DATAN VISUALISOINTI JA VÄÄRISTYMÄT

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tuotantotalouden diplomityö

2025

Sampsa Kihlakaski

Tarkastajat: Tutkijatohtori, dosentti Lasse Metso

Yliopisto-opettaja, DI Leena Tynninen

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tuotantotalous

Sampsa Kihlakaski

Datan visualisointi ja vääristymät

Tuotantotalouden diplomityö

2025

71 sivua, 25 kuvaa, 8 taulukkoa ja 2 liitettä

Tarkastaja: Tutkijatohtori, dosentti Lasse Metso & Yliopisto-opettaja, DI Leena Tynninen

Avainsanat: Datan visualisointi, visualisointien vääristymät

Hallitus- ja johtoryhmätyöskentelyllä on avainrooli yrityksen menestyksekkäässä ohjaamisessa. Ohjaamisen on perustuttava oikea-aikaiseen ja laadukkaaseen raportointiin, jossa visualisoinnit on tehty tarkoituksenmukaisesti kuvaamaan kohteena olevia ilmiöitä. Tutkimuksen tavoitteena on analysoida case yrityksen liiketoiminta-alueen liiketoimintakatsausta - kuvaavatko visualisoinnit tarkasteltavaa ilmiötä vääristämättä sitä?

Tutkimus on toteutettu tapaustutkimuksena. Teoreettinen viitekehys perustui datan oikeelliseen esitystapaan ja vääristymien välttämiseen. Empiirinen osuus sisälsi datan tuottajien näkökulmasta datan luomat rajoitteet, liiketoimintakatsauksen teoriaan pohjautuvan analyysin sekä tiedon hyödyntäjien näkemyksen katsauksen nykytilasta.

Analyysi osoitti, että vaikka visualisoinnit olivat pääosin informatiivisia ja tarkoituksenmukaisia, havaittiin kehityskohteita läpinäkyvyyden, mittarien yhdenmukaisuuden ja värien käytön osalta. Työn tuloksena laadittiin kehitysehdotuksia raportoinnin eheyden, ymmärrettävyyden ja visuaalisen johdonmukaisuuden parantamiseksi.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

School of Engineering Science

Industrial Engineering and Management

Sampsa Kihlakaski

Data visualization and distortions

Master's thesis

2025

71 pages, 25 figures, 8 tables and 2 appendices

Examiners: Post-doctoral Researcher, Docent Lasse Metso & University Lecturer, DI Leena Tynninen

Keywords: Data visualization, visualization distortions

Board and management play a key role in the successful steering of a company. Effective steering must be based on timely and high-quality reporting, in which visualizations are designed purposefully to represent the phenomena under examination. The objective of this study is to analyze the business review of a case company's business area and to determine whether the visualizations depict the examined phenomena without distortion.

The research was conducted as a case study. The theoretical framework focused on the accurate presentation of data and the avoidance of visual distortions. The empirical section included three perspectives: data-related constraints from the viewpoint of data producers, a theory-based analysis of the business review, and insights from information users regarding the current state of the business review.

The analysis revealed that while the visualizations were generally informative and appropriate, areas for improvement were identified in terms of transparency, consistency of metrics, and use of colors. As a result, development proposals were formulated to enhance the integrity, comprehensibility, and visual coherence of the reporting.

ALKUSANAT

Diplomityö on usein viimeinen etappi pitkässä opintopolussa – niin myös tässä tapauksessa. Sen valmistuminen merkitsee paitsi opintojen päättymistä, myös eräänlaista yhteenvetoa vuosien varrella kertyneestä osaamisesta, kokemuksesta ja kasvusta. Opiskelu työn ohessa vaatii venymistä, ja erityisesti lapsen syntymä kesken prosessin toi uuden näkökulman ajankäyttöön ja merkityksellisyyteen – ja samalla muistutti, että tärkeimmät asiat elämässä eivät mahdu aikatauluihin.

Suurin kiitos kuuluu kuitenkin vaimolleni Kaisalle. Hänen päivittäinen tukensa, ymmärryksensä ja kannustuksensa ovat olleet tämän työn hiljainen selkäranka. Ilman häntä tätä tutkielmaa ei olisi koskaan saatettu valmiiksi – eikä koko opiskelu olisi ollut mahdollista. Kiitos.

Espoossa 18. marraskuuta 2025.

Sampsa Kihlakaski

Sisällysluettelo

Tiivistelmä

Abstract

Alkusanat

1 Johdanto ... 8
1.1 Tausta ja motivaatio työlle ... 8
1.2 Tutkimuksen tavoitteet ja rajaukset ... 9
1.3 Työn rakenne ... 10
2 Datan visualisointi - määritelmä, vääristäminen ja työkalut ... 11
2.1 Visualisointien lyhyt historia ... 11
2.2 Visualisoinnin määritelmä ja paikka data-analytiikan prosessissa ... 15
2.3 Datan esikäsittely ... 20
2.4 Visualisointikeinot ja värit ... 22
2.5 Tiedon vääristäminen visualisointien avulla ... 31
2.6 Käytetyimmät visualisointityökalut ... 40
3 Tutkimusmenetelmät ja aineisto ... 44
3.1 Tapauksen kuvaus ... 44
3.2 Tapaustutkimus ... 44
3.3 Aineisto ja tiedon keräämisen menetelmät ... 45
4 Liiketoimintakatsauksen analyysi ... 49
4.1 Tiedon käsittelyn haasteet ja prosessit ... 49
4.2 Taulukot ja donitsit ... 52
4.3 Pylväs- ja viivadiagrammit ... 56
4.4 Työn keskeiset tulokset ... 67
5 Johtopäätökset ... 69
5.1 Vastaukset tutkimuskysymyksiin ... 69
5.2 Tutkimuksen merkitys, rajoitteet ja jatkotutkimus ... 70
Lähteet ... 72

Liitteet

Liite 1. Datan tuottajien työpajamuotoiset teemahaastattelut

Liite 2. Tiedon hyödyntäjien näkökulma

ILMOITUKSET

Turnitin

Tämän opinnäytetyön alkuperäisyys on tarkistettu Turnitin – samankaltaisuuden tarkastuspalvelun avulla.

Tekoäly

Tämän opinnäytetyön kirjoittaja Sampsa Kihlakaski on käyttänyt seuraavia tekoälytyökaluja opinnäytetyön tekemisessä:

ChatGPT

Työkalun käytön tarkoitus: Löytää yleisiä työkaluja laajemmin kirjallisia lähteitä laajentamaan työn teoriakatsantoa. Etsitty myös sekä ChatGPT:llä että perinteisin keinoin löytyneiden lähteiden avoimia versioita luettavaksi. Summattu myös artikkeleita, jos tiivistelmä ei ole ollut tarpeeksi kattava.

Selvitys työkalun käytöstä: ChatGPT:tä on pyydetty etsimään avainsanoilla kirjallisuutta. Pyritty löytämään lähteistä versiot, jotka eivät olisi maksumuurin takana. Syötetty PDF-versio artikkelista työkaluun ja pyydetty tiivistämään.

Vastuu

Kirjoittaja, Sampsa Kihlakaski, ottaa täyden vastuun tämän opinnäytetyön sisällöstä ja on tarkistanut ja muokannut mahdollisen tekoälytyökalujen käytön myötä syntyneen sisällön.

1 Johdanto

Opinnäytetyössä tutkitaan datan visualisoinnin oikeellisuutta, hyödyntäen laajaa ammatillista kokemusta ja aiheeseen liittyvää kirjallisuutta. Tavoitteena on rakentaa teoriakehys, joka varmistaa, että datan visualisoinnit ovat sekä tarkkoja että tarkoituksenmukaisia, edistäen siten luottamusta ja läpinäkyvyyttä liiketoiminnassa. Teoriakehystä tullaan hyödyntämään kohdeyrityksen liiketoimintajohdon raportoinnin arviointiin ja parantamiseen.

1.1 Tausta ja motivaatio työlle

Nykypäivän yhä monimutkaistuvassa liiketoimintaympäristössä datan visualisointi on noussut korvaamattomaksi työkaluksi ymmärryksen syventämisessä ja päätöksenteon helpottamisessa. Kehittyneiden data-analytiikan ja visualisointityökalujen ansiosta eri alojen ammattilaiset pystyvät aiempaa paremmin tulkitsemaan monimutkaisia kokonaisuuksia, tunnistamaan trendejä ja tekemään perusteltuja päätöksiä. Internetin aikakaudella visualisoinneista on tullut kriittinen väline, jonka avulla raakadata muunnetaan yksinkertaisempaan, helposti tulkittavaan muotoon mahdollistamaan nopeampi ja valistuneempi päätöksenteko. (Alhadad 2018; Nguyen et al. 2021; Lisnic et al. 2023; Lisnic et al. 2024)

Visualisointien tekijöiden vallan kasvaessa heidän vastuunsa datan eettisestä esittämisestä korostuu. Harhaanjohtavat tai manipuloidut visuaaliset esitykset voivat vääristää todellisuutta ja johtaa vakaviin seurauksiin paitsi yrityksille ja sidosryhmille, myös laajemmin yhteiskunnalle. Yhteiskunnan näkökulmasta medialla tulisi olla korostunut vastuu visualisointien neutraalista esittämisestä. Eturyhmien edut ja ennakkoasenteet voivat paistaa niistä läpi, kuten Welhausenin (2022) tutkimus osoittaa - Yhdysvaltojen opioidikriisi esitettiin mediassa kansanterveysongelmana, kun taas crack-epidemiat kriminologisena ongelmana. Esityksillä pyritään usein edistämään esittäjän tai hänen edustamansa tahon etua, erilainen narratiivi vaikuttaa ihmisten saamaan apuun ja jakaa yhteiskuntaa entisestään (Lisnic et al 2023, 2; Wellhausen 2022). Koska päätöksentekijöillä ei useinkaan ole aikaa, halua tai osaamista perehtyä alkuperäiseen raakadataan, kantaa visualisoinnin laatija merkittävää vastuuta siitä, että esitetty tieto on oikeellista, tasapuolista ja eettisesti kestävää (Alhadad 2018, 60-61). (Nguyen et al. 2021; Lisnic et al. 2024; Lisnic et al. 2023)

Johdon tai asiakkaiden kouluttaminen on asiantuntijatyössä eräs suurimmista tavoista luoda lisäarvoa ymmärryksen lisääntymisen kautta. Asiantuntijan toimiessa esittelijänä myös esittelijän vastuu on tärkeä ymmärtää ja muistaa, esittelijän asiantuntemukselle annetaan usein suuri painoarvo ja voi usein olla niinkin, ettei yleisöllä ole kykyä tai haluakaan ymmärtää läpikäytävää asiaa kovin syvällisesti. Esittelijän näkökulmaa ja näkemystä oikeista toimenpiteistä yleensä pyydetään joko etukäteen tai osana esitystä.

Opinnäytetyön motivaatio kumpuaa tarpeesta käsitellä mahdollisia vääristymiä visualisoinneissa. Yhdistämällä teoreettiset näkemykset käytännön kokemuksiin, tutkimus pyrkii tarjoamaan vankan kehyksen oikeelliselle datan visualisoinnille. Datan visualisointien tulee olla totuudenmukaisia, läpinäkyviä ja luotettavia, parantaen siten liiketoiminnan käytäntöjen ja päätöksenteon eheyttä.

1.2 Tutkimuksen tavoitteet ja rajaukset

Tutkimusongelma on datan visualisointi ja visualisointien käyttäminen ja niissä mahdollisesti esiintyvät vääristymät. Tutkimus on rajattu koskemaan visualisointikeinojen tarkoituksenmukaisuutta. Tutkimuksessa ei oteta kantaa datan oikeellisuuteen eikä keskitytä analysoimaan mittareiden ja visualisointien oleellisuutta liiketoiminnan johtamisessa, mutta tarkastellaan datan esikäsittelyä sen merkittävän roolin vuoksi. Tutkimuksen ulkopuolelle rajataan visualisointien käytännön toteuttaminen.

Tutkimuksen avulla pyritään vastaamaan seuraaviin tutkimuskysymyksiin:

1. Mitä asioita yrityksen tulisi huomioida datan visualisoinnissa?
2. Miten datan visualisoinnin avulla voidaan vääristää vahingossa tai tarkoituksellisesti visualisoinnin antamaa informaatiota?
3. Kuvaavatko yrityksen raportoinnissa käytetyt visualisoinnit esitettävää dataa tarkoituksenmukaisesti ja neutraalisti tarjoten pohjan oikeaan tietoon perustuvalle päätöksenteolle?

Tutkimusta varten haastateltiin henkilöitä kohdeyrityksen business control -osastolta sekä analysoidun liiketoimintakatsauksen hyödyntäjiä. Tutkimuksen avulla pyritään selvittämään yrityksessä käytössä olevien visualisointitapojen tarkoituksenmukaisuutta ja antavatko ne

oikeanlaisen kuvan esitettävästä asiasta. Tarkoituksena on luoda arvoa kohdeyritykselle parantamalla raportoinnin ymmärrettävyyttä ja laatua.

1.3 Työn rakenne

Teorialuku jakautuu kirjallisuuskatsaukseen, visualisointien historiakatsantoon, käytettyihin visualisointityökaluihin sekä muutamaan case-tutkimukseen. Tutkimusmenetelmät ja aineisto -luvussa käsitellään käytettyjä menetelmiä sekä kuvaillaan aineiston keruuta. Liiketoimintakatsauksen analyysissä on yhdistetty kolme näkökulmaa tarkastelukohteittain selkeyden lisäämiseksi. Lopussa kokonaisuus vedetään yhteen ja annetaan yritykselle kehitysehdotuksia tutkimuksen pohjalta. Työn rakennetta selventää input/output -kaavio taulukossa 1.

Taulukko 1. Input/output -kaavio, esitys työn aineistosta ja tuloksista.

Input (aineisto)	Käsittely	Output (tulokset ja johtopäätökset)
Tutkimusongelma, motivaatio, aiheen merkitys	Luku 1. Johdanto	Tutkimuskysymykset, rajaukset ja työn rakenne
Kirjallisuuskatsaus: Visualisoinnit osana data-analytiikan prosessia, visualisointien vääristäminen, värit, työkalut	Luku 2. Datan visualisointi - määritelmä, vääristäminen ja työkalut	Teoreettinen viitekehys visualisointien arviointiin
Case-yrityksen kuvaus, haastattelut ja aineiston keruu	Luku 3. Tutkimusmenetelmät ja aineisto	Tutkimusasetelma, aineiston luonne ja keruumenetelmät
Liiketoimintakatsauksen visuaalit, teoriasta nousseet näkökulmat	Luku 4. Liiketoimintakatsauksen analyysi	Visualisointien arviointi ja tulokset - havaitut vääristymät ja kehityskohteet
Tutkimuksen tulokset	Luku 5. Johtopäätökset	Tutkimuskysymyksiin vastaaminen, kehitysehdotukset, tutkimuksen merkitys ja jatkotutkimuskohteet

2 Datan visualisointi - määritelmä, vääristäminen ja työkalut

Vanhan sanonnan mukaan kuva kertoo enemmän kuin tuhat sanaa. Elgendi et al. (2020) ovat tutkineet väitettä ja todenneet kuvan kertovan itseasiassa saman verran kuin 30 000 sanaa, joka kuvaa osuvasti kuvien voimaa. Näinpä luvussa tullaan esittelemään visualisoinneille määritelmiä sekä hyviä ja erityisesti huonoja tapoja, joita tulisi välttää mahdollisimman edustavien visualisointien luomiseksi. Luvussa tullaan antamaan katsaus yleisimmin käytettyihin visualisoinnin työkaluihin, joiden mahdollisuudet toki ulottuvat syvemmällekin data-analytiikan ketjuun kuin vain visualisointien piirtämiseen.

2.1 Visualisointien lyhyt historia

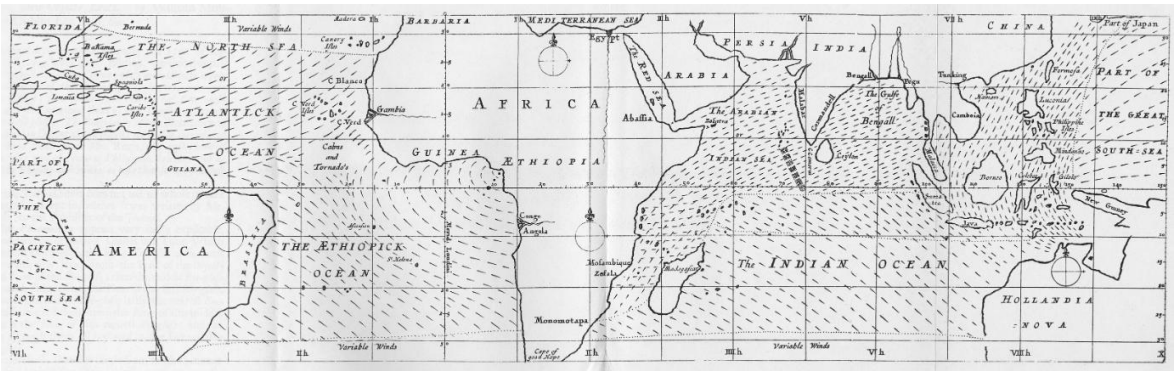
Historian ymmärtäminen antaa syvyyttä ja erilaisia harmaan sävyjä nykyhetken analysointiin. Ihmiset ovat historian alkuhämäristä asti pyrkineet viestittämään asioita visualisoinneilla, joskin alkuun käytössä olivat matkan varrelta löytyneet kepit ja kävyt, sittemmin tekoälyllä luodut kokonaiset maailmat. Tässä aliluvussa käydään läpi lyhyesti visualisointien historiaa.

Tietojen visualisoinnin varhaisimmat muodot voidaan tunnistaa jo esihistoriallisilta ajoilta, jolloin ihmiset käyttivät luolamaalauksia ja kivikaiverruksia tiedon esittämiseen. Maalaaminen sekä kuvilla tai symboleilla viestiminen on edeltänyt laajemmin käyttöön otettua kirjoitettua tekstiä ja sillä on voitu viestiä monimutkaisiakin asioita. Tietoja erilaisista reiteistä on varmasti merkitty luontoon nuolilla tai erilaisilla viilloilla puiden rungoissa – visualisoitua tiedon välittämistä sekin, vaikkakin pääasiassa kadonnut aikojen saatossa. (Yellowfin, 2023, Wainer & Friendly (2022, 460-463)

1600-luvulta alkaa fyysisen mittaamisen kulta-aika, oli kyseessä sitten ajan, etäisyyden tai avaruuden mittaaminen. Friendly (2006) on kasannut erinomaisesti läkähdyttävän listauksen vuosisadan saavutuksista: analyyttisen geometrian ja koordinaattisysteemit (Descartes ja Fermat), mittausvirheen ja arvioinnin teoriat (pohjautuen Galileon analyysiin Tycho Brahen tähdestä 1572), todennäköisyysteorian synty (Pascal ja Fermat) ja demograafisten statistiikkojen alku (John Graunt 1662). Alkunsa sai William Pettyn (1665) vapaasti suomennettuna

poliittinen aritmetiikka tai laskenta, joka tutkii väestöä, maa-alueita, veroja sekä tavaroiden arvoa, joiden avulla ymmärretään paremmin valtion varallisuutta. Seuraavassa esitellään 1600-luvun tärkein kaavio, joka yhdisti monta aikakauden keksintöä.

Edmond Halleyn henkilöhistoriaa ja saavutuksia on kuvattu monessa teoksessa (Armitage, 1966; Cook, 1998; Ipsen, 2004), mutta hänen ehdottomasti tärkein jälkipolville jäänyt saavutuksensa on kuvassa 1 esitetty kuvaus vallitsevista tuulista valtamerillä. Välttämättömänä perustana Halleyn saavutuksille toimivat Galileon vuosina 1593 keksimä lämpömittari ja 1638 barometri, jossa lämpötila ja ilmanpaine ovat keskeisinä meteorologisina komponentteina. Halley uskoi Galileon tavoin vahvasti siihen, että suora empiirinen havainnointi on antiikkisia uskomuksia parempi tie kehittää maailmantuntemusta. Halley käytti sen verran laajaa repertuaaria erilaisia keksintöjä ja datapisteitä, että lopputuloksena syntynyttä karttaa voi pitää lähes ensimmäisenä big datan käyttötapauksena. On käsittämätöntä ajatella, että 1600-luvulla on pystytty havaitsemaan maan pyörimisliikkeen vaikutus tuuliin, auringon lämmön epätasainen jakautuminen maapallolla sekä vuodenaikojen vaihtelun vaikutus tuulien suuntiin.



