

Kalle Karlsson

HÄMEEN ALUEEN AVOIMEN DATAN ESIKÄSITTELY, VISUALISOINTI JA VIESTINTÄ

Opinnäytetyö

Tradenomi (AMK)

Liiketoiminnan data-analytiikan ja visualisoinnin koulutus

2021



**Kaakkois-Suomen
ammattikorkeakoulu**



Kaakkois-Suomen
ammattikorkeakoulu

Tutkintonimike	Tradenomi (AMK)
Tekijä/Tekijät	Kalle Karlsson
Työn nimi	Hämeen alueen avoimen datan esikäsittely, visualisointi ja viestintä
Toimeksiantaja	Smart & Lean Hub Oy
Vuosi	2021
Sivut	67
Työn ohjaaja(t)	Päivi Hurri

TIIVISTELMÄ

Tämän opinnäytetyön tarkoituksena oli tutkia, miten ja millä työkaluilla voitiin toteuttaa avoimesta datasta interaktiivisia visualisointeja toimeksiantajan verkkoviestintää varten.

Opinnäytetyön toimeksiantaja oli lahtelainen yritys, joka toimii konsortion kumppanina Euroopan Unionin Horizon2020 -ohjelman rahoittamassa PoliRural TKI-hankkeessa. Hankkeen kohdealue on Häme, mikä pitää sisällään Päijät - ja Kanta-Hämeen maakunnat. Toimeksiantaja halusi sosiaalisessa mediassa helposti jaettavia interaktiivisia datan visualisointeja Hämeen tai koko Suomen datasta. Visualisointien tarkoitus oli tuoda esille demograafinen, taloudellinen tai muu muutos pitkällä aikavälillä.

Opinnäytetyön teorian tavoitteena oli esittää datan visualisointia yleisesti ja datan visualisoinnin hyviä käytäntöjä. Opinnäytetyön prosessuaalisessa vaiheessa tutkittiin, mistä datasta visualisointi toteutettiin, miksi ja miten.

Opinnäytetyön lopputuloksena syntyi viisi erilaista datan visualisointia sekä prosessuaalinen kuvaus niiden toteuttamisesta. Opinnäytetyön loppuun on laadittu visualisointi verkkoviestinnän saavuttamasta huomiosta sosiaalisessa mediassa.

Asiasanat: datan visualisointi, avoin data, Power BI

Degree	Bachelor of Business Administration
Author (authors)	Kalle Karlsson
Thesis title	The editing, visualization, and communication of open data from the Häme region
Commissioned by	Smart & Lean Hub Oy
Time	2021
Pages	67
Supervisor	Päivi Hurri

ABSTRACT

The purpose of this thesis was to find out how and with what tools could interactive data visualizations be produced from open data for a company's social media communication. The thesis was commissioned by a Lahti-based company that is a consortium partner in the PoliRural project funded by the European Union's Horizon2020 program. The PoliRural project aims to improve the attractiveness of Europe's rural areas. One pilot in the project is the Häme area, which includes the regions of Päijät-Häme and Kanta-Häme. The client wanted interactive data visualizations of the Häme area and/or the entire Finland that could be easily shared on social media. The purpose of the visualizations was to highlight demographic, economic, or other long-term changes.

The aim of the thesis theory was to go through data visualization in general and good data visualization practices. In the procedural phase of the thesis, the discussion focuses on which data the visualization was carried out, why and how.

The result of the thesis was five different data visualizations and a procedural description of their implementation. At the end of the thesis, visualization was made about attention received in LinkedIn and Twitter through media communication.

Keywords: data visualization, open data, Power BI

SISÄLLYS

1	JOHDANTO.....	5
1.1	Opinnäytetyön tarkoitus ja tavoitteet.....	5
1.2	Opinnäytetyön määrittely	6
2	DATAN VISUALISOINTI.....	7
2.1	Mitä datan visualisointi on?.....	7
2.2	Datan visualisoinnin lyhyt historia ja nykytilanne	13
2.3	Datan visualisoinnin toteuttamisen suunnittelu	16
2.3.1	Onko visualisointi konseptuaalinen vai datapainotteinen?.....	17
2.3.2	Visualisoinnin tavoite ja kohderyhmä.....	18
2.3.3	Visualisoinnin ulkomuoto	20
2.4	Loppupäätelmä visualisaation ulkomuodosta	27
2.5	Esimerkki datan visualisoinnista	28
3	TOIMEKSIANTAJA.....	30
4	VISUALISOINTIEN TOTEUTTAMINEN.....	30
4.1	Python-visualisointi: Eläkeläiset ja väkiluku kunnittain Päijät-Hämeessä.....	31
4.2	Power BI-visualisointi: Puukartta Hämeen alueen yritysten toimipaikoista	34
4.3	Power BI-visualisointi: Animoitu viivadiagrammi väestön muutoksesta	40
4.4	Power BI-visualisointi: Viivadiagrammi Hämeen alueen maakuntien populaation koulutusasteesta	44
4.5	Mapbox-visualisointi: Mediaanitulot postinumeron mukaan Suomessa	49
5	TULOKSET JA JOHTOPÄÄTÖKSET	58
5.1	Toimeksiantajan mielipiteet	58
5.2	Työntekijän mielipiteet	59
5.3	Kehittämiskohteet	60

LÄHTEET

KUVA- JA TAULUKKOLUETTELO

1 JOHDANTO

1.1 Opinnäytetyön tarkoitus ja tavoitteet

Tämän opinnäytetyön tarkoituksena on tutkia, miten interaktiivisia datan visualisointeja voidaan toteuttaa. Opinnäytetyössä tehtyjä interaktiivisia avoimen datan visualisointeja hyödynnetään toimeksiantajan, Smart & Lean Hub Oy:n verkkoviestinnässä.

Opinnäytetyön teoriaosuudessa (luku 2) perehdytään interaktiivisten visualisointien toteuttamiseen. Opinnäytetyön empiirisessä osassa (luku 4) testataan toimeksiantajan määrittelemien interaktiivisten visualisointien toteuttamista sekä lopuksi analysoidaan visualisointien saamaa huomiota.

Toimeksiantajalle tehtävien visualisointien tärkeä piirre on interaktiivisuus eli visualisoinnin katsoja kykenee interaktiivisesti muuttamaan, muokkaamaan ja porautumaan visualisointiin. Interaktiivisuuden tarve datan visualisoinnissa rajaa toimeksiannon aikana käytettäviä ohjelmia ja työkaluja. Toimeksiantaja käyttää visualisointeja sosiaalisessa mediassa sekä yrityksen verkkosivuilla. Tämä tarkoittaa, että datan visualisoinnit pitää toteuttaa sellaisessa muodossa, että visualisoinnin näkijän ei tarvitse ladata visualisointia tai käyttää minkäänlaisia ohjelmia visualisoinnin näkemiseksi interaktiivisena.

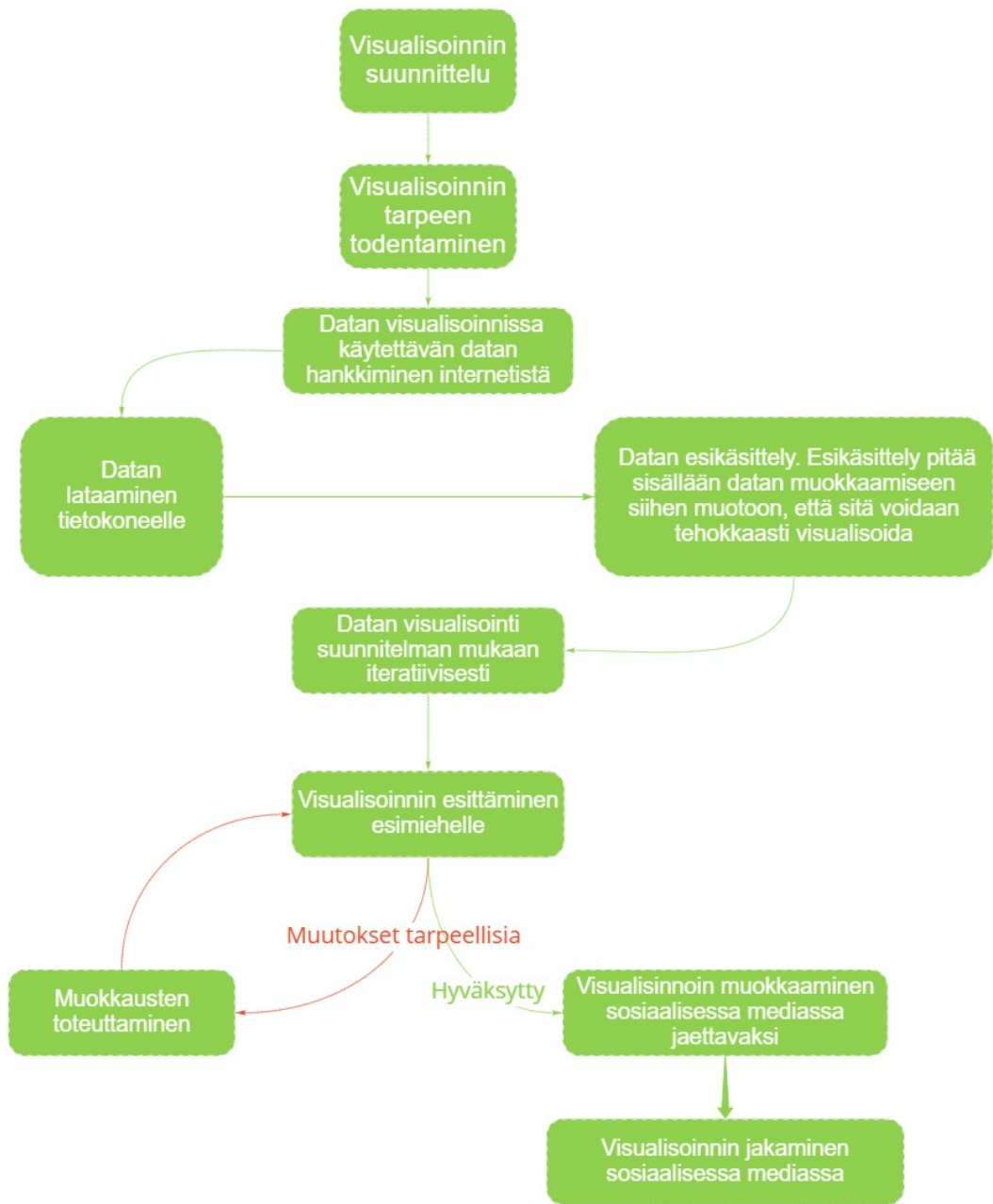
Visualisointien toteuttamisessa hyödynnetään monipuolisesti erilaisia ohjelmia ja muita tietoteknisiä työkaluja kuten Python- ja DAX-kieliä, MapBox-studiota sekä Microsoftin Power BI-ohjelmaa ja sen erilaisia lisäosia. Esimerkkinä visualisoinneista ovat kartat, puukartat, pylväskaaviot sekä pistekaaviomalliset diagrammit. Visualisaatioista osa toteutetaan animoituna, mikä mahdollistaa datan visualisoinnin liikkuvan kuvan muodossa. Opinnäytetyön aikana tehtävissä visualisoinneissa tärkein muuttuja on ajallinen muutos, joka pitää tulla esiin suurimmassa osassa visualisoinneista.

1.2 Opinnäytetyön määrittely

Toimeksianto sisältää useita toimenpiteitä alkaen datan etsimisestä, esikäsittelystä ja visualisoinnista aina visualisointien jakamiseen ja viestinnän tehokkuuden arviointiin asti. Opinnäytetyön prosessuaaliset vaiheet ovat:

1. Visualisoinnin tarpeen todentaminen
2. Visualisoinnin suunnittelu
3. Datan visualisoinnissa käytettävän datan hankkiminen internetistä
4. Datan lataaminen tietokoneelle
5. Datan esikäsittely, joka sisältää datan muokkaamiseen sellaiseen muotoon, että sitä voidaan tehokkaasti visualisoida
6. Datan visualisointi suunnitelman mukaan iteratiivisesti
7. Visualisoinnin esittäminen esimiehelle
8. Muokkauksen toteuttaminen
9. Visualisoinnin muokkaaminen sosiaalisessa mediassa jaettavaksi
10. Visualisoinnin jakaminen sosiaalisessa mediassa

Opinnäytetyön prosessuaalinen tutkimusosuus toteutetaan vaiheittain tutkimmalla jokaisen edellä esitetyn 10 vaiheen toteutusta ja merkitystä kokonaisprosessissa. Datan visualisointia kuvaillaan hyödyntämällä edellä listattuja vaiheita. Visualisointien toteuttamista kuvaillaan prosessuaalisesti vaihe vaiheelta ja pyritään toteuttamaan useita erilaisia visualisaatioita erilaisilla ohjelmilla. Opinnäytetyön tuloksena on erilaisilla visualisointeja, joiden toteuttamiseen on käytetty samaa prosessikuvausta. Kuva 1 kuvaa datan visualisoinnin toteuttamista tässä opinnäytetyössä hyödyntämällä prosessikuvausta sekä Miro-nimistä verkosta saatavaa ohjelmaa. Prosessikuvasta hyödynnetään toimeksiantajan haluamien visualisointien toteuttamisessa (kuva1). Prosessin lopputuloksena syntyy verkossa jaettavia interaktiivisia visualisointeja, joita toimeksiantaja voi hyödyntää viestinnässään.



miro

Kuva 1. Prosessikuvaus prosessuaalisista vaiheista Miro-esityksenä

2 DATAN VISUALISOINTI

2.1 Mitä datan visualisointi on?

Datan visualisointi on datan graafista esittämistä yksinkertaisemmassa ja tiedollisesti rikkaassa muodossa (TableAu, 2020). Datan visualisoinnin lopputulos voi olla kuva, kartta tai vaikkapa sovellus.

Visualisointeja voidaan toteuttaa lähes jokaisesta aiheesta, kunhan visualisoinnin hyödyntämä data on hallittavassa rakenteellisessa muodossa. Visualisoinnit hyödyntävät ihmisen näköaistia monimutkaisen tai lukumääräisesti vaikeasti hahmotettavan asian yksinkertaistamiseksi (Few, 2021). Visualisoinnin päämääränä on ottaa vaikeasti hahmotettava asia, aihe tai datapaketti ja muokata se helposti ymmärrettävään ja selkeään visuaaliseen muotoon. Visualisoinnin avulla voidaan tuoda esille datasta hahmotettavia aihealueita eli trendejä, toistuvia malleja ja poikkeavaisuuksia (Bush, 2018).

Näiden aihealueiden hahmottaminen datasta voi olla vaikeaa tai jopa mahdotonta riippuen datan laajuudesta tai sen monimutkaisuudesta. Hyvin toteutettu visualisaatio tuo esille suuren joukon dataa yksinkertaisessa ja kompaktissa muodossa eli visualisaatiossa on paljon dataa suhteessa käytettävään tilaan (Tableau, 2021). Visualisointien avulla voidaankin tarkastella ihmiselle vaikeasti hahmotettavaa ja lukumääräisesti massiivista dataa. On myös mahdollista, että dataa ei voida tarkastella ilman usean erilaisen datan yhdistelyä toisiinsa ja yhteyksien luomista. Esimerkiksi datassa tapahtuvien muutosten havaitseminen raakadatasta on usein vaikeaa tai esimerkiksi valtion populaatiomuutoksen tarkka tarkastelu vaatii muita muuttujia tekemään datasta tietellisempää. Näiden aihealueiden ja yhteyksien hahmottamiseksi hyödynnetään visualisointeja tärkeiden lukujen esille tuomiseksi.

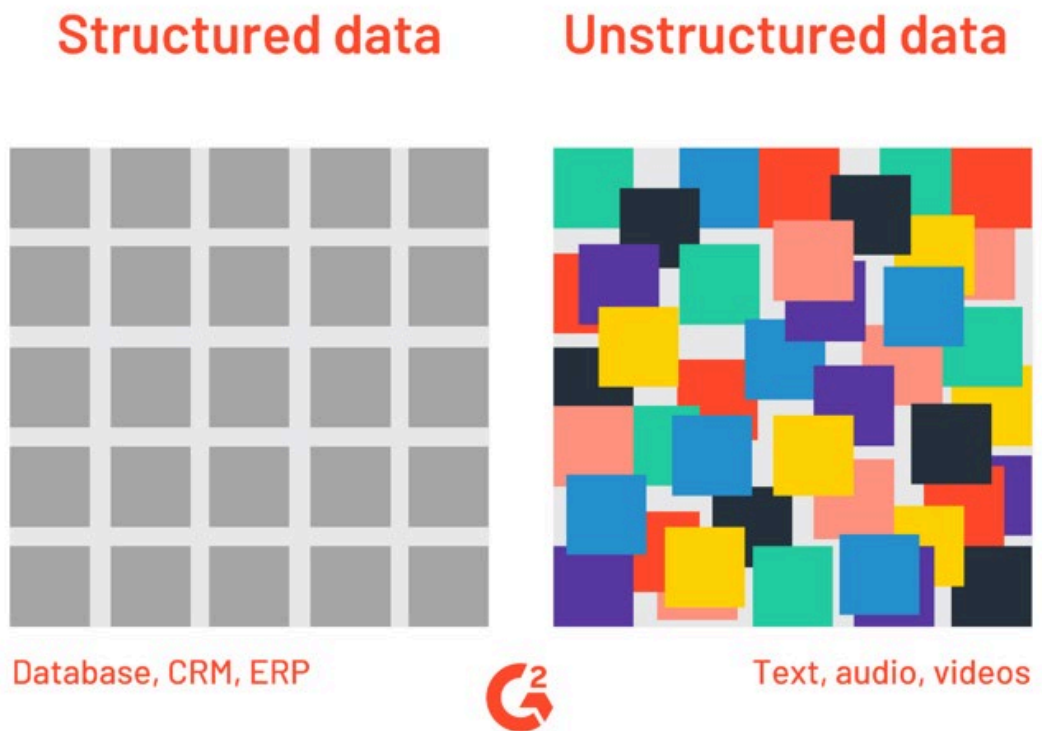
Datan visualisointi mahdollistaa myös datan tunnuslukujen, datassa tapahtuvan muutoksen tai datan eriävyyksien tehokkaan graafisen korostamisen sekä vertailun keskenään. Visualisoinnin tarve korostuu, kun dataa halutaan yksinkertaistaa, mutta datan kertoma tarina sekä muoto halutaan pitää ennallaan. Datan visualisointi mahdollistaa tällaisen datan esille tuomisen suuremmalle yleisölle. Jotta dataa voidaan hyödyntää opettamiseen, viestimiseen, asioiden selittämiseen tai tutkimiseen, datan pitää olla esitettävässä ja hahmotettavassa muodossa. Visualisoitavalla datalla usein halutaan tuoda jotain esille, eli datalla on tarkoitus ja sillä pyritään ilmaisemaan jotain. Visualisointi mahdollistaa datan tarkoituksen muuttumattomuuden, mutta datan ulkomuodollisen esittämisen visualisointina. Visualisoinnissa näin ollen yhdistyy data-analytiikka, graafinen suunnittelu sekä tietotekninen osaaminen.

Datan tietoteknillinen rakenne on tärkeä huomioon otettava seikka datan visualisoinnissa. Visualisoitavan datan pitää olla rakenteellisessa muodossa, eli datalla pitää olla selkeä ennestään asetettu rakenne. Word-dokumentissa oleva data ei ole rakenteellista, mutta Excel-tiedostossa tai CSV-tiedostossa oleva data on. Muita esimerkkejä rakenteellisesta datasta ovat JSON, List ja SQL. Esimerkki osittain rakenteellisesta datasta on HTML-tyyppinen tiedosto. Rakenteellisen eli määrällisen data ja rakenteettoman datan ero ovat datamallituksen puute ja dataan tehtävien hakujen mahdollisuus (Roshani & Albena 2021, 167).

Rakenteellisesta datasta voidaan hakea helposti tiettyjä sanoja tai arvoja, mutta rakenteettomasta datasta näin ei voida tehdä helposti. Tietomalli on datan eri osa-alueita ja elementtejä yhdistelevä rakenne, mikä mahdollistaa datan eri osa-alueiden helpon tarkastelun sekä luo osa-alueille selkeän järjestyksen. Hyvänä esimerkkinä tietomallista on Excelin sarakkeet (columns). Sarakkeet ovat rakenteita, joissa data sijaitsee. Sarakkeiden vierekkäisyys on taas se rakenteen piirre, joka ennalta määrittää datan suhteet datan rakenteessa. Rivit taas erottelevat arvot toisistaan. Rakenteettoman datan lukeminen on mahdotonta suurimmalle osalle konekielistä, mutta se on usein vaivatonta ihmiselle. Yleisesti ottaen voidaankin väittää, että rakenteellinen data on helposti luettavaa tietokoneelle, kun taas rakenteeton data on helposti luettavaa ihmiselle (Gramlich 2020).

Kuva 2 havainnollistaa rakenteellisen ja rakennettaman datan eroja. Rakenteellinen data on järjestelmällistä, järjestyksessä olevaa ja siististi lueteltua dataa, kun taas rakenteettomalta datalta puuttuvat nämä ominaisuudet. Esimerkkeinä rakenteettomasta datasta voidaan pitää videoita, MP3 tyyppisiä äänitiedostoja, artikkelissa olevaa tekstiä ja kuvia. Jotta rakenteetonta dataa voitaisiin visualisoida, pitää siitä muokata rakenteellista. Rakenteettoman datan tyyppistä riippuen tämä saattaa olla helppoa tai mahdotonta. Esikäsittelyn aikana toteutettava datan muokkaus pitää sisällään datan muokkaamisen rakenteelliseksi sekä muita toimenpiteitä. Dataa voidaan siistiä, muokata tai ryhmitellä muiden toimenpiteiden lisäksi. Datan muokkaus usein pitää sisällään datan ul-

komuodon, kirjoitus- tai visuaalisen muodon muokkaamisen. Esikäsittelyyn liittyy myös datatyyppien, ajan, päivämäärien sekä välimerkkien muokkaaminen oikeanlaisiksi.



Kuva 2. Structured vs Unstructured Data – What's the Difference (Pickell, 2018)

Jos data ei ole valmiiksi rakenteellisessa muodossa, pitää sille toteuttaa esikäsittely. Esikäsittelyssä datasta voidaan työkalujen avulla tehdä rakenteellista ja muokata data visualisoimisen mahdollistamiseksi. Datan visualisointeja voidaan toteuttaa kaikesta rakenteellisesta datasta, mutta datan visualisoimiseen käytettävien ohjelmia välillä on eroja siinä, minkälaista rakenteellista dataa ohjelma hyväksyy. Useat visualisointiin tarkoitetut ohjelmat vastaanottavat kaiken muotoista dataa vaivattomasti, mutta yksinkertaisemmilla ohjelmilla monimutkaisen rakenteellisen datan visualisointi voi osoittautua vaikeaksi tai mahdottomaksi. Datan visualisoinnin iso haaste onkin datan rakenteellisuus, koska suurin osa datasta, noin 80–90 %, on rakenteetonta (Davis 2019).

Taulukko 1. Metsäteollisuuden ulkomaankauppa, Tuonti ja vienti kuukausittain. (Luke, 2020)

Maan- osa	Met- säse ktorin tuot- teet	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
EU- ROOP PA	Arvo, milj. e	7899, 6	8071, 4	7597, 2	7686, 2	7718, 4	7764, 0	7571, 1	7670, 9	8294, 5	7832, 7
AASIA	Arvo, milj. e	1419, 7	1696, 2	1786, 8	1948, 5	1862, 6	2044, 0	2144, 8	2620, 9	2828, 5	2653, 3
AF- RIKKA	Arvo, milj. e	482,4	448,4	559,4	567,2	627,8	627,0	603,9	610,1	760,2	689,3
POH- JOIS- AME- RIKKA	Arvo, milj. e	581,6	538,5	587,1	630,6	627,3	772,0	770,1	722,3	846,1	860,8
LATI- NALAI- NEN AME- RIKKA	Arvo, milj. e	315,2	310,9	325,1	338,1	323,2	337,0	285,0	319,6	391,3	385,0
OSEA- NIA	Arvo, milj. e	150,5	195,4	191,8	184,5	160,6	137,0	116,9	138,4	175,4	153,3

Vanha sanonta ”kuva kertoo enemmän kuin tuhat sanaa” ilmaisee datan visualisoinnin päämäärän hyvin. Visualisointi mahdollistaa halutun asian, aiheen tai idean välittämiseksi käyttämällä hyväksi ihmisen näköaistia sekä kykyä visuaaliseen hahmottamiseen. Datan esittäminen visuaalisesti onkin tärkeä keino tiedon, aiheen tai trendien markkinoimiseksi sekä viestimiseksi. Esimerkkinä taulukko 1 kuvaa rakenteellista dataa suomalaisen metsäteollisuuden vuosittaisesta vientimäärästä miljoonina euroina. Taulukko on ladattu Suomen Luonnonvarakeskuksen (LUKE) verkkosivuilta, ja sille on toteutettu esikäsittely Excelissä, missä tarpeeton tieto poistettiin taulukosta. Taulukko 1 kuvaa vuosittaista Metsäsektorin tuotteiden vientiä eri maaosiin vuosien 2010–2019 välillä.

