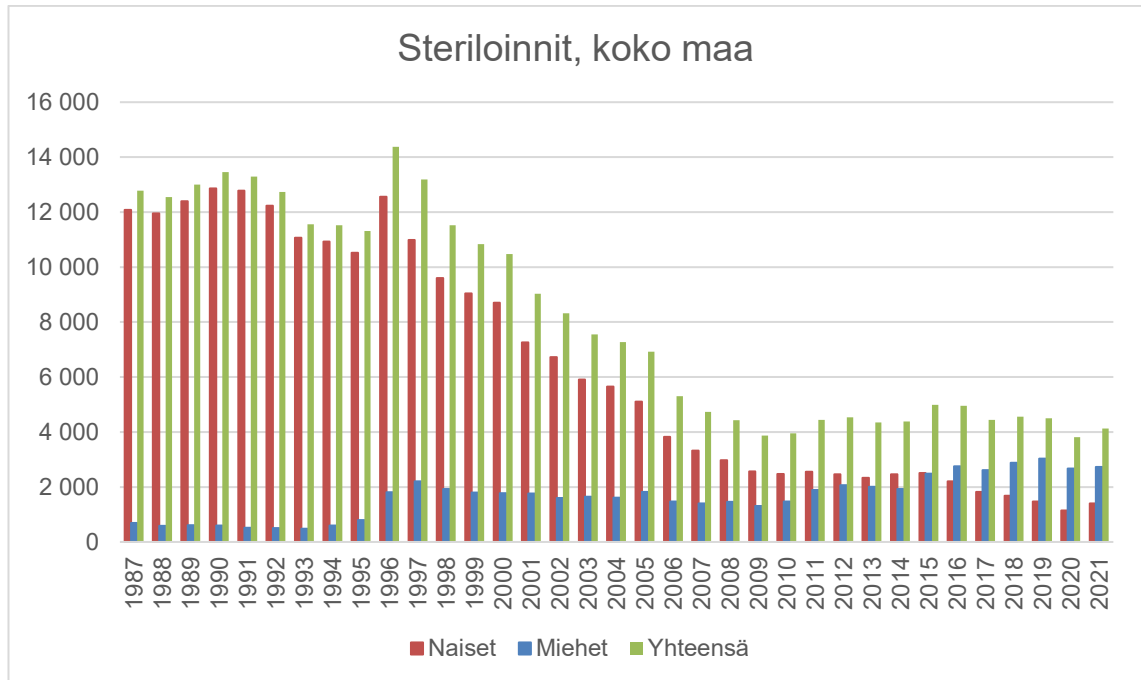
kuvaamaan tällaista dataa, sillä siitä selviää sekä steriloitujen lukumäärä että vuosi.



Kuvio 3. Sukupuolten väliset steriloinnit Suomessa. Viivakaavio.

4.3.2 Pylväskaavio

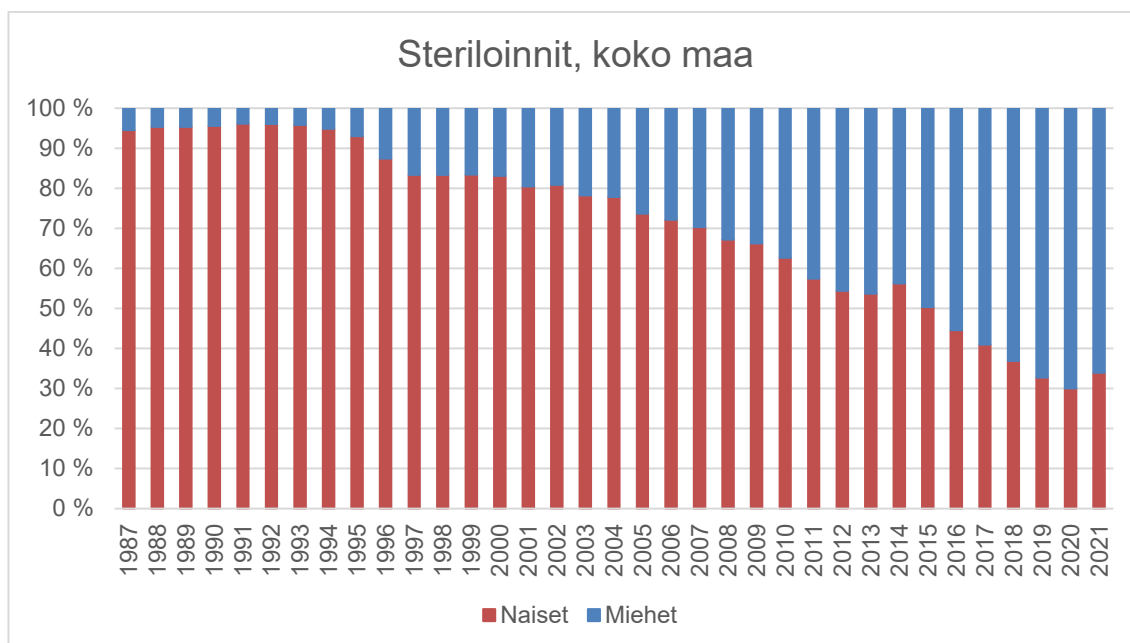
Pylväskaavio sopii hyvin havainnollistamaan dataa, jonka vaaka-akselilla on aikaa kuvaavaa dataa, esimerkiksi vuosia. Pylväskaavio sopii käytettäväksi kaavioksi hyvin, jos kummallakin akselilla on yksittäinen jatkuva asteikko. Kuvio 4 nähdään, että usean arvon asettaminen samalle vaaka-akselin arvolle tekee kaaviosta hyvin sekavan näköisen. Kaaviosta saa kyllä selville saman kuin kuvio 3, mutta pylväskaavio ei ole välttämättä selkein tällaisen datan visualisoimiseen.