Kuva 1. Edmond Halleyn esitys vallitsevista tuulista valtamerillä (Halley 1686).

1800-luvun alussa nähtiin ennen näkemättömän räjähtävä kasvu tilastollisten graafien ja teemakarttojen maailmassa 1700-luvun lopun keksintöjen innoittamana. Kaikki statistiikkaa kuvaavat kuvaajat kehiteltiin 1800-luvun alussa, mukaan lukien aiemmin mainitut pylväs- ja piirakkakaaviot, histogrammit, viivakuvaajat, aikasarjakuvat, käyräkuvat, scatter-plotit tai hajontakaaviot. Teemakartografiassa kartoitus eteni yksittäisistä kartoista kattaviin atlas-

tyyppisiin kokoelmiin, jotka sisälsivät tietoa kattavasti monista aiheista kuten taloudellisista, sosiaalisista, moraalisista ja lääketieteellisistä asioista. (Friendly 2005, 2006)

Atlasten perustana toimi vuonna 1825 Ranskan oikeusministeriön perustama keskitetty kansallinen rekisteri rikosten raportoinnista, jota kerättiin kaikilta alueilta neljänneksittäin. 1833 André-Michael Guerry hyödynsi tarjolla olevia rikos-, lukutaito-, itsemurha-, lahjoitus- ja muita tietoja rakentaakseen ensimmäisen tietoja yhdistelevän eheän analyysin, jota voidaan pitää nykyisen yhteiskuntatieteen pohjana. Mielenkiintoista oli, että rikos- ja lukutaitodataa yhdistelemällä Guerry onnistui todistamaan vääräksi näkemyksen niiden negatiivisesta korrelaatiosta. Lukutaidon lisääntyminen ei vähentänyt rikollisuutta, kuten ajateltiin. Guerry laajensi näkemystään valtioiden välille ja vertaili Englannin ja Ranskan rikollisuutta, josta hän sai Ranskan tiedeakatemian Moynton-palkinnon statistiikassa. (Friendly 2005, 2006; Guerry, 1833, 1864).

Jos aikakausi 1600-luvulta 1800-luvun lopulle oli esitysgrafiikan kehityksen kulta-aikaa, niin Friendly ja Denis (2000) esittelee 1900-luvun alkupuoliskon grafiikan moderniksi pimeäksi ajaksi, jolloin uusia innovaatioita ei juuri nähty, mutta aiemmat innovaatiot saivat jalansijaa julkisessa keskustelussa ja koulujärjestelmissä. Sinällään viittaus ihmiskunnan pimeisiin aikoihin ei osu täysin kohdalleen, koska aiemmat innovaatiot levisivät ja kehitys kulki eteenpäin. Suuri vaikutus grafiikan alhoon oli varmasti Fisherin 1925 julkaisemalla Statistical Methods for Research Workers, vapaasti suomennettuna Tilastolliset menetelmät tutkijoille, joka toi mahdollisuuden laskea tarkkoja todennäköisyyksiä ja virheen todennäköisyyksiä (Friendly & Denis 2000). Toki kaksi maailmansotaa tuskin helpottivat kehitystä.

1960-luvun tienoilla alkoi graafisen esityksen uusi tuleminen John. W. Tukeyn ja Jacques Bertin alustamana. Tukeyn (1962) The Future of Data Analysis (data-analytiikan tulevaisuus) pyrki nostamaan data-analyysin omaksi tutkimusalueekseen ja Bertinin (1967) Semiologie Graphique pyrki järjestämään graafisten esitysten visuaaliset ja havaintoelementit datan ominaisuuksien ja suhteiden mukaan. Tietokoneiden kehitys mahdollisti uusia suuntia, vaikkakin korkearesoluutioiset grafiikat ottivat aikansa ennen yleiseen käyttöön leviämistä. Tässäkin alkuun varmasti oli haasteena tietokoneiden koko ja käyttäjäjoukon rajallisuus, joka helpotti vähitellen tietokoneiden koon ja hinnan kutistuessa. (Friendly, 2006) Tukeyn ja Bertinin rinnalla aikakaudelta on mainittava Edward Tufte, jonka kirjat The Visual Display of Quantitative Information (1983) ja Visual Explanations (1997) voi nähdä visuaalisen kommunikaation perusteoksina. Tuften (1983) pyrkimys selkeyteen, tarkkuuteen ja

tehokkuuteen sekä tarpeettomien elementtien poistamiseen voidaan nähdä osana Bertinin aloittamaa linjaa asioiden järjestämisestä ja kategorisoimisesta.

Friendly (2005) summasi periodin 1950-1975 loppupuolen kehityksen kolmella merkittävällä yhteistyöalueella: tietojenkäsittelytieteen tutkimus Bell Laboratoriossa (ohjelmistotyökalut, C-kieli, UNIX), data-analyysin kehitys (EDA, psykometriikka) sekä näyttö- ja syöttöteknologian kehitys (kynäpiirturit, graafiset päätelaitteet, hiiret). Edellä mainitut tarjosivat uusia mahdollisuuksia tilastollisten ideoiden ilmaisemiseen ja datan esittämiseen, joka johti räjähdysmäiseen kasvuun visualisointimenetelmissä ja -tekniikoissa.

2000-luvulla tietojenkäsittelykyvykkyyden kasvu ja joka kodista löytyvät tietokoneet ovat moninkertaistaneet mahdollisuudet luoda erilaisia staattisia visualisointeja, mutta luoneet myös kyvykkyyden dynaamisten visualisointien luomiseen, joka on nostanut visualisointien esityskyvyn täysin uusille tasoille. Tekninen kehitys on laajentanut visualisointien kehitystä ja tekemistä yritysten teknillisiltä osastoilta massoille, joka on luonut täysin uusia käyttötapauksia ja helpottanut monen ihmisen elämää. Power BI:n ja Tableaun yksinkertaiset käyttöliittymät ja datan sisäänlukemisen helppous yhdistettynä vaivattomaan grafiikan luomiseen ovat tehneet niistä nykyaikaisen organisaation raportoinnin kulmakiven. Tulevaisuudessa tullaan hyödyntämään entistä enemmän tekoälyä, josta ensivauhdit on saatu viime vuosina erilaisten generatiivisten tekoälysovellutusten, kuten ChatGPT:n muodossa. Generatiivinen tekoäly on juuri rantautumassa vahvasti eri organisaatioihin ja käyttötarkoituksia etsitään kuumeisesti.

Asioissa on valitettavasti usein kaksi puolta. Visualisointien mahdollisuudet ovat räjähtäneet, mutta samalla on ilmaantunut entistä selvempi haaste grafiikan käytön etiikasta ja oikeellisuudesta. Edellä esiteltyjen, historiallisesta tarkastelusta selvinneiden suurmiesten grafiikoissa etiikka oli sisäänrakennettuna - oli halu ymmärtää ja kuvata paremmin maailman ilmiöitä tuoden ratkaisuja aikakauden ongelmiin parhaan ymmärryksen mukaisesti. Nykyään, kun kuka tahansa voi luoda grafiikkaa helposti, kuvaan astuu mahdollisuus vääristellä ja esittää dataa omaa tarkoitusta tukemaan. Tätä on varmasti tapahtunut kautta historian, kuten Darrell Huffin 1955 julkistamasta How to Lie with Statistics (Miten valehdella statistiikalla) voi päätellä, mutta mahdollisuudet eettisesti arveluttavan toiminnan ympärillä ovat moninkertaistuneet teknisen kehityksen myötä. (Pandey et al. 2015; Best 2015; Welhausen 2022)

2.2 Visualisoinnin määritelmä ja paikka data-analytiikan prosessissa

Data-analytiikka on laaja ja moniportainen aihe, josta tutkimuksessa valikoitu tiedon tai datan visualisointi on pitkän prosessin lopputulos. Tiedon visualisoinnissa on yleensä kaksi päätapaa, toimia analyysin apuna tai viestintävälineenä. Tarkoituksena on herättää mielenkiintoa ja opastaa lukijaa kohti parempaa ymmärrystä käsiteltävästä asiasta. Visualisoinneissa tarkoituksenmukaista onkin, että viestitty päätelmä on sama riippumatta käytetystä ohjelmasta tai sen luoneesta käyttäjästä (Evergreen & Metzner 2013, 5). Liiketoiminnan päätöksenteon ohella politiikassa ja journalismissa datan visualisointien perusteella tehdyt analyysit ohjaavat päätöksentekoa ja niiden avulla pyritään luomaan datapisteistä vetoavia viestejä ja tarinoita (Correl 2019). Erilaisten visualisointien käyttö lisääntyy merkittävästi koko ajan käyttäjäkunnan sekä kohdeyleisö laajentuessa jatkuvasti perinteisemmästä liiketoimintajohdon päätöksenteosta kohti jokaista kansalaista. Erilaiset visualisointimallit ovat mahdollistaneet jopa suhteellisen merkityksettömän tiedon muovaamisen vakuuttavaksi oivallukseksi (Ragnathan 2022, 1-2).

Kirjallisuudesta on löydettävissä useita määritelmiä datan visualisoinnille. Lisäksi datan visualisoinnista käytetään muita ilmauksia, joita ovat esimerkiksi tiedon visualisointi, visuaalinen analytiikka ja tieteellinen visualisointi. Datan visualisoinnin määritelmiä on koottu taulukkoon 2.

Taulukko 2. Datan visualisoinnin määritelmiä.

Datan visualisoinnin määritelmän esittäjä	Määritelmä datan visualisoinnille
Azzam et al. (2013)	"Prosessi, joka perustuu laadullisiin tai kvantitatiivisiin tietoihin ja johtaa raakadataa edustavaan kuvaan, joka on katsojien luettavissa ja tukee tietojen tulkintaa, tutkimista ja viestintää."
Bertin (1983), Lee et al. (2013) mukaan	"Datan visualisointi on algoritminen prosessi, jossa tiedot kartoitetaan kahteen tasomaiseen muuttujaan (x- ja y-pisteet) ja rajoitettuun määrään muuttujia, kuten väriin, kokoon, arvoon, rakenteeseen, muotoon ja suuntautumiseen."
Fisher et al. (2014)	"Visualisointi tarjoaa tehokkaan tavan datan muuttamiseksi ymmärrettävään muotoon."
Melocon & Warner (2017)	"Datan visualisointi tarkoittaa visuaalisten keinojen käyttöä suurien tietomäärien esittämiseksi tiettyjen parametrien tai luokkien mukaan."
Post (2011)	"Datan visualisointi on sovellusvetoinen kenttä, joka pyrkii aina tyydyttämään asiakkaitaan ja mukautumaan monien sovellusalueiden vaatimuksiin, kulttuureihin ja työnkulkuihin."
SAS Institute Inc. (n.d.)	"Datan visualisointi on tietojen esittämistä kuva- tai graafisessa muodossa."

Azzam, Evergreen, Germuth ja Kistler (2013) esittävät datan visualisoinnin olevan kvalitatiivisiin tai kvantitatiivisiin tietoihin pohjautuva prosessi, joka on sellaisessa muodossa, että käyttäjä pystyy ymmärtämään sen. Datan visualisointi tukee datan tulkintaa, tutkimista sekä kommunikointia (Azzam et al. 2013). Datan visualisointia on kuvattu myös algoritmisena prosessina, jossa data luokitellaan tasomaisiin muuttujiin (x- ja y-pisteisiin) sekä tiettyyn määrään muuttujia, joista esimerkkejä ovat väri, koko, arvo, rakenne ja muoto (Bertin 1983, Lee, Sips & Seidel 2013 mukaan). Tässä on ehkä hyvä huomata, että visualisointeja voidaan luoda kolmiulotteisina x-, y- ja z-pisteisiin, mutta se ei juuri muuta määritelmän sanomaa. Esitellyistä muuttujista väri on moniulotteisin, sillä värisokeus on yleistä ja värisävyjen huono käyttö haastaa normaalinkin värinäön omaavia. Myös muiden attribuuttien hyviä ja huonoja käyttötapoja käydään läpi.

Fisherin et al. (2014) mukaan datan visualisointi mahdollistaa tehokkaan tavan datan muuttamiseksi ymmärrettävään muotoon. Datan visualisoinnin on esitetty olevan sovellusvetoinen ala ja pyrkivän tyydyttämään asiakkaiden tarpeet sekä mukautumaan useiden eri sovellusalueiden asettamiin vaatimuksiin, prosesseihin sekä kulttuureihin (Post 2011). Tämä on itseasiassa mielenkiitoinen näkökulma, sillä suurin sovellusriippuvuus on varmastikin data-

analyysin vaiheissa, jotka ovat ennen visualisointien luomista. Itse visualisoinnin luominen, kun tiedetäänsen pohjana olevat arvot, on suoraviivaista, joskin helpompaa sovellusavusteisesti. Niin ikään datan visualisoinnin on kuvattu tarkoittavan datan esittämistä graafisessa muodossa (SAS Institute Inc. n.d.). Meloncon ja Warner (2017) ovat esittäneet määritelmän datan visualisoinnista, ja he kuvaavat sen tarkoittavan erilaisten visuaalisten keinojen käyttöä suurien tietomäärien esittämiseksi joidenkin tiettyjen luokkien tai parametrien mukaisesti. Analysoitavana oleva data voidaan esittää erilaisten visualisointikeinojen avulla kaavioina tai kuvaajina (Meloncon & Warner 2017).

Edellä esitellyissä määritelmissä on kaikissa omat vahvuutensa. Melocon & Warner (2017) painottavat enemmän tilastollista puolta, Fisher et al. (2014) ehkä pykälää idealistisempaa näkökulmaa, Bertin (1983) kuvailee itse visualisointia, Azzam et al. (2013) on prosessilähtöinen ja sisältää oletuksen läpinäkyvyydestä, Post (2011) ottaa sovellusvetoisen näkökulman ja SAS Institutella kuvaus on yksiselitteinen. Moninaisuus kuvaa, kuinka hankalasti määriteltävästä asiasta on kyse tuoden erinomaisesti esille eri näkökulmia, joista visualisointeja voi tutkia. Monista kulmista ehkä tärkeimpänä nousee tarve ymmärtää tutkinnan kohteena olevaa ilmiötä ja päästä kiinni prosessiin, jolla tarkasteltava kuva on tuotettu. Näin Azzam et al. (2013) läpinäkyvä prosessi on kuvauksista ehdottoman tärkeä ja SAS Instituten konstailematon toteamus visualisointien sinällään rajaamattomista tyypeistä tai variaatioista täydentää prosessilähtöistä kuvausta. Yhdistelmän voisi kirjoittaa seuraavasti

”Datan visualisointi on tietojen esittämistä kuva- tai graafisessa muodossa niin, että visualisoinnin luontiprosessi raakadatasta kuvaan on jäljitettävissä ja visualisointi tukee esitettävien tietojen tulkintaa, tutkimista ja viestintää.”

Datan visualisoinnilla on todettu olevan merkittävä rooli yrityksen raportointitoimenpiteissä ja datan analysoinnissa. Business intelligence -työkaluilla ja niiden sisältämien visualisointiominaisuuksien hyödyntämisellä voidaan edesauttaa suurien datamäärien ymmärtämistä. Datan visualisointia hyödynnetään sisäisen raportoinnin lisäksi viestittäessä tietoa yrityksen ulkoisille sidosryhmille. Visualisointi voi olla kriittisessä roolissa ulkoisessa raportoinnissa, ja se sitouttaa sekä tukee sidosryhmien toimintaa lisäämällä niiden kykyä ymmärtää yrityksen tuottamaa tietoa, jolloin ne voivat osaltaan osallistua tulosten arviointiin. (Azzam et al. 2013.) Datan visualisoinnilla on siis tärkeä rooli datan tulkinnassa ja sitä voidaan hyödyntää erilaisten mallien tunnistamiseen tai tulosten viestimiseen (Goldstone, Pestilli & Börner 2015).

On hyvä muistaa, että kaikki BI-työkalujen ominaisuudet eivät välttämättä paranna datan ymmärrettävyyttä. Tästä on muutamia mielenkiintoisia näkökulmia. Meloncon ja Warner (2017) esittävät BI-työkalujen vuorovaikutteisuus- ja animaatio-ominaisuuksien jopa heikentävän tiedon ymmärrettävyyttä. Heer & Robertson (2007) taas tarkemmassa tutkimuksessaan löysivät argumentteja puolesta ja vastaan, vähemmän yllättävästi käyttötarkoitus ja tietynlainen samankaltaisuus visualisointien kesken paransi ymmärrettävyyttä. Visualisointien luojien onkin hyvä muistaa tehdä niistä mahdollisimman ohjaavia. Joskus voi olla tarkoituksenmukaista rajata käyttäjän muutosmahdollisuudet minimaalisiksi, jotta tieto välittyy halutulla tavalla käyttäjille. Datan visualisointi ei välttämättä aina osoita tietojen välisiä suhteita oikein, vaan visualisointi saattaa vääristää niitä, jolloin dataa mallintavien kuvioiden tulkitsemisessa voi tulla väärinymmärryksiä. Riskinä datan visualisoinnissa on siis se, että käytetään sellaisia datan visualisointikeinoja, jotka vääristävät dataa eivätkä esitä todellista tilannetta, jolloin visualisoinnin pohjalta tehtävät johtopäätökset voivat olla vääriä. (Azzam et al. 2013., Heer & Robertson 2007, Meloncon & Warner (2017)

Datan visualisoinnin paikka data-analytiikan prosessissa

Datan visualisointi on kapeahko, mutta nimensä mukaisesti näkyvä data-analytiikan prosessin osa. Prosessille on löydettävissä toisistaan eroavia tulkintoja, mutta pääviestin tai idean voidaan arvioida olevan sama eri lähteissä. Muun muassa Runkler (2016) on esittänyt näkemyksensä siitä, minkälaisen prosessin data käy läpi sen keräämisestä visualisoinnin tai tulkinnan tasoille. Muskan et al. (2022) keskittyy vaiheiden kuvauksessa ensimmäiseen kolmeen Runklerin vaiheeseen, jotka johtavat jo valmiiseen visualisointiin. Tunnistettuja neljää data-analyysin vaihetta havainnollistetaan kuvassa 2.

Valmistelu	Datan esikäsittely	Analyysi ja visualisointi	Jälkivaiheet
Suunnitteleminen	Siivoaminen	Visualisointi	Tulkinta
Datan kerääminen	Täydentäminen	Ennustaminen	Arviointi
Datan valinta	Virheiden korjaaminen	Luokittelu	Dokumentointi
	Suodattaminen	Korrelaatio	
	Standardointi	Regressio	
	Transformointi	Klusterointi	
	Integrointi		

Kuva 2. Data-analyysin vaiheet (mukaillen Runkler 2016, 3; Muskan et al. 2022)

Data-analyysin ensimmäisessä vaiheessa suunnitellaan päämäärä ja kerätään tarvittava data (Runkler 2016, 3; Muskan et al 2022, 1711). Usein dataa kerätään monista lähteistä, joten järjestelmien on pystyttävä käsittelemään eri lähteistä saatavaa dataa (Chawla et al. 2018, 37). Data voi olla strukturoitua, strukturoimatonta tai semi-strukturoitua (Chawla et al. 2018; Gamage 2016). Strukturoitu data on helposti tallennettavissa relaatiotietokantoihin ja esimerkkinä tällaisesta datasta on laskentataulukko riveineen ja sarakkeineen (Gamage 2016). Esimerkkejä strukturoimattomasta datasta ovat videot, erilaiset tekstit (kuten sähköpostit) ja sosiaalisen median sisältö (Chawla et al. 2018; Gamage 2016).