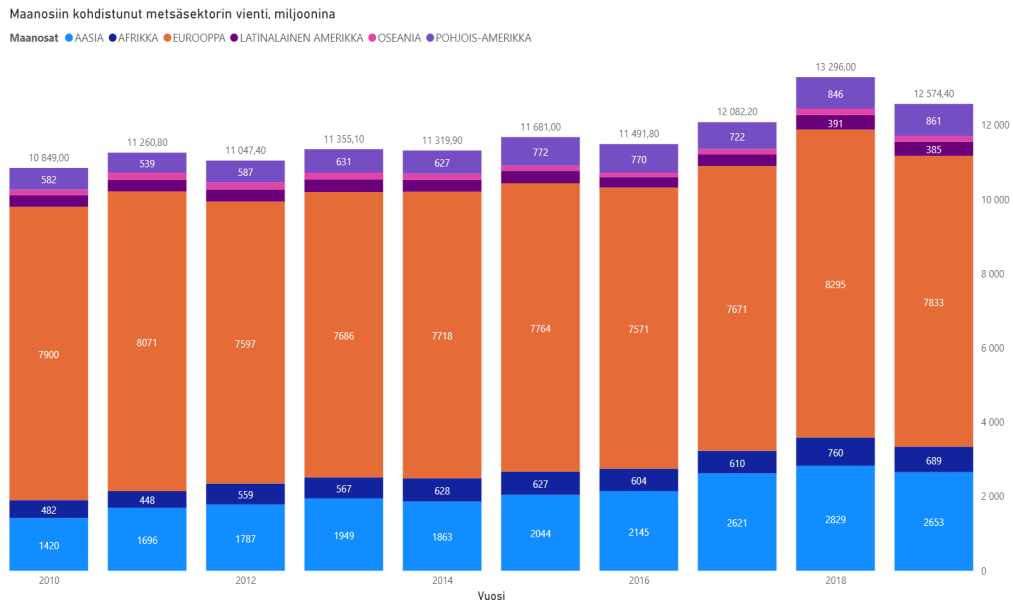
Taulukko on tehokas tapa tarkastella numeerista dataa vuosittain, mutta datasta saattaa olla vaikeaa poimia trendejä tai toistuvia malleja. Taulukko ei anna nopeasti katsottuna käsitystä metsäteollisuuden viennin kokonaiskuvasta eikä ilman paneutumista taulukon numeroihin ole mahdollista tarkastella vuosittaista muutosta helposti. Näiden seikkojen takia taulukon käyttäminen datan ilmaisemiseksi saattaa olla hankalaa. Taulukon heikkous on myös maanosien välisten erojen tarkastelun vaikeus sekä viennin kokonaisluvun puuttuminen. Jotta taulukon avulla voitaisiin tarkastella viennin kokonaismäärää, pitäisi se laskea ja luoda kyseiselle laskutoimitukselle uusi rivi.

Maanosien erojen tarkastelu vaatii taulukon ekstensiivistä tarkastelua ja kokonaiskuvan hahmottaminen on vaikeaa. Datan visualisointi mahdollistaa myös useiden erilaisten tietojoukkojen yhdistelyn sekä näiden tietojoukkojen vertailun. Datan visualisoinnissa voidaan myös tuoda esille matemaattisia malleja, ja visualisoida tiedolle tehtäviä matemaattisia laskuja. Toisin kuin taulukossa visualisoinnin avulla voidaan pureutua tarkemmin datasta esille tuotaviin aihepiireihin, kuten esimerkiksi muutokseen tai yhtäläisyyksiin.

Datan visualisoinnin tehokkaan havainnollistamiseksi, toteutettiin datan visualisointi (kuva 3) taulukosta yksi. Visualisoinnin ideana on tuoda esille datassa tapahtuvaa vuosittaista muutosta käyttämällä minimaalisesti tekstiä ja tuomalla muutoksen trendejä esille kasatun pylväsdiagrammin avulla. Visualisointi tuo nopeasti esille datassa tapahtuvat vuosittaiset muutokset ja trendit. Visualisoinnista voidaan havaita Metsäsektorin tasaisesti kasvanut ulkomaan vienti vuodesta 2010 eteenpäin, mutta 2019 tapahtunut ulkomaan viennin notkahdus alaspäin kertoo omaa tarinaansa metsäteollisuudesta. Kuvasta 3 voidaan myös huomata hitaasti mutta varmasti kasvanut vienti Euroopan ulkopuolelle niin Aasiaan, Pohjois-Amerikkaan kuin Afrikkaan.

Visuaalisina keinoina kuvasta voidaan huomata värien hyödyntäminen muutujien eli maanosien erottelussa sekä kuvailevan tekstin minimaalisuus. Kasatun pylväsdiagrammin korkeus kertoo kokonaisviennin määrän nopeasti. Kokonaisvientiä ei tarvitse laskea kuten taulukossa 1 vaan kokonaisviennin määrän voidaan nopeasti hahmottaa tarkastelemalla pylvään korkeutta. Kuvan 3 kaltaisia visualisointeja voidaan hyödyntää havainnollistavina kuvina.

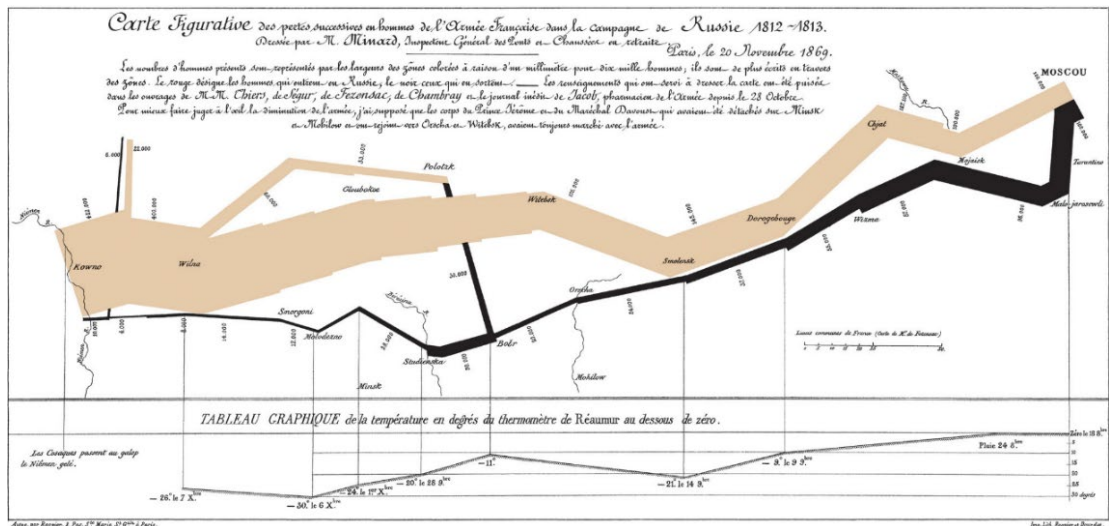
Havainnollistavien kuvien tehtävänä on korostaa datan aihealueita ja näitä ovat kuvassa 3 on maanosien ero viennissä ja vuosittainen muutos.



Kuva 3. Metsäsektorin vienti suomesta vuosittain ja maanosittain. Power BI.

2.2 Datan visualisoinnin lyhyt historia ja nykytilanne

Datan visualisoinnilla on pitkä historia. Datan visualisointia on esiintynyt historiallisesti valtioiden toiminnassa, jotka hyödynsivät datan analyysia verojen keräämisessä tai demografisen datan analysoimisessa (Miyazaki 2020). Visualisointien tavoite on historiallisesti ollut samankaltainen kuin nykyisin eli vaikeasti hahmotettavien asioiden esille tuominen ymmärrettävämmässä muodossa. Hyvänä esimerkkinä datan visualisoimisesta on ranskalaisen siviili-insinöörin virtaus-karttadiagrammi Napoleonin sotaretkestä Venäjälle vuonna 1812 (kuva 4). Diagrammi kuvaa Napoleonin armeijan liikkeitä sotaretken aikana hyödyntäen värejä sekä virtausdiagrammia visuaalisesti, mikä kuvailee Ranskan armeijan lukumääräistä kokoa sen edetessä ja vetäytyessä Venäjältä. Diagrammi korostaa visuaalisesti armeijan koon pienenemistä etenemisen aikana ja kuvasta voidaan helposti ymmärtää sotaretken aiheuttamat tappiot Ranskan armeijalle.



Kuva 4. Kartta Napoleonin sotarekkestä Venäjälle. (Minard, 1869)

Datan visualisointia on historiallisesti rajoittanut tiedon puute, sen huono saatavuus sekä graafiseen suunnitteluun ja toteuttamiseen vaadittu käsillä tehtävä työ. Vaaditun työn laajuuden takia datan visualisointia on hyödynnetty aikakaudellisesti tärkeiden aiheiden kuvailemiseksi eikä samalla tavalla kuin modernina aikakautena, jolloin datan visualisointeja voidaan hyödyntää lähes missä tahansa tilanteessa. Datan visualisoinnin tarkoituksena ovat pääsääntöisesti olleet kytköksissä tieteeseen, navigointiin ja tähtitieteeseen (Chun-Houh ym. 2008, 18). Visualisointien avulla kuvattiin tähtien liikettä taivaalla tai havainnollistettiin taivaankappaleiden liikkeitä kartalla merellä navigointia varten. Modernin datavisualisoinnin aikakauden alkuna voidaan pitää 1800-luvun alkupuoliskoa, jolloin maataloudessa alettiin käyttämään niin pylväs- kuin hajontakaavioita (Chun-Houh ym. 2008, 25). Tätä aikakautta voidaan pitää myös työelämässä käytettävän datan visualisoinnin ensimmäiseksi aikakaudeksi.

Datan visualisoinnin tärkeys hiipui 1800-luvun ajanjakson jälkeen siirryttäessä uudelle vuosisadalle, mutta 1950-luvulta alkaen datavisualisointi koki niin sanotun renessanssin. Niin akateemiset tahot kuin teknologiset mullistukset tietotekniikassa ja koodikielessä kasvattivat datavisualisoinnin tärkeyttä ammattikuntana sekä datan visualisoinnin arvoa liiketoiminnassa. Suuret teknologiset harppaukset, kuten UNIX-käyttöjärjestelmä, graafiset terminaalit sekä 2D- ja 3D-teknologia mahdollistivat data-analytiikan sekä data visualisoinnin tietoteknisillä työkaluilla (Chun-Houh ym. 2008, 40). En-

simmäisiä data-analyysiin ja datan visualisoimiseksi käytettäviä ohjelmia alkoi tulla markkinoille jo 1970-luvulla. SAS ("Statistical Analysis System") oli ensimmäinen datan visualisoinnin mahdollistava ohjelma (Wikipedia 2021). SAS:n avulla dataa kyettiin muokkaamaan, visualisoimaan sekä sille kyettiin toteuttamaan laskutoimituksia tietokoneen avulla.

Datan käsittely, muokkaaminen ja visualisointi kokivat suuria mullistuksia tietotekniikan kehittymisen myötä. Datan visualisointi tuli ohjelmien sekä tietokoneiden mahdollistaman grafiikan myötä suuren yleisön tietoon. Tietokone on ollut mullistava voima datan visualisoinnissa. Tietokone sallii tehokkaan datan visualisoinnin, sillä lähes jokainen vaihe datan visualisoinnin kokonaisuudessa voidaan toteuttaa tietokoneella. Historiallisesti visualisointien tekeminen on ollut käsillä tehtävää työtä, mikä on tuonut visualisointeihin tietyn esteettisyyden (Friendly & Weiner 2020, 6). Erilaiset graafiset ohjelmat antavat käyttäjälle kyvyn toteuttaa visualisointeja ilman historiallisesti tärkeää taiteellista kykyä.

Tietotekniikan yleistymisen ansioista datan analysoimisesta on tullut tärkeä osa liiketoimintaa ja yritysmaailmaa. Internetin ja tietokoneen mahdollistama tehokas datan hankinta, tallentaminen sekä käsittely mahdollistavat lukumääräisesti massiivisten datajoukkojen tarkastelun ja visualisoinnin. Lukumääräisesti suuresta data joukosta on vaikeaa tuoda esille tärkeää tietoa kirjallisesti tai vokaalisesti mikä korostaa data visualisoinnin tärkeyttä (Ossama 2018, 86). Datan lukumäärän takia voi olla siis vaikeaa tuoda datan tärkeitä aihepiirejä esille. Internetin luomat kommunikaation mahdollisuudet sekä tiedonjakamiseen tarkoitetut alustat kuten eurostat.com ovat olleet myös suuressa roolissa kehittämässä data visualisoinnin mahdollisuuksia.

Avoin data eli ilmainen, rakenteellinen ja helposti saatavilla oleva data on kasvattanut visualisointiin käytettävien datan saatavuutta suuresti. Avoin data on tuonut aikaisemmin vaikeasti saatavilla olevan datan julkiseksi sekä helposti saatavaksi. Avoimen datan toimintaperiaatteen on adoptoinut niin Euroopan Unioni kuin Suomen hallitus mikä on mahdollistanut valtiollisten toimielimien tuottaman datan visualisoinnin (Eduskunta 2018). Vaikka datan visualisoin-

nissa nousee esille suurien datajoukkojen visualisoiminen voi datan visualisointeja toteuttaa hyvinkin pienistä datajoukoista tai aiheista. Datan visualisoinnin toimintaperiaatteet ja vaiheet pätevät samalla tavalla niin pienien datajoukkojen kuin isojen datajoukkojen visualisoimiseksi. Suomesta avointa dataa voidaan löytää valtion ja organisaatioiden verkkosivuilta. Avointa dataa opin- näytetyössä etsitään muun muassa Tilastokeskuksen Stat.fi-sivustolta ja Maanmittauslaitoksen verkkosivuilta. Esimerkkinä datalähteestä toimeksian- nolle voisi olla Tilastokeskuksen sivuilta löytyvä datasetti, joka sisältää Suo- men kuntien ja maakuntien populaation iän sekä sukupuolen mukaan.

Tietokoneen mahdollistama datan kerääminen, luominen sekä säilyttäminen on räjähdysmäisesti kasvattanut saatavilla olevan tiedon määrää. Tietokone on sallinut tiedon käsittelyn massiivisissa määrissä, mikä olisi ihmiselle tai ih- misille mahdotonta. Tietokoneen lisäksi datan määrää on kasvattanut inter- netti sekä internetin mahdollistama kommunikaatio. IMB oli arvioinut vuonna 2017, että 90 % kaikesta datasta mitä ihmiskunnan aikana on luotu, luotiin kahden vuoden periodin aikana, 2015–2017 (IBM 2016). Datan määrä on kas- vattanut datan visualisoinnin tärkeyttä ja tehnyt datasta resurssin, jota pysty- tään hyödyntämään yrityksen, valtion tai organisaation toiminnassa. Jotta lu- kumääräisesti massiivista dataa voidaan hyödyntää ihmisten välisessä kans- sakäymisessä datan ulkomuotoa pitää muokata yksinkertaisemmaksi sekä helposti lähestyttävämmäksi datan visualisointia hyödyntäen. Voidaan todeta, että datan lukumäärä ja datan muokkaamisen mahdollistavat työkalut kasvat- tavat datan visualisoinnin tarvetta. Jos lähivuosina koettu tehdyn datan massii- vinen kasvu internetin ja tietoteknisten laitteiden yleistymisen ansioista jatkuu, datan visualisoinnin tarve liikemaailmassa ja sen ulkopuolella kasvaa tasai- sesti internetin ja tietokoneen suosion vanavedessä.

2.3 Datan visualisoinnin toteuttamisen suunnittelu

Datan visualisoinnin suunnittelussa korostuu tarve, tavoite sekä visuaalinen ulkomuoto. Visualisoinnille on aina jonkinlainen tarve. Tarve voi olla esimer- kiksi muutoksen, trendien tai poikkeuksien korostaminen tai jonkin aihealueen esille tuominen markkinointia varten. Tarvetta voidaan pitää visualisoinnin to- teuttamista pääpiirteisesti ohjaavana voimana, joka asettaa visualisoinnille

tiettyjä tekijöitä. Huomioon otettaviksi tekijöiksi visualisointia tehdessä voidaan luokitella seuraavat asiat:

1. Onko visualisointi konseptuaalinen vai datapainotteinen?
2. Visualoinnin tavoite ja kohderyhmä
3. Visualisoinnin ulkomuoto
 - Diagrammi eli kaavio
 - Kontrasti sekä väri
 - Havainnollistava teksti
 - Kaavioroska

Kaikki nämä tekijät vaikuttavat oleellisesti visualisoinnin toteuttamiseen. Visualisointien toteuttaminen kannattaa aloittaa keskittymällä visualisoinnin haluttuihin tavoitteisiin ja visualisointiin vaikuttaviin tekijöihin. Visualisoinnin tavoitteiden ymmärtäminen mahdollistaa visualisoinnin objektiivisen tarkastelun visualisoinnin kohderyhmän näkökulmasta, kun ymmärretään kenelle visualisaatiota ollaan toteuttamassa. On myös helpompaa lähestyä visualisoinnin toteuttamiseksi käytettäviä graafisia menetelmiä objektiivisesti, kun visualisoinnin kohderyhmä ja tavoitteet ovat selkeät. Visualisointeja toteuttaessa pitää aina huomioida, että ne muodostavat kokonaisuuden. Jokaisen visualisoinnin tekeminen vaatii omanlaisensa suunnittelun, resurssit sekä työkalut. Onkin tärkeää, että visualisointien tekemistä lähestytään projektipohjaisesti, ja jokainen projekti on erilainen. Datat visualisoimisessa on tärkeää suunnitella ja pohjustaa omaa tekemistään, ennen kuin visualisointeja aletaan toteuttamaan.

2.3.1 Onko visualisointi konseptuaalinen vai datapainotteinen?

Visualisointia suunniteltaessa on tärkeää kysyä itseltään, onko visualisaatio konseptuaalinen vai datapainotteinen ja yritetäänkö visualisoinnilla selittää tai tuoda esille jotain vai pyritäänkö sillä tutkia jotain (Berinato 2016). Konseptuaalisen ja datapainotteisen visualisaatioiden eroavat toisistaan suuresti. Konseptuaalinen visualisaatio voi olla esimerkiksi visualisaatio yrityksen sijoittajakalenterissa olevista tapahtumista. Konseptuaalinen visualisointi perustuu käsitteen tai idean esittelemiseksi graafisessa muodossa, sen tarvitse sisältää empiiristä dataa (Clark & Matthews 2000, 92). Konseptuaalisen visualisoinnin tärkeä piirre on konseptin eli aiheen tai idean esille tuominen ilman dataa, eli visualisaation käsittäminen ei tällöin vaadi empiiristä dataa. Konseptuaalinen visualisaatio pyrkii opettamaan tai yksinkertaistamaan monimuotoisia ja

vaikeasti hahmotettavia konsepteja, ideoita tai aiheita visuaalisin keinoin. Konseptuaaliset visualisaatiot yleisesti sisältävät enemmän havainnollistavaa tekstiä, ja ne eriyvät datapainotteisista visualisaatioista käytettyjen värien tai diagrammien perusteella. Konseptuaalisia visualisaatioita voidaan havaita paljon koulukirjoissa, lääketieteellisessä kirjallisuudessa ja tutkimuksessa.

Datapainotteinen visualisaatio taas pyrkii visualisoimaan raakaa dataa graafisessa muodossa (Clark & Matthews 2000, 92). Datapainotteinen visualisointi sisältää usein matemaattista tai empiiristä dataa. Datapainotteisten visualisointien lähtökohtana on visualisoida datan trendejä, muutoksia ja poikkeamia eli tutkia datan suhteita keskenään (Clark & Matthews 2000, 92). Esimerkkinä datapainotteisesta visualisaatioista on viivadiagrammi yrityksen liikevaihdon kasvusta vuosittain tai valtion vuosittainen väestön kasvu. Datapainotteisen visualisaation pääpiirre on tuoda esille datassa tapahtuvia muutoksia vertailemalla datajoukon eri muuttujia keskenään.

2.3.2 Visualisoinnin tavoite ja kohderyhmä

Toinen suuri vaikuttava tekijä visualisoinnin suunnittelua varten on visualisoinnin tavoite: miten ja mihin visualisoinnilla pyritään vaikuttamaan? Kuva 5 esittää minkälaisiin kategorioihin visualisointeja voidaan luokitella.

”Hyvä esimerkki tutkivasta visualisoinnista on yliopistoissa tapahtuva akateeminen tutkimustyö. Tai jos firmasi pomo tulee vetämään sinua hihasta ja pyytää selvittämään syyn laahanneisiin myyntilukuihin, niin silloin tutkiva visualisointi auttaa parhaiten selvittämään syitä”, kirjoittaa Kanerva (2016). Tutkivalla visualisoinnilla pyritään edistämään, selittämään tai havainnollistamaan dataa, kun taas selittävän visualisoinnin tavoitteena on visualisoinnin kohderyhmän vaikuttaminen tai heidän auttamisensa johtopäätökseen pääsemiseen.



Kuva 5. TIEDON VISUALISOINTI – PARHAAT KÄYTÄNNÖT (Kanerva 2016)

Aikaisemmassa kappaleessa viitattu visualisoinnin kohderyhmä näyttelee yleisesti myös suurta osaa visualisointien toteuttamisessa. Kohderyhmä pitää olla kirkaasti esillä visualisointia suunniteltaessa, sillä kohdeyleisön teknilliset taidot sekä vaatimukset visualisoinnin suhteen eriävät suuresti. On tärkeää ymmärtää kohdeyleisön tarve sekä se, miten visualisoinnilla voidaan tuoda arvoa kohderyhmälle. (Treehouse Technology Group 2021) Datan visualisointien suunnittelussa kohdeyleisö ja se, mitä visualisoinnilla pyritään saavuttamaan, eli visualisoinnin tarve on tärkeä huomioonotettava seikka. Hyvässä visualisatiossa tarve ja tarkoitus tulevat selkeästi esille. Usein visualisointia suunniteltaessa keskitytään paljon visualisoinnin kohderyhmään eli siihen, kenelle tai keille visualisointia toteutetaan ja kuka visualisointia tarkastelee. Tämä vaikuttaa visualisoinnin ulkomuotoon, visualisoinnissa hyödynnettäviin väreihin sekä siihen, miten havainnollistavaa tekstiä hyödynnetään visualisoinnissa. Kohderyhmä vaikuttaa paljon visualisoinnin lopulliseen ulkomuotoon ja halutun tiedon esille tuomiseen.

Jos visualisoinnin tarkoituksena on esitellä yrityksen johdolle tai vaikkapa sijoittajille yrityksen taloudellisia lukuja, on visualisointia tehdessä kiinnitettävä huomiota enemmän datan yksinkertaiseen esittämistapaan sekä mahdollisimman tehokkaaseen viestimiseen kuin visualisoinnin esteettiseen ulkomuotoon.

Toteutettaessa julkiseen viestintään meneviä data painotteisia visualisointeja visualisoinnin ulkomuoto saattaa olla suuremmassa osassa visualisointia, kuin visualisoinnin sisältö. Kohderyhmänä visualisoinnille voi olla lähes mitä tahansa, yrityksen työntekijöistä yrityksen johtoon tai vaikkapa tietty ikäryhmä tai

sukupuoli. Datan visualisoinnin kohderyhmää mietittäessä voidaan käyttää samoja periaatteita kuin esimerkiksi markkinoinnissa. Jos visualisaation kohderyhmänä ei hallitse tiettyjä teknillisiä taitoja, kannattaa visualisoinnin ulkomuotoa miettiessä keskittyä yksinkertaisuuteen (Pierson 2020).