Kuvio 4. Sukupuolten väliset steriloinnit Suomessa. Pylväskaavio.

4.3.3 Pinottu pylväskaavio

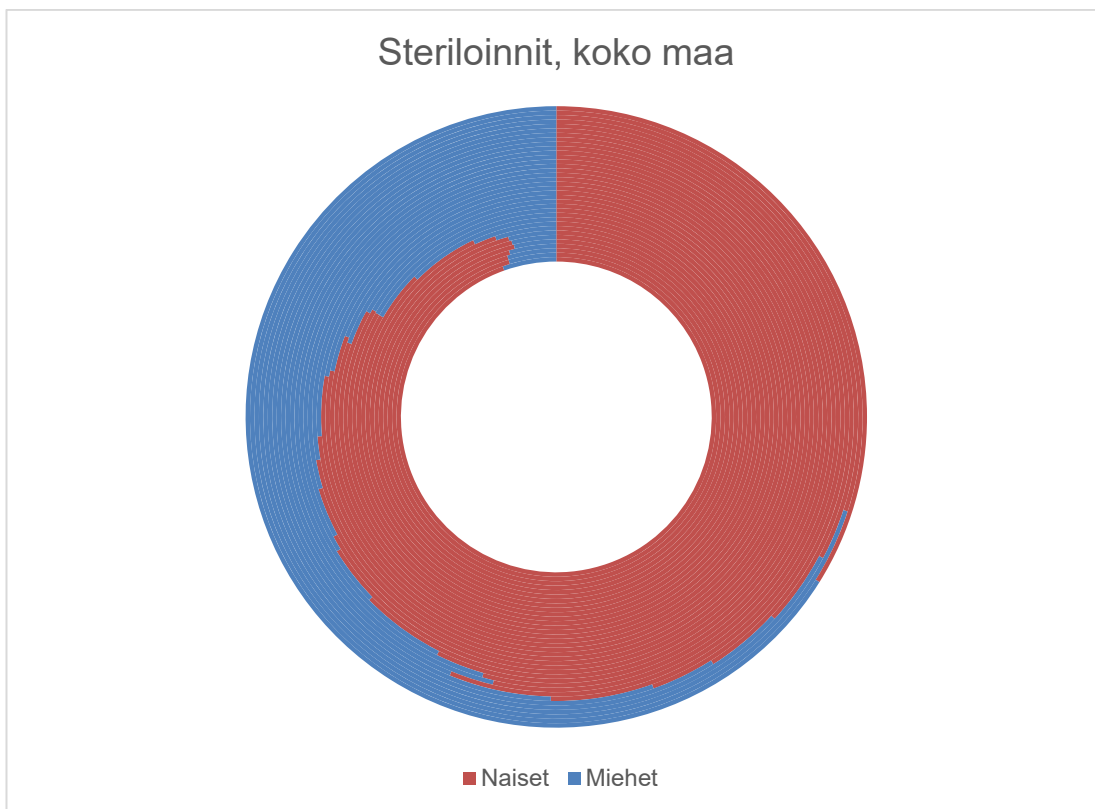
Kuviossa 5 esitetystä kaaviosta on poistettu naisten ja miesten yhteenlaskettu määrä. Kuviossa 5 ei käy ilmi kummankaan sukupuolen kokonaismäärää, mutta niiden väliset suhteet on helppo hahmottaa. Vuosien mittaan miesten osuus steriloinneissa kasvaa, vaikka steriloitujen miesten määrä ei kasva huomattavasti, mikä johtuu tietysti siitä, että steriloitujen naisten määrä on vuosien myötä laskenut huomattavasti. Pinottu pylväskaavio on hyvä kaavio selvittämään kahden arvon välistä suhdetta.