Datan keräämisen jälkeen dataa täytyy esikäsitellä, johon kuuluu muun muassa datan siivoaminen, suodattaminen ja mahdollisten virheiden korjaaminen (Runkler 2016, Famili et al 2011; Muskan et al 2022). Tällaisia virheitä voivat olla poikkeavat datapisteet. Kyseisessä vaiheessa data muokataan haluttuun muotoon, useimmiten strukturoituun muotoon. (Runkler 2016.) Käsitelty data tallennetaan tyypillisesti tietovarastoon, jolloin data on helposti saatavilla ja hyödynnettävissä eri analyysi- ja BI-työkaluilla työskenneltäessä ja joissakin tapauksissa datalähteet voidaan suoraan yhdistää raportteihin (Famili et al. 1997, 11). Data-analyysin prosessissa datan esikäsittely on usein eräs eniten aikaa vievistä vaiheista, sillä pohjadata on usein huonolaatuista ja dataa joudutaan keräämään erilaisissa muodoissa eri lähteistä. Datan esikäsittelyä kuvataan tarkemmin luvussa 2.2.

Kolmas data-analyysin vaihe on datan analyysi ja visualisointi. Dataa on mahdollista analysoida erilaisilla metodeilla, kuten regressiomalleilla ja korrelaatiomenetelmillä eikä perusarvot keskiarvosta, mediaanista ja määristä ole yleensä pahitteeksi (Runkler 2016, 59, 67; Muskan et al 2022, 1711). Datan visualisointien analysointi on työn keskiössä ja siihen on olemassa lukuisia tapoja, joskin tärkeää on hyödyntää käyttötarkoitukseen sopivaa visualisointia (Muskan et al 2022, 1711–1712; Hill 2024).

Viimeisessä data-analyysin vaiheessa tulkitaan, arvioidaan ja dokumentoidaan tuloksia, jotka datasta on analyysin avulla tuotettu (Runkler 2016). Datan tulkinta, usein visualisointien pohjalta, on monitahoinen asia, sillä tulkintoja voi olla lähes yhtä paljon kuin tulkitsijoita. Visualisointiosuudessa onkin tärkeää huolehtia mahdollisimman yksiselitteisistä ja selkeistä malleista, jotta useat eri käyttäjät päätyisivät saman suuntaisiin johtopäätöksiin niitä luettuaan. Toki samat luvut voivat tarkoittaa hieman eri asioita eri taustoista tuleville (Pandey et al. 2014) ja erilaiset näkökulmat tuovat syvyyttä analyysiin, mutta täysin vastakkaisia johtopäätöksiä ei pitäisi olla tehtävissä.

2.3 Datan esikäsittely

Datan esikäsittely on ohutta neljänneksen Runklerin (2016) mallista vievää palasta suurempi ja aikaa vievempi kokonaisuus. Kokonaisuuden laajuutta kuvaa lainaus Famili et al. (1997):

> *" Data preprocessing consists of all the actions taken before the actual data analysis process starts",*

joka vapaasti suomennettuna olisi Datan esikäsittely käsittää kaikki tehdyt toimenpiteet ennen datan analyysiprosessin alkua. Esimerkin ajankäytöstä tai kuluista saa Adolfsson et al. (2010, 5) tutkimuksesta, jossa tilastollisista haastattelututkimuksista tehdyssä analyysissä editoinnin, joka tässä yhteydessä tarkoittaa datan käsittelyä, kustannus tutkimusten kokonaiskustannuksista oli noin kolmannes.

Datan esikäsittelyn haasteet tai vaiheet voi jakaa Runklerin (2016) mukaan hienojakoisesti seitsemään osaan: siivoaminen, täydentäminen, virheiden korjaaminen, suodattaminen, standardointi, transformointi sekä integrointi. Runkler (2016) katsoo esikäsittelyä yleisestä näkökulmasta, kun taas Diehl (2004) käsittelee enemmän big dataa ja laajoja datamassoja ja esittelee oman näkemyksensä neljästä osasta: siivous, integraatio, transformaatio ja reduktointi. Samat problematiikat esiintyvät Muskan et al. (2022) hieman eri nimillä, joten vaikka eri lähteissä käsitteet hieman eroavat, puhutaan yleensä samasta asiasta.

Siivous on esitellyistä viheliäisin ja usein aikaa vievin toimenpide ja sisältää osaltaan sekä täydentämisen että virheiden korjaamisen (Muskan et al 2022; Runkler 2016; Diehl 2024). Virheiden korjaamiseksi ne on ensin löydettävä, ja Runkler (2016, 23–25) jakaakin ne stokastisiin virheisiin, kuten mittaus- tai siirtovirheisiin, ja deterministisiin, systemaattisista ongelmista johtuviin virheisiin. Näiden lisäksi Runkler tunnistaa poikkeamat, jotka voivat olla joko stokastisia tai deterministisiä, mutta myös inhimillisiä virheitä. Monet yrityksissä käytetyt tietokannat rakentuvat lähdejärjestelmistä tulevien tietovirtojen varaan. Yleisen kokemuksen mukaan lähdejärjestelmissä on usein sekä puutteellisia määrittelyjä käytettävissä kentissä, lukuarvokenttään voi usein syöttää tekstiä, että myös ajassa eläviä ohjeistuksia, jolloin tietokantaan päätyy monella eri logiikalla tallennettua tietoa. Tiedon muokkaamiseen voidaan kehittää puoliautomatisoituja ratkaisuja (Hoogland et al. 2011, 6), mutta harmillisen usein tieto joudutaan korjaamaan käsin, mikä on aikaa-vievää ja kallista. Diehl (2004, 11) esittelee mahdollisuuden täydentää dataa perustuen korrelaatioon muihin datapisteisiin. Aina suuretkaan erot tai merkittävästi poikkeavat havainnot (eng. outlier) eivät ole virheitä

datassa, vaan antavat tärkeitä signaaleja (Runkler 2016, 25). On tunnistettavissa pilkkuvirheitä, joita muodostuu järjestelmäkohtaisesta pilkun tai pisteen käytöstä desimaalierottimena, jotka voisi automaattisesti korjata (Hoogland et al. 2001, 17). Esimerkki poikkeavasta havinnosta, joka ei ollut ns. tuhatvirhe, on Trumpin Liberation day osakekaupankäynnissä, joka sai aikaan voimakasta markkinaliikehdintää (Vantage 2025). Virheiden korjaamisen kanssa onkin hyvä olla tarkkana, jottei oikeellista dataa tule korjattua ja väärin korjatut virheet saattavat tuottaa oman päänvaivansa ja virhelähteensä analyysivaiheessa. Tärkeintä olisikin dokumentoida virheiden korjaamisen prosessi, jotta käyttäjät olisivat siitä tietoisia.

Suodattaminen on virheiden korjaamista järeämpi siivoamiskeino, joka sopii datan tasoittamiseen ja kohinan poistamiseen ja sitä kautta ymmärtämiseen (Runkler (2016, 27–32; Famili et al. 1997, 11). Yleisesti käytetään liukuvaa mediaania tai liukuvaa keskiarvoa. Liukuvien määreiden käyttö toimii myynnin mittaamiseen. Suomessa on yleistä lomailla heinäkuussa ja jouluna, jolloin usein myynnissä nähdään tiputukset kahtena kuukautena. Tätä voi ehkäistä 12 kuukauden liukuvalla myynnillä, josta saa verkkaisesti muuttuvan indikaation myynnin kehityksen suunnasta.

Transformointi ja standardointi tekevät datasta yhteismitallista, auton tehokkuus voi olla hevosvoimia, kilowatteja tai watteja, mutta olisi hyvä valita yksi yhteinen vertailtava muuttuja ja usein jatkokäytössä datan vaaditaankin olevan tietyssä muodossa analyysiä varten (Runkler 2016, 33; Diehl 2004, 13).

Integrointi on yleinen ongelma, kun data tuodaan monesta eri lähteestä (Diehl 2004, 12–13; Runkler 2016, 35). Usein datan transformointiin ja integraatioon hyödynnetään ETL (extraction, transformation, loading) -prosessia, jolla dataa saadaan yhteismitallistettua (Runkler 2016, 35). Integroinnin perusongelma on sirpaloituneet datalähteet, joita halutaan yhdistellä visualisoinnissa ja raportoinnissa. Usein nämä järjestelmät ovat eri aikakausilta ja kyvykkyydet tuottaa dataa tietovarastoihin vaihtelevat merkittävästi, joka tuottaa haasteita datan tuonnissa ja esikäsittelyssä. Osa haasteista voi heijastella visualisointiin asti, kun tarpeellisia dimensioita ei saada käyttöön.

Reduktointi on ehkä moniulotteisin ja tarkoittaa käytännössä oikeellisen analysoinnin tekemistä pienemmästä edustavasta dataotannasta (Diehl 2004, 14; Famili et al. 1997, 13). Reduktointi on tärkeää suurten datamassojen kanssa, joilla koko datasta analyysin tekeminen

saattaa olla jopa mahdotonta. Tästä perinteisempi ongelma ja ratkaisu on päivätason datan aggregointi kuukausitasolle, jolloin käytettävyys ja tiedonhaun nopeus oleellisesti paranee.

2.4 Visualisointikeinot ja värit

Oikeaoppisten visualisointien luomiseen voi koota kolmen kokonaisuuden alle: läpinäkyvyys, tarkkuus ja vastuullisuus. Näille kolmelle lähtökohdalle voidaan esittää kaksi lyhyempää tiivistelmää. Laajalle yleisölle tunnettu William of Ockhamin mukaan nimetty ”Occam’s razor”, joka esitetään yleensä joko ”entities must not be multiplied beyond necessity” tai tutummin ”The simplest explanation is usually the best one”, joista molemmat korostavat yksinkertaisuuden tärkeyttä. Toinen, hieman edelliseen pohjautuva, on Tuften (1997, 73) esittelemä ”…design strategy of the smallest effective difference: Make all visual distinctions as subtle as possible, but still clear and effective”, joka vapaasti suomennettuna olisi pienimmän mahdollisen eron tavoitteleminen. Tavoite on tehdä visualisointien erot niin pieniksi kuin mahdollista, mutta silti selkeiksi ja puhutteleviksi (Tufte 1997; Meloncon & Warner 2017). Toki tavoite on pyrkiä esittämään mahdollisimman vähällä mahdollisimman tarkasti esiin tuotava tieto. Occamin partaterä on näistä kahdesta laajemmin sovellettavissa eri elämänaloille, kun taas Tuften versio sukeltaa hieman lähemmäs visualisointien maailmaa.

Läpinäkyvyys on näistä Jobin et al. (2019) tutkimuksen mukaan tulevaisuutta ennakoivassa AI-kirjallisuudessa yleisimmin mainittu eettinen periaate. Läpinäkyvyyden kulmakivenä toimii käytettyjen menetelmien, lähteiden ja rajoitteiden selkeä esittäminen visualisointien yhteydessä. Selkeästi tehdyt lähdemerkinnät vähentävät mahdollista epäselvyyttä ja antavat näkyvyyden lähteenä toimivalle materiaalille. Visualisointien usein pohjautuessa monimutkaisiin tietokantarakenteisiin ja niiden taustalla oleviin liiketoimintajärjestelmiin, täydellisen läpinäkyvyyden saavuttaminen on luonnollisesti hankalaa. Yleensä yrityksissä tunnetaan jollain tasolla yleisesti käytettyjen tietolähteiden sisältö, joka helpottaa lähdeviittausten tekemistä. Neutraalin visualisoinnin luomiseksi olisi tärkeää näyttää läpinäkyvästi ketju datan poiminnasta ja laskennasta tulosten valintaan sekä kommunikointiin. Mieluiten toki niin, että prosessi on toistettavissa toisen henkilön toimesta. Usein visualisointien tekijän harteille jää kattava vastuu tehdä relevantit huomiot pohjautuen kohdeyleisön tietämykseen eikä yksiselitteistä oikeaa tapaa ole. (Jobin et al. 2019; Best 2015; Mansoor & Harrison 2018)

Tarkkuuden määreen voi tässä yhteydessä jakaa kolmeen osaan: selkeyteen (clearness), täsmällisyyteen (precision) ja itse tarkkuuteen (accuracy). Selkeyden määreeseen päästään parhaiten luomalla yksinkertaisia ja mahdollisimman riisuttuja visualisointeja, jotka toteuttavat edellä esiteltyä Tuften (1997) esittelemää pienimmän mahdollisen eron tavoitetta. Täsmällisyydessä ja tarkkuudessa voidaan nojata samaan määritelmään, mutta laajennettuna pitää huolta visualisointien tarkoituksenmukaisuudesta kuten akseliston kuvaavuudesta ja sopivuudesta juuri kyseiseen esitystarkoitukseen. Esitettävää dataa ei tulisi manipuloida jonkun tavoitteen saavuttamiseksi, vaan pyrkiä esittämään se mahdollisimman totuudenmukaisesti. Esiteltävän tiedon on oltava tarkasti rajattu ja lähdetietojen tarkoituksenmukaisesti mainittu esitystarkoituksen mukaisesti riisuen kaikki ylimääräinen, jotta päästään tarvittavaan täsmällisyyden tasoon lopputuloksessa. Tarvittaessa näin visualisoinnin toistaminen samoilla oletuksilla on mahdollista.

Vastuullisuus on monimutkainen käsite, joka tarkoittaa vielä monesti eri asiaa eri ihmisille. Työn yhteydessä vastuullisuus jaetaan tekijän vastuuseen saattamiseen (accountability) ja visualisoinnin tekijän pyrkimykseen toteuttaa läpinäkyvyyttä ja tarkkuutta visualisointien luontiprosesseissa. Ferrell et al. (2024) osuvasti kuvaavat accountabilityä viittauksella oikeellisuuden vaateeseen ja siitä huolehtimiseen, palautteen huomioon ottamiseen sekä huomattujen virheiden tai vinoumien korjaamiseen. Yritysmaailmassa vastuullisuuden pyrkimystä voi toteuttaa selkeällä omistajuudella käytetyistä visualisoinneista, jolloin tekijällä on kannustin pyrkiä parhaaseen lopputulokseen.

Edellä esiteltiin, kuinka luoda oikeellinen visualisointi. Taulukko 3 esittelee yleisimpiä datan visualisointiin soveltuvia visualisointikeinoja. Taulukkoon on koottu lyhyet kuvaukset yleisimmistä visualisointikeinoista.

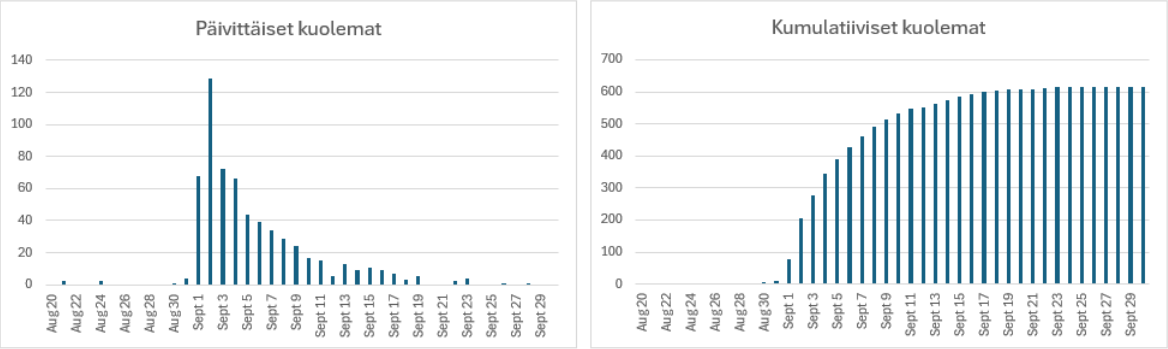
Taulukko 3. Datan visualisointikeinoja ja niiden kuvauksia (Goldstone et al. 2015; Meloncon & Warner 2017; Hill 2024; Kosslyn 2006)

Visualisointikeino	Visualisointikeinon kuvaus
Taulukko	Rivien ja sarakkeiden järjestelmällinen esitysmuoto ruudukkona. Ruudukon solut voivat sisältää esimerkiksi kielellisiä, kuvallisia tai geometrisia symboleja.
Kaavio	Esittää kvalitatiivisen tai kvantitatiivisen datan ilman määriteltyä viitejärjestelmää.
Diagrammi	Esittää kvalitatiivisen tai kvantitatiivisen datan muuttujat hyvin määritellyssä viitejärjestelmässä, kuten x- ja y-akselien koordinaateissa.
Kartta	Datan visualalinen esitysmuoto, joka perustuu datapisteiden fyysisiin (spatiaalisiin) suhteisiin sekä alueelliseen jakautumiseen.
Verkostomalli	Käytetään solmuja kuvaamaan datapisteiden joukkoja aineistossa sekä solmuja yhdistäviä linkkejä esittämään datapisteiden välisiä suhteita.
Pylväskuvaaja	Sopii datapisteiden suhteellisen tarkkaan vertailuun, varsinkin pylväiden keskinäiseen kokoon ja arvioon absoluuttisesta tasosta.
Piirakka- ja donitsikuvaaja	Sopii erityisen hyvin isoimman viipaleen tunnistamiseen, mutta myös useamman elementin vertailuun. Vaatii kuitenkin enemmän keskittymistä.

Kaavioiden on kuvattu olevan hyvä tapa visualisoida dataa. Pylväskaaviot soveltuvat useiden datapisteiden vertailuun ja ihmissilmä erottaa niistä suhteellisen tarkasti tason, johon pylväs kohoaa (Hill 2024). Pylväskaaviot ovat yleisiä ja ihmiset ovat niihin tottuneet, joten ymmärtäminen ei vaadi kummoista vaivaa. On esitetty, että pylväskaavioiden käyttämistä visualisoinnissa tulisi harkita huolella yleisö ja käyttötarkoitus huomioiden, vaikka ne voivat olla hyödyllisiä datan visualisoinnissa. Pylväskaaviota on verrattu viivakaavion kanssa, mutta tutkimuksissa on saatu hieman ristiriitaisia tuloksia siitä, kumpaa kaaviotyyppiä käyttäjät pitävät helpommin tulkittavana ja molemmilla on omat käyttötarkoituksensa. (Meloncon & Warner 2017; Hill 2024)

Erottelu pylväs- tai viivakaavioihin ei ole ainoa mahdollinen erottelutekijä, vaan pylväskaavioiden sisälläkin voidaan esittää tietoja eri tavoilla kuvan 3 mukaisesti. Vasemmalla on esitetty kuolemat koleraan päivittäin ja oikealla kumulatiivisesti (Tufte 1997, 29). Molemmilla tavoilla on vahvuutensa, vasemmassa versiossa nähdään helpommin epidemian

laantuminen, kun taas oikealla olevassa kuvassa nähdään helpommin kuolemien kokonaismäärä. Kumulatiivisuutta käytetään liiketoiminnassa usein esimerkiksi vuoden sisäiseen laskentaan budjettikauden mukaisesti. Vastaavia erilaisia käyttötarkoituksia on muillakin kaaviotyypeillä ja niitä esitellään tarkemmin aliluvussa 2.5 Tiedon vääristäminen.