2.3.3 Visualisoinnin ulkomuoto

Datan visualisoinnin tärkein piirre on visualisoinnin ymmärrettävyys sekä visualisoinnin yhteensopivuus datan kanssa. Jokainen visualisointi on erilainen, joko visualisoitavan datan takia, visualisoinnin kohderyhmän vuoksi tai jonkin muun huomioon otettavan seikan vuoksi (Kirk 2016). Visualisoitava data ei aina sovellu viivadiagrammiksi tai pylväsdiagrammiksi, joten datan rakenteellinen muoto ja se, data edustaa pitää aina ottaa huomioon visualisointeja tehdessä.

Visualisoinnin ulkomuotoa suunniteltaessa nousee esille kolme osa-aluetta, jotka pitää ottaa huomioon visualisointia toteutettaessa. Visualisoinnissa hyödynnetään diagrammeja (Eng. Chart) visualisoinnin väripaletti sekä havainnollistava teksti. Näitä kolmea osa-aluetta yhdisteltäessä kyetään luomaan uniikkeja visualisointeja, joiden avulla visualisaation tekijä kykenee vastaamaan visualisoinnin tarpeisiin niin ulkomuodollisesti kuin ymmärrettävyyden kannalta.

Diagrammi eli kaavio

Visualisoinneissa yleisesti käytetään erilaisia diagrammeja eli kaavioita. Diagrammi tulee kreikan sanasta *diagraphēin*, joka tarkoittaa viivoilla merkitsemistä (etymonline 2021). Diagrammeista yleisempiä ovat viiva- ja pylväsdiagrammit sekä maantieteellistä dataa hyödyntävät kartat. Diagrammit ovat integraalisen osa datan visualisointia, ja ne mahdollistavat datan graafisen kuvaamisen muotojen avulla. Datan visualisoinnissa yhdistyvät diagrammit, kuvaava teksti sekä värien hyödyntäminen muuttujien erottelussa. Dataa visualisoitaessa on tärkeää lähestyä visualisointia kokonaisuutena ja ottaa kaikki osiot huomioon kokonaisuutta luodessa. Visualisaatioista hyödynnettävää diagrammia tai grafiikkaa valittaessa. Yleisesti puhutaan niin sanotusti perusdiagrammeista. Perusdiagrammeihin kuuluvat seuraavat kaaviot:

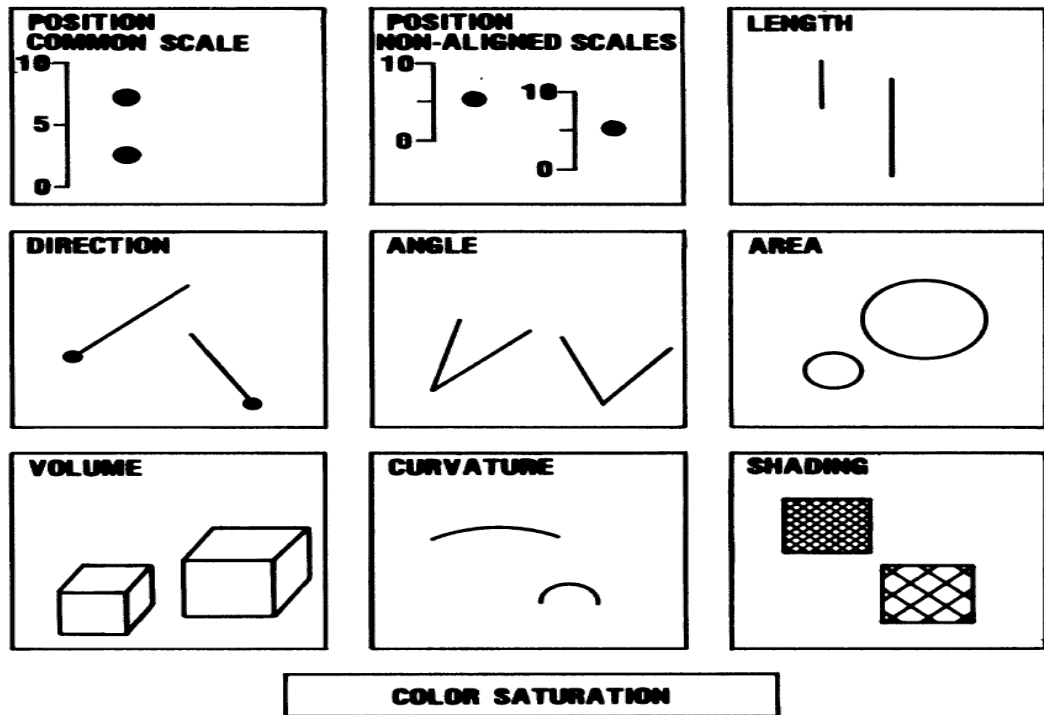
- Pylväsdiagrammi
- Piirakkadiagrammi
- Pistekaavio
- Lämpödiagrammi
- Kartat

Perusdiagrammien etuna on niiden tuottamisen helppous, ja perusdiagrammit ovat suurelle yleisölle helppoja ymmärtää sekä hahmottaa (Magnuson 2016). Vaikkakin perusdiagrammit ovat helposti ymmärrettävissä, tämä ei tietenkään tarkoita sitä, että niiden ymmärrettävyys pätee kaikkiin ihmisiin ja kaikkiin visualisointeihin. Jotta visualisaatio on datasta, aiheesta ja katsojasta riippumatta olisi helposti ymmärrettävissä, pitää visualisaatioita toteutettaessa ottaa huomioon useita ulkomuotoon liittyviä seikkoja sekä lisäksi visualisaation kohderyhmän tietotekniset taidot. Erilaisia visualisoinneissa käytettäviä diagrammeja on satoja, joten visualisointia toteutettaessa pitää aina pyrkiä valitsemaan se yhteensopivin ja objektiivisesti paras diagrammi visualisoimaan dataa. Kuten edellä todettiin, perusdiagrammit ovat yleisesti ymmärrettävyyden kannalta parhaita, mutta tämä ei tietenkään tarkoita sitä, että perusdiagrammi on aina paras vaihtoehto visualisaation diagrammiksi.

Diagrammien etuna on asian hahmottaminen visuaalisesti, esittelytapa sekä suurien datamäärien esittäminen kompaktisti. Diagrammit mahdollistavat monimuotoisen ja suuren datajoukon yksinkertaistaminen vaikkapa pylväskaa- vioksi. Yksinkertaistamisen ideana on ottaa suuri määrä dataa ja muokata siitä helposti ymmärrettävä kokonaisuus diagrammiksi. Tämä yksinkertaistaminen mahdollistaa datan, aiheen tai idean viestinnän, markkinoinnin ja suuremman katsojankunnan. Lähes jokaisesta aihepiiristä, datasetistä tai ajatuksesta voidaan toteuttaa diagrammi. Työkalujen ja mahdollisten diagrammien määrä mahdollistaa lähes kaiken datan tehokkaan visualisoinnin.

Oikeanlaisen diagrammin valitseminen on iso osa visualisaation suunnittelua, ja se asettaa esteettisen pohjan, jonka päälle lopullinen visualisaatio rakennetaan. Oikeanlaisen diagrammin valinta on tärkeää visualisoinnin ulkomuodon kannalta ja oikeanlaisen diagrammin valitsemiseksi pitää ottaa useita seikkoja huomioon, kuten visualisaation kohderyhmä sekä käytettävän datan muoto (Evergreen 2019). Kolmas varteenotettava seikka on ihmisen kyky havaita muotoja ja näin ollen diagrammeja. William S. Clevelandin ja Robert McGillin

vuonna 1984 tekemä tutkimus pyrkii asettamaan erilaisia diagrammeja hierarkiaan perustuen ihmisten kykyyn hahmottaa graafisessa muodossa esitettyä kvantitatiivista dataa. Heidän tutkimuksensa havaitsi, että tietyntyyppiset muodot ovat helpompia hahmottaa kuin toiset (Cleveland & McGill 1984).



Kuva 6. Graafinen käsitys: Teoria, kokeilu, ja toteuttaminen modernin grafiikan kehityksessä (Cleveland & McGill 1984)

Kuva 6 listaa hierarkkisessa järjestyksessä kuvia muodoista sen perusteella miten tehokkaasti ja tarkasti ihminen pystyy niitä hahmottamaan. Kuvan 6 hierarkiassa ylhäällä olevat diagrammit ovat täsmällisesti hahmotettavissa keski-verta ihmiselle Clevelandin ja McGillin (1984) mukaan, kun taas kuvassa alempana on vaikeammin hahmotettavissa olevia muotoja. Kuva 6 kuvaa muotoja, joita käytetään lähes kaikissa diagrammeissa. Esimerkiksi ensimmäinen kuva hierarkiassa, (Kuva 6: Position common scale) kuvaa hyvin piste-kaaviota. Tällaisessa diagrammissa yhteisellä akselilla sijaitsee useita arvoja pisteinä mitkä riippuvat yhteisen akselin lisäksi Y-akselissa olevasta arvosta. Kolmannen kuvan muotoa length eli pituus on taas pylväsdigrammissa sekä viivadiagrammissa käytettäviä muotoja.

Kuvasta 6 voidaan havaita kaikissa perusdiagrammeissa olevia muotoja. Kuvassa 6 hierarkian loppupäässä on "Volume" eli tilavuus, "Curvature" eli kaarevuus, "Shading" eli varjostus ja aivan viimeisenä on "Colour saturation" eli värikylläisyys. Clevelandin ja McGillin (1984) mukaan nämä muodot ovat vaikeasti hahmotettavissa ihmisille. Cleveland ja McGill tulivat siihen johtopäätökseen, että ihmiset kykenevät hahmottamaan erilaisia muotoja paremmin tai huonommin (Cleveland & McGill 1984, 553). Tämä tarkoittaa datan visualisoinnin kannalta, että jotkin diagrammit mahdollistavat tarkemman hahmottamisen kuin toiset. Diagrammia valitessa kannattaakin ottaa huomioon Clevelandin ja McGillin havainnot ihmisen graafisesta havaintokyvystä. Jos visualisaation päätarkoitus on datan tarkka viestintä mahdollisimman suurelle yleisölle, kannattaa valita diagrammi Clevelandin ja McGillin luoman hierarkian yläpäästä.

Kontrasti ja väri

Visualisaation toteuttamiseksi käytettävän diagrammin lisäksi, ulkomuodon suunnittelussa korostuvat myös väri sekä kontrasti. Diagrammin muodon lisäksi ne muodostavat visualisoinnin yleisen ulkomuodon. Yleisen ulkomuodon osa-alueiksi voidaan luokitella väri, kontrasti sekä varjostus. Visualisoinneissa värit ovat merkittävä osa visualisoinnin ulkomuotoa, mutta väri on myös tärkeä työkalu datan ilmaisemiseksi sekä halutun asian korostamiseksi. Visualisoinnissa käytettävä väri on tärkeä osa visualisoinnin ymmärrettävyyden sekä lähestyttävyyden kannalta. Värit tuovat esille visualisaation osa-alueita sekä korostavat tärkeäksi määritettyä dataa. Värien avulla voidaan tuoda esille datan suhteiden eriävyyksiä, trendejä sekä muuttujia visuaalisesti. Datan visualisointia suunniteltaessa kannattaa päättää minkälaista väripalettia hyödyntää visualisoinnin toteuttamisessa. Väreillä on suuri vaikutus visualisoinnin toimivuuteen, ja ne ovatkin olennainen visualisoinnin kokonaisuutta. Värien käyttö visualisoinnissa helpottaa datan tulkitsemista sekä mahdollistaa tarkkojen johtopäätöksien tekemisen datasta (Magnuson 2016).

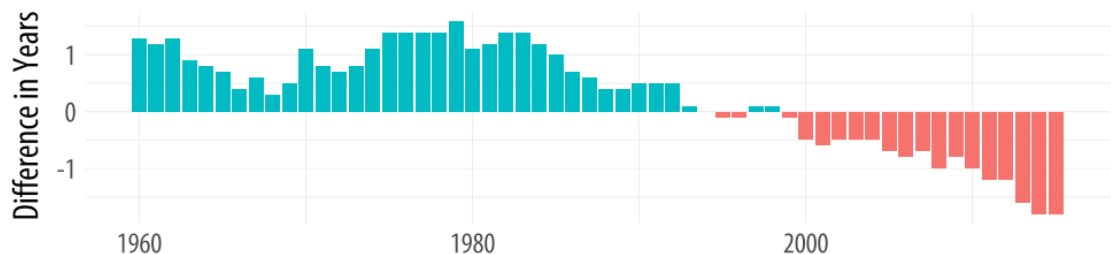
Ehkä tärkein värin käyttötarkoitus visualisoinnissa on värin hyödyntäminen kontrastin luomiseksi (Magnuson 2016). Väriä voidaan hyödyntää datassa olevien eriävyyksien esille tuomiseksi kontrastin avulla. Esimerkiksi isoa lukua ja pientä lukua voidaan visualisoida esittää kahdella eri värillä. Kontrastia

voidaan hyödyntää niin datassa, kuin visualisoinnin yleisessä ulkomuodossa. Kontrastia hyödynnetään myös datan eri elementtejä esille tuotaessa. Kuvassa 3 voidaan havaita, että visualisoitu data on kasatussa pylväässä maanosittain. Värien avulla erilaiset elementit voidaan erotella toisistaan väriä hyödyntämällä. Värittömään tai yksiväriseen visualisaatioon törmääminen on harvinaista, joten voidaankin väittää, että värejä hyödynnetään melkein kaikissa visualisaatioissa.

Kuva 7 tuo omasta mielestäni kontrastin tärkeyden hienosti esille. Kontrasti värien positiivisten ja negatiivisten arvojen välillä luo helposti hahmotettavan trendin. Kuvasta 7 voidaan huomata kontrastin luominen niin värien kuin muodon avulla. Kontrasti mahdollistaa erilaisten dataelementtien tarkastelun erilaisina hyödyntämällä värejä erilaisuuden korostamiseksi. Kontrastin ideana on tuoda esille se, että visualisaatioista olevat elementit ovat erilaisia, toisistaan riippumattomia tai riippuvaisia (Magnuson 2016).

The US Life Expectancy Gap

Difference between US and OECD average life expectancies, 1960-2015



Data: OECD. After a chart by Christopher Ingraham, Washington Post, December 27th 2017.

Kuva 7. Datan visualisointi, käytännöllinen perehdyttäminen (Healy 2018)

Kuva 7 tuo omasta mielestäni kontrastin tärkeyden hienosti esille. Kontrasti värien positiivisten ja negatiivisten arvojen välillä luo helposti hahmotettavan trendin. Kuvasta 7 voidaan huomata kontrastin luominen niin värien kuin muodon avulla. Kontrasti mahdollistaa erilaisten data elementtien tarkastelun erilaisina, hyödyntämällä värejä erilaisuuden korostamiseksi. Kontrastin ideana on tuoda esille, että visualisaatioista olevat elementit ovat erilaisia, toisistaan riippumattomia tai riippuvaisia (Magnuson 2016).

Havainnollistava teksti

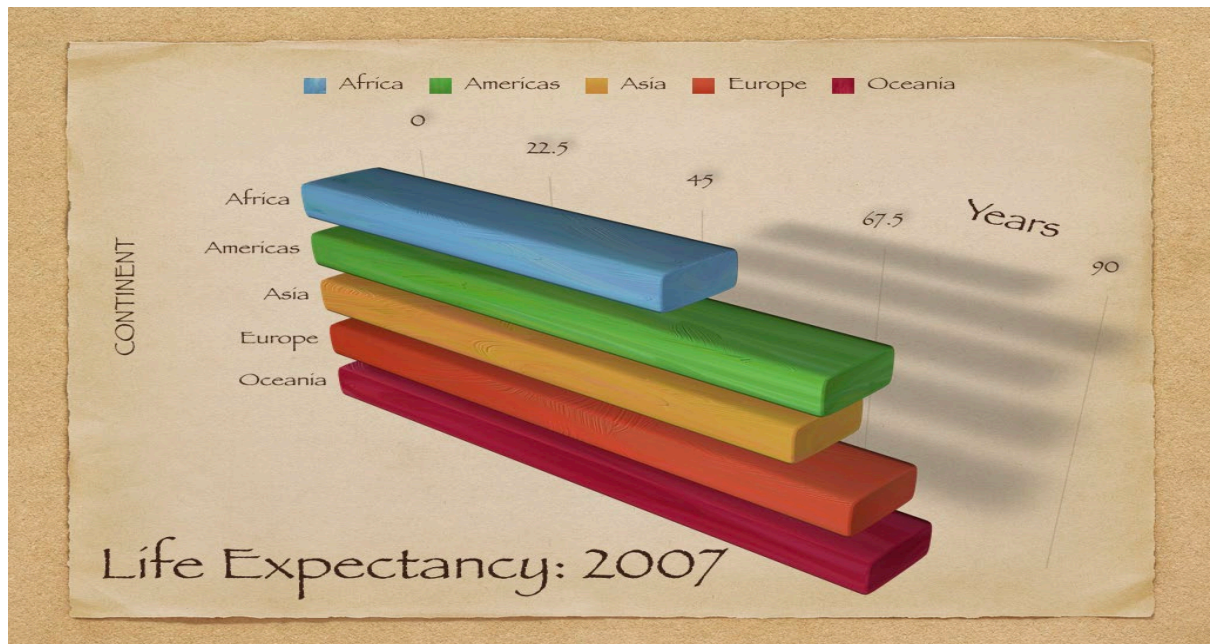
Kolmas visualisaation ulkomuodon osa-alue on teksti. Tekstiä mitä visualisaa-tiossa hyödynnetään, kutsutaan havainnollistavaksi tekstiksi. Havainnollistava teksti voi olla visualisaation otsikko, alaotsikko, leipäteksti tai kuvateksti. Ehkä yleisin tekstimuoto visualisaatioista on otsikko ja selite. Selitettä käytetään vi-sualisoinnin dataelementtien merkkäämiseksi. Teksti visualisaatioista mahdol-listaa visualisoinnin nopean tulkitsemisen ja parantaa visualisaation ymmärret-tävyyttä. Tekstissä on tärkeää muistaa, että vähemmän on enemmän. Liian suuri määrä tekstiä saattaa pilata visualisaation ja tehdä siitä vaikeasti hahmo-tettavan (Krystian 2017).

Visualisaatioissa käytettävä teksti on useasti havainnollistavaa, eikä tekstin ulkoasusta pyritä tekemään visuaalisesti näyttävää. Havainnollistava teksti kannattaa pitää yksinkertaisena sekä yksivärisenä, jottei se kiinnitä katsojan huomiota harhaanjohtavasti. Yksinkertainen fontti sekä mahdollisimman pieni mutta visualisaation sopiva kirjainkoko on tärkeä seikka visualisoinnin ulko-muotoa toteutettaessa (Treehouse Technology Group 2021). Kuten aikaisem-min todettiin, visualisaatiot ovat useiden osa-alueiden luoma kokonaisuus. Käytettävä diagrammi luo pohjan visualisaation ulkomuodolle, kun taas värit ja havainnollistava teksti ovat suuressa roolissa määrittämässä visualisoinnin ulko-muotoa. Vaikka visualisoinnin ulkomuoto perustuukin useaan visuaaliseen keinoon diagrammista väripalettiin, usein voidaan kuitenkin todeta, että vähän on paljon.

Kaavioroska

Visualisointien esteettisyyteen vaikuttaa suuresti myös niin sanottu kaavio-roska (Tuft 2006). Kaavioroska on visualisaatioissa oleva informaatioon tai dataan liittymätön visuaalinen elementti, mikä tekee visualisoinnista sotkuisen ja vaikeasti luettavan. Kaavioroska on yleistä, kun visuaalisointien toteuttaja ei ota visualisoinnin luettavuutta ja yksinkertaisuutta huomioon. Kaavioroska voi olla tarpeettomia muotoja, liiallisuuksiin meneviä väriefektejä tai visualisoinnin datasta riippumaton grafiikka, joka ei rikastuta visualisoinnin sanomaa ollen-kaan. Visualisoinneissa on tärkeää, että se tuo esille mahdollisimman paljon dataa kokoonsa nähden. Kaavioroska on lähes arkkityypillinen vastakohta

kompaktille ja tehokkaalle visualisoinnille, sillä se ei rikastuta visualisointia mitenkään. Jos visualisaatioissa on paljon kaavioroskaa, visuaalisatiosta tulee sotkuinen ja vaikeasti hahmotettava, joka vaikuttaa katsojan kykyyn käsitellä visualisoinnin sanomaa tai lukea siinä olevaa dataa. Voidaankin sanoa, että kaavioroska on niin sanotusti sisällötöntä, joka ei tee visuaalisatiosta tiedon kannalta rikkaampaa.



Kuva 8. Datan visualisointi, käytännöllinen perehdyttäminen (Healy 2018)

Kuva 8 on hyvä esimerkki visualisoinnista, jossa kaavioroska loppujen lopuksi pilaa visualisaation. Visuaalisten efektien määrä ei paranna kuvassa 8 visualisoinnin sanomaa, eikä tee siitä datan osalta rikasta. Kieran Healy kirjoittaa kuvasta 8 seuraavasti: *“Let us start with the bad taste. Figure 1.4 shows [Viittaa kuvaan 8] a chart that is both quite tasteless and has far too much going on in it, given the modest amount of information it displays. The bars are hard to read and compare. It needlessly duplicates labels and makes pointless use of three-dimensional effects, drop shadows, and other unnecessary design features.”* (Healy 2018). Kaavioroskan välttäminen visualisointeja toteutettaessa on tärkeää, sillä sen lisäksi että se pilaa visualisoinnin, se myös ärsyttää katsojaa. Kaavioroskan välttäminen vaatiikin loppujen lopuksi hyvää visuaalista ymmärrystä ja hyvää makua.

Hyvällä maulla kuitenkin pääse pitkälle ja Kieran Haley kirjoittaakin, että” *It is not hard to jettison tasteless junk, and if we look a little harder we may find that the chart can do without other visual scaffolding as well*” (Healy 2018).

2.4 Loppupäätelmä visualisaation ulkomuodosta

Visualisaatiota toteuttaessa kannattaa pitää mielessä vanha sanonta: ”Vähemmän on enemmän”. Tätä sanontaa voidaankin lähestyä ohjeistuksena visuaalisatiota toteutettaessa. Yleisesti ottaen visualisaatioiden toteuttaminen on melko yksinkertaista, mutta virheiden, kuten kaavioroskan, välttämiseksi visuaalisointien tekijä saattaa kääntyä niiden toteuttamiseksi ohjeistuksen puoleen. Tuften (Tuft 2000, 177) mukaan ei ole olemassa koottuja periaatteita täydellisen visualisaation toteuttamiseksi. Hän väittää kuitenkin, että on olemassa useita ohjeita kehittämään arkipäiväisten ja rutiininomaisten eli perusdiagrammi tyyppisten visualisaatioiden laatua. Tällaisia piirteitä ovat seuraavat:

- Oikein valittu malli sekä formaatti
- Käytetään johdonmukaisesti tekstiä, numeroita ja kuvia
- Visualisaatio heijastaa tasapainoa, osuutta, merkityksellisen mittakaavan tunnetta
- näyttää helposti saatavilla olevan yksityiskohtien monimutkaisuuden
- Omaavat laadukkaan kertomuksen, kertovat tarinan datasta
- Ovat piirretty ammattitaitoisesti, teknilliset seikat tehty huolella
- Vältetään sisältöä sisältämätöntä koristelua, mukaan lukien kaavion roskaa.

Yllä mainitut ohjenuorat pyrkivät opastamaan visualisointien tekijää toteuttamaan yksinkertaisempia visualisaatioita, mitkä ottavat kaikki visualisaatioiden osa-alueet huomioon. Kohta ”Käyttävät tekstiä, numeroita ja kuvia yhdessä” korostaa tässä opinnäytetyössäkin esille tuotua lähestymistapaa eli visuaalisointien kokonaisuuden tärkeyttä. Nämä ohjeistuksena pidettävät kohdat ovat omasta mielestäni hyvä käsikirja visualisaatioiden toteuttamiseksi, ja ne helpottavat sekä kehittävät niin aloittelevan kuin kokeneenkin henkilön taitoja visuaalisatiota toteutettaessa.