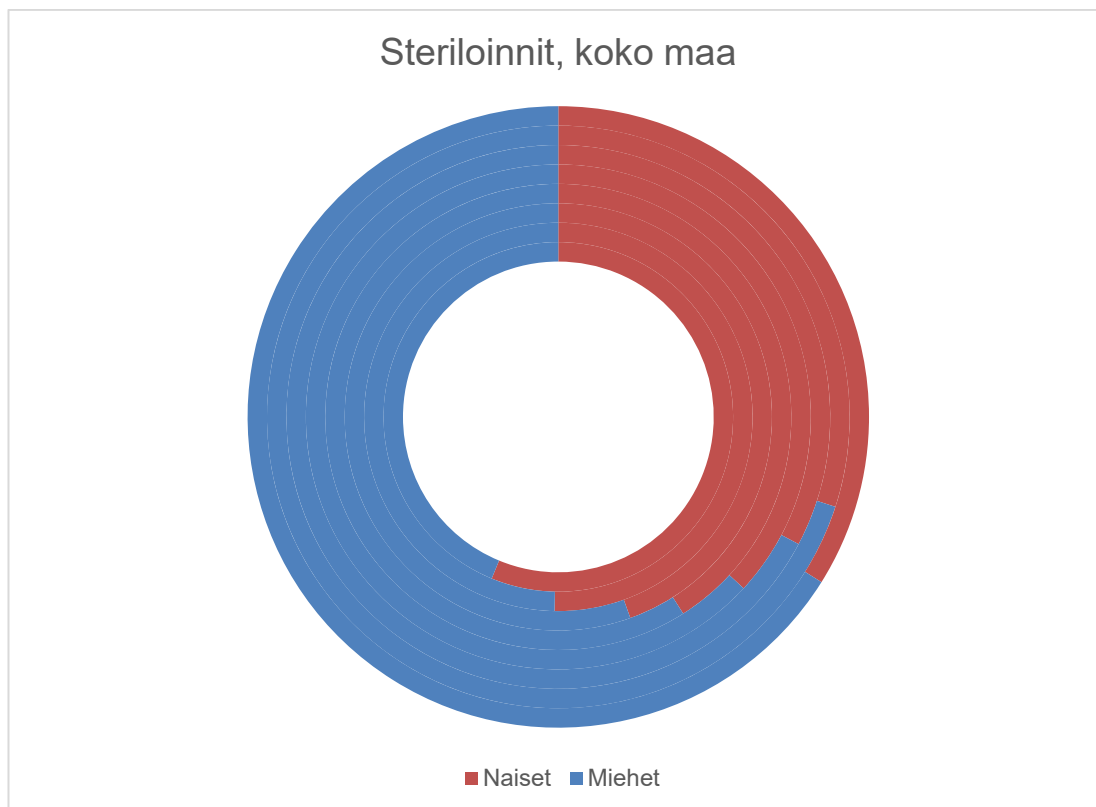
Kuvio 5. Sukupuolten väliset steriloinnit Suomessa. Pinottu pylväskaavio.

4.3.4 Rengaskaavio

Perinteinen ympyräkaavio tai piirakkakaavio ei sovellu tässä luvussa käsiteltävän datan visualisoimiseen. Ympyräkaavion hyvänä puolena on, että siitä pystyy helposti hahmottamaan ympyrän puolikkaan, kolmanneksen tai neljänneksen ja vertaamaan osioiden kokoja. Ympyräkaaviosta on kuitenkin hyvin vaikea hahmottaa osuuksien välisiä eroja tarkasti. Kuviossa 6 käytetty rengaskaavio on hyvin saman tapainen kuin ympyräkaavio. Rengaskaaviossa luetaan renkaita ympyräkaavion piirakan palan tapaisten sektorien sijaan. Rengaskaavio ei ole ehkä yleisin visualisointiin käytettävistä kaavioista, mutta se ajaa asiansa. Kaaviosta näkee kohtalaisen helposti, kuinka suuri osuus naisilla ja miehillä on jokaisella yksittäisellä renkaalla. Rengaskaavio sopisi vielä paremmin, jos datassa olisi vähemmän vuosia visualisoitavana. Kuviossa 7 on esimerkki samasta datasta, mutta vain viimeisen kahdeksan vuoden ajalta. Tässä kaaviossa näkyy selkeästi, että miesten osuus on ohittanut naisten osuuden ja viimeistä vuotta lukuun ottamatta jatkaa kasvamista vuosi vuodelta.