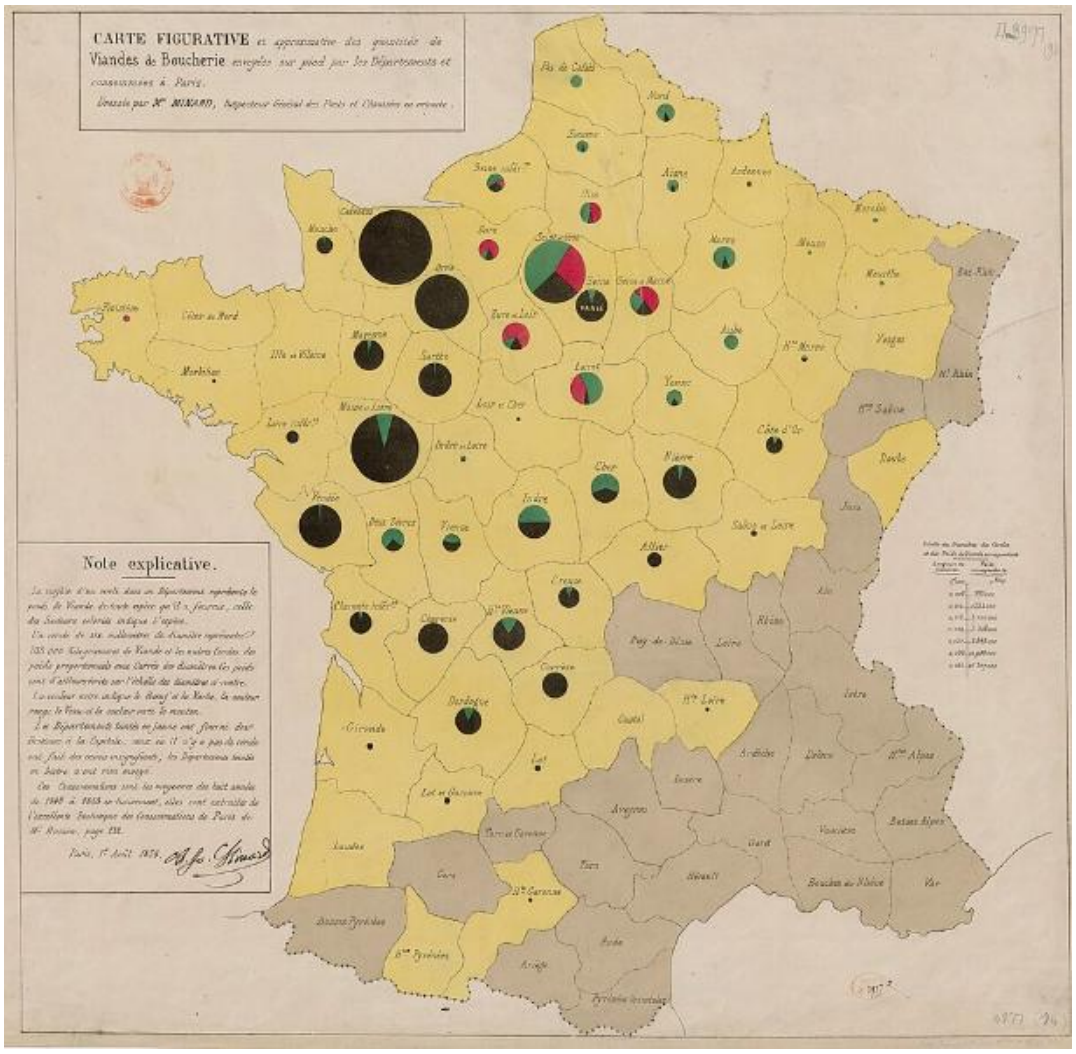


Kuva 3. Kolerakuolemat 1854 elo-syyskuussa (Tufte 1997, 29).

Yleisesti ottaen visualisoinnin tulkitsijan kannalta voi olla hyödyllistä käyttää koordinaattiviivoja tai datapisteiden tunnisteita. Kuitenkaan molempia ei tulisi sisällyttää samaan visualisointiin, vaan tarkkojen arvojen ollessa tärkeitä on hyödylliseksi visualisointikeinoksi esitetty taulukon käyttäminen. Himmeitä koordinaattiviivoja on hyvä käyttää, jos visualisoinnin tavoitteena on havainnollistaa trendejä tai arvioita. (Evergreen & Metzner 2013)

Tutkimuksessa karttavisualisointien käytöstä on huomattu esitettävän alueen tuntemisen voivan vaikuttaa datan tulkintaan ja grafiikan ymmärtämiseen (Meloncon & Warner 2017). Karttavisualisointeja on monenlaisia perinteisistä kartoista innovatiivisempiin ratkaisuihin. Kuvassa 4 on esitelty kartta, jossa esitetään piirakoilla lihan määrän ja eri tyyppien toimitukset Pariisiin. Keltainen väri erottaa lihaa toimittavat alueet haaleammalla olevista alueista, jotka eivät toimita lihaa ollenkaan. Piirakkadiagrammin koko viittaa toimitetun lihan määrään, eri väreillä kerrotaan lihan tyyppi – musta on naudanlihaa, punainen vasikanlihaa ja vihreä lampaanlihaa. Kuvasta näkeekin yhdellä silmäyksellä paljon informaatiota, kuten merkittävimmät lihan toimittajat ja niiden toimittama lihatyyppi. Karttojen vahvuus on alueellisen datan kokoaminen yhdellä silmäyksellä nähtäväksi, joka kohdeyrityksen tapauksessa voisi olla alueellinen tuotemyynti. Giering (2008) on osuvasti tutkinut alueellisen demografian ja myynnin yhteyttä yksittäisen kaupan tasolla ja todennut, että se toimii

huomattavasti paremmin kuin kokonaisuuden tasolla. Kohdeyrityksen ollessa Suomessa laajasti levittäytynyt, tarkemman tason ennusteet voisivat parantaa myyntiä. Melocon & Warner (2017) tiivistivät, että karttavisualisointia käytettäessä olisi hyvä sisällyttää karttaan vain ne tiedot, jotka ovat välttämättömiä viestin välittämiseksi. Minardin (1858) kuvassa lähestytään jo liiallista informaatiota.



Kuva 4. Minardin 1858 piirakkakartta teurastamoiden toimittamista lihoista Pariisiin (Wikicommons 2025; Myös Friendly, 1999, 4).

Minard (1858) esitteli hyvän tavan käyttää piirakkadiagrammia muutaman asian esittämiseen, johon sen on sittemmin nähty soveltuvan. Hill (2024) mukaan pylväs-, piirakka-, ja donitsidiagrammien kohdalla 5–7 kategoriaa on maksimi. Siirtola et al. (2019) tutkivat

ihmissilmän toimintaa ja havaintokykyä piirakkadiagrammilla sekä kolmella erikokoisen keskellä olevan reiän omaavalla donitsidiagrammilla. Tutkimuksessa havaittiin, että käyttäjät arvioivat kulmaa, pinta-alaa sekä donitsin reunan pituutta lähes yhtäläisesti. Havainto eroaa yleisestä oletuksesta pelkän kulman analysoimisesta. Hill (2024, 17–18) nostaa esille, että vaikka pylväskuvat ovat usein tarkempia, niin tärkeämpää olisi ymmärtää välitettävä viesti, joka voi välittyä tehokkaammin yksinkertaisuudessaan visuaalisesti näyttävämmistä piirakka- ja donitsikuvista.

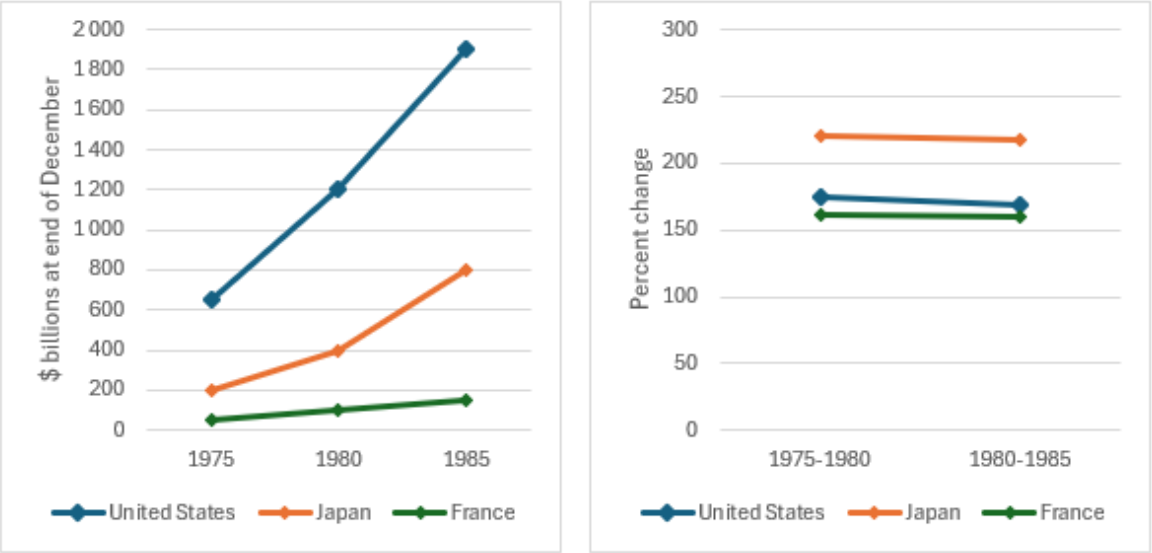
Erilaisten visualisointien lisäksi on tärkeää tarkastella, mitä tarkalleen ottaen visualisoidaan. Dougherty & Ilyankou (2021) puhuvat datan normalisoinnista, tekemisestä paremmin vertailukelpoiseksi. Absoluuttiset luvut eivät aina kerro meille kovin paljoa ilman mahdollisuutta suhteuttaa niitä kokonaisuuteen. Taulukko 4 esittääkin datan, josta voidaan tehdä erilaisia johtopäätöksiä riippuen valitusta näkökulmasta. Pelkkiä absoluuttisia kuolemia katsoessa Rhode Island vaikuttaa turvalliselta paikalta. Kun verrataan asukasmäärään, muuttuu Rhode Island enemmän keskimääräiseksi ja ajokilometreillä mitattuna vaaralliseksi paikaksi. Esimerkissä toki on kyseessä pienet luvut, joita yksikin lisäkuolema heiluttaa paljon, joten hedelmällisempää on katsoa Massachusettsin ja New Yorkin vertailua, joissa näkökulman valinnalla saadaan erilainen lopputulos. Näkökulmat on korostettu taulukkoon. Mielenkiintoinen taulukosta johdettava näkökulma on keskimääräinen ajomatka per henkilö, joka Massachusettsissa on 9 700 mailia (15 500 kilometriä) ja New Yorkissa vain 6 300 mailia (10 000 kilometriä), joka tarkoittaa 50 % enemmän ajoa keskimäärin per henkilö.

Taulukko 4. Liikennekuolemat eri osavaltioissa (Mukaillen Dougherty & Ilyankou 2021).

Osavaltio	Kolarikuolemia	Asukasluku	Ajetut mailit (miljoonia)	Kuolemat per 100 000	Kuolemat per 100 miljoonaa ajomailia
Rhode Island	59	1 057 315	8 009	5,6	0,74
Massachusetts	360	6 902 149	66 772	5,2	**0,54**
New York	943	19 542 209	123 510	**4,8**	0,76
New Jersey	564	8 908 520	77 539	6,3	0,73

Toinen esimerkki suhteuttamisesta on Kosslynin (2006, 212) esitys prosenttikasvusta eri markkinoilla ja markkinoiden koko. Kuvassa 5 Osakemarkkinoiden koko ja %-kasvu 1975–1985 (Mukaillen Kosslyn 2006, 212) näyttää, että puhuttaessa vain

prosenttikasvuista saadaan virheellinen käsitys kokonaisuudesta. Dougherty & Ilyankoun (2021) sekä Kosslynin (2006) esimerkeistä huomataan, että asioiden suhteuttaminen kokonaisuuteen on erinomaisen tärkeää oikeellisen näkökulman muodostamiseksi.



Kuva 5. Osakemarkkinoiden koko ja %-kasvu 1975-1985 (Mukaillen Kosslyn 2006, 212).

Kun edellä tarkasteltiin tarkoituksenmukaisten visualisointien käyttöä käyttötarkoituksesta riippuen, Meloncon ja Warnerin (2017) tutkimuksen tulokset osoittavat, että visualisointityökalujen suunnitteluominaisuudet voivat vaikuttaa siihen, miten visualisointien antamaa informaatiota tulkitaan ja ymmärretään. Tällaisia suunnitteluominaisuuksia ovat otsikot, selitteet, kuvatekstit sekä muut niin kutsutut ”genreominaisuudet” kuten Kittivorawongin (2024) tutkimat arvopisteet sarjoilla. Todellisten lukujen sisällyttäminen käytettävään visualisointiin on tärkeää oikean tarkkuuden hahmottamiseksi ja käytetyn visualisointikeinon oikeanlaisen tulkinnan edesauttamiseksi. Lukujen sisällyttäminen visualisointeihin voi osoittautua haasteelliseksi ja usein visualisointiohjelmissa on algoritmit, jotka päättävät käyttäjän puolesta, mitä numeroita näytetään (Kittivorawong 2024). Numeroiden näyttämiseen koodatut algoritmit eivät yleensä ymmärrä sinällään sisältöä ja niiden asettelu voi olla satunnaista. Oikein sijoiteltuna ne tuovat merkittävästi lisäymmärrystä kuvan tulkintaan. (Meloncon & Warner 2017; Kittivorawong 2024)

Värien käyttäminen datan visualisoinnissa

Värit ovat tärkeässä roolissa datan visualisoinnissa, mutta ne ovat eräs monitulkintaisimpia visuaalisia elementtejä, jonka takia niitä on hyvä käsitellä syvällisemmin. Värit ovat yksi yleisimmistä ongelmista visualisoinneissa (Nguyen et al. 2021). Värejä voidaan käyttää tiedon merkitsemiseen, elävöittämiseen sekä mittaamiseen. (Lin, Fortuna, Kulkarni, Stoner & Heer 2013). Värikoodausta voidaan hyödyntää erilaisten hälytysten ja varoitustilojen osoittamiseen, jos tietyt raja-arvot alittuvat tai ylittyvät (Glisic, Yarnold, Moon & Aktan 2014, 386–387). Usein valitut värit ovat ns. liikennevalo-skaalalla, joka on omiaan aiheuttamaan ongelmia puna- ja vihernäköhaasteellisen väestön joukossa. On hyödyllistä, että värin lisäksi näkyvillä on käsillä oleva arvo ja sen vertailukohta. Eroja on sävyissä ja tarkoituksenmukaisesti valitut värisävyt mahdollistavat dataan pohjautuvien kategorioiden tunnistamisen ja erottamisen. (Lee et al. 2013, 1746; Lin et al. 2013). Kategoriset värikoodaukset auttavat kategorioiden tunnistamista, jos ne ovat konsistentit läpi visualisointiperheiden. (Evergreen & Metzner 2013). Järjestelmällistä ja kvantitatiivista vertailua voidaan helpottaa värien kylläisyyttä hyväksikäyttämällä (Lin et al. 2013).

Värien erottuvuudessa on merkittäviä eroja, joka korostaa oikeaa värivalintaa (Goldstone et al. 2015, 464–465.) Epäloogisesti hyödynnetyt värit tai niiden kirkkaus ja kylläisyys voivat johtaa virheelliseen käsitykseen datan kategorisoinnista ja kategorioiden järjestyksestä (Lin et al. 2013), ja kääntäen oikein käytettynä voidaan korostaa varmempia datapisteitä voimakkaalla sävyllä epävarmempien ollessa hailakampia (Goldstone et al. 2015). Värisävyjä oikein käyttämällä saadaan siis tuotua helposti samaan kuvioon lisää tietoa. Yleisesti värit on hyvä valita niin, että ne erottuvat toisistaan (Lin et al. 2023) ja yllä todetusti huomioida värisokeuden mahdollisuus, 8 % miehistä ja 0,5 % naisista on puna-vihersokeita (Duodecim Terveyskirjasto, 2021). Värisokeuden rinnalla nykyisenä tekniikan aikakautena on hyvä kiinnittää eroa värien erottuvuuteen erilaisilla pääte- tai esityslaitteilla. Tätä haastaa merkittävästi yhtään suuremmissa yrityksissä käytössä olevat brändiväriavarauudet, joista erottuvien ja tarkoitustenmukaisten värien löytäminen voi usein olla haastavaa.

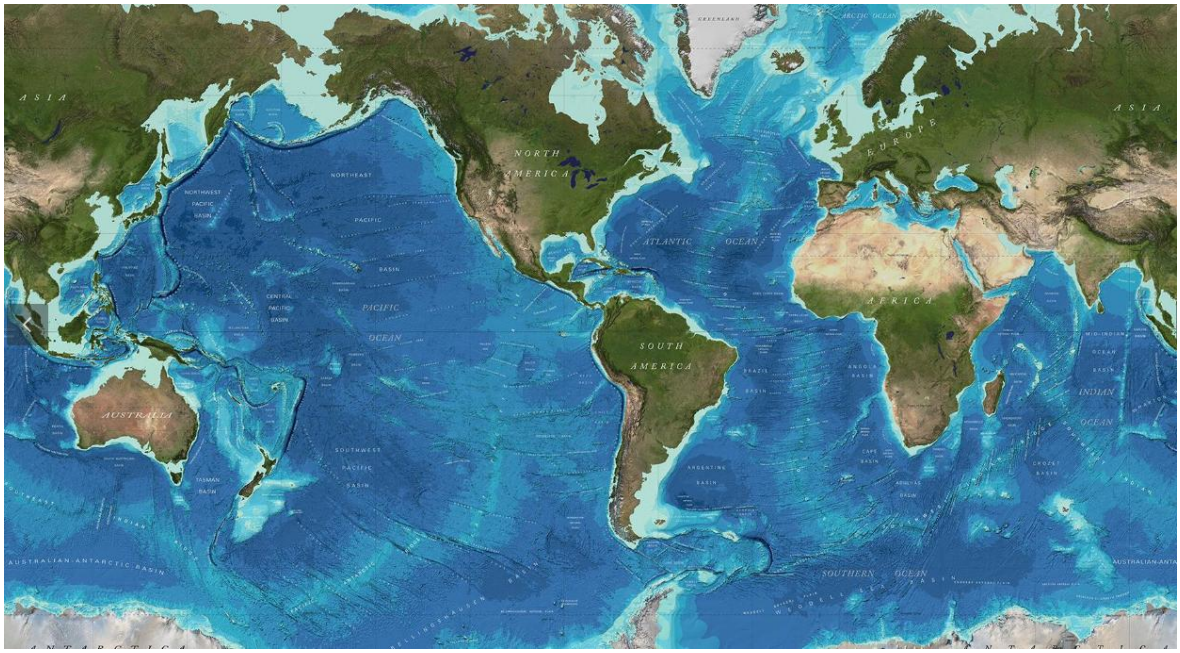
Väripaletteja koostaessa onkin siis noudatettava useaa reunaehtoa aina ihmisten rajoitteista päätelaitteiden ja yritysten markkinointiosastojen asettamiin rajoitteisiin. Tärkeintä olisi pyrkiä siihen, että visualisointiin valitut värit eivät johtaisi virheelliseen tai harhaanjohtavaan tulkintaan (Lee et al. 2013; Nguyen et al. 2021). Linin et al. (2013) ohjaa käyttämään BI-työkaluista usein löytyviä väripaletteja harhaanjohtavuuden välttämiseen, joskin siinä voi

piillä konflikti yrityksen brändivärien kanssa. Tässäkin on hyvä olla tarkkana, sillä päätyminen sateenkaariväreihin usein häiritsee tiedon tulkintaa (Nguyen et al.2021). Yleisesti värien merkitys tulee esittää raportilla selkeästi, jotta lukija voi sisäistää raportilla viestittävän ajatuksen. Kaavioiden ja kuvien tekstien ja numeroiden on oltava mielellään selkeästi erotettavissa taustaväristä luettavuuden varmistamiseksi, joka usein vaatii mieluummin suurta kuin pientä kontrastia. On hyvä kiinnittää huomioita taustan neutraaliuteen, käytännössä usein valkoiseen taustaan, jotta raportin luettavuus on hyvä. (Evergereen & Metzner 2013)

Värisokeuden ja brändivärirajoitteiden lisäksi väreillä voi olla kulttuureihin liittyviä mielleyhtymiä, joista osa saattaa olla vahvoja. Lin et al. (2013) esittelee esimerkkeinä Yhdysvalloista vihreän assosioituvan rahaan ja punaisen vihaan. Luonnollisesti mielleyhtymät vaihtelevat kulttuureittain, joka vaikeuttaa kansainvälisesti toimivien yhtiöiden visualisointien toteuttamista. Joissain kulttuureissa tiettyjen arvojen korostamiseksi tai huomion kiinnittämiseksi yleinen keltainen väri liitetään pelkuruuteen, joka voi aiheuttaa tulkintaongelmia (Evergreen & Metzner 2013). Lin et al. (2013) esittelevät hienon termin tällaiselle tietylle käsitteelle miellettyyn värivalintaan, semanttisesti resonoiva värivalinta.

Yleisesti voidaan todeta, että visualisointien pitäminen mahdollisimman neutraaleina ja rauhallisina auttaa kiinnittämään katselijan huomion tärkeimpiin elementteihin. Evergreen & Metznerin (2013, 11-13) mukaan hyvä tapa on esittää rivit taulukossa tai kuvaajassa harmaana ja korostaa huomiovärillä rivi tai pylväs, johon halutaan kiinnitettävän huomiota. Halutessa voidaan esittää väreillä kynnysarvojen alittumista ja ylittymistä (Glisic et al. 2014, 386–387), mutta silloin on hyvä muistaa edellä esitetyt rajoitteet värien käytölle.