2.5 Esimerkki datan visualisoinnista

Kappaleessa ”2.4 Visualisoinnin ulkomuoto” käydään läpi visualisoinnissa varteen otettavia seikkoja ja myös esitetään Edward R. Tuften (2000) ohjeita arkipäiväisten visualisointien toteuttamiseksi. Tässä luvussa pyritään toteuttamaan esimerkiksi visualisointi, missä hyödynnetään aikaisemmissa luvuissa käytyjä asioita.

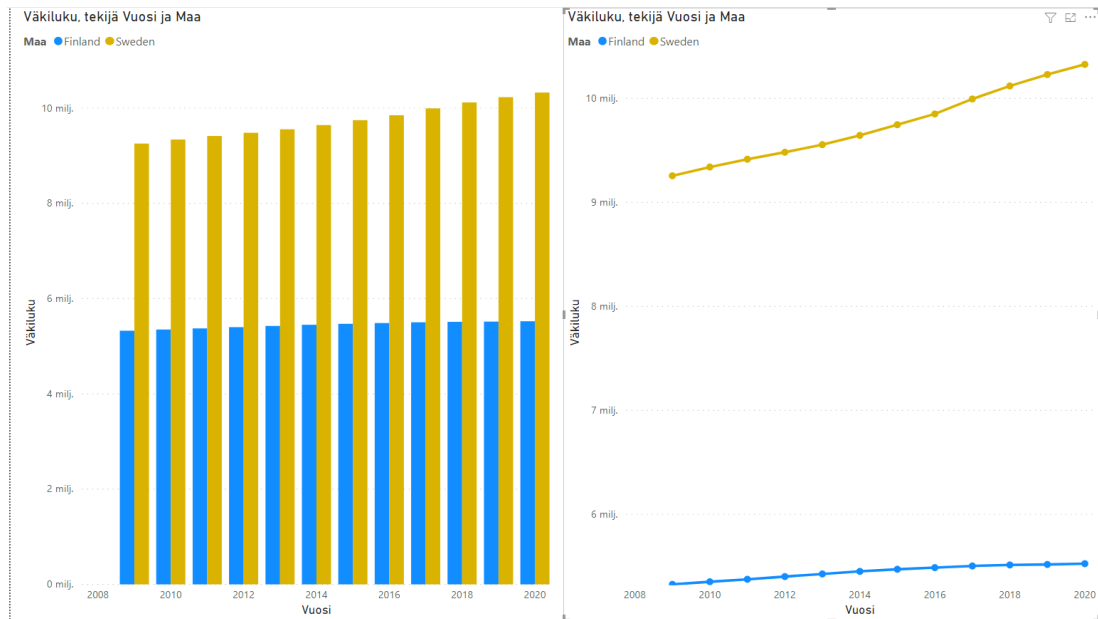
Kuvassa 9 kuvataan Suomen ja Ruotsin väkiluvun muutosta vuosittain hyödyntämällä Power BI-ohjelmaa. Data haettiin Eurostat-nimiseltä sivustolta. Data pitää sisällään Suomen ja Ruotsin vuosittaisen väestönkasvun maittain. Käyttämällä pylväs- ja viivadiagrammeja vuosittaista muutosta voidaan kuvata tehokkaasti. Pylväs- ja viivadiagrammit ovat molemmat niin sanottuja perusdiagrammeja ja ne sopivat todella hyvin kuvaamaan aikavälillä tapahtuvaa muutosta, jos kuvattava data on jatkuvaa sekä numeraalista. Tässä visualisaatiossa käytettävä data on vuosittaisessa muodossa, joten visualisointi onnistuu hyvin käyttämällä vuosia x-akselina.

Vaikka molemmat kuvat ovat diagrammeina erilaisia, ne molemmat yksinkertaistavat datan ymmärrettäväksi visualisaatioksi sekä luovat helposti ymmärrettävän trendin molempien muuttujien eli Ruotsin ja Suomen kohdalla. Tässä visualisoinnissa hyödynnetäänkin datapainotteista visualisaatiota, joka pyrkii selittämään maiden väkiluvun vuosittaista muutosta. Kuvasta 9 voidaan huomata heti, että muuttujia eli Ruotsia ja Suomea kuvataan eri värein.

Diagrammin y-akselin arvo on väkiluku, ja x-akselissa taas on vuosi. Molemmat visualisoinnit kuvaavat dataa hyvin, mutta vuosittaisen muutokset kahden muuttujan välillä pystytään havainnoimaan viivadiagrammista paremmin. Viivadiagrammi myös mahdollistaa Suomen ja Ruotsin väestönkasvun nopeuden eron, ja siitä voidaan myös havaita tulevaisuuden trendejä paremmin.

Kuvista voidaan myös huomata värin ja havainnollistavan tekstin tärkeys. Suomelle ja Ruotsille annettu väri, keltainen ja sininen, ovat muuttujan erottava parametri. Vasemmassa yläkulmassa oleva ”Maa”-legenda taas mahdollistaa

maan tunnistamisen väriparametrin perusteella ilman ylimääräistä tekstiä. Yksinkertainen muoto niin viiva- kuin pylväsdiagrammissa myös mahdollistaa datan esille tuomisen ilman visualisaatioita roskaavia datapisteitä, tekstiä tai muita hyödyttömiä visuaalisia efektejä. Kuvan 9 visualisaatioissa onkin pyritty toteuttamaan mahdollisimman yksinkertainen visualisaatio hyödyntämällä Tuf-ten (2000) oppeja.



Kuva 9. Suomen ja Ruotsin väkiluvun muutos. Power BI

Kuvassa 9 esitetään kokonaisluvullista dataa eli miljoonia sekä tuhansia. Visualisaatioiden toteuttamiseksi pitää ottaa aina huomioon datan rakenne. Jos datassa olevat arvot ovat esimerkiksi prosentuaalisia, pylväs- tai viivadiagrammi saattaa olla huono vaihtoehto datan visualisoimiseksi. Muita huomioon otettavia seikkoja ovat visualisaatioiden muuttujien lukumäärä, interaktiivisuus sekä tarpeellisten arvojen lukumäärä. Kuva 9 tuo esille hyvin visualisaatioiden oleellisimman piirteen, eli eri osa-alueiden tehokkaan yhdistämisen. Niin diagrammi, väripaletti sekä havainnollistava teksti ovat kaikki tärkeässä roolissa hyvän kokonaisuuden luomiseksi. Myös kaavioroska on minimissään. Visualisaatioiden ulkomuotoa miettiessä onkin tärkeää huomioida kaikki ulkomuodolliset sekä esteettiset seikat. Kuvasta 9 voidaan myös helposti tunnistaa visualisoinnin tavoite eli populaation muutoksen kuvaaminen kahdessa valitussa maassa. Hyvä visualisaatio onkin lopulta melko yksinkertainen kokonaisuus, ja hyvän visualisaation ei tarvitse olla monimutkainen.

3 TOIMEKSIANTAJA

Opinnäytetyön toimeksiantaja on Smart & Lean Hub Oy (S&L), joka on lahtelainen tutkimuksen, kehityksen ja innovaatioalan (TKI) pienyritys. Yrityksessä työskentelee tällä hetkellä kolme ihmistä, joista opinnäytetyöntekijä on yksi. S&L toimii Euroopan Unionin rahoittamissa Horizon2020 TKI-projekteissa, toteuttaa kansallisia pienprojekteja ja toimii ELY:n palveluntarjoajana pk-yrityksille.

S&L toimii konsortion partnerina EU:n Horizon2020-ohjelman rahoittamassa PoliRural-projektissa. PoliRural pyrkii parantamaan Euroopan laajuisesti kahdessatoista pilotissa maaseudun houkuttelevuutta. Näiden pilottialueiden tehtävänä on tutkia ja testata poliittisten toimien vaikutusta pilottialueiden houkuttelevuuteen. Kanta- ja Päijät-Hämeen maakunnista muodostuva Hämeen alue on Suomen pilottialue. Yksi PoliRuralin tavoite on tuoda näkyvyyttä niin hankkeen toimintaan kuin maaseudun nykytilanteeseen. Tästä syystä toimeksiantaja halusi useita interaktiivisia visualisaatioita Hämeen alueesta. Niitä voidaan hyödyntää yrityksen viestinnässä. S&L halusi mielenkiintoisia ja interaktiivisia visualisaatioita sellaisessa muodossa, että niitä voidaan vaivattomasti jakaa yrityksen verkkoviestinnässä, Twitterissä ja LinkedInissä.

Yritys haluaa nostaa Hämeen alueen demografista статистиikkaa esille sosiaalisessa mediassa. S&L haluaa myös saada Hämeen alueen demografiselle ja taloudelliselle muutokselle lisää näkyvyyttä. Visualisointien pitää kuvata Hämeen alueen tai koko Suomen demografista ja taloudellista muutosta yli kymmenen vuoden ajalta.

4 VISUALISOINTIEN TOTEUTTAMINEN

Opinnäytetyön aikana toteutettiin 5 erilaista visualisaatiota avoimesta datasta. Data pääsääntöisesti haettiin Tilastokeskuksen sivuilta, mutta myös muita lähteitä hyödynnettiin esimerkiksi Maamittauslaitoksen dataa. Visualisointeja tehtiin kolmella erilaisella ohjelmalla: Python ohjelmointikielellä, Power BI-visualisointi ohjelmalla sekä Mapbox-karttatyökalun avulla. Seuraavassa 5 luvussa kuvailaan, miten ja minkälaisia visualisointeja luotiin. Visualisointien toteuttaminen eteni prosessuaalisesti. Visualisaatioita laadittiin vaiheittain ja pyrittiin

toteuttamaan kaikki visualisaatiot hyödyntämällä samaa prosessikuvausta (kuva 1, s. 7).

4.1 Python visualisointi – Eläkeläiset ja väkiluku kunnittain Päijät-Hämeessä

Python visualisointi ”Eläkeläisten prosentuaalisesta muutoksesta vuosittain Päijät-Hämeen maakunnassa” pyrkii nostamaan esille pienemmissä kunnissa tapahtuvaa vanhusen lukumäärän nopeaa kasvua. Idea visualisoinnin toteuttamisesta tästä aihepiiristä syntyi esimiehen kanssa käydyn keskustelun jälkeen väestön vanhenemisesta.

Visualisoinnissa käytetty data löytyi helposti Tilastokeskuksen tilastotietokannasta ([Linkki tietokantaan](#)). Visualisoinnissa käytettävä data oli valmiiksi rakenteellisessa muodossa ja datan esikäsittely oli tämän tietokannan avulla tehty helpoksi. Visualisoinnin toteuttamiseksi käytin kolmea eri tietokenttää. Suomen väkiluku kunnittain, vuosittain sekä Suomen kuntien eläkeläisten prosentuaalinen määrä vuosittain. Jotta tämä data olisi rakenteellisesti valmiista visualisointia varten, piti ainoastaan ladatusta XLSX-tiedostosta poistaa ylimääräinen lähdetieto.

Esikäsittelyn jälkeen tiedosto luettiin Pythoniin Pandas-kirjaston avulla. Kuva 10 näyttää miltä Pythonin Pandas-kirjaston avulla luotu datakehys näyttää. Datakehysten avulla voin helposti muokata sekä hyödyntää datakehyksessä olevaa dataa visualisoinnissa tai datan muokkauksessa. Pythonissa käytin Plotly-nimistä kirjastoa visualisointien toteuttamiseksi.

	Kunta	Vuosi	Taajama-aste, %	Väkiluku	Väkiluvun muutos edellisestä vuodesta, %	Alle 15-vuotiaiden osuus väestöstä, %	15-64-vuotiaiden osuus väestöstä, %	Yli 64-vuotiaiden osuus väestöstä, %	Ruotsinkielisten osuus väestöstä, %	Ulkomaan kansalaisten osuus väestöstä, %	Alueella olevien työpaikkojen lukumäärä
0	Asikkala	1995	62.3	8782	0.2	18.1	64.4	17.5	0.3	0.6	2745
1	Asikkala	1996	62.7	8728	-0.6	17.8	64.4	17.8	0.2	0.6	2646
2	Asikkala	1997	62.2	8694	-0.4	17.5	64.8	17.7	0.3	0.7	2707
3	Asikkala	1998	62.5	8609	-1.0	17.2	64.9	17.8	0.3	0.8	2773
4	Asikkala	1999	62.4	8636	0.3	17.3	64.8	17.9	0.3	0.8	2668

Kuva 10. Data pythonissa

Plotly mahdollistaa interaktiivisten visualisointien toteuttamisen sekä niiden lataamisen verkkosivustolle, jotta niitä voidaan jakaa interaktiivisina eteenpäin hyödyntämällä ainoastaan hyperlinkkiä.

Käytin alla olevaa koodia visualisoinnin luomiseksi. Koodista voidaan huomata X – ja Y-akselissa käytettävän tiedon sarakkeen nimi sekä visualisoinnin ulko-
muodon muokkaamisessa käytetty koodi.

```
fig = px.line(df, x="Yli 64-vuotiaiden osuus väestöstä, %", y="Väkiluku",
color="Kunta",line_shape="spline", render_mode="svg", title="Population and
percentage of elderly by year",
line_group="Kunta", hover_name="Vuosi", text="Vuosi")
fig.update_layout(title='Percentage of elderly by year and county in Päijät-
Häme',
xaxis_title='Percentage of elderly',
yaxis_title='Population')
fig.update_traces(mode='lines + markers', marker=dict(
size=9,
line=dict(
width=2))),
fig.update_layout(xaxis_ticksuffix = '%')
```

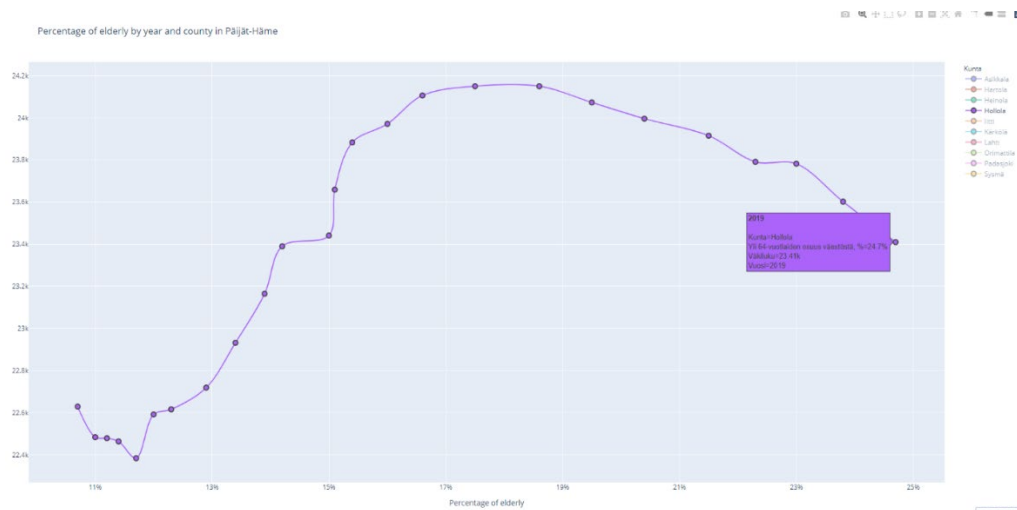
Esimerkkinä kohta "fig.update_layout" - komennon avulla visualisoinnin teksti-
kenttiä voidaan muokata sekä "fig.update_layout(xaxis_ticksuffix = '%')"
avulla x-akselin arvoille annetaan prosenttimerkintä arvojen loppuun. Visuali-
sionissa käytettävä diagrammi on pistekaavio, koska X- ja Y- arvot ovat jatkuvia.
Tämä mahdollistaa trendin ja muutoksen esille tuomisen pistekaavion avulla.
Visualisaation värimaailma on taas mahdollisimman yksinkertainen, mutta
kunnat esitetään eri värein. Tämä tekee visuaalisatiosta helposti luettavan
(kuva 11).



Kuva 11. Python visualisointi

Visualisoinnin avulla voidaan vaivattomasti tarkastella kunnan väkiluvun sekä eläkeläisten määrän kasvua. Pistekaavio tuo esille väkiluvun sekä eläkeläisten joka vuosi kasvavan määrän. Visualisaatioissa vuodet eivät ole ajallisessa järjestyksessä. Tämä tarkoittaa sitä, että visualisaatioissa olevat datapisteet eivät etene vasemmalta oikealla ajallisessa järjestyksessä. Jokainen datapiste eli pallo kuvaa yhtä vuotta ja pallon väri viittaa taas kuntaan. Viivan pituus taas kuvaa muutoksen suuruutta.

Jos visualisoinnissa halutaan tarkastella yhtä kuntaa, voidaan kahdesti klikata oikealla olevasta legendasta haluttua kuntaa. Jos halutaan nähdä tietyn vuoden X – ja Y-akselin arvot, pitää hiiren kursori viedä pallon päälle. Kuvassa 12 voidaan nähdä miltä visualisointi näyttää, kun visualisointi suodatetaan yhden kunnan perusteella sekä hiiren kursori viedään yhden pallon päälle.



Kuva 12. Suodatettu kuva

Kuvista 11 ja 12 voidaan huomata, että eläkeläisten määrä kasvaa melko ra-
justi vuosittain joissain kunnissa. Kuvasta voidaan myös huomata väkiluvun
minimaalinen kasvu.

Linkki visualisaation, <https://chart-studio.plotly.com/~KalleKK/31/#>

4.2 Power BI visualisointi – puukartta Hämeen alueen yritysten toimi- paikoista

Tämän visualisoinnin tarkoituksena on esittää Hämeen alueen maakuntien,
Päijät – ja Kanta-Häme, yritysten taloudellista dataa toimialojen mukaan.
Wikipedian mukaan toimiala Suomessa on ”*Taloudellista toimintaa harjoitta-
van yrityksen tai ammatinharjoittajan pääasiallisen toiminnan perusteella mää-
rättyvä elinkeinoluokka.*” (Wikipedia, 2018). Suomessa toimialat luokitellaan
numeraalisen toimialaluokituksen mukaan. Toimialaluokitusta ylläpitää suo-
messä Tilastokeskus.

Idea visualisoinnin toteuttamiseksi syntyi halusta tarkastella Hämeen alueen
yritysten taloudellisia tunnuslukuja maakunnan sekä toimialan perusteella.
Tieto täydentää opinnäytetyön toimeksiantajan näkemystä maakunnan elinvoi-
maisuudesta. Visualisoinnissa käytettävä data löytyi Tilastokeskuksen StatFin
tilastotietokannasta ([Linkki tietokantaan](#)). Data on valmiiksi rakenteellisessa
muodossa. Data sisältää neljä erilaista taloudellista tunnuslukua (Yritysten

toimipaikat, Yritysten toimipaikkojen henkilöstö, Yritysten toimipaikkojen liikevaihto sekä yritysten toimipaikkojen liikevaihto/henkilö) eroteltuna toimialan, mittausvuoden sekä maakunnan perusteella. Dataa tietokannasta löytyy vuodesta 2013 eteenpäin, aina vuoteen 2019 asti. Tämän datan rakenteesta sekä arvojen ulkomuodosta johtuen datalle piti toteuttaa esikäsittely. XLSX-Excel tiedoston lataamisen jälkeen tehtiin seuraavat muutokset dataan Excel-ohjelman avulla. Muutettiin vuoden ajallista muotoilua vastaamaan PP.KK.VV muotoa, sillä visualisoinnin toteuttamiseen käytetty ohjelma Power BI vaatii kyseisen muodon ajallisen datan hyödyntämiseksi. Poistettiin myös tyhjiä rivejä sekä lähdetiedot. Näiden toimenpiteiden lisäksi tehtiin kaksi uutta saraketta mihin erittelin maakunnan nimen sekä mittausvuoden. Kuvasta 13 voidaan nähdä miltä taulukko näyttää ennen esikäsittelyn toteuttamista. Kuvasta 13 voidaan myös huomata, että kaikissa sarakkeissa ei ole arvoja, vaan tyhjiä (null) arvoja kuvaavat pisteet. Esikäsittelyn aikana korvattiin tyhjät arvot luvulla nolla, sillä Power BI ei osaa lukea tyhjiä arvoja kuvaavia pisteitä oikein.

Yritysten toimipaikat toimialoittain ja maakunnittain muuttujina Vuosi, Maakunta, Toimiala (TOL 2008) ja Tiedot

	Yritysten toimipaikat (km)	Yritysten toimipaikkojen henkilöstö (ntv)	Yritysten toimipaikkojen liikevaihto (1 000 euroa)	Yritysten toimipaikkojen liikevaihto/henkilö (1 000 euroa)
2013				
MK05 Kanta-Häme				
A Maatalous, metsätalous ja kalatalous (01-03)	2 596	1 995	44 592	22
01 Kasvinviljely ja kotieläintalous, riistatalous ja niihin liittyvät palvelut	2 152	1 612
02 Metsätalous ja puunkorjaus	439	381	43 799	115
03 Kalastus ja vesiviljely	5	2	792	344
B Kaivostoiminta ja louhinta (05-09)	24	93	27 907	299
07 Metallimalmin louhinta
08 Muu kaivostoiminta ja louhinta	21
09 Kaivostoimintaa palveleva toiminta	3
C Teollisuus (10-33)	860	11 552	3 266 499	283
10 Elintarvikkeiden valmistus	79	1 958	674 502	345
11 Juomien valmistus	3
12 Tupakkatuotteiden valmistus
13 Tekstiilien valmistus	28	77	11 696	152
14 Vaatteiden valmistus	34	123	3 171	26
15 Nahan ja nahkatuotteiden valmistus	6
16 Sahatavaran sekä puu- ja korkituotteiden valmistus (pl huonekalut), olki- ja punontatuotteiden valmistus	78	620	168 488	272
17 Papein, paperi- ja kartonkutuotteiden valmistus	8	602	182 775	304
18 Painaminen ja tallenteiden jäljentäminen	35	346	57 123	165
19 Koksain ja jalostettujen öljytuotteiden valmistus
20 Kemikaalien ja kemiallisten tuotteiden valmistus	11	97	35 122	363

Kuva 13. Yritysten toimipaikat taulukko StatFin tilastotietokannassa

Esikäsittelyn jälkeen ladattiin tiedoston Power BI:hin. Kuva 14 näyttää miltä data näyttää Power Bissä esikäsittelyn jälkeen.