Kuvio 6. Sukupuolten väliset steriloinnit Suomessa, vuodet 1987–2021. Rengaskaavio.



Kuvio 7. Sukupuolten väliset steriloinnit Suomessa, vuodet 2014–2021. Rengaskaavio.

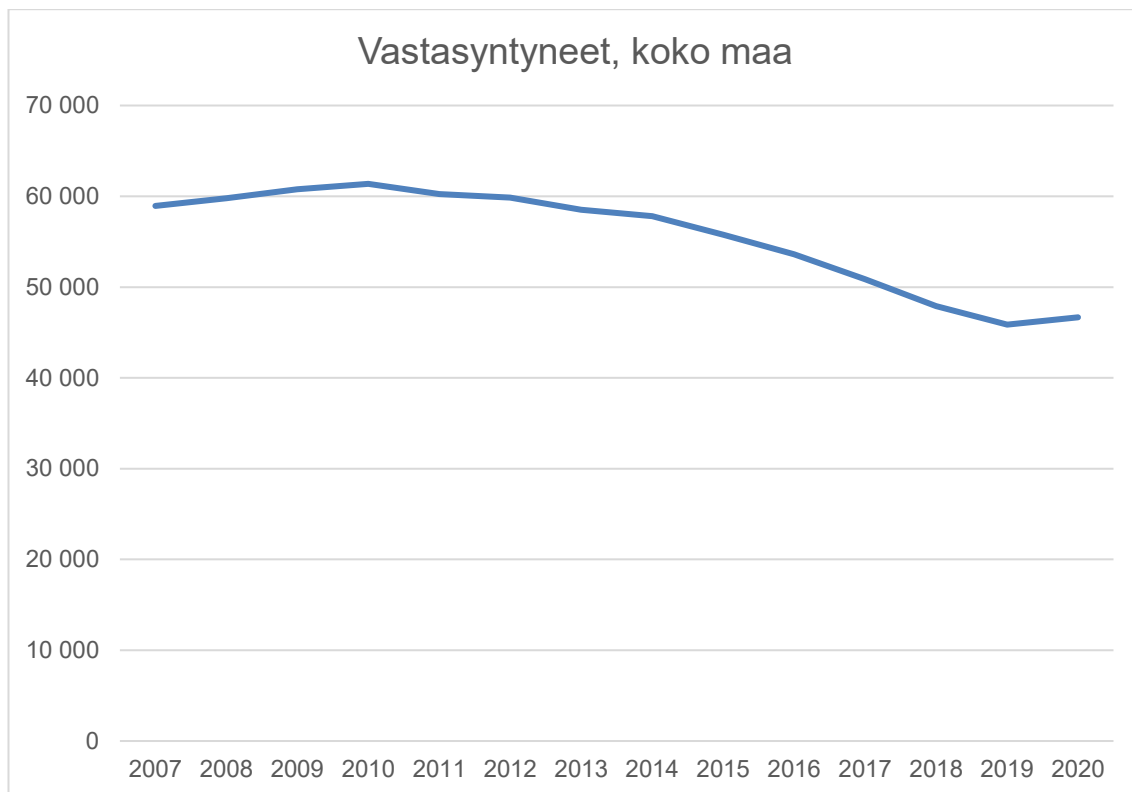
4.4 Esimerkki THL:n avoimen datan visualisoinnista: Vastasyntyneet

Näiden visualisointien luomiseen käytetty data löytyi THL:n avoimen datan tilastotietokannasta (THL 2021). Data osoittaa vastasyntyneiden lasten määrät Suomessa vuodesta 2007 alkaen vuoteen 2020 asti. Tässä luvussa toistetaan edellisen luvun visualisoinnit ja vertaillaan kaavioista luettavia tuloksia. Kaikissa luvun kaavioissa käytetty data on samasta tietokannasta.