Kuvassa 6. Väriskaalojen käyttö korkeuden ja syvyyden havainnollistamisessa (Gebco 2014) on erinomainen esimerkki värien tuomista mahdollisuuksista esittää lisäinformaatiota samassa kuvassa. Kuvassa on esitetty erilaisella väriskaalalla meren syvyyttä ja maanpinnan korkeutta, sekä tuotu mukaan aavikoituneiden alueiden korostusta. Karttavisuaalit yleisesti ovatkin nerokkaita värien käytössä, ja ihmisten kokemus karttojen lukemisesta tekevät niistä intuitiivisia.



Kuva 6. Väriskaalojen käyttö korkeuden ja syvyyden havainnollistamisessa (Gebco 2014)

Seuraavassa aliluvussa tutkitaan tiedon vääristämistä visualisointien avulla.

2.5 Tiedon vääristäminen visualisointien avulla

Data-analytiikan ja visualisointien jokapäiväistyessä niiden väärinkäyttö ja väärinymmärrykset ovat yleistyneet (Pandey et. al. 2015, 1469; Weissgerber et al. 2015). Tosiasioihin perustumattoman kaavion tai visualisoinnin luominen on suhteellisen helppoa ja vahinkojen määrä luonnollisesti korreloi vahvasti tekijöiden määrän lisääntymisen kanssa. Hankalaa tilanteesta tekee luonnollisesti se, että läheskään kaikki visualisoinnit eivät ole tarkoituksella harhaanjohtavia, vaan visualisointien tekijätkään eivät aina itse tiedä kaavion välittävän virheellistä tai jopa valheellista tietoa (Ranganathan 2022, 1–2). Mahdollista on, että lukijalle ei tarjota oikeanlaisia visualisointeja arvioida esitettyä dataa (Weisgerber et al. 2015). Tiedon vääristäminen voi siis johtua asiantuntijuuden puutteesta, mutta mahdollisuudet vaikuttaa suureenkin yleisöön voivat luoda kannustimia vääristää visualisoitavaa dataa (Pandey et al. 2015, 1469–1470; Wellhausen 2022).

Vanhan vitsin mukaan on kolmenlaisia valheita: valheita, kirottuja valheita ja tilastoja. Statistiikan ja visualisointien avulla harhaanjohtamisesta ja valehtelusta löytyykin merkittävästi

tutkimustietoa. Best (2015, 211) on ansiokkaasti listannut lukuisia teoksia tiedon vääristämisestä alkaen Darrell Huffin vuoden 1954 klassikosta How to lie with Statistics jatkuen hieman uudempaan listaukseen kaiken todistamisesta oikeaksi:

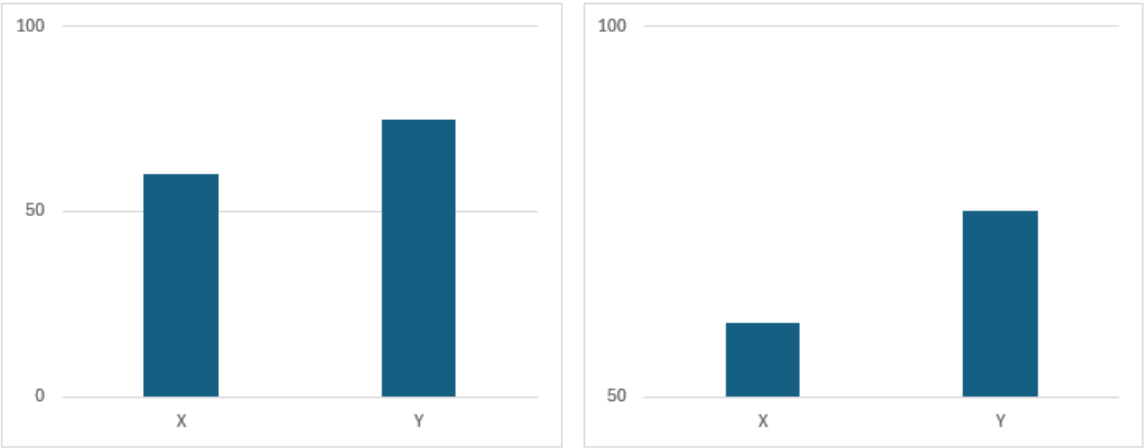
- How to Lie with Statistics (Jones, 1995)
- How to Lie with Maps (Mon monier, 1996)
- How to Tell the Liars from the Statisticians (Hooke, 1983)
- The Honest Truth about Lying with Statistics (Homes, 1990)
- How Numbers Lie (Runyon, 1981)
- Thicker than Blood: How Racial Statistical Lie (Zuberi, 2001)
- Damned Lies and Statistics (Best, 2001)
- More Damned Lies and Statistics (Best, 2004)

Ihmiset ovat alttiita ennakkoluuloille ja asenteille, eikä ominaisuus poistu visualisointeja tehdessä. Ennakkoluulot voivat vääristää tahattomasti taustadatan poimintaa ja lopullista visualisointia. Henkilölle itselleen mieluisat lopputulokset tulevat usein helpommin raportoitua kuin ennakkoasenteiden vastaiset, joiden todistamiseen tarvitaan merkittävästi enemmän tukevaa dataa. Esimerkkejä tiedon vääristämisestä löytyy mediassa, televisiossa ja internetistä. Narratiivia voidaan ohjata kohti terveysongelmaa tai rikollisuutta (Welhausen 2022) tai keinoja vääristää tietoa ovat akselin suunnan tai asteikon manipulointi, suhteettoman suurien kokojen käyttö, väärä esitystapa sekä epälineaariset mittakaavat. (Pandey et al. 2014; Tufte 2001, 55-59; Best 2015; Mansoor & Harrison 2018)

Pandey et al. (2014) käsittelivät artikkelissaan ansiokkaasti kolmea erilaista aihetta, yritysverotusta, vankilajärjestelmän vaikutuksia sekä videopelejä satunnaistetulla tutkimuksella. Tuloksena huomattiin graafisten esitysten olevan taulukkoja vakuuttavampia asioihin neutraalisti suhtautuvilla. Voimakkaammin polarisoituneille osallistujille taulukot olivat graafista esitystä vaikuttavampia, jonka epäiltiin pohjautuvan kriittisempään tapaan käsitellä esiintymää ja taulukko tarjoaa usein tarkempaa dataa. Erilainen suhtautuminen tuokin mielenkiintoisen näkökulman datavisualisointien luomiseen, jonka työn kirjoittajakin on huomannut työelämässä – joskus kuvion avulla esitetty kehityskulku on vastoin lukijan käsitystä asiasta ja tällöin usein pyydetään dataa taulukkomuotoisena eri ajanjaksoilta, jotta voidaan

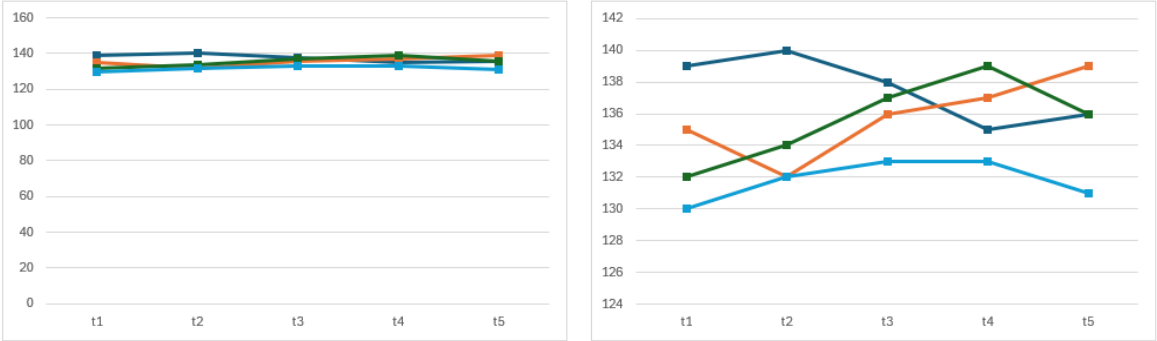
varmistaa kuvion oikeellisuus. Varmistaminen toki on erinomainen asia, koska se kertoo lukijoiden kriittisestä ajattelusta ja hyvällä tavalla haastaa visualisointien tekijöiden ajattelua. Tiedon visualisoijalla onkin suuri eettinen vastuu esittää asiat mahdollisimman totuudenmukaisesti ja olla avoin tarkoitusperistään yleisön luottamuksen säilyttämiseksi. (Pandey et al. 2014; Best 2015; Mansoor & Harrison 2018; Welhausen 2022)

Datan visualisointia voi vääristää liioittelemalla tai vähättelemällä viestiä muokkaamalla kaavioiden akseleita, dataa kuvaavaa pinta-alaa tai kuvasuhdetta. Katkaistun akselin visuaalisessa vääristymässä yksi tai useampi kaavion akseleista muutetaan muuttamalla minimi- tai maksimiarvoa. Kuvassa 7 voidaan nähdä, miten akselin minimiarvoa muuttamalla vaikutetaan X- ja Y-arvon välillä nähtävään eroon (Pandey et. al. 2015, Kosslyn 2006). Pylväskaavioiden y-akselin olisi aina hyvä alkaa nollasta (Ranganathan 2022, 23–25). Kuvan 7 oikeanpuoleisesta kaaviosta voidaan nopeasti tehdä päätelmä, että Y:n määrä on moninkertainen suhteessa X:ään. Kuitenkaan näin ei ole, kun tarkastellaan kaaviota kokonaisuudessaan niin, että asteikko alkaa nollasta kuten vasemmanpuoleisessa kaaviossa.



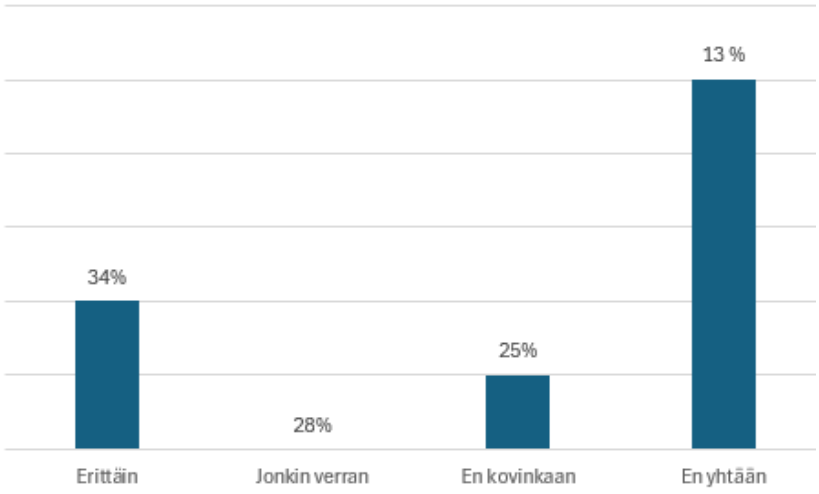
Kuva 7. Kaavion akselin arvojen muuttaminen (Pandey et. al. 2015; Koslyn 2006, 204-205)

Asteikon katkaiseminen voi olla poikkeuksellisesti mielekästä viivakaavioissa, joissa on tarpeen tarkastella pieniä eroavaisuuksia. Kuvassa 8 voidaan nähdä, että mikäli asteikko alkaa nollasta, ei arvot erotu toisistaan. Mikäli asteikkoa hieman suurennetaan, voidaan selkeämmin havaita eroja pisteiden välillä. Viivakaavio muuttuu huomattavasti luettavammaksi, kun sitä skaalataan säädetyn data-alueen mukaan. (Ranganathan 2022, 32.)



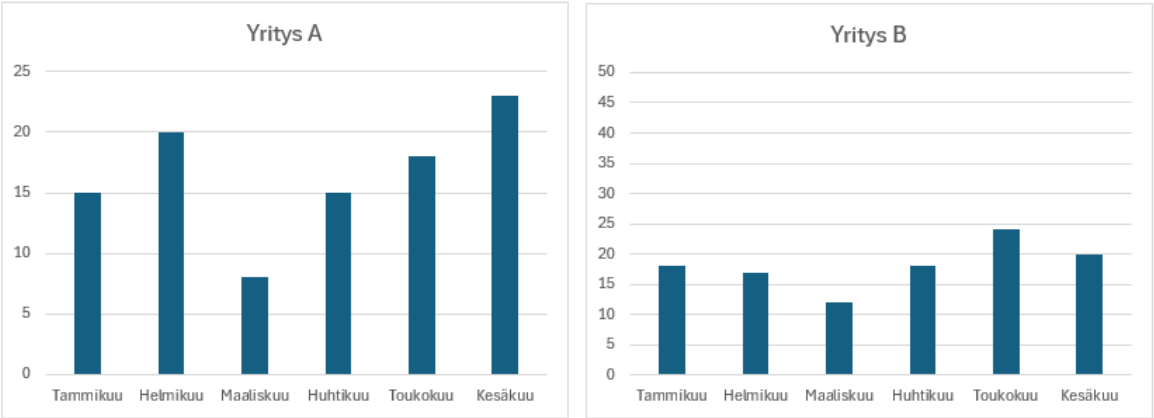
Kuva 8. Viivakaavion y-akselin skaalaus (Ranganathan 2022, 34; Myös Kosslyn 2006, 204–205)

Visualisointeja tehdessä tulee kiinnittää huomiota siis siihen, että kaavio on skaalattu oikein. Pylväskaavioissa pylväiden pituuden tulee aina olla suoraan verrannollisia niiden edustamiin suureisiin. Lisäksi akseleiden avulla voidaan hämätä eron näyttävän suuremmalta kuin mitä se todellisuudessa on. Tai vastaavasti akselit voivat olla täysin mielivaltaisia. Nguyen et al. (2021) löysivät esimerkin, jossa arvot 0,3 % ja 48,4 % oli esitetty rinnakkain niin, että 0,3 % oli puolet jälkimmäisestä. Ensin oli siis luotu visualisointi, ja sitten lisätty numerot ilman, että graafit tukivat niitä. Katastrofaalinen esimerkki oli NBC:n uutisissa esiintynyt kuva 9, jossa pylväillä ei ollut mitään tekemistä numeron kanssa (Nguyen et al. 2021).



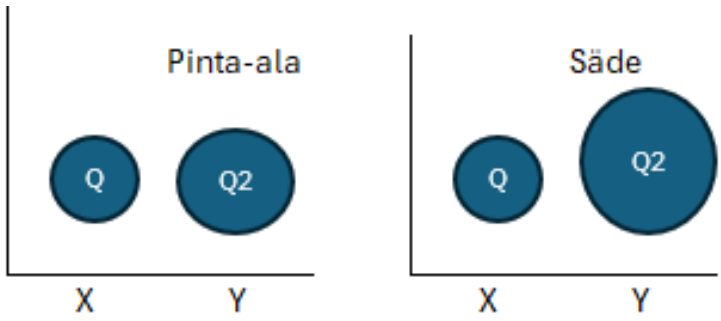
Kuva 9. NBC:n esitys "Kuinka huolestunut olet Zika-viruksesta?" (Nguyen et al. 2021, 11)

Kahden kaavion tapauksessa, jossa vertaillaan kahta muuttujaa, olisi kaaviot hyvä tehdä saman asteikon mukaan. Epäjohdonmukaiset y-akselin asteikot kahdessa vierekkäisessä kaaviossa tekevät niiden vertailusta haastavaa (Krok 2021, 3937; Ranganathan 2022, 155). Kuva 10 esittää kahden yrityksen irtisanomiset tammi-kesäkuussa. Kaaviota tarkemmin tarkasteltaessa voidaan huomata, että niiden asteikot poikkeavat paljon toisistaan. Alkuun saa käsityksen, että yritys A on irtisanonut reilusti enemmän työntekijöitä, mutta tarkemmalla tarkastelulla itseasiassa yritys B:stä on irtisanonut 10 % enemmän. Molempien kuvaajien y-akselien tulisi olla siis samat, jotta se ei johda lukijaa harhaan. (Krok 2021, 3937)



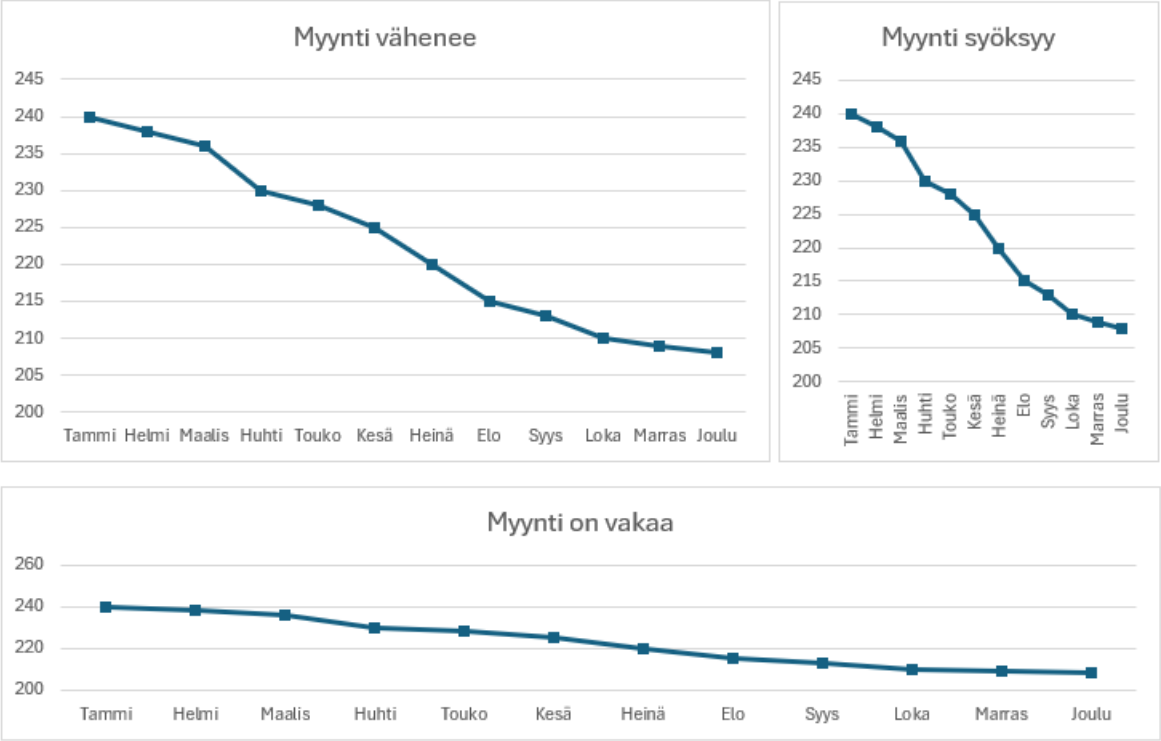
Kuva 10. Pylväskaavioiden vertailu toisiinsa, yritysten A ja B irtisanomiset kuukausittain (Krok 2021; vastaava myös Ranganathan 2022, 155.)