Vuosi	Maakunta	Toimiala (TOL 2008)	Yritysten toimipaikat (lkm)	Yritysten toimipaikkojen henkilöstö (htv)	Yritysten toimipaikkojen liikevaihto	Yritysten toimipaikkojen liikevaihto/henkilö (1 000 euroa)
maanantai 1. tammiukuuta 2018	MK07 Päijät-Häme	C Teollisuus (10-33)	1025	13365	€ 4 150 371 000	311
sunnuntai 1. tammiukuuta 2017	MK07 Päijät-Häme	C Teollisuus (10-33)	1040	13427	€ 3 911 869 000	291
tiistai 1. tammiukuuta 2019	MK07 Päijät-Häme	C Teollisuus (10-33)	1017	13193	€ 3 909 341 000	296
perjantai 1. tammiukuuta 2016	MK07 Päijät-Häme	C Teollisuus (10-33)	1061	13247	€ 3 734 898 000	282
torstai 1. tammiukuuta 2015	MK07 Päijät-Häme	C Teollisuus (10-33)	1084	13479	€ 3 643 399 000	270
kesäviikko 1. tammiukuuta 2014	MK07 Päijät-Häme	C Teollisuus (10-33)	1105	13629	€ 3 618 407 000	265
tiistai 1. tammiukuuta 2013	MK07 Päijät-Häme	C Teollisuus (10-33)	1127	14878	€ 3 607 554 000	242
maanantai 1. tammiukuuta 2018	MK05 Kanta-Häme	C Teollisuus (10-33)	772	10755	€ 3 503 906 000	326
tiistai 1. tammiukuuta 2019	MK05 Kanta-Häme	C Teollisuus (10-33)	765	10416	€ 3 356 607 000	322
sunnuntai 1. tammiukuuta 2017	MK05 Kanta-Häme	C Teollisuus (10-33)	794	10587	€ 3 293 948 000	311
tiistai 1. tammiukuuta 2013	MK05 Kanta-Häme	C Teollisuus (10-33)	860	11552	€ 3 266 499 000	283
maanantai 1. tammiukuuta 2018	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1958	7501	€ 3 192 215 000	346
kesäviikko 1. tammiukuuta 2014	MK05 Kanta-Häme	C Teollisuus (10-33)	835	10798	€ 3 142 488 000	291
tiistai 1. tammiukuuta 2019	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1911	7438	€ 3 129 677 000	345
tiistai 1. tammiukuuta 2013	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	2192	8454	€ 3 070 405 000	321
sunnuntai 1. tammiukuuta 2017	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	2025	7684	€ 3 065 649 000	339
perjantai 1. tammiukuuta 2016	MK05 Kanta-Häme	C Teollisuus (10-33)	807	10664	€ 3 022 570 000	283
torstai 1. tammiukuuta 2015	MK05 Kanta-Häme	C Teollisuus (10-33)	818	10881	€ 3 021 042 000	278
perjantai 1. tammiukuuta 2016	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	2031	7816	€ 2 991 187 000	325
kesäviikko 1. tammiukuuta 2014	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	2147	8067	€ 2 984 188 000	302
torstai 1. tammiukuuta 2015	MK07 Päijät-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	2059	7940	€ 2 909 415 000	329
tiistai 1. tammiukuuta 2019	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1691	7569	€ 2 610 062 000	421
sunnuntai 1. tammiukuuta 2017	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1797	7515	€ 2 545 800 000	399
maanantai 1. tammiukuuta 2018	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1696	7336	€ 2 537 399 000	426
perjantai 1. tammiukuuta 2016	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1808	7715	€ 2 510 399 000	383
torstai 1. tammiukuuta 2015	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1841	7601	€ 2 498 122 000	366
tiistai 1. tammiukuuta 2013	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1904	7640	€ 2 448 602 000	363
kesäviikko 1. tammiukuuta 2014	MK05 Kanta-Häme	G Tukku- ja vähittäiskauppa, moottoriajoneuvojen ja moor	1856	7649	€ 2 311 571 000	370
tiistai 1. tammiukuuta 2013	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	1124	5139	€ 1 505 618 000	259
sunnuntai 1. tammiukuuta 2017	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	1036	4262	€ 1 469 462 000	292
maanantai 1. tammiukuuta 2018	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	980	4021	€ 1 464 716 000	327
kesäviikko 1. tammiukuuta 2014	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	1100	4830	€ 1 454 789 000	254
tiistai 1. tammiukuuta 2019	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	942	4128	€ 1 448 421 000	315
perjantai 1. tammiukuuta 2016	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	1016	4353	€ 1 431 411 000	270
torstai 1. tammiukuuta 2015	MK07 Päijät-Häme	47 Vähittäiskauppa (jäl moottoriajoneuvojen ja moottori	1044	4541	€ 1 419 678 000	274
tiistai 1. tammiukuuta 2019	MK07 Päijät-Häme	F Rakentaminen (41-43)	1656	6132	€ 1 282 250 000	395

Kuva 14. Esikäsitelty data Power BI ohjelmassa

Toimiala (TOL 2008)-sarakkeen avulla kyetään luomaan visualisoinnin, mikä esittää Hämeen-alueella toimivien yritysten taloudellisia lukuja mittausvuoden sekä toimialan perusteella. Kuvasta 14 voidaan huomata neljä erilaista taloudellista tunnuslukua:

- Yritysten toimipaikat
- Yritysten toimipaikkojen henkilöstö
- Yritysten toimipaikkojen liikevaihto
- Yritysten toimipaikkojen liikevaihto/henkilö (1000 euroa)

Koska halutaan luoda visualisoinnin minkä avulla kyetään tarkastelemaan kaikkien toimialojen vuosittaista dataa, ei pystytä hyödyntämään perusdiagrammeja (katso 2.3.1) visualisoinnin toteuttamiseksi. Toimialoja on suuri määrä, toimialojen taloudellisia tunnuslukuja on neljä sekä vuosia on kuusi. Tämä tekisi visualisoinnista liian sekavan, jos hyödyntäisin jotakin perusdiagrammeista. Jotta visualisointi olisi helposti ymmärrettävissä, luodaan painikkeen, minkä avulla diagrammin esittämää arvoa voidaan vaihtaa interaktiivisesti painikkeen avulla.

Koska halutaan esittää neljää erilaista arvoa eli tässä tapauksessa taloudellista lukua, joudutaan luomaan Power BIn DAX-koodikielellä uuden taulukon, jonka avulla pystytään luomaan interaktiivisen painikkeen. Taulukko (kuva 15)

sisältää viittauskoodin ja halutun muuttajan eli taloudellisen arvon sarakkeen nimen.

Nimi	code
Yritysten toimipaikat LKM	1
Yritysten toimipaikkojen henkilöstö	2
Yritysten toimipaikkojen liikevaihto/henkilöt (1000 euroa)	3
Yritysten liikevaihto	4

Kuva 15. Taulukko painiketta varten

Alla oleva koodi hyödyntää kuvan 15 taulukkoa viittauskoodin eli sarakkeen "code" perusteella. Hyödyntämällä taulukkoa sekä DAX-koodia, pystytään luomaan Power BI:n slicer-työkalun avulla painike, mikä vaihtaa interaktiivisesti klikattaessa visualisoinnin muuttujaa eli taloudellista arvoa. Vasemmalla oleva kuva 16 näyttää miltä painikkeet näyttävät visualisoinnissa. Painikkeiden luomisen jälkeen luodaan itse visualisaation.

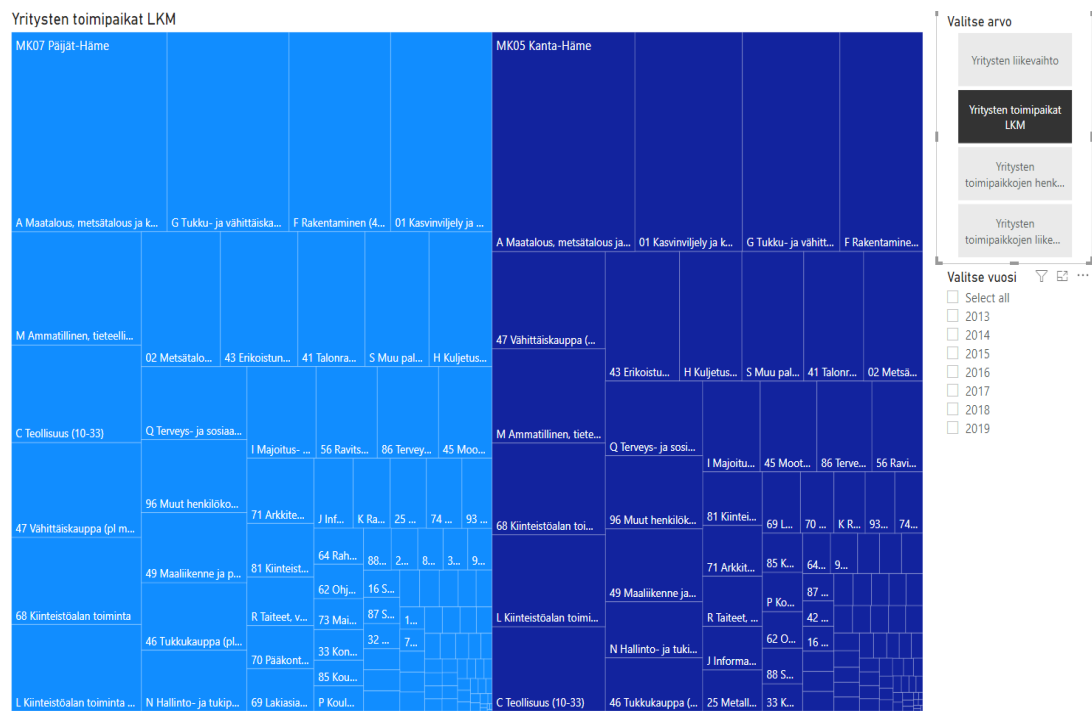
```
Valittu arvo =
SWITCH(
    [Selected measure],
    1,[Yritysten toimipaikat lkm],
    2,[Yritysten toimipaikkojen henkilöstö],
    3,[Yritysten toimipaikkojen liikevaihto/henkilö 1 000 euroa],
    4,[Liikevaihto])
```



Kuva 16. Painike Power BI:ssä

Koska toimialoja on niin monta (101 kpl) ja halutaan esittää toimialoja kolmen eri muuttujan perusteella: taloudellinen arvo, maakunta sekä vuosi, päädyttiin käyttämään puukaaviota. Puukaavio näyttää kaikki toimialat samaan aikaan kahtena diagrammina maakunnan perusteella. Visualisoinnissa näytetään va-semmalla Päijät-Häme ja oikealla Kanta-Häme.

Visualisoinnin tarkastelijalla haluaa myös muokata visualisaatiota interaktiivisesti vuoden perusteella. Tämän vuoksi tehtiin toinen painike slicer-työkalun avulla. Lopullisen visualisoinnin toteuttaminen onnistuu Power Bln puukaavio diagrammia hyödyntämällä. Puukaavion toteuttaminen vaatii kolme erilaista muuttujaa. Ryhmä, informaatio sekä arvo. Ryhmä on tässä visualisaatiossa maakunta, informaatio on toimiala sekä arvo on kuvassa 16 näkyvä valittu painike. Näiden kolmen muuttujan lisäksi puukaavio tässä tapauksessa vaatii sen asetusten muuttamista. Luotiin uusi titteli visualisoinnille, muutettiin tekstin fonttia, kokoa sekä väri. Annettiin myös ryhmä muuttujalle uudet vaalean – ja tummansiniset värit kuvaamaan maakuntia eri värein.



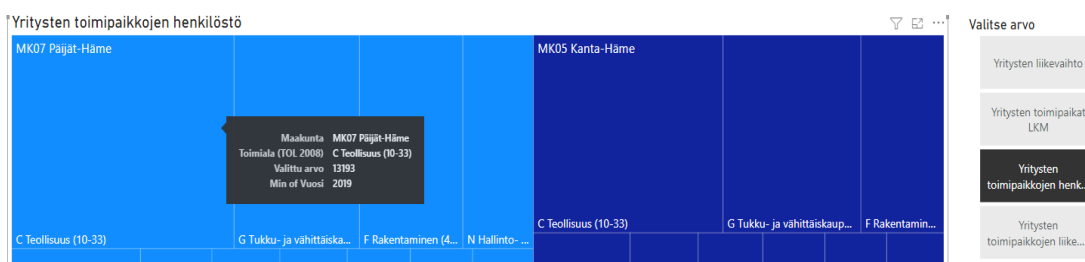
Kuva 17. Visualisointi Power Bissä.

Kuva 17 näyttää lopullisen visualisoinnin. Puukaavion oikealla olevat kaksi painiketta mahdollistavat visualisoinnin interaktiivisen muokkaamisen taloudellisten arvojen sekä vuoden perusteella. Itse puukaavio esittää toimialat koon

ja toimialan nimen perusteella. Kuvasta 17 voidaan myös huomata, että puukaavio erittelee toimialat maakunnan perusteella hyödyntämällä sijaintia sekä värejä. Painamalla oikealla olevista painikkeista, puukaavio muuttuu interaktiivisesti ja nopeasti.

Kuvasta 17 nähdään, että oikealla olevasta vuosipainikkeesta ei ole valittu vuotta. Tämä tarkoittaa sitä, että kaikkien mittausvuosien (2013–2019) data on otettu visualisoinnissa huomioon. Kuvasta 17 voidaan todeta, että molemmissa maakunnissa maa - ja metsätaloudessa on eniten toimipaikkoja, kun tarkastellaan puukaaviossa olevien laatikoiden suuruutta. Taloudellisten arvojen suuruutta esitetään laatikon koon sekä laatikon sijainnin perusteella. Se laatikko eli toimiala, missä on suurin arvo, on ensimmäisenä ylävasemmalta katsottuna. Laatikot puukaaviossa eritellään valkoisilla viivoilla ja laatikon nimi eli toimiala lukee alhaalla keskellä valkoisella fontilla. Jos visualisoinnin katsoja haluaa tarkastella yksittäisen laatikon informaatiota tarkemmin, hänen pitää ainoastaan laittaa hiiren kursori laatikon päälle. Kuva 18 näyttää miltä visualisointi näyttää, kun puukaavion laatikon päälle laitetaan hiiren kursori.

Mustalla pohjalla värjätty tekstikenttä antaa enemmän informaatiota siitä laatikosta, minkä päällä kursori on. Kuva 18 näyttää Päijät-Hämäläisten yritysten toimipaikkojen henkilöstön määrän teollisuustoimialalla vuonna 2019. Infokenttä kuvan 18 vasemmalla puolella esittää kuvan oikealla puolella tehdyn valinnan. Tällaisen arvon valitsemiseen käytetyn painikkeen heikko kohta on se, että Power BI ohjelma ei osaa antaa oikeaa nimikettä tekstikentälle painikkeesta valitun arvon perusteella, vaan visualisoinnin tekijä joutuu manuaalisesti antamaan arvolle nimikkeen.



Kuva 18. Visualisointi Power Blissä

Tässä visualisoinnissa annoin nimikkeelle nimen ”Valittu arvo”. Power BI ohjelman avulla pystyn lataamaan tietokoneella tehdyn visualisaation Microsoftin ylläpitämään alustaan hyödyntämällä Embed-toimintoa. Tämän toiminnon avulla kykenen lataamaan ilmaiseksi visualisoinnin verkkoon interaktiivisessa muodossa. Tätä Microsoftin alustalla sijaitsevaa visualisaatioita voidaan sitten jakaa muille käyttäjille ainoastaan URL-linkin avulla ja visualisaation tarkastelunsa vuoksi tarvitaan ainoastaan verkkoyhteys sekä verkkoselain, esimerkiksi Google Chrome. Näin toimeksiantaja kykenee jakamaan interaktiivista visualisaatioita eteenpäin sosiaalisessa mediassa, ilman että visualisoinnin tarkastelija joutuu itse tekemään mitään muita toimenpiteitä kuin klikkaamaan URL-linkkiä.

Linkki interaktiiviseen visualisaation.

<https://app.powerbi.com/view?r=eyJrIjoiaNjFmMWY3MGYtNTU0NS00NDkwLThjZWYtZjYzNWxOTZmMml3liwidCI6ImM4NTBmZTIjLWI0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODA2OCIsImMiOiJh9&pageName=ReportSection737e1512edb5ad03539a>

4.3 Power BI – Animoitu viivadiagrammi väestön muutoksesta

Väestön pieneneminen maaseudulla on yksi PoliRural hankkeen tärkeimmistä teemoista. Toimeksiantaja haluaa tästä johtuen interaktiivisen visualisoinnin, mikä esittää Hämeen-alueen kuntien väestönmuutoksen visuaalisesti. Visualisoinnin päätarpeena on väestönmuutoksen trendin esittäminen mahdollisimman selkeästi.

Visualisoinnissa käytettävä data löytyi Tilastokeskuksen Stat.fi tilastotietokannoista ([Linkki tietokantaan](#)). Tilastotietokannasta saatava data pitää sisällään kunnat, vuodet ja useita erilaisia ”Tiedot” kenttiä, kuten ”Taajama-aste,%” ja ”Väkiluku”. Tilastokeskuksen tietokannoissa on ”valitse muuttajat” asetus, minkä avulla dataa voidaan suodattaa ennen datan lataamista. Tämä visualisointi tarvitsee tilastotietokannasta ainoastaan seuraavat muuttajat, kunnat, eli Päijät – ja Hämeen-kunnat, vuodet 2009–2019 sekä Väkiluku tietokentän. Ladattavaa dataa pitää myös muokata siten, että sen sijaa, että jokainen vuosi olisi oma sarakkeensa vaan vuodet ovat kaikki samassa sarakkeessa. Tämä

onnistuu hyödyntämällä muuta ja laske pudotusvalikosta löytyvää ”käännä manuaalisesti” asetusta.

Tämän asetuksen avulla pystytään dataa muokata. Kuva 19 näyttää miltä ”käännä manuaalisesti” asetuksen kanssa voidaan tehdä. Tämän asetuksen kanssa voidaan valita ovatko muuttujat sarakkeita vai rivejä. Power BI:tä varten on tärkeää, että ohjelmaan syötettävä data on mahdollisimman siistissä muodossa. Ohjelmaan syötettävä data kannattaa pitää minimaalisena sekä sellaisena, että se pitää sisällään mahdollisimman vähän sarakkeita ja rivejä.



Kuva 19. Ruutukaappaus Stat.fi tilastotietokannasta

Säädettyäni asetuksen kuten se on kuvassa 19, voin ladata datan tietokoneelle. Kuva 20 näyttää miltä ladattu data näyttää Excelissä. Kuvasta 20 voidaan huomata, että data pitää sisällään ainoastaan kolme saraketta. Ilman Stat.fi tilastotietokannassa tehtyä muokkausta, ladattu data pitäisi sisällään kolme-toista saraketta, jokaisen vuoden ollessaan oma sarakkeensa.

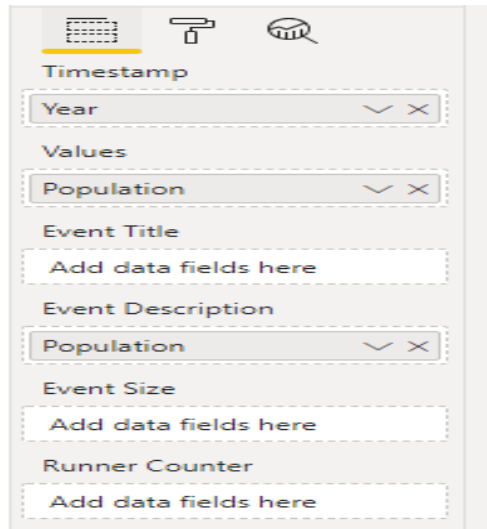
	A	B	C	D
1	Kuntien avainluvut 1987-2019			
2				
3	Alue 2021	Vuosi	Väkiluku	
4	Asikkala	2009	8551	
5	Asikkala	2010	8552	
6	Asikkala	2011	8498	
7	Asikkala	2012	8461	
8	Asikkala	2013	8405	
9	Asikkala	2014	8374	
10	Asikkala	2015	8287	
11	Asikkala	2016	8323	
12	Asikkala	2017	8248	
13	Asikkala	2018	8149	
14	Asikkala	2019	8083	
15	Forssa	2009	17807	
16	Forssa	2010	17904	
17	Forssa	2011	17833	
18	Forssa	2012	17727	
19	Forssa	2013	17667	
20	Forssa	2014	17521	
21	Forssa	2015	17422	
22	Forssa	2016	17332	
23	Forssa	2017	17185	
24	Forssa	2018	17028	
25	Forssa	2019	16901	

Kuva 20. Data excelissä

Ladattu data vaatii esikäsittelyn, sillä ladattu data pitää sisällään ylimääräistä informaatiota kuten tyhjiä rivejä ja otsikon. Näiden poistotoimenpiteiden jälkeen vuosi sarakkeen dataa pitää muuttaa vastaamaan pp.kk.vvvv muotoa. Tämä onnistuu helposti liittämällä "&" komennolla 31.12. ennen jokaista vuotta. Power BI vaatii ajallisen informaation, kuten vuodet pp.kk.vvvv muodossa. Muuten Power BI ei ymmärrä aika-arvoja vuosiksi, vaan pitää niitä ainoastaan numeroina. Tämä vaikeuttaa visualisoinnin toteuttamista Power BI ohjelmalla, sillä se ei osaa laittaa visualisoitavia arvoja aikajärjestykseen. Viimeisenä vaiheena on sarakkeiden nimien muuttaminen englanniksi. Näiden toimenpiteiden jälkeen data on valmis Power BIhin luettavaksi.

Koska visualisoitava data pitää sisällään ajallista dataa ja muuttujan eli väkiluvun ja muuttuja on jatkuva (Väkiluku data on numeerista ja sillä on selkeä alku ja loppu, vuosi 2009 on alku ja vuosi 2019 on loppu) päädyttiin käyttämään tässä visualisoinnissa viivadiagrammia. Esittääkseen väkiluvun muutoksen trendin, päädyttiin myös hyödyntämään tässä visualisoinnissa animaatiota. Animoidun viivadiagrammin luomiseen tarvitaan Pulse Chart-nimisen lisäläajennuksen Power BIhin. Lisäläajennuksia voi ladata ilmaiseksi ohjelmaan sen application-kaupasta.

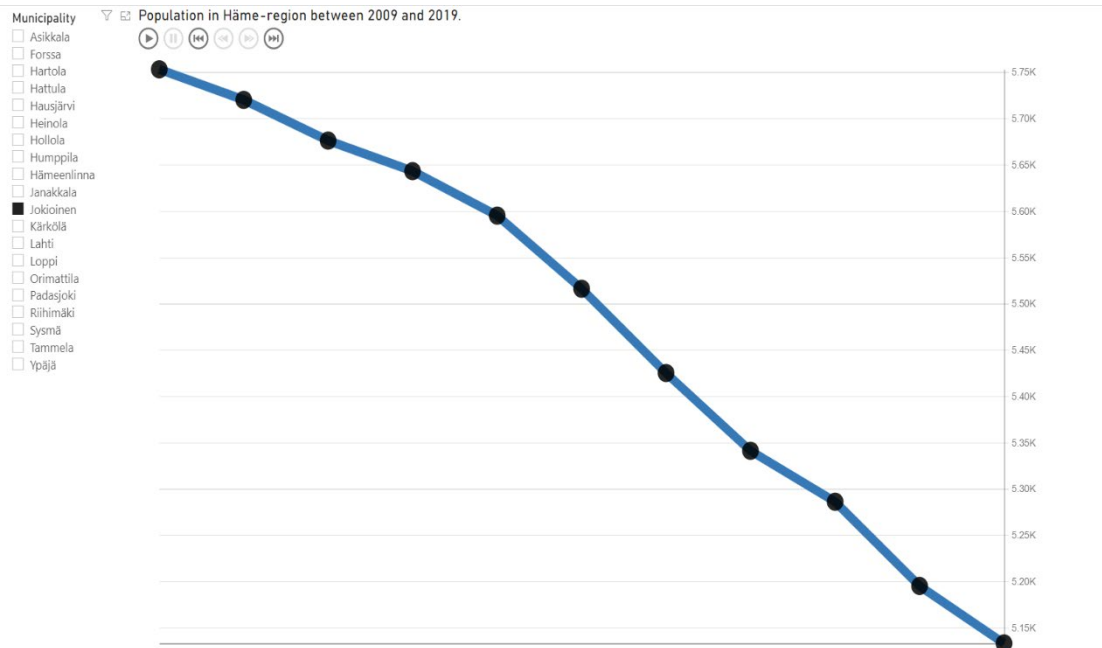
Pulse Chartin avulla pystyn tekemään yksinkertaisen, mutta tehokkaan animoidun viivadiagrammin. Viivadiagrammi toteuttaminen onnistuu helposti lisäämällä datasta sarakkeita Power BIin vaadittuihin datakenttiin. Kuva 21 näyttää mitä datakenttiä tarvitsen viivadiagrammin toteuttamiseksi. "Timestamp" datakenttä luo X-akselin sekä animoitavan ajanjakson, tässä tapauksessa vuodet 2009–2019. "Values" datakenttä taas tarvitsee sen muuttujan, minkä perusteella viivadiagrammi piirretään X- ja Y- akselille. Tässä tapauksessa se on populaation eli väkiluku.



Kuva 21. PowerPoint datakentät

”Event Description” on taas datapiste, mikä piirretään viivalle tietyn arvon perusteella. Tämä datapiste havainnollistaa paremmin mikä on jokaisen vuoden arvo, sillä datapisteet piirretään viivalle vuoden ja väkiluvun perusteella. Muutin myös viivadiagrammin asetuksia, sillä en ollut tyytyväinen oletusasetuksiin. Muutin viivan värin siniseksi, tein datapisteistä suurempia sekä tein Y-akselista väkiluvun mukaan skaalautuvan. Visualisoinnin lisäksi luon Power BI raporttiin vielä osittajatyökalun avulla osittajan, minkä avulla visualisoinnin tarkastelija valitsee yhden kunnan animoitavaksi.