4.4.1 Viivakaavio

Kuvion 8 kuvaaja kertoo riittävän hyvin vastasyntyneiden lasten määrän jokaisena vuonna. Neljä ensimmäistä vuotta on tasaista kasvua vastasyntyneiden määrässä, minkä jälkeen on yhdeksän vuotta laskua. Viimeisenä vuonna vastasyntyneiden määrä on taas kasvussa. Kuviossa 3

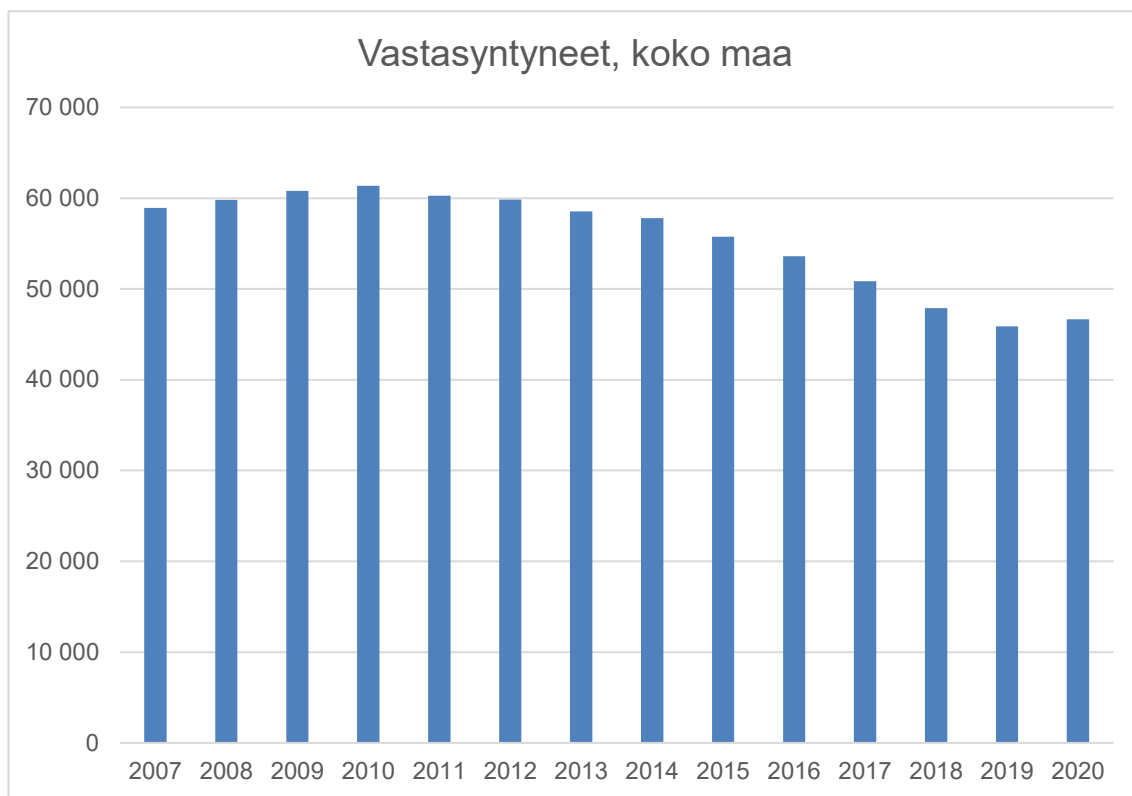
esitetty kuvaaja kertoo paljon enemmän siihen käytetystä datasta kuin kuvio 8 vastaavasti sen datasta. Viivakaavion käyttö tämän datan visualisointiin ei ole väärin, mutta parempiakin vaihtoehtoja on.



Kuvio 8. Vastasyntyneet Suomessa. Viivakaavio.

4.4.2 Pylväskaavio

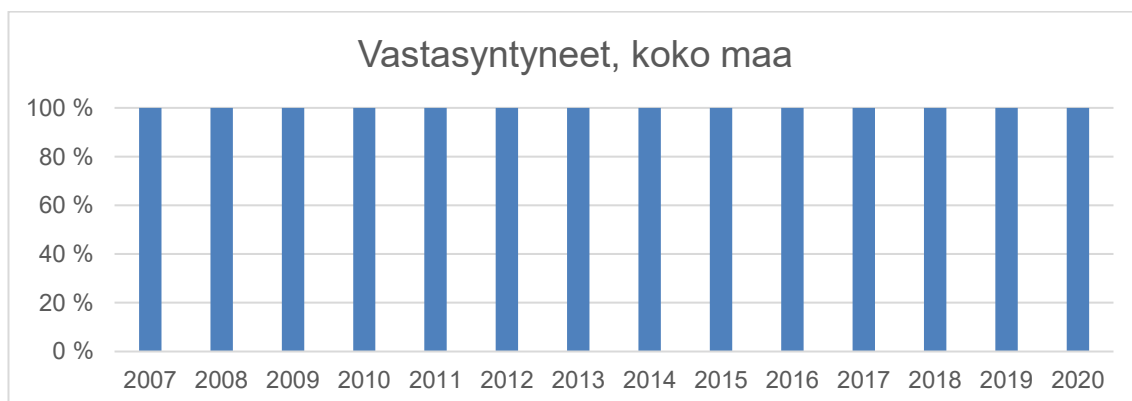
Pylväskaavio on tämänkaltaisen datan visualisoimiseen paras vaihtoehto. Kuvion 9 kaavio on hyvin yksinkertainen ja siitä näkee jokaisen vuoden kohdalla selkeästi, kuinka monta lasta sinä vuonna on syntynyt. Pylväskaaviosta saa helposti selville myös vastasyntyneiden määrän kasvut ja laskut.