Visualisointien tekeminen koon mukaan on kohdannut paljon kritiikkiä. Koon mukaan tiedon visualisointi vaatii huolellista perehtymistä ja tietojen kartoittamista visualisointien avulla (Kosslyn 2006; Krok 2021; Ranganathan 2022). Vaikkei datan esittämiseen pinta-alan avulla ole selkeitä ohjetta, pitävät tutkijat yhden suhde yhteen (1:1) kartoitusta datan ja graafisen alueen välille vähiten alttiina vääristymille. Koon avulla dataa vääristettäessä datan esittämiseksi voidaan valita yksi kahdesta grafiikan muuttujasta, joka vaikuttavaa visualisointiin eksponentiaalisesti. Kuvan 11 vasemmanpuoleisessa kaaviossa koko määräytyy pinta-alana, oikeanpuoleisella ympyrän säteellä. Kuvasta nähdään selkeästi, miten eri tavat vaikuttavat tiedon vääristymiseen ja oikeanpuoleisessa kuvassa ero X- ja Y-arvon välillä vaikuttaa huomattavasti suuremmalta kuin vasemmanpuoleinen esitys, joka vastaa enemmän todellisuutta. (Pandey et. al. 2015, 1472–1473; Ngueyn et al. 2021)



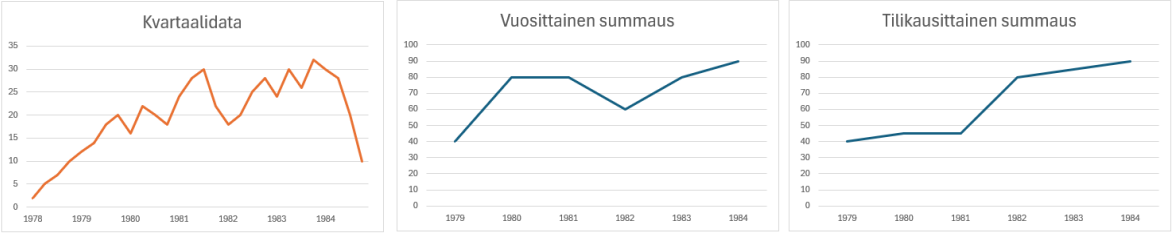
Kuva 11. Visualisoinnin vääristäminen koon avulla (Pandey et. al. 2015, 1472–1473; Nguyen et al. 2021, 11)

Kuvasuhteen avulla voidaan vääristää tietoa. Kuvasuhde määritellään sen leveyden ja korkeuden suhteena. Tämän tyyppinen vääristyminen vaikuttaa ensisijaisesti viivakaavioihin, koska se vaikuttaa yhden suureen kasvu- tai laskunopeuteen. Vaikutusta voi tarkastella kulman kaltevuuden kautta ja kaavion kuvasuhteen muokkaaminen voi muuttaa käsitystä kuvattavasta trendistä. Kuvan 12 esittämät kolme kaaviota poikkeavat toisistaan siinä, millaisen viestin ne välittävät yleisölleen. Sisältö kuvissa on identtinen, mutta akseleita on venytetty tai tiivistetty eri suuntiin. Tietoja olisi mahdollista esittää logaritmisella asteikolla, joka pienentää merkittävästi kasvua, mitä ylemmäs akselia noustaan. Kosslyn (2006, 208) esittää mahdollisuuden muuttaa pylväiden lähtökohtaa, jolloin pylvään pituus kertoo oikean tason, mutta pylväät alkavat eri tasoilta. Kuvattu esitystapa on harvinainen. (Pandey et. al. 2015; Ranganathan 2022, 68, Kosslyn 2006)



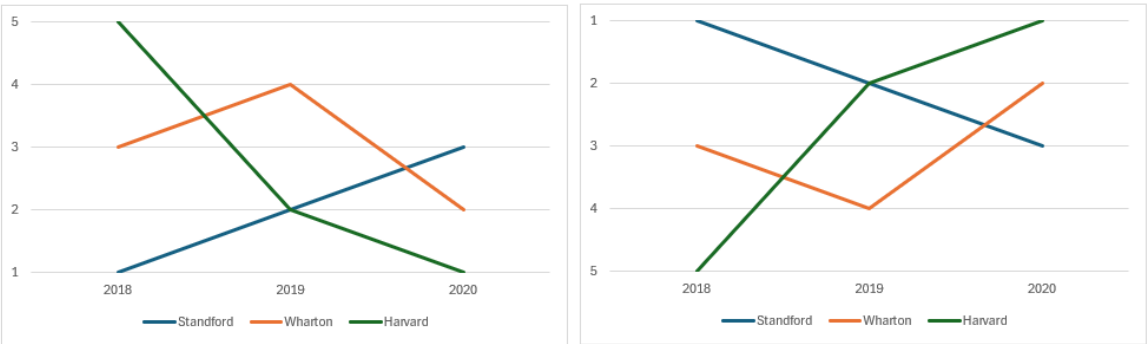
Kuva 12. Sama asia esitettynä kuvasuhdetta muuttamalla / kuvasuhteen vääristäminen (Ranganathan 2022, 87; myös Kosslyn 2006, 207; Pandey et al 2015, 1472–1473)

Kuva 13 on toinen näkökulma kuvan 12 rinnalle tulojen vaihtelun esittämisestä. Tässä sama data summataan eri tavalla, kun edellisessä kuvassa akseleiden kuvasuhteita muokattiin. Kuvassa esitetään kolme vaihtoehtoa esittää yrityksen tuloja vuosien 1978 ja 1984 välillä. Ensimmäinen esitys perustuu vuosineljänneksittäin esiintyviin datapisteisiin, toisessa vuosineljännekset on summattu mukailemaan yrityksen heinäkuusta kesäkuuhun kestävää tilikautta ja kolmannessa summaaminen on tehty kalenterivuosittain. Eri esitystavoista voi helposti vetää erilaiset johtopäätökset. Neljänneksittäin esitetty data näyttää useita laskuja tai kuoppia tulovirrassa. Keskimmäisessä esitystavassa näkyy tilikauden 1982 tulojen lasku ja sitä edeltänyt merkittävä kasvuhuippu. Kolmannessa versiossa myyntitulot näyttävät jatkavan 1982 kasvuaan ja edellinen vuosi on ollut poikkeus kasvuun. Jos kyseistä yritystä olisi myymässä tai listaamassa pörssiin, olisi kolmas esitystapa varmasti myyjälle mieluisin useamman vuoden kasvutarinana. (Joseph, 1992)



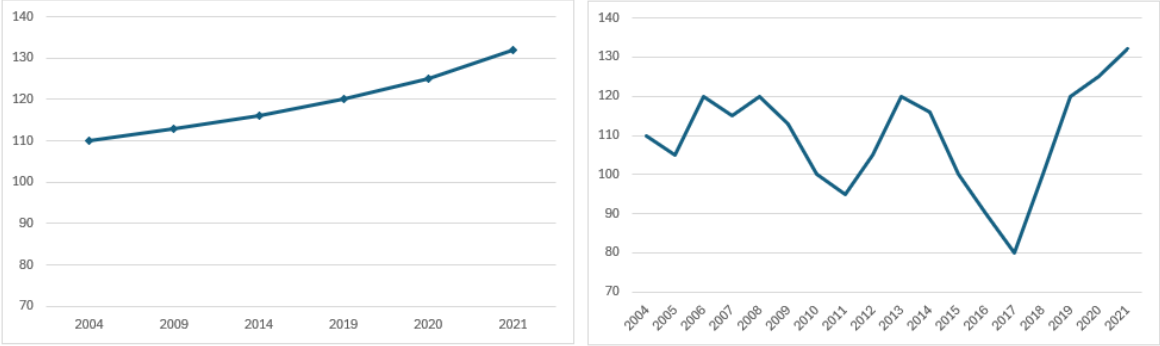
Kuva 13. Datan summaamisen vaikutus tulosvaihtelun ymmärtämiseen mukaillen Gregory P. Josephin teosta Modern Visual Evidence (1992)

Käänteisten akselien avulla voidaan pyrkiä vääristämään tietoa, sillä ihminen yhdistää yleensä suunnat trendeihin. Viivakaavion yhteydessä ylöspäin menevä viiva tarkoittaa nousua ja lisäystä, kun taas alaspäin menevä kuvastaa laskua. Suuntatulkinta tekee käänteisestä akselista yhden yleisimmistä vääristämistekniikoista. (Pandey et. al 2015, 1472–1473) Kuvassa 14 on esitetty dataa ensin niin kutsutulla normaalilla tavalla sekä käännetyn akselin avulla. Ihmisen silmät ovat tottuneet siihen, että y-akselin asteikon, joka alkaa alhaalta on linjassa käsitykseemme siitä, että arvot ovat suuremmat mitä korkeammalle akselilla mennään. Kuten kuvassa 14 voidaan huomata, että y-akselia kääntämällä voidaan selkeästi muokata ihmisen käsitystä yliopistojen paremmuusjärjestyksestä. X-akselin suhteen taas on totuttu, että kaavio skannataan vasemmalta oikealle -periaatteella, jolloin vuosiluvut on hyvä aloittaa vasemmalta oikealle niin, että oikealla on kuluva vuosi tai ennusteet tulevasta (Ranganathan 2022, 87). Y-akselin kääntelyä vastaava ongelma saadaan siis aikaan kääntämällä vuosiluvut kasvamaan epäloogisesti oikealta vasemmalle (Ranganathan 2022, 88).



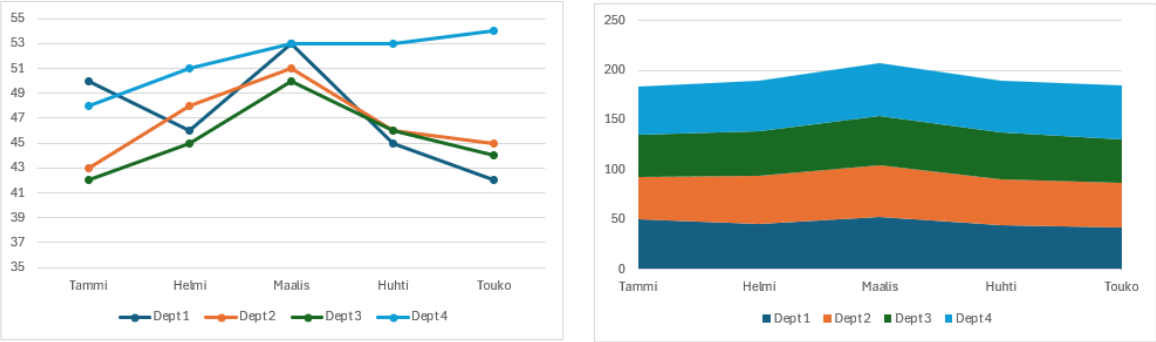
Kuva 14. Tiedon vääristyminen käännetyn akselin takia (Ranganathan 2022, 88)

Lisäksi kaavion mittakaavan muuttuminen osittain jommallakummalla akselilla voi vääristää esitettävää dataa. Vääristymä on erityisen tärkeää huomata erityisesti tarkastellessa aikasarjallista dataa. Kuvan 15 vasemmanpuoleisessa esityksessä nähdään hitaassa kasvussa oleva trendi, joka on datapisteistä valinnalla tuotettu harha, kun esitetään vain osan vuosien kehitys. Kun kaaviota tarkastelee tarkemmin, voidaan huomata, että x-akselin vuodet menevät viiden vuoden välein vuoteen 2019 asti, minkä jälkeen data esitetään vuoden välein. Todellisuudessa kaavio näyttäisi huomattavasti erilaiselta, kun otettaisiin kaikki vuodet huomioon. (Ranganathan 2022, 74; Krok 2021, 3940)



Kuva 15. Datan valinta ja/tai epälooginen aikajakso (yhdistelty Krok 2021, 3940(Cherry picking); Ranganathan 2022, 74)

Summaviiva- ja summapylväskuvioissa elementit tulisi pyrkiä järjestelemään niin, että alimmaisessa kuviossa on se sarja, jossa on vähäisempää vaihtelua (Koponen et. al. 2016). Pinottu aluekaavio sekä pinottu pylväskuvaaja voivat olla ongelmallisia, mikäli niitä ei osata lukea oikein. Kuva 16 on aluekaavio, joka näyttää eri osastojen kuluja viimeisen kuuden kuukauden osalta neljältä eri osastolta. Nopealla vilkaisulla kuvassa 16 voidaan olettaa, että kaikilla osastoilla on ollut tasaisia kuluja koko kuuden kuukauden aikana. Olettamus osoittautuu täysin virheelliseksi, kun samaa muuttujaa tarkastellaan viivakaavion avulla. Tästä voidaan todeta, että pinotun aluekaavion käyttöä kannattaa harkita ja käyttää vain sellaisissa tapauksissa, jossa se on merkityksellistä. (Ranganathan 2022, 82–83)



Kuva 16. Sama asia esitettynä pinotussa aluekaaviossa sekä viivakaaviossa (Ranganathan 2022, 82–83)

Seuraavassa aliluvussa käsitellään käytetyimpiä visualisointityökaluja, jotka vaikuttavat osaltaan visualisointien tuottamiseen.

2.6 Käytetyimmät visualisointityökalut

Datan visualisointityökalut ovat kriittisiä nykyaikaisessa datalähtöisessä ympäristössä, sillä ne mahdollistavat yrityksille, tutkijoille ja analyytikoille monimutkaisten tietoaineistojen tehokkaan tulkinnan. Erilaiset työkalut on suunniteltu vastaamaan eri tarpeisiin ja osaamistasoihin, käyttäjäystävällisistä käyttöliittymistä edistyneisiin räätälöityihin visualisointimahdollisuuksiin. Seuraavassa tarkastellaan laajimmin käytettyjä datan visualisointityökaluja Tableauta, Microsoftin Power BI:tä, Qlik Senseä, D3.js, SAS Viyaa ja Googlen Data Studiota. Näiden viiden työkalun yhteenlaskettu markkinaosuus on reilu puolet datan visualisoinnin kokonaismarkkinasta, sillä neljän ensimmäisen markkinaosuus on jo 47 % luokkaa. (SNS Insider (2024); GMI (2024))

Tableau

Tableau on yksi merkittävimmistä datan visualisointityökaluista, tunnettu tehokkaista ominaisuuksistaan ja käyttäjäystävällisestä käyttöliittymästään. Tableaun vedä ja pudota -toiminnallisuus mahdollistaa monimutkaisten visualisointien luomisen helposti, mikä tekee siitä saavutettavan sekä teknisille että ei-teknisille käyttäjille. Työkalu tukee erilaisia tietolähteitä, kuten taulukoita, tietokantoja ja pilvipalveluita, mahdollistamalla saumattoman dataintegraation ja reaaliaikaisen analyysin. (Tableau 2024) Tableaun laaja käyttö eri

toimialoilla johtuu sen vahvoista analytiikkaominaisuuksista ja interaktiivisista raporteista (Mordor Intelligence, 2024). Sen pilvipohjaiset ratkaisut, kuten Tableau Online ja Tableau Public, parantavat sen saavutettavuutta ja yhteistyöominaisuuksia, tehden siitä monien organisaatioiden suosiman valinnan (Grandview Research, 2024).

Microsoft Power BI

Microsoft Power BI on toinen johtava työkalu datan visualisointimarkkinoilla, erityisesti tunnettu integroitumisestaan Microsoft Office- ja Azure-palveluihin. Koneoppimisen integrointikyvyt lisäävät sen suosiota yrityksissä. Power BI tarjoaa tutun käyttöliittymän Microsoft-tuotteisiin tottuneille käyttäjille sekä edistyneitä toimintoja, kuten DAX (Data Analysis Expressions) monimutkaisia laskelmia varten ja Power Query datan muokkaamiseen. (Power BI 2024)

Työn empiriaosuudessa käsitellyssä yrityksessä Microsoftin tuotteet ovat pääosassa ja suuri osa raportoinnista on historiallisesti ollut Microsoft Excelin avulla tehtyä. Tätä raportointia on konvertoitu vähitellen Power BI:n puolelle, jotta raportointiin pääsee helpommin käsiksi muutkin kuin raportoinnin toteuttajat. Power BI:n, kuten muiden visualisointityökalujen datamalleissa ja visualisoinneissa on rajoitteita, jotka pakottavat raportointia vieläkin vanhoihin työkaluihin.

Qlik Sense

Qlik Sense tunnetaan assosiatiivisesta datamallistaan ja tehokkaista muistinvaraisista käsittelyominaisuuksistaan, jotka mahdollistavat käyttäjille vapaan datan tutkimisen ja piilossa olevien suhteiden löytämisen ilman ennalta määriteltyjä kyselyitä. Qlik Sensellä on kyky käsitellä suuria datamääriä ja tehdä monimutkaisia analyysejä, joka tekee siitä sopivan suuremmille yrityksille (Mordor Intelligence, 2024). Joustavuus tekee Qlik Sensestä erityisen hyödyllisen tutkivassa data-analyysissä ja yllättävien oivallusten löytämisessä. (Qlik 2024)

D3.js

D3.js (Data-Driven Documents) on tehokas JavaScript-kirjasto, jota käytetään räätälöityjen, interaktiivisten datavisualisointien luomiseen verkkoselaimissa. Toisin kuin muut mainitut työkalut, D3.js vaatii koodausosaamista, mikä tekee siitä sopivamman kehittäjille ja teknisesti taitaville käyttäjille. D3.js tarjoaa tarkkaa hallintaa datan visuaalisesta esittämisestä, mikä mahdollistaa räätälöityjen visualisointien luomisen erityistarpeisiin. (D3.js 2024)

SAS Viya

SAS on tunnetusti vahva toimija analytiikkatuotteissa, jotka integroituvat nyt sujuvasti visualisointeihin SAS Viya -työkalun kautta. SAS:in toimiala-spesifit valmiit mallit tehostavat data-analyysiä ja sitä kautta nopeuttavat visualisointien toimintaa. SAS kehuu Futurum Groupin testissä todettua vahvaa skaalautumista. (SAS 2025)

Google Data Studio

Google Data Studio on ilmainen, pilvipohjainen datan visualisointityökalu, joka integroituu saumattomasti muihin Googlen palveluihin, kuten Google Analytics, Google Sheets ja BigQuery. Se tarjoaa helposti lähestyttävän alustan interaktiivisten ja jaettavien raporttien luomiseen. (Google Data Studio 2024)

Google Data Studion yksinkertaisuus ja helppokäyttöisyys yhdistettynä reaaliaikaisiin datapäivityksiin tekevät siitä houkuttelevan valinnan yrityksille, jotka etsivät kustannustehokasta ratkaisua. Sen markkinaosuus kasvaa erityisesti pienten ja keskisuurten yritysten keskuudessa, jotka hyötyvät sen integroitumisesta Googlen ekosysteemiin (Coherent Market Insights, 2023).

Työkalujen markkina ja yhteenveto

Arviot datan visualisointityökalujen markkinasta liikkuvat 7–10 miljardin euron haarukassa lähteestä riippuen, ja kasvun uskotaan olevan nopeaa, reilu 11 % vuodessa. Markkinaosuuksien laskeminen kasvualalla, jossa osa käyttäjistä saattaa olla usean eri visualisointityökalun käyttäjä, on haastavaa ja eri lähteistä saadaan erilaisia arvoja. Pelkästään tarkan toimialan määrittely visualisointityökaluihin lähestyy mahdotonta, sillä kuten edellä on esitelty, suurin osa työkaluista sisältää merkittävästi enemmän ominaisuuksia kuin pelkkää visualisointia. (SNS Insider (2024); GMI (2024); Grandview Research (2024))

Datan visualisointityökalujen kenttä on monipuolinen, tarjoten eri tarpeisiin sopivia ratkaisuja helppokäyttöisyydestä ja integrointimahdollisuuksista edistyneisiin analyysitoimintoihin ja räätälöintimahdollisuuksiin. Tableau, Microsoft Power BI, Qlik Sense, D3.js, SAS Viya ja Google Data Studio ovat laajimmin käytettyjä työkaluja, joista jokainen tarjoaa ainutlaatuisia ominaisuuksia, jotka edistävät niiden suosiota. Monesti tosin yrityksen muut teknologiavalinnat voivat ohjata visualisointityökalun valintaa, Googlen ympäristössä toimivana yrityksen Data Studio on suoraviivainen ottaa käyttöön, ja Microsoftilla on saman

kaltainen etulyöntiasema Power BI:n integroituessa eri Microsoftin työkaluihin saumattomasti. Yleensä kaikista työkaluista löytyy tarvittavat integraatiorajapinnat hyödynnettäväksi eri yhteyksissä, joten täysin yksioikoista sopivan työkalun valinta ei ole. Valitsemalla sopivan työkalun erityistarpeiden perusteella organisaatiot voivat parantaa data-analyysikykyjään, tehostaa päätöksentekoa ja edistää datalähtöistä kulttuuria. Työkalut ovat välttämättömiä monimutkaisen datan muuntamisessa merkityksellisiksi oivalluksiksi, tehden datan visualisoinnista olennaisen osan nykyaikaista data-analytiikkaa.

3 Tutkimusmenetelmät ja aineisto

Tässä luvussa esitellään tutkimuksen kohteeksi valitun tapauksen kuvaus, tiedonkeruun menetelmät sekä käytetty aineisto.