Kuva 22 näyttää miltä valmis visualisointi näyttää. Klikkaamalla toistonappulaa vasemmassa yläkulmassa, visualisointi animoi viivadiagrammin ensimmäisestä datapisteestä, vuodesta 2009 viimeiseen datapisteeseen, vuoteen 2019. Visualisoinnin vasemmalla puolella on osittaja, minkä avulla katselija voi valita visualisoitavan kunnan. Vaikka visualisointi on yksinkertainen, se tuo hyvin esille väestönmuutoksen ja esittää väestön muutoksen trendin hyvin. Visualisoinnin animointi ei sinänsä tuo enempää statistista arvoa visualisoinnille, mutta se tukee visualisointia ulkomuodollisesti sekä tuo vähän dramatiikkaa visualisoinnin tavoitteen esille tuomiseksi.



Kuva 22. Väkiluku kunnittain, animoitu.

Linkki interaktiiviseen visualisointiin.

[https://app.powerbi.com/view?r=eyJrl-](https://app.powerbi.com/view?r=eyJrl-joiYTVmNDY2MTMtOWY4Ni00OWQzLTg1OWEtOW-RhYmE1NDQ5OTc4IiwidCI6ImM4NTBmZTljLWI0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODA2OCIsImMiOjh9&pageName=ReportSection)

[joiYTVmNDY2MTMtOWY4Ni00OWQzLTg1OWEtOW-](https://app.powerbi.com/view?r=eyJrl-joiYTVmNDY2MTMtOWY4Ni00OWQzLTg1OWEtOW-RhYmE1NDQ5OTc4IiwidCI6ImM4NTBmZTljLWI0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODA2OCIsImMiOjh9&pageName=ReportSection)

[RhYmE1NDQ5OTc4IiwidCI6ImM4NTBmZTljLWI0NmMtNGlyZC1iODYzLTAx](https://app.powerbi.com/view?r=eyJrl-joiYTVmNDY2MTMtOWY4Ni00OWQzLTg1OWEtOW-RhYmE1NDQ5OTc4IiwidCI6ImM4NTBmZTljLWI0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODA2OCIsImMiOjh9&pageName=ReportSection)

[ZmEyYTg5ODA2OCIsImMiOjh9&pageName=ReportSection](https://app.powerbi.com/view?r=eyJrl-joiYTVmNDY2MTMtOWY4Ni00OWQzLTg1OWEtOW-RhYmE1NDQ5OTc4IiwidCI6ImM4NTBmZTljLWI0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODA2OCIsImMiOjh9&pageName=ReportSection)

4.4 Power BI – Viivadiagrammi Hämeen alueen maakuntien populaation koulutusasteesta

Tämä visualisoinnin tarkoituksena on tarkastella Päijät-Hämeen maakuntien populaation koulutusasteen vuosittaista muutosta. Idea populaation koulutusasteen tarkastelemiseksi tuli esimieheltä. Koulutusaste on tärkeä mittari väestön hyvinvoinnin tarkastelemiseksi, sillä voidaan väittää, että korkeammalla koulutusasteella on positiivinen vaikutus tuottavuuteen, talouteen sekä ihmisten hyvinvointiin (Asplund & Maliranta 2006, 10). Visualisoinnin päämääränä on esitellä vuosittaista muutosta väkiluvussa koulutusasteen mukaan. Koulutusaste luokittelee tutkinnot suomalaisen koulutusjärjestelmän luoman hierarkian mukaan. Koulutusasteeseen kuuluvat esiaste, perusaste, toinen aste sekä alin – ja ylempiaste. Esimerkiksi toiseen asteeseen kuuluvat suomessa lukio, ammattikoulu sekä erikoisammattikoulu. Visualisaatio kuvaa viivadiagrammia hyödyntäen henkilöiden lukumäärän koulutusasteen, sukupuolen, vuoden ja maakunnan perusteella. Visualisaation otettiin myös mukaan ne

henkilöt, jotka eivät omaa peruskoulun jälkeistä tutkintoa eli he eivät ole saaneet tutkintoa esiastetta korkeammasta koulutuksesta. Visualisoitava data löytyi Tilastokeskuksen Stat.fi tietokannasta nimellä ”115d -- Väestö alueen, pääasiallisen toiminnan, koulutusasteen, sukupuolen, iän ja vuoden mukaan, 1987–2019”. Tietokannasta dataa haettaessa valittiin seuraavat kentät ja latsin datan tietokoneelle Excel-tiedostona.

- Alue
- Pääasiallinen toiminta
- Koulutusaste
- Sukupuoli
- Ikä
- Vuosi

Ladatulle datalle piti tehdä esikäsittely, jotta data olisi mahdollisimman siistiä sekä optimaalisessa muodossa, jotta Power BI osaisi sitä lukea. Excelin avulla suodatettiin Alue-sarakkeesta kaikki muut maakunnat pois paitsi Kanta – ja Päijät-Hämeen. Pääasiallinen toiminta sarakkeesta poistettiin muut kuin arvon ”Yhteensä”. Koulusteesta poistettiin ”Yhteensä” arvot, sillä visualisoinnissa haluan esittää kaikki koulutusasteet tai tutkinnot eri viivoina diagrammissa ja yhteensä arvo ei sinänsä ole oleellinen visualisoinnin tarvetta ajatellen. Ikä-kentän taas rajattiin arvoon ”18–64” sillä työikäisten koulutusaste on omasta mielestäni mielenkiintoisempi visualisoinnin kontekstia mielessä pitäen kuin eläkeläisten tai alle 18-vuotiaiden. Vuodet kentän rajattiin vuosiin 2000–2019. Vuodet olivat merkittynä datassa ilman päivää tai kuukautta, mutta tilastotietokanta tietää kertoa, että datassa esiintyvä väestö arvo on aina mitattuna aina samana päivänä. Koska Power BI tarvitsee ajallisen tiedon pp.kk.vvvv muodossa, lisäsin 1.1. ennen jokaista vuosi arvoa. Näin Power BI ymmärtää, että arvot kuvaavat vuotta eivätkä ole ainoastaan numeerisia arvoja.

Kuva 23 näyttää datan ulkomuodon Excelissä esikäsittelyn jälkeen. Kuvasta 23 voidaan huomata, että esikäsitelty data pitää sisällään kuusi saraketta ja kaikki mahdolliset muuttuvat arvot, kuten koulutusaste ja vuosi ovat yhdellä sarakkeella. Tämä tekee datasta kompaktin ja helposti käsiteltävän.

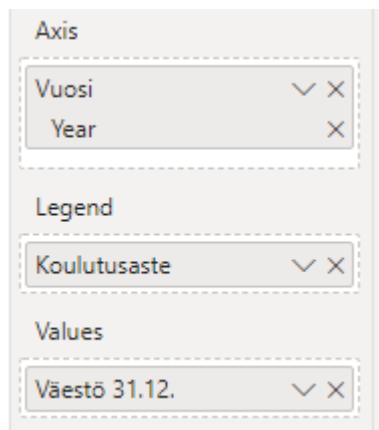
A	B	C	D	E	F	G	H
Alue	Pääasiallinen toim	Koulutusaste	Vuosi	Ika	Sukupuoli	Väestö 31.12.	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2000	18 - 64	Miehet	23307	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2000	18 - 64	Naiset	20946	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2001	18 - 64	Miehet	23609	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2001	18 - 64	Naiset	21126	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2002	18 - 64	Miehet	23845	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2002	18 - 64	Naiset	21403	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2003	18 - 64	Miehet	24187	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2003	18 - 64	Naiset	21793	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2004	18 - 64	Miehet	24548	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2004	18 - 64	Naiset	22099	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2005	18 - 64	Miehet	24928	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2005	18 - 64	Naiset	22387	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2006	18 - 64	Miehet	25326	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2006	18 - 64	Naiset	22665	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2007	18 - 64	Miehet	25800	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2007	18 - 64	Naiset	23084	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2008	18 - 64	Miehet	26234	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2008	18 - 64	Naiset	23325	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2009	18 - 64	Miehet	26657	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2009	18 - 64	Naiset	23632	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2010	18 - 64	Miehet	27031	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2010	18 - 64	Naiset	23732	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2011	18 - 64	Miehet	27233	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2011	18 - 64	Naiset	23769	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2012	18 - 64	Miehet	27263	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2012	18 - 64	Naiset	23721	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2013	18 - 64	Miehet	27328	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2013	18 - 64	Naiset	23603	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2014	18 - 64	Miehet	27487	
MK05 Kanta-Häme	Yhteensä	Toinen aste tai erikois	1.1.2014	18 - 64	Naiset	23474	

Kuva 23. Koulutusaste data Excelissä

Kuvassa 23 oleva data on valmis Power BI:n luettavaksi. Datan lukeminen Power BI:n onnistuu yksinkertaisesti hyödyntämällä Power BI:n Get data -toimintoa, mikä lukee tiedoston Power queryä hyödyntämällä. Power queryn avulla voidaan myös tehdä datalle esikäsittely ja muokata dataa DAX-ohjelmointikielillä avulla sekä hallinnoida useita datalähteitä samaan aikaan. Tässä visualisoinnissa on vain yksi datalähde, eli Excel-tiedosto.

Power BI:ssä viivadiagrammin tekeminen onnistuu mutkitta ja nopeasti. Diagrammien toteuttaminen Power BI:ssä onnistuu lisäämällä luetun datan sarakkeita vaadittuihin tietokenttiin. Kuva 24 näyttää minkälaisia tietokenttiä Power BI vaatii viivadiagrammin tekemiseksi. "Axis" -tietokenttä on X-akselille sijoitettavat arvot, eli tässä tapauksessa sarakkeessa Vuosi olevat arvot.

Kuva 24. Power BI:n viivadiagrammissa käytettävät kentät

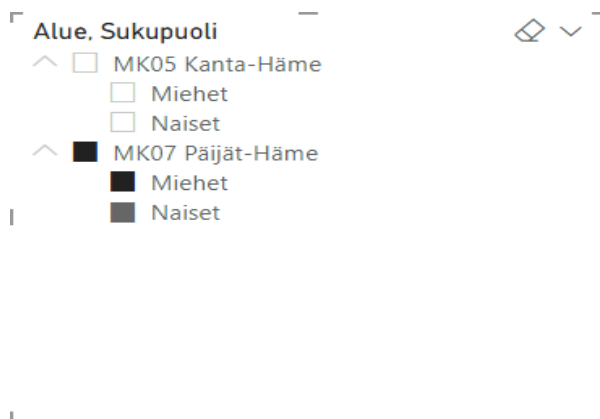


”Legend” tietokenttä eli selite taas yksilöi viivadiagrammissa näytettävät arvot halutun sarakkeen arvojen perusteella mikä on tässä tapauksessa Koulutusaste. ”Values” tietokenttä taas lisää arvoja viivadiagrammin Y-akseliin tietokenttään lisätyn sarakkeen perusteella. Tässä diagrammissa hyödynsin ”Väestö 31.12.” saraketta mikä pitää sisällään numeerisen arvon, mikä viittaa mistäkin koulutusasteesta valmistuneeseen väkilukuun vuoden, sukupuolen sekä maakunnan perusteella.

Jotta visualisointia voitaisiin suodattaa sukupuolen ja maakunnan perusteella luotiin vielä suodatin käyttämällä slicer-työkalua. Slicer-työkalun avulla pystytään tekemään painikke, mitä painamalla visualisointi muuttuu interaktiivisesti.

Painikkeen luominen onnistuu lisäämällä Power BIhin luetusta datasta sarakkeita slicer-työkalun tietokenttään, mitä halutaan käyttää visualisoinnin suodattamiseksi. Tässä tapauksessa ne ovat ”Alue” ja ”Sukupuoli” (katso kuva 23). Sarakkeita voidaan lisätä useampi slicer-työkaluun, mikä tekee painikkeista hierarkkisia. Riippuen sarakkeiden järjestyksestä työkalussa ja lukumäärästä slicerissä painikkeiden järjestys muuttuu.

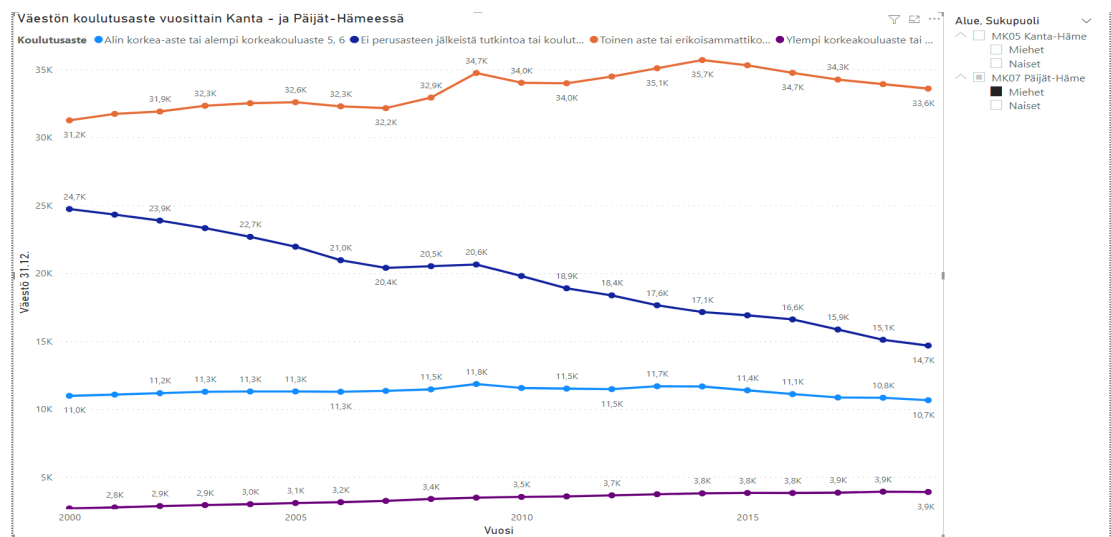
Kuva 25 näyttää miltä slicer-työkalu näyttää tässä visualisoinnissa. Slicer-työkalussa on kaksi saraketta, Alue ja Sukupuoli ja Alue on järjestyksessä ensimmäisenä ja Sukupuoli toisena. Klikkaamalla nuolta voidaan porautua tarkemmin Alueen eli maakunnan muihin suodatettavissa oleviin arvoihin, tässä tapauksessa sukupuoleen. Musta täyttö valintaruudussa kertoo, että kyseinen arvo on valittu ja visualisointia suodatetaan sen arvon perusteella.



Kuva 25. Slicer-työkalu Power Blissä

Seuraavaksi muutettiin visualisoinnin oletusasetuksia. Muutettiin datapisteiden eli vuosittaisia arvoja kuvaavien pallojen kokoa. Tehtiin Y-akselista skaalautuvan sekä muutin Y- ja X-akselin nimiä havainnollistavimmaksi. Tämän jälkeen muutettiin datapisteiden arvoja kuvaaman niitä tuhansien perusteella sekä yhden desimaalin tarkkuudella.

Kuva 26 on valmis visualisointi. Visualisoinnin oikealla puolella huomataan slicer-työkalu, minkä avulla pystytään kuvaa suodatamaan sukupuolen sekä maakunnan perusteella. Viivadiagrammissa olevat eri väriset viivat taas kuvaavat koulutusasteelta valmistuneiden henkilöiden määrää vuosittain. Koulutusasteet merkitään eri värein. Visualisoinnin avulla voidaan nopeasti havaita Päijät-Hämeen miesten kouluttamattomien vuosittainen pudotus ja pieni kasvu ylemmältä korkeakouluasteelta valmistuneiden määrässä.



Kuva 26. Koulutusaste maakunnan ja sukupuolen mukaan.

Visualisointi toimii oman ja esimieheni mielipiteen perusteella mainiosti ja kyky interaktiivisesti suodattaa visualisointia teki siitä helposti ymmärrettävän. Visualisointi tuo mielenkiintoisesti esille trendin sekä suuret eriävyydet miesten ja naisten koulutusasteessa.

Linkki interaktiiviseen visualisointiin

<https://app.powerbi.com/view?r=eyJrIjoiMzI1MDM2Y2YtNm->

[QxNS00MDFhLTg1ZGI0NWExNDUzY2Y2VjZjhmliwidCI6ImM4NTBmZTljLWl0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODAzOCIsImMiOiJh9&pageName=ReportSection](https://app.powerbi.com/view?r=eyJrIjoiMzI1MDM2Y2YtNm-QxNS00MDFhLTg1ZGI0NWExNDUzY2Y2VjZjhmliwidCI6ImM4NTBmZTljLWl0NmMtNGlyZC1iODYzLTAxZmEyYTg5ODAzOCIsImMiOiJh9&pageName=ReportSection)

4.5 Mapbox visualisointi – Mediaanitulos postinumeron mukaan Suomessa

Tässä visualisoinnissa toteutan mediaanituloja kuvaavan kartan Suomesta postinumeroalueittain. Visualisointi näyttää jokaisen suomalaisen postinumeroalueen mediaanitulon koropleettikartalla. Koropleettikartta esittää arvojen vaihtelua maantieteellisellä alueella (kartalla) hyödyntämällä muotoja sekä arvoja. Tässä visualisoinnissa muoto on suomalainen postinumeroalue ja arvo on mediaanitulos.

Visualisoinnin tarve ja idea syntyi omasta halustani esittää arvoja maantieteellisellä kartalla visuaalisesti. Aikaisemmat visualisoinnit ovat olleet diagrammeja ja karttoja ennen tätä visualisointeja ei ole hyödynnetty ollenkaan. Visualisoinnin tarve on esittää arvoja kartalla, jotta niitä on helppoa verrata keskenään ja jotta katsoja kykenisi hahmottamaan koko Suomen tilanteen mediaanitulojen osalta.

Visualisoinnin ohjelmaksi valittiin selaimella toimivan Mapbox studion. Mapbox studio vastaanottaa dataa useissa erilaisessa tiedostomuodossa sekä valmiita visualisointeja voidaan jakaa hyperlinkkinä toisille henkilöille, ilman kirjautumista tai muita toimenpiteitä. Kaiken lisäksi ohjelma on ilmainen.

Visualisoinnin toteuttamiseksi tarvitsen kaksi erilaista taulukkoa, jotka pitää liittää yhdeksi tiedostoksi ennen kuin niitä voidaan hyödyntää visualisoinnin toteuttamiseksi Mapbox studiossa. Mediaanitulos postinumeroittain sekä postinumeroiden maantieteellisen sijainnin Suomessa. Mediaanitulos postinumeroittain löytyy Paavo-tietokannasta nimikkeellä ”3. Asukkaiden käytettävissä olevat rahatulos, 2019”. Maantieteellisen sijainnin postinumero alueille löytyy ArcGIS Hub nimiseltä verkkosivustolta nimellä Postinumeroalueet (ArcGIS Hub, 2015) Tämä data pitää sisällään postinumeroalueeseen liittyvää dataa, kuten postinumeroalueen nimen, postialueen, pinta-alan sekä sen maantieteelliset koordinaatit. Tämän postinumeroalue datan avulla kyetään lisäämään Mapboxin kartalle kaikki Suomessa sijaitsevat postinumeroalueet koordinaatteja hyödyntäen. Maantieteelliset koordinaatit tässä tapauksessa ovat SHP-muodossa. SHP-tiedosto eli shapefile on vektoreihin perustuva paikkatiedon

tallennusmuoto. SHP-tiedosto hyödyntää vektoreita muodostamaan muotoja, pisteitä ja koordinaatteja maantieteellisellä kartalla. Tässä tapauksessa SHP-tiedosto pitää sisällään vektorit, mitkä muodostavat postinumeroalueen muodon kartalle. SHP-tiedosto on niin sanottu tiedosto kokoelma, mikä pitää sisällään vektorit sisältävän SHP-tiedoston lisäksi muita tiedostoja. Nämä muut tiedostot pitävät sisällään paikkatiedon vektoreihin liittyvää muuta informaatiota, kuten numeraalista dataa. SHP-tiedoston lisäksi kokoelma vaatii kaksi muuta tiedosta, .SHX – ja .DBF loppuiset tiedostot. Onkin yleistä, että Shapefile tiedostoista luodaan zippejä, eli tiedostopakkauksia.

Paavo-tietokannasta löytyvä taulukko esittää erilaisia taloudellisia arvoja postinumeroalueen perusteella vuonna 2021. Taulukko pitää sisällään postinumeroalueen sekä useita erilaisia taloudellisia arvoja, kuten mediaani – ja keskitulot. Taulukko ladattiin CSV-muodossa. Lataamisen jälkeen CSV-tiedostolle toteutettiin esikäsittelyn Excel-ohjelmalla. Turhien rivien sekä sarakkeiden poistamisen lisäksi jouduttiin muokkaamaan itse mediaanituloja esittävää saraketta. Joissakin postinumeroalueissa ei ole arvoa ollenkaan tai arvo nolla.

2968	99170 Pulju (Kittilä)	...	
2969	99180 Rautusjärvi (Kittilä)		17768
2970	99190 Hanhimaa (Kittilä)		15906
2971	99195 Pokka (Kittilä)	...	
2972	99230 Jeesiöjärvi (Kittilä)		17361
2973	99240 Kuivasalmi (Kittilä)	...	
2974	99250 Kiistala-Jeesiöjärvi (Kittilä)		21595
2975	99270 Hormakumpu (Kittilä)		15418
2976	99280 Tepsa (Kittilä)		21328
2977	99290 Kelontekemä (Kittilä)		16451
2978	99300 Muonio Keskus (Muonio)		21958
2979	99310 Yli-Muonio (Muonio)		22697
2980	99320 Kätkäsuvanto (Muonio)		21309
2981	99330 Pallastunturi (Muonio)		0
2982	99340 Raattama-Pallastunturi (Kittilä)		17494

Kuva 27. Mediaanitulo data Excelissä

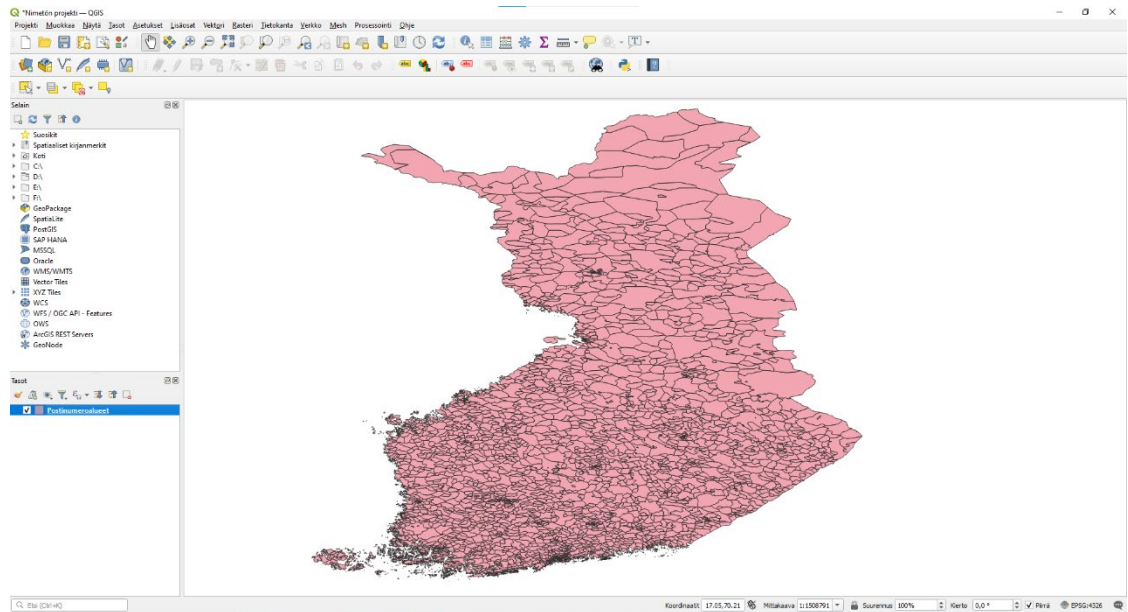
Kuva 27 näyttää miltä mediaanitulo data näyttää Excelissä esikäsittelyn tässä vaiheessa. Muut sarakkeet ovat poistettu taulukosta ja siinä on vain kaksi saraketta jäljellä. Ensimmäinen sarake on Postinumeroalue ja toinen sarake on mediaanitulon arvo euroina. Kuvasta 27 voidaan huomata, että arvoja, joita ei ole olemassa kuvataan kolmella pisteellä. Mapbox-ohjelma ei ymmärrä näitä kolmeen pisteen arvoja, joten ne pitää korvata arvolla 0. Tämä onnistuu vattomasti Excelin haku – ja korvaus toiminnolla. Jotta Excelissä olevan datan

yhdistäminen SHP-tiedostoon olisi mahdollista pitää CSV-tiedostoon luoda vielä yksi sarake, joka numeroi kaikki rivit. Koska Postinumeroalueita on 3027, pitää sarake sisällään numerot 1–3027 järjestyksessä ensimmäisestä viimeiseen numeroon riveittäin. Näiden arvojen perusteella voidaan CSV-tiedosto yhdistää SHP-tiedostoon. Viimeiseksi tallennetaan CSV-tiedoston CSV UTF-8 (Comma delimited) muodossa, jotta tiedostojen yhdistämiseen käytettävä ohjelma ymmärtää Ä sekä Ö kirjaimet. CSV-tiedoston esikäsittelyn jälkeen siirryttiin yhdistämään visualisoinnissa käytettävät tiedostot keskenään. CSV- ja SHP-tiedostot pitää yhdistää keskenään, jotta Mapboxin avulla dataa voitaisiin visualisoida. Mapbox vaatii visualisoitavien numeraalisten arvojen olevan samassa tiedostossa kuin maantieteellinen tieto. Eli tässä tapauksessa SHP-tiedosto (maantieteellinen tieto) pitää olla samassa taulukossa kuin numeraaliset arvot (CSV-tiedostossa olevat mediaanitulot). Esikäsittelyn toteuttaminen postinumeroalueita kuvaavalle datalle ei ollut tarpeellinen.