Kuvio 9. Vastasyntyneet Suomessa. Pylväskaavio.

4.4.3 Pinottu pylväskaavio

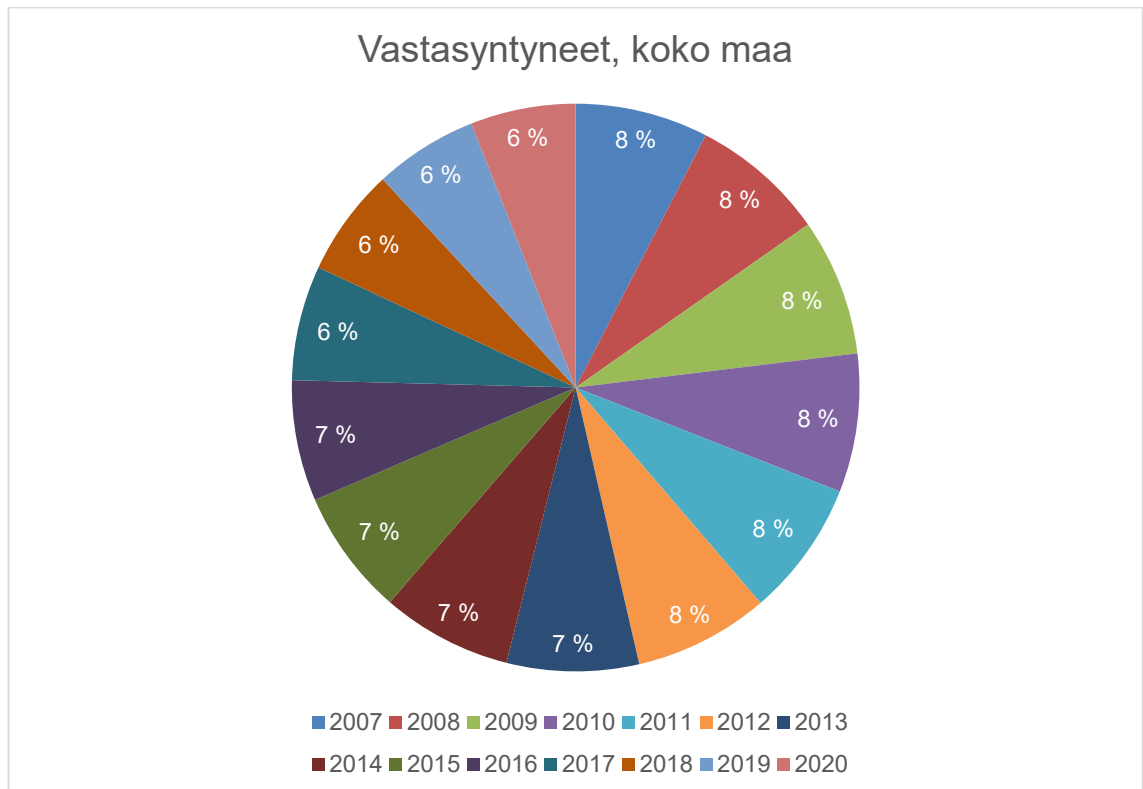
Yhden arvon sisältävästä datasta on turhaa tehdä pinottua pylväskaaviota, sillä se näyttää muuten täysin samalta kuin kuviossa 5, mutta jokainen palkki on yksivärinen.



Kuvio 10. Vastasyntyneet Suomessa. Pinottu pylväskaavio.

4.4.4 Ympyräkaavio

Rengaskaaviota tästä datasta ei kannata tehdä, sillä siitä ei voisi lukea minkäänlaista dataa. Rengaskaavio olisi samanlainen kuin kuvion 10 kaavio, mutta se olisi ympyrän muotoinen. Ympyräkaavio on miellyttävän näköinen, mutta se ei kerro paljoa. Kuvioista 11 huomataan, että kaikki ympyräkaavion osat näyttävät hyvin samankokoisilta. Kaavioon voidaan lisätä jokaisen osan prosenttiosuus kokonaisuudesta, mutta se kertoo siitä, että datan visualisoimiseen olisi pitänyt valita jokin toinen kaavio. Ympyräkaaviosta ei myöskään näe, kuinka monta vastasyntynyttä minäkin vuonna on, ellei erikseen mainita kokonaismäärää.