3.1 Tapauksen kuvaus

Tarkasteltavaksi kokonaisuudeksi valikoitui erään liiketoiminta-alueen liiketoimintakatsaus, jolla on tarkoitus saada kokonaiskäsitys liiketoiminnan menestyksestä ja sujumisesta eri mittareilla. Kohderyhmä katsauksella on laaja liiketoiminnan esittelyistä aina hallituskäsittelyihin. Hallitusnäkökulmasta on erityisen tärkeää, että johdon tuottama raportointi antaa todenmukaisen kuvan yhtiön kehityksestä ja teoriaosuudessa esiin nostetut potentiaaliset ongelmat visualisointien rakentamisessa vältetään. Tarkastelussa pääroolissa on liiketoimintakatsauksen analyysi, jota pohjustetaan datan esikäsittelyn sekä rajoitteiden ymmärtämisellä. Liiketoimintakatsausta tullaan analysoimaan esitettyjen kuvien kautta teoriapohjaa sekä haastattelujen tuloksia vasten. Analyysi on jaettu kahteen alilukuun, joista jälkimmäisessä analysoidaan pylväs- ja viivadiagrammeja ja ensimmäisessä muita tapoja. Yrityksessä on ollut pidempään käytössä visualisoinneissa Microsoftin Power BI, joka edellä tunnistettiin käytetyksi työkaluksi ja integroituu sujuvasti Office -perheen työkalujen kanssa.

3.2 Tapaustutkimus

Tapaustutkimus valikoitui pääasialliseksi tutkimustavaksi, sillä tutkimuksessa on keskiössä selittää miten ja miksi tietyt graafiset esitykset sopivat esittämään ilmiötä halutulla tavalla. Tapaustutkimusten aineisto on tyypillisesti laadullista (Eriksson & Koistinen 2014), niin kuin tässäkin tapauksessa. Yin (2014, 187-190) esittelee erilaisia tapaustutkimustyyppejä tai rakenteita: lineaaris-analyyttinen, vertaileva, kronologinen, teoriaa kehittävä, käänteinen sekä vapaasti järjestetty. Tutkimuksessa tullaan käyttämään näistä ensimmäistä, lineaaris-analyyttistä rakennetta, jota diplomityön runko lähtien lähtökohdista ja aineistosta päätyen johtopäätöksiin tukee. Eriksson & Koistinen (2014, 41) toteavat tyypin sopivan ekstensiiviseen, selittävään, kuvailevaan tai eksploratiiviseen tapaustutkimukseen.

Liiketoimintakatsauksen tarkastelussa on katsottu sen sopivan yksittäistapaustutkimuksen ja intensiivisen tapaustutkimuksen yhdistelmään, jossa tarkoituksena on ymmärtää yksittäistä ilmiötä eikä niinkään luoda yleistä teoriaa tai monistettavuutta. Eriksson & Koistinen (2014, 18) kuvaavat intensiivistä tapaustutkimusta, sopien yksittäistapaustutkimukseen, liiketoimintakatsauksen analyysin käyttötarkoituksessa sanoilla (korostus lisätty):

" Tarkoituksena ei ole niinkään tehdä tapausta koskevia yleistyksiä, vaan selvittää, ***millä logiikalla juuri tarkasteltava ainutlaatuinen ja erityinen tapaus toimii****."*

Tapaustutkimuksessa eettisyys ansaitsee oman pohdintansa. Tapaustutkimuksessa tutkimuksen kohteet voivat olla tunnistettavissa, olivat ne sitten organisaatioita, henkilöitä tai henkilöryhmiä ja onkin hyvä varmistaa, halutaanko tutkimuskohde pitää anonyyminä (Eriksson & Koistinen 2014, 43). Diplomityön osalta analyysin kohteet tullaan pitämään anonyymeinä ja tutkimuksen kohteena olevan katsauksen luvut tullaan anonymisoimaan. Anonymisointi toteutetaan siten, että liiketoimintakriittistä tietoa ei vahingossa paljasteta, mutta analyysi voidaan tehdä oikeellisesti.

3.3 Aineisto ja tiedon keräämisen menetelmät

Aineisto voidaan jakaa kolmeen erilliseen osaan, joilla on keskenään selkeä yhteys. Liiketoimintakatsaus omana dokumenttinaan on näistä selkein, käsittelyssä oli joulukuun 2024 versio katsauksesta. Liiketoimintakatsaus oli ollut käytössä noin puolitoista vuotta pienin muutoksin ja tarve uudistukselle liiketoimintamuutosten jäljiltä ilmeinen. Liiketoimintakatsauksen muodostamiseksi tarvitaan dataa, joka asettaa rajoitteet katsauksessa käsiteltäville näkökulmille, ja hyödyntäjien näkökulma sille, mikä raportointi parhaiten tukee päätöksentekoa ja antaa kattavan kuvan liiketoiminnasta.

Aineiston keruu toteutettiin teemahaastatteluna. Tutkimuksen tutkimuskysymyksien vastauksien tullessa sekä teoriasta että haastatteluista voidaan nähdä tutkimuksessa olevan induktiivisia sekä deduktiivisa piirteitä, joihin molempiin temaattinen analyysi sopii erinomaisesti. Tai kuten Ahmed et al. (2025) osuvasti toteavat

> *" Thematic analysis is not bound to any specific theoretical framework, making it a versatile tool suitable for both inductive and deductive approaches to data analysis."*

Taulukko 5 kerää yhteen tiedon keruuseen osallistuneet henkilöt ja roolit organisaatiossa. Taulukosta ja tarkastelusta on rajattu ulos henkilöt, joiden kommentit voitiin katsoa olevan päällekkäisiä laajemmin kommentoiden henkilöiden kanssa. Tällä pyritään selkeyttämään ulkopuoliselle mahdollisesti hieman sekavaa roolitusta ja vähentämään muistettavien henkilöiden ja roolien määrää. Liiketoiminta-alueen yhtiöissä rooleista on valittu henkilön tarkastelun osalta merkitsevin rooli, johtajalla C voidaan katsoa olevan kolmekin relevanttia eri roolia.

Taulukko 5. Haastatellut tiedon hyödyntäjät ja tuottajat.

Henkilö	Rooli	Rooli liiketoiminta-alueen yhtiöissä	Työkokemus vuosina
Johtaja A	Hyödyntäjä	Yhtiö A:n hallituksen puheenjohtaja	30
Johtaja B	Hyödyntäjä	Yhtiö A:n hallitus, Päällikkö A ja B esihenkilön esihenkilö	30
Johtaja C	Hyödyntäjä	Yhtiö A:n toimitusjohtaja	30
Johtaja D	Hyödyntäjä	Yhtiö B:n toimitusjohtaja	40
Johtaja E	Hyödyntäjä	Yhtiö C:n toimitusjohtaja	40
Päällikkö A	Tuottaja ja Hyödyntäjä	Liiketoiminta-alueen Business Control osaston päällikkö	15
Päällikkö B	Tuottaja ja Hyödyntäjä	Liiketoiminta-alueen Business Control vetäjä, yhtiöt A-E	15
Asiantuntija A	Tuottaja ja Hyödyntäjä	Yhtiöiden A, B ja C Business Controller	15
Asiantuntija B	Tuottaja ja Hyödyntäjä	Yhtiön D Business Controller	35
Asiantuntija C	Tuottaja ja Hyödyntäjä	Yhtiön E Analyytikko, myöhemmin myös A, B ja C	10
Asiantuntija D	Tuottaja ja Hyödyntäjä	Yhtiön F Business Controller	40
Asiantuntija E	Tuottaja ja Hyödyntäjä	Yhtiön E Business Controller, myöhemmin myös A	10

Datan rajoitteiden osalta käsitys on muodostunut pidemmältä aikajaksolta vuosien 2021–2024 aikana perustuen keskusteluihin kahden tietovaraston analyytikoiden kanssa ja olemassa olevaan dokumentaatioon. Kuten usein ongelmana on, dokumentaatio osoittautui monelta osin riittämättömäksi ja vuosien varrella on käyty lukuisia keskusteluja tietokantojen ylläpitäjien kanssa kenttien tietosisällöistä ja datavirroista. Tieto on keskittynyt vahvasti eri yhtiöiden business control -osastoille, ja yhtiöiden vastaavat business controllerit tietävät syvällisesti oman alueensa rajoitteet. Tällä taustalla järjestettiin uuden johtaja A:n toiveesta läpikäyntejä, joissa pyrittiin yhdistelemään eri tuotealueilta löytyvää dataa eheäksi kokonaisuudeksi hyödyntäen business controllereille kertynyt tieto nykyisistä rajoitteista. Työpajat järjestettiin 15.8.2024, 22.8.2024 ja 28.8.2024, joiden lisäksi käytiin lukuisia keskusteluja datan tarjoamista mahdollisuuksista näiden välissä. Työpajoissa pyrittiin datan rajoitteet huomioiden käymään keskustelua mukaillen kolmatta tutkimuskysymystä - onko meillä mahdollisuutta luoda raportointiin visualisointeja, jotka ovat tarkoituksenmukaisia sekä

neutraaleja ja mahdollistavat oikeaan tietoon perustuvan päätöksenteon? Näissä työpajoissa piirtyi toisin sanoen konkreettisemmin käsitys tällä hetkellä datan mahdollistamista raportointinäkökulmista ja keskityttiin olennaisiin, raportoinnin eheyttä tai yhtenäisyyttä vaarantaviin ongelmakohtiin. Työpajoihin osallistuivat päällikkö A sekä asiantuntijat A, B, C, D ja E, joiden työkokemus alalta vaihteli 5–40 vuoden välillä. Herätteenä näihin tarkempiin tarkasteluihin toimi huhtikuussa 2024 aloittanut uusi liiketoimintajohto. Tiedon haasteita ja rajoitteita yrityksen osalta kuvataan tarkemmin luvussa 4.1 ja tarkemmin yksittäisiin kuviin aiheutuneita rajoitteita luvuissa 4.2 ja 4.3

Hyödyntäjiltä tietoa saatiin tietoa eri forumeilta, joissa kaikissa oli hieman eri näkökulma asian lähestymisessä. Liiketoimintakatsaus toimii yhtiö A:n hallituksessa toimitusjohtajan katsauksena ja yhtiössä aloitti sopivasti uusi hallitus, jolloin saatiin suoraa ja tuoretta palautetta oikeastaan kaikkien tutkimuskysymysten näkökulmasta – antaako liiketoimintakatsaus kattavan ja oikeellisen kuvan liiketoiminnasta juuri aloittaneelle hallitukselle ja huomataanko mahdollisesti visualisointeja, jotka eivät ole tarkoituksenmukaisia? Yhtiö A:n hallituksen puheenjohtaja on huhtikuussa aloittanut liiketoimintajohtaja A ja yhtiö A:n toimitusjohtaja on johtaja C. Kuusihenkisessä hallituksessa istuu johtaja B, joka on päällikkö A:n ja B:n esihenkilön esihenkilö, joka tosin käy suoraa keskustelua päälliköiden kanssa välissä olevasta johtajasta huolimatta. Palautetta saatiin yhtiö A:n hallituksen kokouksista 25.10.2024 ja 13.12.2024 toimitettuihin materiaaleihin perustuen. Toinen näkökulma oli liiketoiminta-alue A:n johtoryhmä, jossa istuu muiden johtajien lisäksi yhtiöiden A, B ja C toimitusjohtajat. Yhtiöt A, B ja C toimivat liiketoiminta-alueen sydämenä vastaten noin 90 % liikevaihdosta. Liiketoimintakatsausta käsiteltiin kvartaalikokouksissa 24.10.2024 ja 6.2.2025. Liiketoimintakatsausta käytiin kalvo kalvolta läpi erikseen järjestetyssä läpikäynnissä johtaja C:n kanssa 24.1.2025, jossa tarkasteltiin datan asettamia rajoitteita läpi johtoryhmän ja hallituksen kokouksia tarkemmin.

Liiketoiminnassa aloitti uusi päällikkö B 1.1.2025, jolla ensimmäinen tehtävä huhtikuussa 2024 aloittaneelta johtaja A:lta ja yhtiö A:n hallituksessa istuvalta johtaja B:lta (joka siis päällikkö B:n esihenkilön esihenkilö) oli uudistaa liiketoimintakatsausta kuvaamaan paremmin huhtikuussa ja lokakuussa 2024 tehtyjä liiketoimintamuutoksia. Tässä uudistuksessa toimi apuna pääosin 2024 syksyllä valmistunut diplomityön teoriatausta, johon uusia visualisointeja luodessa oli mahdollista tukeutua. Liiketoimintakatsausta lähdettiin siis uudistamaan liiketoiminnassa 2025 alussa aloittaneen päällikkö A:n kanssa vanhan materiaalin

läpikäynnillä ja teoriasta nousseilla näkökulmilla, ensimmäinen läpikäynti oli työpajassa 5.2.2025. Seuraavaksi uudistettua materiaalia käytiin läpi yhtiö A:n toimitusjohtajan (johtaja C) kanssa 26.2.2025. Vastauksia saatiin mukailtuna kolmanteen tutkimuskysymykseen, jossa pyrittiin ymmärtämään ei niinkään kuvakusen neutraaliutta vaan kattavuutta liiketoiminnan ohjauksen tukena: ”Kuvaavatko yrityksen raportoinnissa käytetyt visualisoinnit tarkoituksenmukaisesti ja **kattavasti liiketoiminnan kokonaisuutta** tarjoten pohjan oikeaan tietoon perustuvalle päätöksenteolle?” Tässä tutkimuksessa ei ole mahdollisuutta tai tarkoituksenmukaistakaan salassapitonäkökulmien vuoksi ottaa kantaa kattavasti liiketoiminnan kokonaisuuden kuvaukseen, mutta kysymykseen saadut vastaukset auttoivat merkittävästi liiketoimintakatsauksen jatkokehitystä ja vastasivat kysymyksiin tutkimuksen näkökulmasta. Lukuisista läpikäynneistä parhaiten jäi mieleen johtaja C:n osuva toteamus ”*rakennettu tuottaja edellä*” kuvaamaan tuotannollistettua, mutta hieman finalisoimatonta visualisointia.

Käytetty aineisto vastaa erityisesti kolmanteen tutkimuskysymykseen – Ovatko yrityksen nykyisessä raportoinnissa esiintyvät visualisoinnit tarkoituksenmukaisia ja kuvaavat neutraalisti esitettävää dataa, jotta päätöksenteko pohjautuu oikeelliseen tietoon? Ensimmäiseen kysymykseen ”Mitä asioita yrityksen tulisi huomioida datan visualisoinnissa?” ja toiseen ”Miten datan visualisoinnin avulla voidaan vääristää vahingossa tai tarkoituksellisesti visualisoinnin antamaa informaatiota?” saadaan vastaukset vahvasti teorian pohjalta, toki huomioiden toisen ja kolmannen kysymyksen osittainen päällekkäisyys. Selkeyden vuoksi sekä teorian että työpajojen näkökulmat esitellään kuvien yhteydessä seuraavassa luvussa.

4 Liiketoimintakatsauksen analyysi

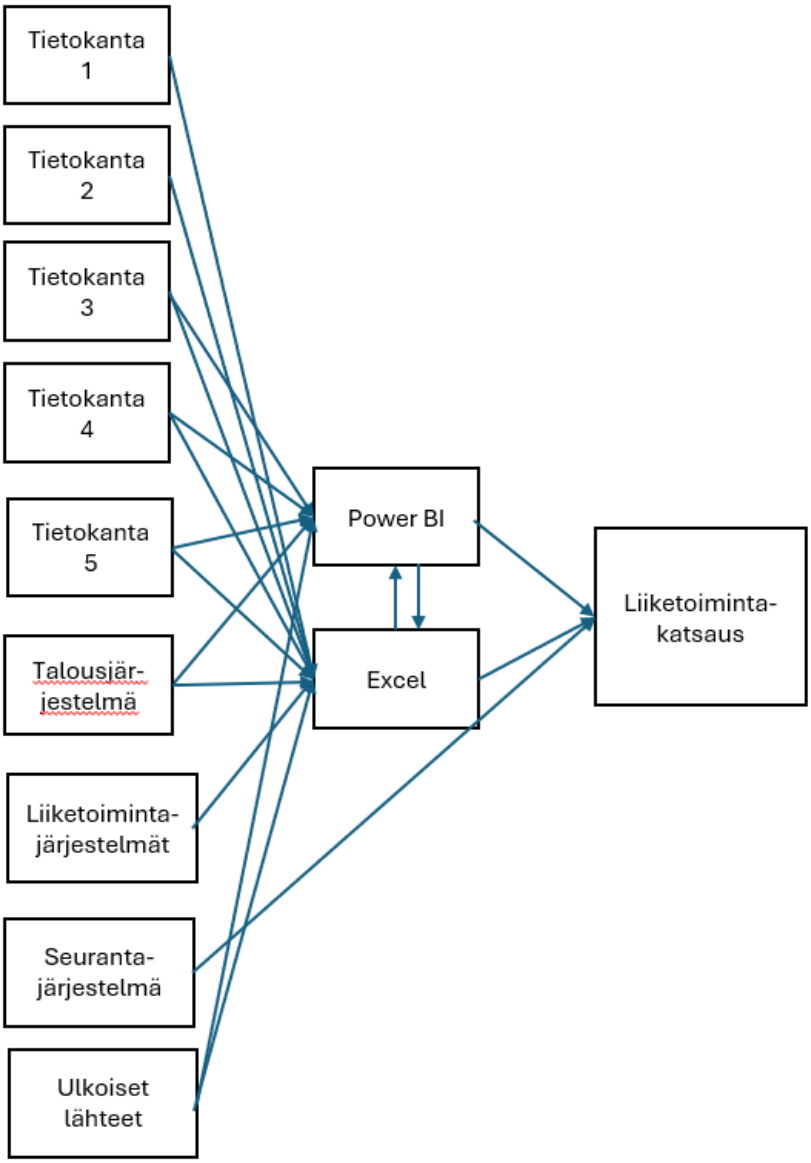
Tässä luvussa tullaan ensin käsittelemään yleisesti tiedon käsittelyssä havaittuja haasteita. Haasteiden jälkeen siirrytään käsittelemään itse katsauksessa löytyviä visualisointeja. Visualisointien käsittelyssä on syytä kiinnittää huomiota rakenteeseen, lyhyt kuvaus, havainnot tiedon rajoitteista, teoriaan perustuva analyysi, hyödyntäjien näkökulma, yhteenveto. Selvyyden vuoksi käsiteltävät kuvat on jaettu kahteen alilukuun, joista ensimmäisessä käsitellään taulukoita ja donitseja, toisessa muita esitystapoja. Lopuksi esitetään työn keskeiset tulokset.

4.1 Tiedon käsittelyn haasteet ja prosessit

Usein kehitystoiveiden kanssa tasapainoillaan rajoitteiden ja suoranaisten purkkaviritysten välimaastossa ja käydään narunvetoa ylläpidettävyydestä. Kommentti, mitä monesti ylläpidollisesti helpommista graafeista tulee eri sanoin, mutta voi tiivistää mielestäni nerokkaaseen lainaukseen ”rakennettu tuottaja edellä”. Lainaus tarkoittaa lyhyesti sitä, että graafin tai esityksen ylläpito on tehty helpoksi ja vähätöiseksi, mutta hyödyntäjä toivoisi hieman erilaista toteutusta. Haasteena on hyödyntäjien vähäinen tuntemus rajoitteista, joita käytetyt visualistointityökalut ja data asettavat toteutukselle. Joskus toivottu pienen oloinen muutos saattaa vastata viikon työtä. Toinen teema, josta usein eri muodoissa hyödyntäjät tiedustelevat, on eri yhtiöiden tai tuotteiden tietojen yhdistäminen. Yhdistämättömyys juontaa tietokantojen rajoitteisiin, joiden kiertäminen manuaalisesti on ylläpidollisesti täysin kestämätöntä. Hyödyntäjien kanssa dialogia käyden on löydetty muutamia erityisen tärkeitä datoja, joita on lähdetty yhdistämään käytännössä käsin ja saatu parempi kokonaiskuva tekemisestä. Näiden käsin tehtyjen ratkaisujen määrä on pyritty pitämään mahdollisimman pienenä.

Hyödyntäjien istuessa korkealla päättävissä elimissä on heillä mahdollisuus usein vaikuttaa resursointiin ja kehityseuroihin. Yhteistyöllä ja rajoitteiden läpikäynnillä on mahdollisuus poistaa raja-aitoja ja kehittää yhtenäisempää datapohjaa. Tähän onkin jo ryhdytty ja liiketoimintakatsauksen uusi versio, jossa on poistettu rajoitteita ja huomioitu tässäkin työssä nousseita parannusehdotuksia, on jo käytössä.