Mediaanituloja kuvaava CSV-tiedoston sekä postinumeroaluita kuvaavaan informaation yhdistäminen onnistuu QGIS paikkatieto-ohjelmiston avulla. Ensimmäiseksi tiedostot lisätään QGIS-ohjelmaan. Tämä onnistuu ohjelman käynnistämisen jälkeen lisäämällä Tasoja. Tasoja voidaan lisätä hyödyntämällä ohjelmassa tietolähteiden hallintaa. Tämän toiminnon avulla voidaan yksinkertaisesti lisätä tasoja tietokoneelta avaamalla tiedostot tietolähteiden hallinnan avulla. Tässä visualisoinnissa on kaksi tasoa, postinumeroalue sekä postinumeroaluetta kuvaava numeraalinen data, mediaanitulot vuonna 2021. Ensimmäiseksi lisään postinumeroalue datan QGIS-ohjelmaan. Tämä käy helposti vain lisäämällä ladatun SHP-tiedoston QGIS-ohjelmaan.

Kuva 28 näyttää miltä SHP-tiedosto näyttää QGIS-ohjelmassa. Kuvasta voidaan huomata postinumeroalueita kuvaavat vektorit, eli tässä tapauksessa viivat. Nämä viivat luovat rajat jokaiselle postinumeroalueelle Suomessa ja kuvasta 28 voidaankin huomata mutkitta, että postinumeroalueet muodostavat suomen maantieteelliset rajat. QGIS-ohjelma osaa tuoda SHP-tiedoston vektorit tarkasteltavaksi automaattisesti eikä muita toimenpiteitä tarvita. SHP-tiedostoon liittyvää muuta dataa voidaan tarkastella klikkaamalla vasemmassa alakulmassa olevasta aktiivisesta tasosta eli Postinumeroalueesta. Aktiivisen tason tietää oikeinmerkistä. Tason asetukset voidaan avata hiiren oikealla

näppäimellä ja vektoreihin liittyvää dataa voidaan tarkastella attribuuttitaulukon avulla.



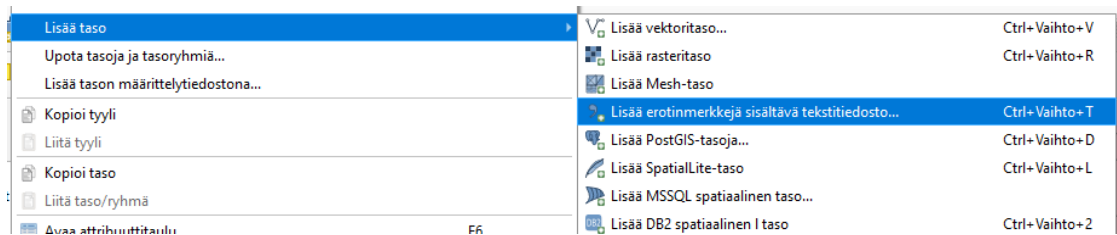
Kuva 28. Postinumeroalue data QGIS-ohjelmassa

Attribuuttitaulukko (Kuva 29) näyttää SHP-tiedoston vektoreihin liittyvän datan. posti_alue sarake pitää sisällään postinumeron ja nimi viittaa postinumeroalueen nimeen.

	OBJECTID	posti_alue	nimi	namn	kunta	kuntanro	vuosi	pinta_ala	Shape_Are	Shape_Len
1	1	00100	Helsinki Keskus...	Helsingfors cen...	091	91	2021	2353278	2353278,469482...	7978,705923088...
2	2	00120	Punavuori	Rödbergen	091	91	2021	414010	414009,5627441...	3936,521840132...
3	3	00130	Kaartinkaupunki	Gardesstaden	091	91	2021	428960	428960,3325195...	2867,317323341...
4	4	00140	Kaivopuisto - U...	Brunnsparken -...	091	91	2021	931841	931840,9553222...	8585,386930507...
5	5	00150	Eira - Hernesaari	Eira - Ärtholmen	091	91	2021	1367328	1367327,614013...	7880,521502161...
6	6	00160	Katajanokka	Skatudden	091	91	2021	541136	541135,6701660...	2923,492348985...
7	7	00170	Kruununhaka	Kronohagen	091	91	2021	713343	713343,2663574...	4130,226548453...
8	8	00180	Kamppi - Ruoh...	Kampen - Gräs...	091	91	2021	1735293	1735293,336914...	8331,468685098...
9	9	00190	Suomenlinna	Sveaborg	091	91	2021	765012	765012,3242187...	7229,011877452...
10	10	00200	Lauttasaari	Drumsö	091	91	2021	2619434	2619434,180664...	10842,92662945...
11	11	00210	Vattuniemi	Hallonnäs	091	91	2021	1478221	1478220,782958...	8090,224290089...
12	12	00220	Jätkäsaari	Busholmen	091	91	2021	684781	684781,2687988...	3654,627944215...
13	13	00230	Ilmala	Ilmala	091	91	2021	171851	171851,1328125...	1681,870295761...
14	14	00240	Länsi-Pasila	Västra Böle	091	91	2021	2439966	2439966,056640...	8009,539492378...
15	15	00250	Taka-Toolö	Bortre Tölö	091	91	2021	3065419	3065419,403076...	14370,33836822...
16	16	00260	Keski-Toolö	Mellersta Tölö	091	91	2021	751805	751804,7734375...	4490,522342097...
17	17	00270	Pohjois-Meilähti	Norra Mejlans	091	91	2021	864732	864732,1467285...	4008,127057284...
18	18	00280	Ruskeasu	Brunskärr	091	91	2021	822672	822671,8488769...	3853,134788288...
19	19	00290	Meilahden saira...	Mejlans sjukhu...	091	91	2021	268739	268739,1105957...	2177,537010846...
20	20	00300	Pikku Huopalahti	Lillhoplax	091	91	2021	851105	851105,4443359...	4410,433120892...
21	21	00310	Kivihaka	Stenhagen	091	91	2021	621966	621965,7795410...	3334,560188913...
22	22	00320	Etelä-Haaga	Södra Haga	091	91	2021	1942280	1942279,824707...	6340,062347460...
23	23	00330	Munkkiniemi	Munksnäs	091	91	2021	1726680	1726679,886962...	6827,499448658...

Kuva 29. Attribuuttitaulukko Qgis ohjelmassa

Jotta voidaan luoda Mapbox-studion avulla visualisoinnin mikä kuvaa kartalla postinumeroalueiden mediaanituloja, pitää aikaisemmin esikäsitelty CSV-taulukko yhdistää kuvassa 29 olevaan taulukkaan. CSV-tiedoston lisääminen onnistuu lisäämällä erotinmerkkejä sisältävän tekstitiedoston (katso kuva 30). CSV-tiedosto eli Comma Separated Values (suom. Pilkulla erotellut arvot) erottelee rakenteellisen datan pilkun avulla, mutta on myös mahdollista lisätä dataa mikä erotellaan muilla erotinmerkeillä.



Kuva 30. Tekstitiedostojen lisääminen.

Kun CSV-tiedosto on lisätty QGIS-ohjelmaan, se ilmestyy tasoksi samaan paikkaan missä Postinumeroalueet taso sijaitsee myös (Kuva 28). CSV-tiedoston arvoja voidaan tarkastella samalla tavalla kuin postinumeroalueen, avaamalla attribuuttitaulukko.

Kuvasta 31 voidaan huomata saman niminen ensimmäinen sarake kuin kuvan 29 postinumeroalue taulukossa. QGIS-ohjelman liitostyökalu tarvitsee saman nimisen sarakkeen, mikä pitää sisällään jokaiselle riville uniikkeja arvoja, jotta taulukkoja voidaan liittää toisiinsa. Hyödyntämällä liitostyökalua lisään CSV-tiedoston sarakkeet (Kuva 31) SHP-tiedostoon (Kuva 29).

OBJECTID	Alue	Asukkaiden mediaanitulot, 2019 (HR)
1	1 Helsinki Keskus...	29706
2	2 Punavuori (Hel...	29816
3	3 Kaartinkaupunk...	32894
4	4 Kaivopuisto - U...	31718
5	5 Eira - Hernesaar...	28478
6	6 Katajanokka (H...	30168
7	7 Kruununhaka (...)	30720
8	8 Kamppi - Ruoh...	27711
9	9 Suomenlinna (...)	27688
10	10 Lauttasaari (Hel...	29703
11	11 Vattuniemi (Hel...	32861
12	12 Jätkäsaari (Helsi...	26016
13	13 Ilmala (Helsinki)	0

Kuva 31. CSV-tiedosto QGIS ohjelmassa

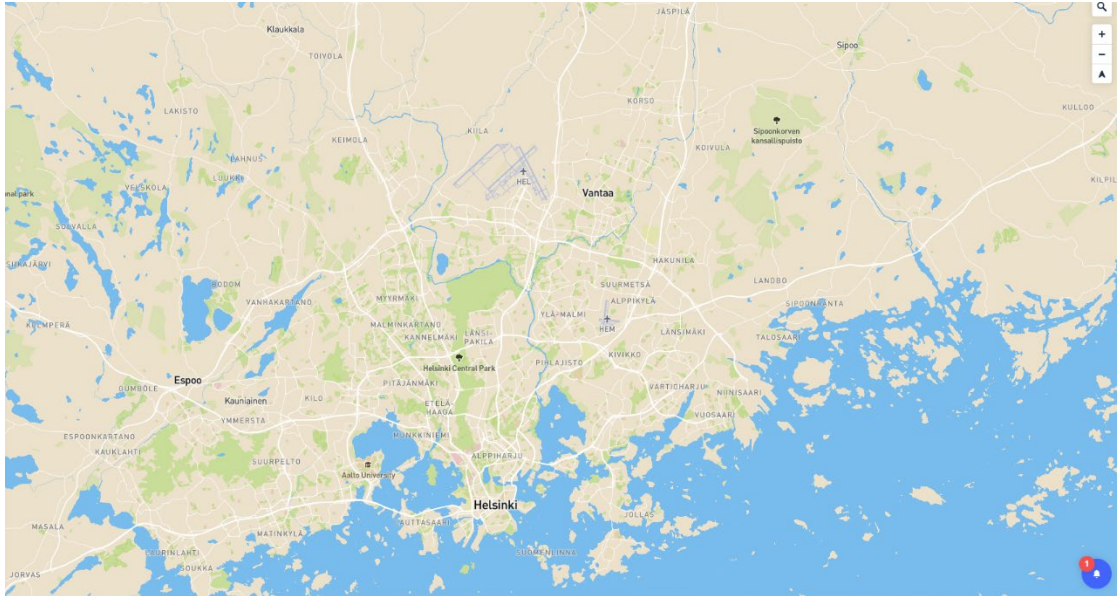
Kuvasta 32 voidaan huomata miltä yhdistetty taulukko näyttää. Taulukkojen yhdistämisen jälkeen klikkaamalla aktiivista tasoa, valitsemalla Export ja sitten valitsemalla ESRI Shape-tiedosto, voidaan taso tallentaa uutena SHP-tiedostona omalle tietokoneelle. SHP-tiedosto tallentuu yksittäisinä tiedostoina eikä QGIS-ohjelma automaattisesti luo SHP-tiedostosta kokoelmaan. SHP-tiedostosta pitääkin luoda pakattu tiedosto eli ZIP ennen kuin se voidaan lisätä Mapbox studioon.

OBJECTID	posti_alue	nimi	namn	kunta	kuntanro	vuosi	pinta_ala	Shape_Are	Shape_Len	Postinumero	Alue	Asukkaiden mediaanitulos_2019 (HR)
1	00100	Helsinki Keskus...	Helsingfors cen...	091	91	2021	235278	235278,469482...	7978,705929088...	100	Helsinki Keskus...	29706
2	00120	Punavuori	Rödbergen	091	91	2021	414010	414009,5627441...	3936,521840132...	120	Punavuori (Hel...	29816
3	00130	Kaartinkaupunki	Gardesstaden	091	91	2021	428960	428960,3325195...	2867,317323341...	130	Kaartinkaupunk...	32894
4	00140	Kaivopuisto - U...	Brunnsparken - ...	091	91	2021	931841	931840,9533222...	8585,386930507...	140	Kaivopuisto - U...	31718
5	00150	Eira - Hemesaari	Eira - Ärtholmen	091	91	2021	1367328	1367327,614013...	7880,521502161...	150	Eira - Hemesaar...	28478
6	00160	Katjanokka	Skatudden	091	91	2021	541136	541135,6701660...	2923,492348985...	160	Katjanokka (H...	30168
7	00170	Kruununhaka	Kronohagen	091	91	2021	713343	713343,2663574...	4130,226548453...	170	Kruununhaka (...)	30720
8	00180	Kamppi - Ruoh...	Kampen - Gräs...	091	91	2021	1735293	1735293,336914...	8331,468605098...	180	Kamppi - Ruoh...	27711
9	00190	Suomenlinna	Sveaborg	091	91	2021	765012	765012,3242187...	7229,011877452...	190	Suomenlinna (...)	27688
10	00200	Lauttasaari	Drumsö	091	91	2021	2619434	2619434,180664...	10842,92662945...	200	Lauttasaari (Hel...	29703
11	00210	Vattuniemi	Hallnäs	091	91	2021	1478221	1478220,782958...	8090,224290089...	210	Vattuniemi (Hel...	32861

Kuva 32. Yhdistetty taulukko

Mapbox studiossa luodaan itse visualisoinnin hyödyntämällä QGIS:in avulla luotua SHP-tiedostoa (kuva 32), mikä pitää sisällään niin maantieteelliset koordinaatit sekä halutun visualisoitavan numeraalisen arvon eli mediaanitulos postinumeroalueittain. Visualisoinnin toteuttaminen Mapboxin avulla alkaa tyylin vallinnasta. Tyylit ovat kartan oletuspohja, minkä päälle visualisointi luodaan.

Tyylit asettavat kartalle ulkomuodon sekä luovat maailman kartan. Tyylin voi luoda kokonaan itse tai valita jokin ohjelman tarjoamista tyyleistä. Tyylit asettavat kartan ulkomuodon, värit ja tekstiin liittyvät seikat. Itse valitsin perustyylin, mikä mielestäni on visuaalisesti miellyttävin. Kuvassa 33 on perustyylistä otettu ruutukaappaus Helsingistä. Perustyyliin on mukana kaupunkien, teiden, kuntien ja maakuntien yms. nimet sekä paikkatiedot, joten tätä informaatiota ei tarvitse tuoda ohjelmaan tietolähteestä.



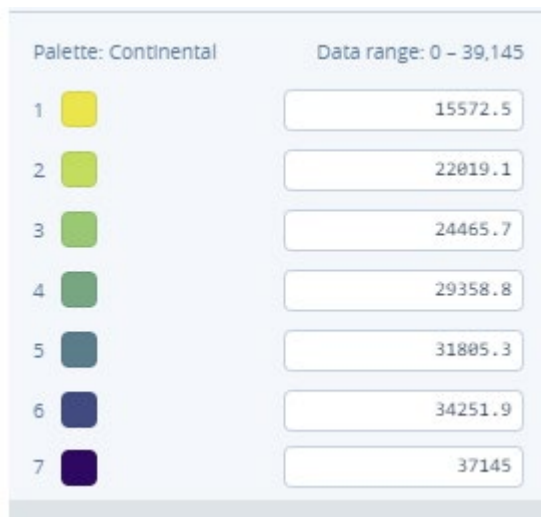
Kuva 33. Ruutukaappaus perustyylistä Mapbox studiossa.

Jotta perustyylin voidaan lisätä omia elementtejä, pitää visualisoitava data ladata Mapboxiin omalta tietokoneelta tai vaikkapa internetistä. Mapbox sitten hyödyntää siihen ladattua dataa luodakseen uusia elementtejä käyttäjän haluamasta datasta. Mapboxissa voi olla useita erilaisia datalähteitä, mutta tässä visualisoinnissa on vain yksi lisätty datalähde sekä kaksi elementtiä, postinumeroaluiden maantieteelliset rajat ja postinumeroalueiden mediaani tulot numerisina arvoina. Postinumeroalueet erotellaan toisistaan vektoreiden eli viivojen avulla, samalla tavalla kuin kuvassa 28. Jokaisella postinumeroalueella on mediaanitulo. Tätä mediaanituloa kuvaan väriasteikon avulla.

Käyttämällä Mapboxin koropleettikartta työkalua voin asettaa SHP-tiedoston avulla paikkatietokohtaiset viivat kuvaamaan postinumeroalueita. Mapbox hyödyntää kuvassa 32 näkyviä Shape_Are ja Shape_Len nimisiä sarakkeita postinumeroalueiden rajoja luodessa. Mapbox studio osaa hyödyntää näitä sarakkeita automaattisesti ja koropleettikarttaa käytettäessä ei tarvitse tehdä mitään muuta kuin antaa koropleettikartta luomiseksi numeraalinen arvo, mikä tässä visualisoinnissa on mediaanitulot.

Seuraavaksi pitää luoda väriasteikon kuvaamaan jokaisen postinumeroalueen mediaanituloja värein hyödyntämällä kuvassa 32 näkyvää Asukkaiden medi-

aanitulo saraketta. Mediaanitulot ovat 0 ja 39145 välissä, joten hyödynnän jatkuvaa väriskaalausta. Annoin itse arvoille värit sekä stopit, milloin väri vaihtuu. Kuva 34 näyttää visualisoinnin väriskaalan.

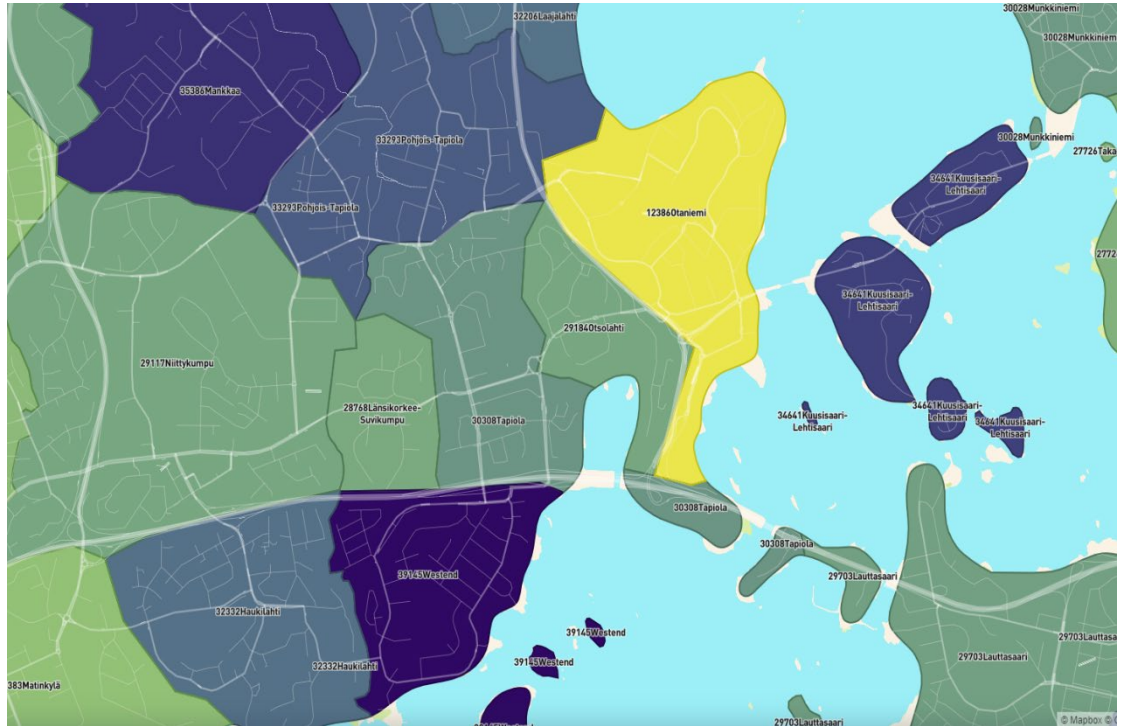


Kuva 34. Väriskaala Mapbox

Keltaiset postinnumeroalueet ovat niitä, missä mediaanitulot ovat alle 15572, kuin taas purppuran väriset ovat niitä missä mediaanitulot ovat yli 37145. Väriskaalassa voi olla lukematon määrä stoppeja, mutta annoin itse sille vain 7. Väriskaalan luomisen lisäksi muutettiin perustyylin asetuksia. Himmennettiin teiden värejä, poistettiin maamerkit sillä ne eivät ole oleellisia tämän visualisoinnin kannalta. Muutettiin myös maakuntien ja kuntien rajoja, jotta ne näkyisivät paremmin. Myös maantieteellisten sijaintien fonttia sekä tein muutoksia havainnollistavan tekstin asetteluun kartalla. Näiden muutosten lisäksi vaihdettiin asetuksista, milloin havainnollistava teksti näkyy. Jos visualisoijan katsoja on zoomannut tarpeeksi kauas, hän ei näe postinnumeroalueita kuvaavaa mediaanitulon arvoa tai postinnumeroalueen nimeä.