Kuvio 11. Vastasyntyneet Suomessa. Ympyräkaavio.

5 Yhteenveto

Opinnäytetyössä selvitettiin, mitä tarkoitetaan avoimella datalla ja datan visualisoinnilla. Työssä käytiin läpi, miten dataa avataan ja mistä avattua dataa löytää. Datan visualisoinnin historiaan perehdyttiin tarkemmin ja lopulta luotiin esimerkkejä datan visualisoinnista.

Opinnäytetyön tarkoituksena oli selventää avointa dataa ja esittää data helposti ymmärrettävässä visualisoidussa muodossa. Työssä toteutetut visualisoinnit tehtiin Suomen kuuden suurimman kaupungin päihde- ja mielenterveyspalveluista sekä kahdesta datasetistä THL:n tarjoaman avoimen datan tietokannasta.

Päihde- ja mielenterveyspalveluiden visualisointien tarkoituksena oli vertailla Suomen kuuden suurimman kaupungin välisiä kustannuksia ja selvittää, onko kaupunkien välillä rahan käyttöön liittyviä painotuksia. Visualisoinnista oli selvästi nähtävissä, että Helsinki käytti muihin viiteen kaupunkiin verrattuna eniten rahaa päihde- ja mielenterveyspalveluihin. Toiseksi eniten käytti Turku, vaikka Turku oli kyseisistä kuudesta kaupungista asukasluvultaan pienin.

Päihde- ja mielenterveyspalveluiden visualisointiin käytettiin pinottua pylväskaaviota ja tavallista pylväskaaviota. Pinotusta pylväskaaviosta on huomattavasti helpommin luettavissa kuin pylväskaaviosta, kuinka suuri osuus kustannuksista menee millekin palvelulle. Datan visualisointi oli onnistunut, sillä kaavioista saa helposti hyvän käsityksen kaupunkien kustannuksista.

THL:n latauspalvelua hyödyntäen ladattiin kaksi eri datasettiä THL:n avoimen datan tietokannasta. Näistä kahdesta datasta tehtiin samat visualisoinnit ja vertailtiin kaavioista selviäviä tuloksia. Visualisoinnin tarkoituksena oli tehdä esimerkkejä erilaisista kaavioista käyttäen kahta eri datasettiä ja vertailla kaavioista selviävää dataa. Kummallekin datasetille saatiin luotua selkeästi paras dataa visualisoiva kaavio.

THL:n datasettien ensimmäistä dataa parhaiten visualisoi viivakaavio ja toista pylväskaavio. Viivakaavio on hyvä vaihtoehto visualisoinnille, jos dataa on

tarpeeksi eikä viivakaaviosta tule liian yksinkertaista. Yksinkertaisen viivakaavion sijaan voi käyttää esimerkiksi pylväskaaviota, jonka vahvuus on erityisesti datassa, jossa on vähän visualisoitavia kategorioita. Visualisoinnissa käytettiin myös pinottua pylväskaaviota, josta saa hyvin selville kahden arvon välisen suhteen. Kummankin datasetin visualisointiin löytyi dataa huonoimmin tai epäselvimmin kuvaava kaavio. Ensimmäistä dataa heikoimmin visualisoi pylväskaavio ja toista ympyräkaavio. Ympyräkaaviolla on hyvä kuvata dataa, jossa ei ole monta kategoriaa ja josta halutaan saada selville useamman arvon väliset suhteet. Jälkimmäisen datasetin tapauksessa pinotusta pylväskaaviosta ei voitu selvittää tarvittavia tietoja. Molempien datasettien pylväskaavioita tarkasteltaessa nähdään, kuinka paljon selkeämpi toisen datasetin kaavio on. Datan visualisoinnin esimerkit onnistuivat, sillä ne ovat hyvin erilaisia toisiinsa nähden.

Lähteet

5stardata.info 2015a. 5 ★ Open Data. Viitattu 6.11.2022

<https://5stardata.info/en/>

5stardata.info 2015b. Temperature forecast for Galway, Ireland. Viitattu 6.11.2022 <https://5stardata.info/en/examples/gtd-4/>

5stardata.info 2015c. Temperature forecast for Galway, Ireland. Viitattu 6.11.2022 <https://5stardata.info/en/examples/gtd-5/>

Avoindata.fi 2020a. Kartoita tietovarannot. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/kartoita-tietovarannot>

Avoindata.fi 2020b. Selvitä datan hyödyntäjien tarpeet. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/selvita-datan-hyodyntajien-tarpeet>

Avoindata.fi 2020c. Selvitä rajoitteet ja riskit. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/selvita-rajoitteet-ja-riskit>

Avoindata.fi 2020d. Valitse avattava data. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/valitse-avattava-data>

Avoindata.fi 2020e. Valitse jakelumuodot. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/valitse-jakelumuodot>

Avoindata.fi 2020f. Valitse tiedostomuodot. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/valitse-tiedostomuodot>

Avoindata.fi 2020g. Julkaise data. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/julkaise-data>

Avoindata.fi 2022a. Viimeistele datan avaaminen. Viitattu 6.11.2022

<https://www.avoindata.fi/fi/opas/viimeistele-datan-avaaminen>

Avoindata.fi 2022b. Kuuden suurimman kaupungin päihde- ja mielenterveyspalvelujen vertailu. Viitattu 10.11.2022

<https://www.avoindata.fi/data/fi/dataset/kuuden-suurimman-kaupungin-paihde-ja-mielenterveyspalvelujen-vertailu>

Brush, K. & Burns, E. 2020. Data visualization. Viitattu 7.11.2022
<https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization>

Creativecommons.fi 2022a. Tietoa lisensseistä. Viitattu 5.11.2022
<https://creativecommons.fi/lisenssit/>

EUR-Lex 2019. Tiivistelmä asiakirjasta: ”Avoimesta datasta ja julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä annettu direktiivi (EU) 2019/1024”. Viitattu 6.11.2022 <https://eur-lex.europa.eu/legal-content/FI/LSU/?uri=CELEX:32019L1024>

Friendly, M. 2006. A Brief History of Data Visualization. Viitattu 7.11.2022
<http://euclid.psych.yorku.ca/datavis/papers/hbook.pdf>

Helsinki Region Infoshare 2017a. Mitä on avoin data? Viitattu 5.11.2022
<https://hri.fi/fi/ohjeet/mita-on-avoin-data/>

Helsinki Region Infoshare 2017b. Datan hyödyntäjälle. Viitattu 5.11.2022
<https://hri.fi/fi/ohjeet/datan-hyodyntajalle/>

Helsinki Region Infoshare 2017c. Opas datan avaamiseen. Viitattu 6.11.2022
<https://hri.fi/fi/ohjeet/datan-avaajalle/opas-datan-avaamiseen/>

Insightsoftware 2019. A Brief History of Data Visualization. Viitattu 7.11.2022
<https://insightsoftware.com/blog/a-brief-history-of-data-visualization/>

Jacobs, F. 2010. The Minard Map – “The best statistical graphic ever drawn”. Viitattu 7.11.2022 <https://bigthink.com/strange-maps/229-vital-statistics-of-a-deadly-campaign-the-minard-map/>

Kim, J. G. 2015a. 5 ★ Open Data. Viitattu 6.11.2022 <https://5stardata.info/en/>

Kim, J. G. 2015b. Temperature forecast for Galway, Ireland. Viitattu 6.11.2022
<https://5stardata.info/en/examples/gtd-4/>

Kim, J. G. 2015c. Temperature forecast for Galway, Ireland. Viitattu 6.11.2022
<https://5stardata.info/en/examples/gtd-5/>

Korhonen, J. & Sore, S. 2017. Lahden kaupungin tuottaman datan avaamisen haasteet – kuinka saadaan tieto hyötykäyttöön? Viitattu 6.11.2022

<https://www.lamkpub.fi/2017/10/11/lahden-kaupungin-tuottaman-datan-avaamisen-haasteet-kuinka-saadaan-tieto-hyotykayttoon/>

Tableau 2022. What Is Data Visualization? Definition, Examples, And Learning Resources. Viitattu 7.11.2022 <https://www.tableau.com/learn/articles/data-visualization>

THL 2021. Vastasyntyneet alueittain. Viitattu 11.11.2022
https://sampo.thl.fi/pivot/prod/fi/synre/vastasyalue/fact_synre_vastasyalue

THL 2022. Steriloinnit, koko maa. Viitattu 11.11.2022
https://sampo.thl.fi/pivot/prod/fi/ster/kokomaa/fact_ster_kokomaa

Tufte, E. R. 1983. The Visual Display of Quantitative Information. Viitattu 7.11.2022
<http://faculty.salisbury.edu/~jtanderson/teaching/cosc311/fa21/files/tufte.pdf>

Tufte, E. R. 1997. Visual Explanations. Viitattu 7.11.2022
<http://www2.jufejus.org.ar/www.jufejus.org.ar/images/doc/ACTIVIDADES/Estadisticas/Jornadas%20de%20Capacitacion/Material/JornadasX/Tufte/Visual%20Explanations.pdf>