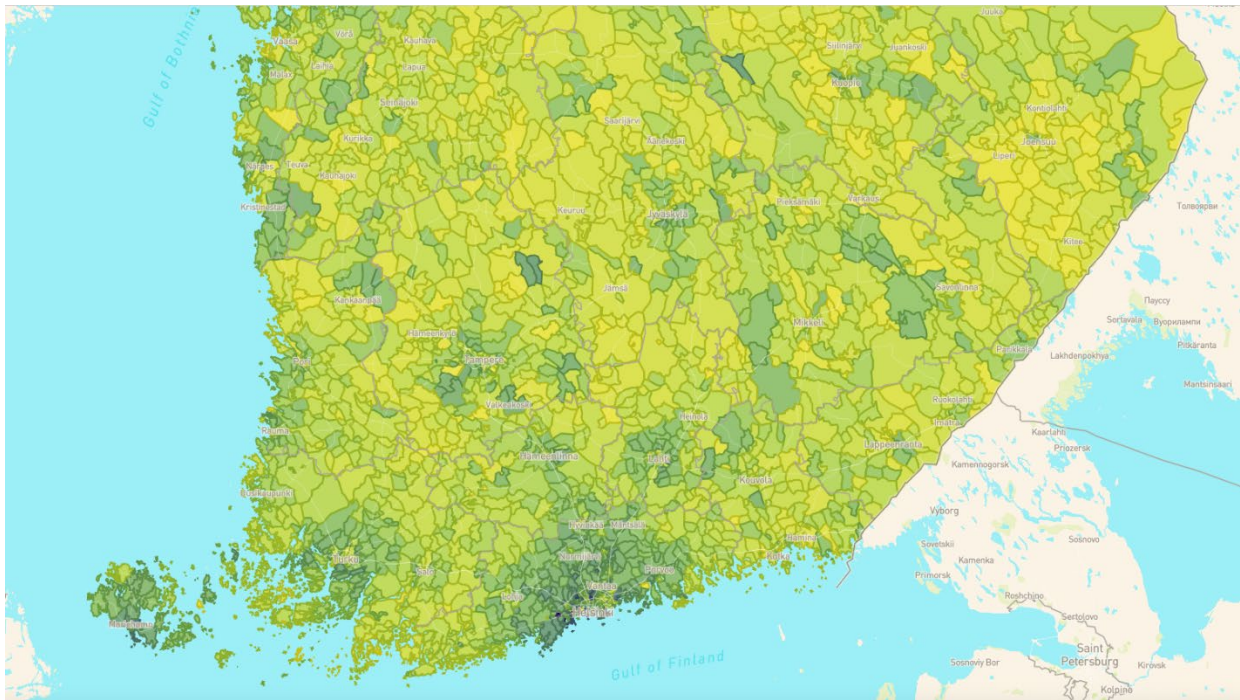
Kuva 35 näyttää valmiin visualisoinnin. Kuva on ruutukaappaus visualisoinnista Helsingistä, Otaniemestä zoomattuna melko lähelle. Kuvasta voidaan huomata kuin tarkasti postinnumeroalueet visualisoidaan kartalle ja kuinka hyvin koropleetin väriskaala toimii tässä tapauksessa. Postinnumeroalue data on maantieteellisesti hyvin tarkkaa ja se toimii hyvin ilman suurempia muutoksia Mapbox ohjelmiston kanssa. Karttaa tarkastellessa ei huomattu yhtäkään sil-

miinpistävää virhettä postinumeroalueiden kanssa. Myöskin valitsemani mediaanitulo data toimii hyvin kartalla. Kuvasta voidaan huomata, että postinumero alue merkitään sen mediaanitulolla ja sen suomenkielisellä nimellä.



Kuva 35. Valmis visualisointi

Kartta tuo hienosti esille kuinka hyvätuloiset suomalaiset ovat pakkautuneet joko rannikolle tai kaupunkeihin. Karttaa on myös omasta mielestäni helppo navigoida sekä sen on tarpeeksi selkeä vaikkapa oman asuinalueen löytämiseksi. Kartan avulla voidaan myös tehdä johtopäätöksiä koko Suomen tilanteesta mediaanitulojen suhteen. Kuten aikaisemmin mainitsin, hyvätuloiset tullaavat pakkaantumaan tiettyihin kaupunkeihin tai länsirannikolle. Kuva 36 näyttää miltä vähemmän tarkennettu kuva näyttää. Kuvasta voidaan huomoida, että mediaanituloja ei näytetä, jos kuvan tarkastelija ei ole zoomannut tarpeeksi lähelle kartassa. Kauemmaksi zoomatun kuvan avulla voidaan helposti tarkastella koko suomen tilannetta mediaanitulojen kanssa postinumeroalueittain. Valmiin visualisoinnin jakaminen onnistuu Mapboxin kanssa vaivattomasti. Ensimmäiseksi visualisointi julkaistaan, jonka jälkeen Mapbox luo hyperlinkin. Hyperlinkin klikkaaja näkee valmiin visualisoinnin ja voi vapaasti tarkastella sitä, mutta ei kykene muokkaamaan visualisoinnin asetuksia.



Kuva 36. Valmis visualisointi

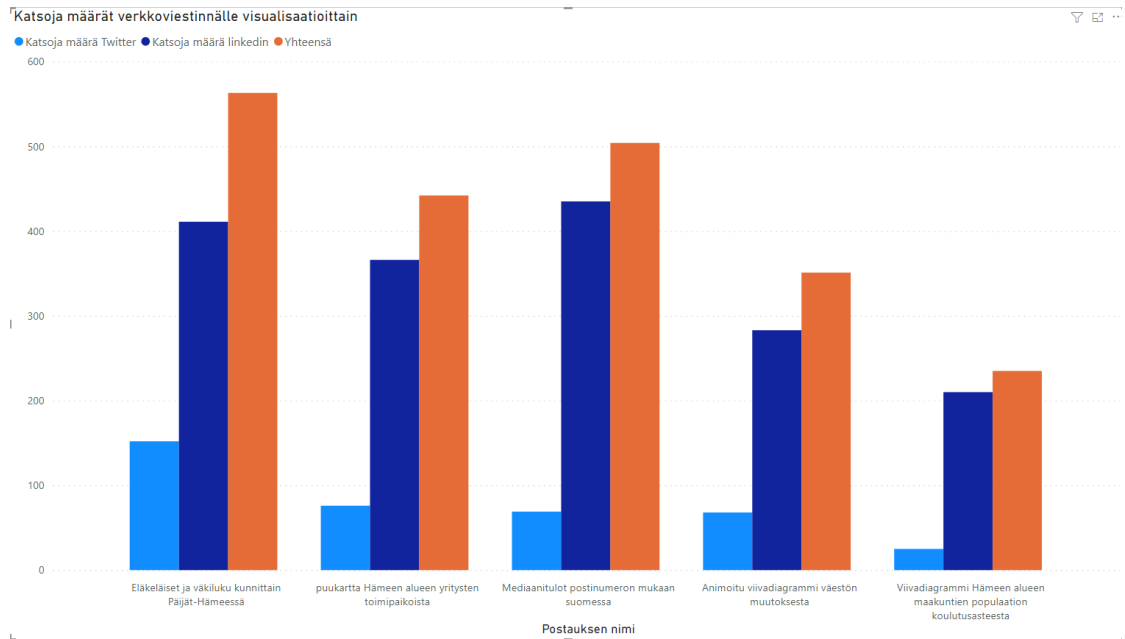
Linkki visualisointiin:

https://api.mapbox.com/styles/v1/naauk/cknd7150g23z717pep36gw1of.html?fresh=true&title=view&access_token=pk.eyJ1IjoibmFhdWsiLCJhIjoieY2trejM2MjJoMG1wMDJ3cDAXbDY2bDN0ZiJ9.8Jtkat_iQ0uNp1kwxM4rIA#6.34/61.488/24.764

5 TULOKSET, JOHTOPÄÄTÖS

5.1 Toimeksiantaja

Opinnäytetyön tuloksena oli viisi interaktiivista visualisointia toimeksiantajan eli Smart & Lean Hub Oyn käyttöön. Toimeksiantaja hyödynsi visualisaatioita niin yrityksen verkkosivuilla kuin sosiaalisessa mediassa tapahtuvassa viestinnässä. Visualisaatiot saavuttivat suurehkon yleisön pääsääntöisesti LinkedInissä, mutta myös Twitterissä. Niin Twitter kuin LinkedIn kokoavat tilastiatkaa henkilöistä, jotka näkivät käyttäjän tekemän viestin tai postauksen. Tätä tietoa hyödyntäen toteutettiin visualisaatio (kuva 37) verkkoviestinnän saavuttamasta katsojamäärästä.



Kuva 37. Verkkoviestinnän saavuttamat katsoja määrät.

Opinnäytetyön toimeksiantaja suhtautui positiivisesti tehtyihin visualisointeihin ja piti niitä hyvinä lisäyksinä yrityksen viestintään. Toimeksiantaja oli erityisen tyytyväinen visualisaatioiden interaktiivisuuteen ja kykyyn porautua tarkastelemaan dataa tarkemmin. Koska onnistuin toteuttamaan toimeksiantajan halusta datasta visualisointeja ja näiden visualisointien jakaminen sosiaalisessa mediassa onnistui vaivattomasti, piti toimeksiantaja toimeksiantoa erittäin onnistuneena. Visualisoinnit tukivat PoliRural-hankkeen kommunikaatio ja viestintä tavoitteita edistämällä hankkeen vaikuttavuutta.

5.2 Työntekijä

Pidän kirjoittamaani teoriaosuutta laajahkona tutkimuksena datan visualisoinnin teoriasta. Ymmärrykseni ja osaamiseni lisääntyi, kun tutustuin aiheeseen ja kirjoitin siitä. Omasta mielestäni pystyin hyvin vastaamaan toimeksiantajan odotuksia visualisaatioiden laadun suhteen. Kykenin myös toteuttamaan interaktiivisia sekä omasta mielestäni tiedollisesti rikkaita visualisaatioita toimeksiantajalle. Erityisen onnistuneena pidän visualisaatioiden jakamiseen käytettyjä toimenpiteitä. Kaikkien visualisaatioiden jakaminen onnistui mutkitta sosiaalisessa mediassa ja tavoite, että visualisaatioita pystyttäisiin tarkastelemaan ilman mitään vaadittavia toimenpiteitä visualisaation katsojalta toteutui.

Visualisointien suunnittelussa tai visualisoinnin aiheen valitsemisessa ei ollut vaikeuksia. Visualisoinneissa käytettävä data löytyi helposti internetistä avoimena datana. Visualisointien ulkomuodollisten osa-alueiden suunnittelu ja toteuttaminen onnistui hyödyntämällä lähdekirjallisuutta sekä aikaisempaa osaamistani.

Visualisaatioiden toteuttamiseen hyödynnetyt ohjelmat (Power BI, Python Plotly ja Mapbox) toimivat moitteettomasti sekä ohjelmien ansioista visualisaatioiden jakaminen interaktiivisina mahdollisti toimeksiantajan vaatimien interaktiivisten visualisointien toteuttamisen. Oma osaaminen datan visualisoinnista ennen toimeksiantoa oli omasta mielestäni hyvä. Toimeksiannon aikana jouduin kuitenkin opettelemaan uusia toimintoja ja käytäntöjä liittyen rakenteelliseen dataan sekä uuden ohjelman mapboxin, jotta toimeksiantajan haluttamia visualisointeja voitaisiin toteuttaa. Kerrytin myöskin omaa kokemustani Power BIn sekä Python käyttämisestä.

5.3 Kehittämiskohteet

Visualisointien ominaisuuksien kannalta toimeksiantajalle oli kaksi vaatimusta, interaktiivisuus sekä saumaton visualisointien jakaminen internetissä. Nämä kaksi jälkimmäistä tarvetta rajoittivat toimeksiannossa käytettävät ohjelmat kolmeen: Power BI, Python ja Plotly-kirjasto sekä Mapbox. Nämä ohjelmat mahdollistavat interaktiivisen visualisoinnin jakamisen URL-linkkinä. Nämä ohjelmat eivät kuitenkaan ole täydellisiä sekä jotkin visualisaatiot olisivat olleet parempia, jos ne olisi toteutettu jollain muulla kuin näillä kolmella ohjelmalla. Visualisointeja olisi voitu toteuttaa esim. Tableau tai R-kielellä ja sitten hyödyntää kolmannen osapuolen ohjelmaa visualisoinnin lataamiseksi verkkoon, jotta sitä voisi jakaa muille henkilöille helposti.

Visualisointia olisin voinut tehdä enemmän ja ehkä visualisointien ulkomuoto ei aina ollut optimaalinen, joten visualisointien esteettisessä ulkomuodossa olisi ollut parantamisen varaa. Myöskin visualisointien osa-alueet, kuten havainnollistava teksti sekä värit eivät aina olleet parhaita mahdollisia. Esimerkiksi visualisoinnissa 4.1 Python visualisointi – Eläkeläiset ja väkiluku kunnittain Päijät-Hämeessä, olisi datapisteiden vuosiluku pitänyt olla selkeämpi.

Lähteet

ArcGIS Hub. 2015. Paavo 2015 - Postinumeroalueittainen avoin tieto. WWW-dokumentti. Päivitetty 4.2.2016 Saatavissa: <https://hub.arcgis.com/data-sets/43692fc1e86b45fca188cc193b6482e1?geometry=-58.990%2C58.119%2C108.266%2C71.042> [viitattu 20.5.2021]

Asplund, R., & Maliranta, M. 2006. Koulutuksen taloudelliset vaikutukset. Media. PDF-dokumentti . Saatavissa: <https://media.sitra.fi/2017/02/27172706/raportti60-2.pdf> [viitattu 20.5.2021]

Berinato, S. 2016. Visualizations That Really Work. Hbr. WWW-dokumentti. Saatavissa: <https://hbr.org/2016/06/visualizations-that-really-work> [viitattu 20.5.2021]

Bush, K. 2018. data visualization. Search business Analytics. WWW-dokumentti. Saatavilla: <https://searchbusinessanalytics.techtarget.com/definition/data-visualization> [viitattu 20.5.2021]

Chun-Houh, C. & Wolfgang, K. & Unwin, A. 2008. Handbook of Data Visualisation. Leipzig: LE-TEX Jelonek, Schmidt & Vöckler GbR.

Clark. A.C & Matthews, B. 2000. A.C. Clark: Scientific and Technical Visualization. Journal for Geometry and Graphics. PDF-dokumentti. Saatavilla: https://www.heldermann-verlag.de/jgg/jgg01_05/jgg0407.pdf [viitattu 20.5.2021]

Cleveland, W., & McGill, R. 1984. GraphicaPerception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association. Saatavilla: www.jstor.org/stable/2288400 [viitattu 20.5.2021]

Davis, D. 2019. AI Unleashes the Power of Unstructured Data. Cio. WWW-dokumentti. Saatavilla: <https://www.cio.com/article/3406806/ai-unleashes-the-power-of-unstructured-data.html> [viitattu 20.5.2021]

Eduskunta. 2018. Eduskunnan avoimen datan palvelu on avattu. Eduskunta. WWW-dokumentti. Saatavilla: https://www.eduskunta.fi/FI/tiedotteet/Sivut/Avoindata_eduskunta_avaus_20180531.aspx [viitattu 20.5.2021]

etymonline. 2021. Diagram. Etymonline. WWW-dokumentti. Saatavilla: <https://www.etymonline.com/word/diagram> [viitattu 20.5.2021]

Evergreen, D. 2019. Effective Data Visualization: The Right Chart for the Right Data. E-kirja London: Sage. Saatavilla: <https://books.google.fi/books?id=ilWKDwAAQBAJ&pg=PA24-IA8&dq=33.+Effective+Data+Visualization:+The+Right+Chart+for+the+Right+Data&hl=fi&sa=X&ved=2ahUKEwigfr-cyrjvAhXIAxAlHVWwCOUQ6AEwAHoECAQQA#v=onepage&q=33.%20Effective%20Data%20Visualization%3A%20The%2> [viitattu 20.5.2021]

Few, S. 2021. Data Visualization for Human Perception. Interaction-design. WWW-dokumentti. Saatavilla: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception> [viitattu 20.5.2021]

Friendly, M., & Weiner, H. 2020. A History of Data Visualization and Graphic Communication. Researchgate. WWW-dokumentti. Saatavilla: https://www.researchgate.net/publication/343696001_A_History_of_Data_Visualization_and_Graphic_Communication_-_Introduction [viitattu 20.5.2021]

Gramlich, M. 2020. What is structured, semi structured and unstructured data? Michael-gramlich. WWW-dokumentti. Saatavilla: <https://www.michael-gramlich.com/what-is-structured-semi-structured-and-unstructured-data/> [viitattu 20.5.2021]

Healy, K. 2018. Data Visualization A practical introduction. E-kirja. Princeton:Princeton University Press. Saatavilla: <https://socviz.co/index.html#preface> [viitattu 20.5.2021]

IBM. 2016. Big Challenge or Big Opportunity? Ibm. PDF-dokumentti. Saatavilla: https://www.ibm.com/watson/infographic/discovery/assets/pdf/data_landscape_infographic_1.pdf [viitattu 20.5.2021]

Kanerva, J. 2016. TIEDON VISUALISOINTI – PARHAAT KÄYTÄNNÖT. In: infograafikko. WWW-dokumentti. Saatavilla: <https://infograafikko.fi/infograafikka/tiedon-visualisointi-parhaat-kaytannot/> [viitattu 20.5.2021]

Kirk, A. 2016. Data Visualisation: A Handbook for Data Driven Design. E-kirja. London: Sage. Saatavilla: <https://books.google.fi/books?id=dXKfDwAAQ-BAJ&printsec=frontcover&dq=Data+Visualisation:+A+Handbook+for+Data+Driven+Design&hl=fi&sa=X&ved=2ahUKEwiNp5XzxrjvAhUNkMMKHb7gBgkQ6AEwAHoECAUQAq#v=onepage&q=Data%20Visualisation%3A%20A%20Handbook%20for%20Data%20Dri> [viitattu 20.5.2021]

Koski, J. 2015. Informaation visualisointi. Medium. WWW-dokumentti. Saatavilla: <https://johanneskoski.medium.com/informaation-visualisointi-e8615483680e> [viitattu 20.5.2021]

Krystian, M. 2017. Why Text Plays a Crucial Role in Data Visualization. Infogram. WWW-dokumentti. Saatavilla: <https://infogram.com/blog/text-plays-crucial-role-data-visualization/> [viitattu 20.5.2021]

Magnuson, L. 2016. Data Visualization: A Guide to Visual Storytelling for Libraries. E-kirja London: Rowman&Littefield. Saatavilla: <https://books.google.fi/books?id=wxrMDAAAQBAJ&printsec=frontcover&dq=Data+Visualization:+A+Guide+to+Visual+Storytelling+for+Libraries&hl=fi&sa=X&ved=2ahUKEwjj05jbyLjvAhWptIsKHUv7AgYQ6AE-wAHoECAQQQAq#v=onepage&q=Data%20Visualization%3A%20A%20Guide%20to%20Vis> [viitattu 20.5.2021]

Miyazaki, M. 2020. A Brief History of Data Analysis. Flydata. WWW-dokumentti. Saatavilla: <https://www.flydata.com/blog/a-brief-history-of-data-analysis/> [viitattu 20.5.2021]

Ossama, E. 2018. Data analysis and Visualization Using Python. Dubai: Apress.

Pierson, L. 2020. Determining Target Audiences for Data Visualizations. Dummies. WWW-dokumentti. Saatavilla: <https://www.dummies.com/programming/big-data/data-science/determining-target-audiences-for-data-visualizations/> [viitattu 20.5.2021]

Roshani, R. & Alben, D. M. 2021. Handbook of Research on the Impact of Deep Learning and IoT on Multi-Industry applications. London: Business Science Reference.

Tableau. 2021. Data visualization beginner's guide: a definition, examples, and learning resources. Tableau. WWW-dokumentti. Saatavilla: <https://www.tableau.com/learn/articles/data-visualization> [viitattu 20.5.2021]

Treehouse Technology Group. 2021. The 9 Most Important Data Visualization Best Practices. Treehouse Technology Group. WWW-dokumentti. Saatavilla: <https://treehousetechgroup.com/the-9-most-important-data-visualization-best-practices/> [viitattu 20.5.2021]

Tufte, E. R. 2000. Visual Display of Quantitative Information. In E. R. Tufte, Aesthetics and Technique in Data Graphical Design. Graphics Press. PDF. Saatavissa: http://www.econ.upf.edu/~michael/visualdata/tufte-aesthetics_and_technique.pdf [viitattu 20.5.2021]

Tufte, E. R. 2006. Beautiful Evidence. Cheshire: Graphics Press.

Wikipedia. 2018. Toimiala. Wikipedia. WWW-dokumentti. Saatavissa: https://fi.wikipedia.org/wiki/Toimiala#cite_note-Suomalainen-1 [viitattu 20.5.2021]

Wikipedia. 2021. SAS (software). Wikipedia. WWW-dokumentti. Saatavissa: [https://en.wikipedia.org/wiki/SAS_\(software\)](https://en.wikipedia.org/wiki/SAS_(software)) [viitattu 20.5.2021]

Kuva- ja taulukko luettelo

Kuva 1. Prosessikuvaus prosessuaalisista vaiheista miro-esityksenä. Karlsson, K. 2021.

Kuva 2. Structured vs unstructured data – what is the difference. Pickell, D. 2018. Saatavissa: <https://fi.pinterest.com/pin/156781630766604038/> [viitattu 18.4.2021].

Kuva 3. Metsäsektorin vienti suomesta vuosittain ja maanosittain. Powerbi. Karlsson, K. 2021.

Kuva 4. Kartta napoleonin sotaretkestä venäjälle. Minard, C.J. 1869. Saatavissa: https://en.wikipedia.org/wiki/charles_joseph_minard#/media/file:minard.png [viitattu 20.3.2021]

Kuva 5. Tiedon visualisointi – parhaat käytännöt. Kanerva, J. 11.9.2016. Saatavissa: <https://infograafikko.fi/infografiikka/tiedon-visualisointi-parhaat-kaytanot/> [viitattu 15.3.2021]

Kuva 6. Graafinen käsitys: teoria, kokeilu, ja toteuttaminen modernin grafiikan kehittälyssä. Cleveland, W.S. & McGill, R. 1984. Saatavissa: https://www.researchgate.net/figure/cleveland-and-mcgills-elementary-perceptual-tasks-all-visual-representations-of_fig2_221513994 [viitattu 20.5.2021]

Kuva 7. Data visualisointi käytännöllinen perehdyttäminen. Healy, K. 2018. Saatavissa: <https://socviz.co/groupfacettx.html> [viitattu 20.5.2021]

Kuva 8. Data visualisointi käytännöllinen perehdyttäminen. Healy, K. 2018. Saatavissa: <https://socviz.co/lookatdata.html> [viitattu 20.5.2021]

Kuva 9. Suomen ja ruotsin väkiluvun muutos. Powerbi. Karlsson, K. 2021.

Kuva 10. Data pythonissa. Karlsson, K. 2021.

Kuva 11. Python visualisointi. Karlsson, K. 2021.

Kuva 12. Suodatettu kuva. Karlsson, K. 2021.

Kuva 13. Yritysten toimipaikat taulukko Stat.Fi tilastotietokannassa. Karlsson, K. 2021.

Kuva 14. Esikäsitelty data Powerbi ohjelmassa. Karlsson, K. 2021.

Kuva 15. Taulukko painiketta varten. Karlsson, K. 2021.

Kuva 16. Painike Powerbi:ssä. Karlsson, K. 2021.

Kuva 17. Visualisointi Powerbi:ssä. Karlsson, K. 2021.

Kuva 18. Visualisointi Powerbi:ssä. Karlsson, K. 2021.

Kuva 19. Ruutukaappaus stat.fi tilastotietokannasta. Karlsson, K. 2021.

Kuva 20. Data Excelissä. Karlsson, K. 2021.

Kuva 21. Powerpoint datakentät. Karlsson, K. 2021.

Kuva 22. Väkiluku kunnittain, animoitu. Karlsson, K. 2021.

Kuva 23. Koulutusaste data Excelissä. Karlsson, K. 2021.

Kuva 24. Powerbi:n viivadiagrammissa käytettävät kentät. Karlsson, K. 2021.

Kuva 25. Slicer-työkalu Powerbissä. Karlsson, K. 2021.

Kuva 26. Koulutusaste maakunnan ja sukupuolen mukaan. Karlsson, K. 2021.

Kuva 27. Mediaanitulo data Excelissä. Karlsson, K. 2021.

Kuva 28. Postinumeroalue data Qgis-ohjelmassa. Karlsson, K. 2021.

Kuva 29. Attribuuttitaulukko Qgis ohjelmassa. Karlsson, K. 2021.

Kuva 30. Tekstiedostojen lisääminen. Karlsson, K. 2021.

Kuva 31. CSV-tiedosto Qgis ohjelmassa. Karlsson, K. 2021.

Kuva 32. Yhdistetty taulukko. Karlsson, K. 2021.

Kuva 33. Ruutukaappaus perustyylistä Mapbox studiossa. Karlsson, K. 2021.

Kuva 34. Väriskaala Mapbox. Karlsson, K. 2021.

Kuva 35. Valmis visualisointi. Karlsson, K. 2021.

Kuva 36. Valmis visualisointi. Karlsson, K. 2021.

Kuva 37. Verkkoviestinnän saavuttamat katsoja määrät. Karlsson, K. 2021.

Taulukko 1, Metsäteollisuuden ulkomaankauppa, tuonti ja vienti kuukausittain.
Luke. 2020.