Yhteistyötä tietokantojen sekä raportoinnin ylläpitoon osallistuvien henkilöiden kanssa on tehty luonnollisesti jatkuvasti jo ennen analysoitua liiketoimintakatsausta, sen rakentamisen aikana ja jatkokehitysvaiheessa sekä tulevan uuden version yhteydessä. Tässä esitetty analyysi kuvastaa jatkuvaa dialogia dataan liittyvistä haasteista, joiden kokonaisuus käytännössä tiivistyi BC-osaston työpajoihin elokuussa 2024. Yleisen problematiikan kuvauksen jälkeen yksityiskohtaisempi, kuvakohtainen analyysi, ilmenee kuvien käsittelyn yhteydessä.



Kuva 17. Yksinkertaistettu kuvaus liiketoimintakatsauksen dataputkesta.

Kuva 17 esittää yksinkertaistetun kuvan liiketoimintakatsauksessa käytetystä datasta ja sen kulkemisesta raportille. Yksinkertaistuksia on tehty tietokantojen esikäsittelyssä, joissa luonnollisesti kaikissa on useita datalähteitä ja tietokantojen mahdolliset näkymäkerrokset sekä muut käsittelyt on rajattu ulos. Excel-työkirjat on niputettu vain yhdeksi, vaikka niitä käytetään lukuisia. Sama pätee Power BI-raportteihin sekä niiden sisältämiin semanttisiin tietomalleihin, jotka kokoavat tietoa useammasta tietokannasta ja excel-työkirjoista. Yhdistämällä edellisessä aliluvussa esitellyn taulukon *5* roolinäkökulman yllä olevaan kuvaukseen dataputkista saa hyvän käsityksen käsittelyssä olevan tapauksen monimutkaisuudesta.

Pääsääntöisesti käytetyt tiedot löytyvät kahdesta tietokannasta (4 & 5). Näiden lisäksi käytössä on kolme muuta tietokantaa, joista löytyy joko aputietoja tai tiettyyn tuotteeseen liittyvää tarkempaa dataa. Yhden päätietokannan ja yhden tukevan tietokannan käyttöön liittyy merkittäviä rajoitteita ja haasteita, ja läpikäynneissä asiantuntijoiden kanssa todettiin kummankin tietokannan olevan alasajovaiheessa ja siten nykyisiin datan käsittelyn rajoitteisiin tuskin on saatavilla apua. Käytössä olevat kaksi päätietokantaa sisältävät suurimman osan käytetyistä datoista, mutta niihin sisältyy merkittäviä rajoitteita muutamien tuotealueiden osalta. Näitä rajoitteita pyritään luonnollisesti ratkomaan ja monia onkin ratkottu, mutta kerran rakennettujen tietokantarakenteiden muuttaminen rajatuilla resursseilla kestää aikansa. Liiketoimintakatsauksessa ongelma näkyykin keskittymisenä tarkempaan analyysiin muutaman tuotealueen osalta, ja rajoitteiden alaiset tiedot tulevat kootummin tai harvemmin päivittyvästi.

Yrityksen datahaasteita voidaan jakaa kolmeen kategoriaan: tiedon näkyvyysongelmat, tietokantaongelmat ja organisaatio-ongelmat. Tiedon näkyvyysongelmat kuvaavat tilanteita, missä käyttäjällä ei ole pääsyä tietoihin joko tarkoituksellisesti tai pääsy pitäisi olla, muttei ole. Yleensä tekniset näkyvyysongelmat saadaan ratkaistua sujuvasti, mutta joskus niidenkin kanssa saattaa yllättävän pitkä tovi vierähtää. Joskus tarkoituksellinen tiedon näkymättömyys on tietyn ajanhetken käsitys tietojen luvituksista, joka on muutettavissa oikean päätäntäelimen päätöksellä. Nämäkin ovat joskus suoraviivaisia, mutta joskus oikean päättäjän löytäminen ja näkyvyysesteiden purkaminen voi olla haastavaa, jos teknisessä toteutuksessa on ajateltu tilanteen olevan pysyvä. Tietokantaongelmissa on hieman päällekkäisyyttä edellä kuvattujen pysyväksi ajateltujen näkyvyysongelmien kanssa, mutta tietojen yhteensopimattomuus tietokantojen tai tietokannan skeemojen välillä on yleistä ja aiheuttaa merkittävän työmäärän datan esikäsittelyvaiheessa. Tietokantaongelmiin lukeutuu yleinen

dokumentaation puutteellisuus. Organisaatio-ongelmat ovat ehkä hieman erikoisempi kokonaisuus, mutta kohdeyritys on monimutkaisesti organisoitu ja asioiden omistajat voivat olla haasteellisia selvittää tai yhteistyötä voi joskus joutua herättelemään, jolloin ongelma on enemmän organisaatiosta kumpuava kuin tietokannasta tai tiedon näkyvyydestä. Läpileikkaavia organisaatio-ongelmia, jotka tuntuvat esiintyvän yrityksessä kuin yrityksessä, ovat määrittelypäätösten dokumentoimattomuus sekä henkilöriskien leverointi ja vääjäämätön laukeaminen.

Tietokantoihin ja organisaatioon liittyvän päätöksenteon dokumentointi on yleensä haasteellista, sillä harvassa organisaatiossa on historiassa ollut dokumentoituna kaikki dataliitännäiset päätökset yhdellä forumilla. Tähän on herätty monessakin yrityksessä perustamalla data office -tyyppisiä toimintoja, joihin pyritään keskittämään datojen määrittelyä ja logiikkaa koskevia päätöksiä sekä tuomaan dataa paremmin jäljitettäväksi. Näin on tehty kohdeyrityksessä osin viranomaisen kannustamana.

Käytetyissä tietokannoissa ja datavirroissa on jonkin verran epätahtisuutta, mikä korostaa business control -funktion roolia raporttien rakentajana ja tiedon yhteensovittajana. Osa tekemisestä voisi hyötyä syvemmästäkin data scientist tai data engineer -tyyppisestä osaamisesta, mutta raportit on toteutettu olemassa olevilla resursseilla ja jatkokehityksen vaatimaa resursointia on lähdetty edistämään.

4.2 Taulukot ja donitsit

Aliluvussa tullaan analysoimaan taulukkoja ja donitseja alussa kuvatun rungolla: kuvaus visuaalista, kuvien tuottajien havainnot, teoriaan perustuva analyysi, hyödyntäjien näkökulmat ja lopuksi napakka yhteenveto.

Taulukko 6. Liiketoiminnan KPI-mittarit.

	Tavoite YTD	Toteuma YTD		Tavoite edellinen vuosi	Tavoitteen toteuma-aste	Tavoite koko vuosi
Tavoitekategoria 1						
Tavoite 1	3,8	3,9		3,9	103 %	3,8
Tavoite 2	>10	15		21	150 %	>10
Tavoite 3	4	3		-	75 %	4
Tavoite 4	2,5	2,48		2,5	99 %	2,5
Tavoitekategoria 2						
Tavoite 5	3	5		7	167 %	3
Tavoite 6	3,4			3		3,4
Tavoite 7	60	55		52	92 %	60
Tavoite 8						0
Tavoite 9	1300	1400		1305	108 %	1300
Tavoitekategoria 3						
Tavoite 10	85	82		57	96 %	85
Tavoite 11	31	30,7		30	99 %	31
Tavoite 12	25	31		34	124 %	25
Tavoite 13	39	52		50	133 %	39
Tavoitekategoria 4						
Tavoite 14	60	55		52	109 %	60
Tavoite 15	48	52		44	92 %	48
Tavoite 16	Aikataulussa					Aikataulussa
Tavoite 17	Tavoitteen mukainen					Tavoitteen mukainen
Tavoitekategoria 5						
Tavoite 18	4	3,9		3,9	103 %	4
Tavoite 19	88	80		95	110 %	88

Taulukko 6 esittää liiketoiminnan ohjaukseen käytettäviä KPI-mittareita, joita raportoidaan kuukausittain omistajille, johdolle ja henkilöstölle. Hallitus, jossa istuu omistajan edustajat, vahvistaa vuosittain käytettävät KPI-mittarit seuraavalle vuodelle.

Tuottajien havainnot: Kuvan tuottamisen näkökulmasta haaste piilee enemmän yksittäisten rivien taustalla olevan datan keräämisessä kuin itse kuvan tuottamisessa. Kestohaasteena läpi liiketoimintakatsauksen menevä eri tuotealueiden myynnin yhdistäminen on myös tässä näkymässä ongelmana. Yhdellä rivillä oleva markkinaosuuksien ulkoinen datalähde tuottaa muutaman kerran vuodessa virheellistä tietoa, jota joudutaan business controllissa oikomaan.

Teoriaan perustuva analyysi: Mittareissa on käytetty toteuman analysointiin liikennevaloja, punainen – keltainen – vihreä -akselistoa. Yhdellä silmäyksellä näkee yleiskuvan tilanteesta, olettaen normaalin värinäön. Tavoitteen toteuma-aste on tuotu %-tasolla viereiseen sarakkeeseen, josta lukeminen on mahdollista, joskin se vaatii hieman pidempää silmäilyä. Eräänä haasteena voi olla, että yleensä mittareissa tavoitellaan tavoitetta suurempaa ja osassa

(tavoitteet 14, 15) pienempää lopputulosta. Osa mittareista päivittyy kuukausittain ja osa vain kerran vuodessa, toki liiketoimintakatsauksen hyödyntäjät ovat tästä tietoisia. Kuvassa on mahdollista vertailla vierekkäisiä tavoite- ja toteumasarakkeita samalta vuodelta ja edellisen vuoden toteumaa, mistä saa suhteellisen hyvän käsityksen suoriutumisen suhteutumisesta tavoitteeseen ja edellisen vuoden toteumaan. Mukana on sanallisia mittareita, joiden värin todentamiseen joutuu käymään keskustelua. Huomioitavaa on tavoite 8, jolla ei ole arvoja tai sanallista tietoa ollenkaan, mutta on merkitty vihreällä. On hankala saada vain värin perusteella kiinni, mitä tässä on ajateltu.

Hyödyntäjien näkökulma: Hyödyntäjät ovat valitsemassa vuosittain mittareita ja sinällään ovat tyytyväisiä omiin valintoihinsa. Taustat mittareille ja niiden epäsuhtainen päivittymissykli olivat periaatteessa tiedossa, mutta unohtuvat välillä ja luovat ihmettelyä. Tavoitteen 16 osalta tunnistettiin aito haaste tavoitteeseen pääsemisestä ja ei koettu tarvetta tunnistaa tavoitteelle tavoitteen ylittämisen mahdollisuutta. Tavoite 17 tunnistettiin subjektiiviseksi, mutta luotettiin johtaja C:n kykyyn arvioida toteumaa objektiivisesti. Myynnin esittämisen haasteet eivät olleet täysin saavuttaneet hyödyntäjiä, joille saatetaan joskus hankalastikin tuottaa haluttuja tietoja. Kuten eräs omistajan edustajista totesi:

> *”En tiedä, kuinka tietoja tuotetaan, mutta olen yleensä saanut pyytämäni tiedot pyydettynä aikana.”*

Yhteenveto: Väriskaala voi aiheuttaa haasteita, mutta on pääsääntöisesti tarkoituksenmukainen. Haasteet kuvan tuottamisessa ei ole täysin saavuttaneet hyödyntäjiä, joka voi aiheuttaa väärinymmärryksiä aikatauluissa.

Taulukko 7. Kvartaalisuunnittelun status.

Painopiste	Tavoite	Status	Omistaja
Painopiste 1	**Tavoite 1**		Henkilö 1 ja 2
	Tavoite 2		Henkilö 3 ja 4
Painopiste 2	Tavoite 3		Henkilö 5
	Tavoite 4		Henkilö 1
Painopiste 3	Tavoite 5		Henkilö 1
	Tavoite 6		Henkilö 1
Painopiste 4	Tavoite 7		Henkilö 6 ja 1
Painopiste 5	**Hanke 1**		Henkilö 5
	Hankkeet 2, 3		Henkilö 6 ja 2

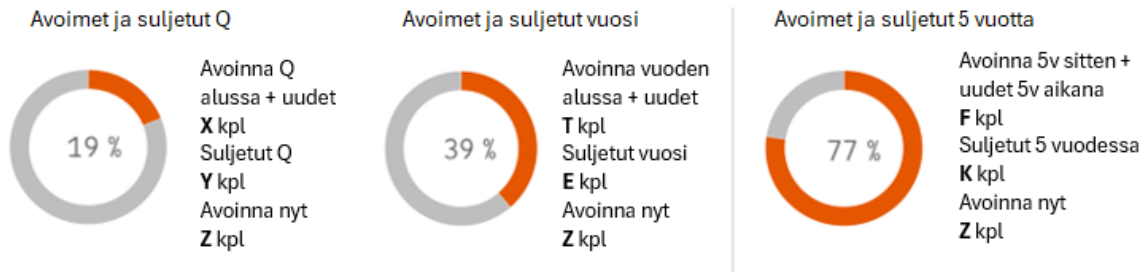
Taulukko 7 esittää liiketoiminnan kvartaalisuunnittelun tilannetta, jota raportoidaan kuukausittain omistajille, johdolle ja henkilöstölle. Seurannasta nähdään lähes reaaliaikaisesti liiketoiminnalle tärkeiden hankkeiden eteneminen.

Tuottajien havainnot: Tiedon tuottamisen rajoitteet ovat kuvan osalta minimaaliset, koska tieto tuodaan sellaisenaan seurantajärjestelmästä. Business controllin näkökulmasta tietojen ajantasaisuus voi olla haaste, mutta yleinen tieto statuksen päätymisestä seurantaan antaa sopivaa painetta ajantasaisuuden varmistamiseen.

Teoriaan perustuva analyysi: Kuvassa käytetään useampaa tehokeinoa. Erityisen tärkeät prioriteetit on tekstissä lihavoitu ja tekemisen status on värikoodattu perinteisellä punainen – keltainen – vihreä -akselistolla. Hyvää on väriassosiaatio, joskin osalle väestöstä valitut värit voivat olla haasteellisia ja ehkä mukana voisi olla toinenkin tapa indikoida statusta. Lihavoidut korostukset näkyvät hieman vaihtelevasti riippuen esitystekniikasta, joka voi aiheuttaa haasteita nopeasti löytää tärkeiden projektien tilanne. Kuva on selkeä ja viestin saa siitä suhteellisen helposti esille.

Hyödyntäjien näkökulma: Hyödyntäjien haastattelussa nousi esille pohdinta listalle priorisoituvien asioiden eri kokoluokasta. Monta pienemmän asian vihreää näyttää valoisammalta kuin yhden suuren asian vihreä tilanne, vaikka työtä voi olla moninkertaisesti. Onkin tärkeää pyrkiä asioiden yhteismitallisuuteen.

Yhteenveto: Kuva on useista visuaalisista tehokeinoistaan huolimatta selkeä. Yksittäisen hankkeen eteneminen selviää selkeästi, mutta hyödyntäjien haastatteluissa noussut yhteismitallisuus on asia, mihin tulee kiinnittää huomioita.



Kuva 18. Käsitellyt tapaukset eri ajanjaksoilla.

Kuva 18 esittää tapausten käsittelyä eri aikajaksoilla ja on liiketoimintakatsauksen ainoa donitsikuva. Esitys on jaettu kolmeen lohkoon, kaikki edustavat tiettyä ajanjaksoa, ja kolmeen tietopisteeseen; avattujen, suoritettujen ja avointen tehtävien määrään.

Tuottajien havainnot: Datanäkökulmasta haaste on tässä hieman samankaltainen kuin edellisessä, tietojen tuleminen erillisestä päivittäisestä liiketoiminnasta irrallisesta seurantajärjestelmästä aiheuttaa haastetta tiedon ajantasaisuuteen. Tämänkin tiedon osalta, kuten kvartaalisuunnittelussa, tieto päätymisestä seurantaan auttaa tietojen ajantasaisuutta.

Teoriaan perustuva analyysi: Nopealla vilkaisulla kuvasta selviää kaksi asiaa, nähdään nopeasti kunkin ajanjakson sen hetken tilanne prosenttiluvusta ja kaavion korostetusta osuudesta ja huomataan säännönmukaisuus suoritettujen osuuden kasvuna, kun mennään vasemmalta oikealle kohti pidempää tarkastelujaksoa. Esitys antaa hyvän ja intuitiivisenkin kuvan tapausten käsittelystä, lisätiedot numeroina grafiikan vieressä syventävät ensisilmäyksellä saatua kokonaiskuvaa. Hallittu värien käyttö lisää kuvassa vaikuttavuutta, merkittävät kappalemäärät on lihavoitu ja donitseissa on esitetty prosentteina ja korostettuna valmiiden tapausten määrä.

Hyödyntäjien näkökulma: Hyödyntäjiltä kuva sai positiivista palautetta juuri intuitiivisuudesta. Tapausten valmistuminen sovitun aikataulun mukaisesti on seurannassa sekä KPI-mittareissa että kvartaalisuunnittelussa, ja niitä raportoidaan vahvasti myös omistajan suuntaan. Nähtiin erittäin tärkeänä, että tiedot ovat ajantasaisia ja että kuvalla on vahva omistajuus.

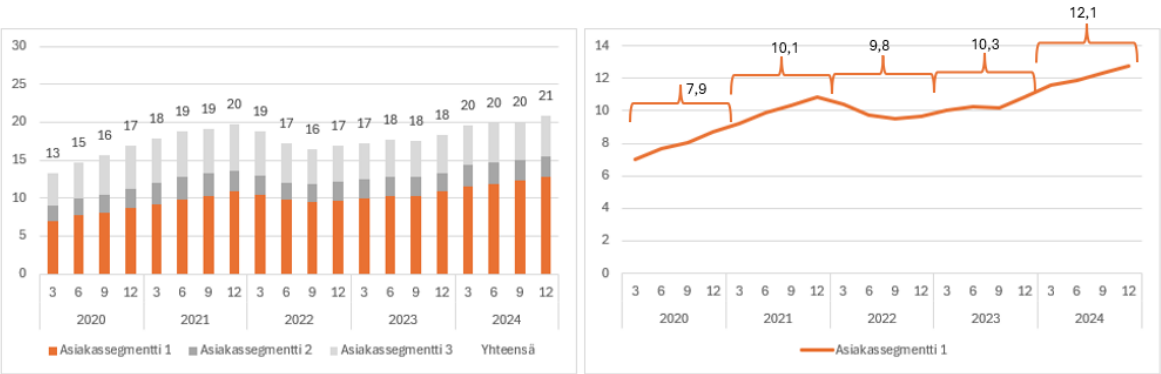
Yhteenveto: Hyödyntäjien näkökulmasta visualisoinnin esittämä sisältö koettiin kriittiseksi ja esitystapa koettiin hyväksi, se oli teoriaa vasten peilaten erinomainen. Tuottajien näkökulmasta liiketoimintajärjestelmistä irralliset järjestelmät aiheuttavat päänvaivaa.

Seuraavassa aliluvussa siirrytään taulukoista ja donitseista pylväs- ja viivadiagrammien tarkasteluun.

4.3 Pylväs- ja viivadiagrammit

Aliluvussa tullaan analysoimaan pylväs- ja viivadiagrammeja alussa kuvatun rungolla: Kuvaus visuaalista, kuvien tuottajien havainnot, teoriaan perustuva analyysi, hyödyntäjien näkökulmat ja lopuksi napakka yhteenveto. Jälkimmäinen aliluku analysoitavista kohteista on

edellistä huomattavasti pidempi, sillä pääasiallinen kuvaustapa analysoitavassa liiketoiminnassa on erilaiset graafit, yleensä joko pylväs- tai viivadiagrammit.



Kuva 19. Omistusten kehitys tuotteessa C.

Kuva 19 esittää tuotteen C omistuksia kootuissa asiakassegmenteissä eri kvartaaleiden lopussa. Oikeanpuoleinen graafi kuvassa kuvaa valitun segmentin omistuksia, kuvaan on lisätty tietyn aikajakson keskiarvo.

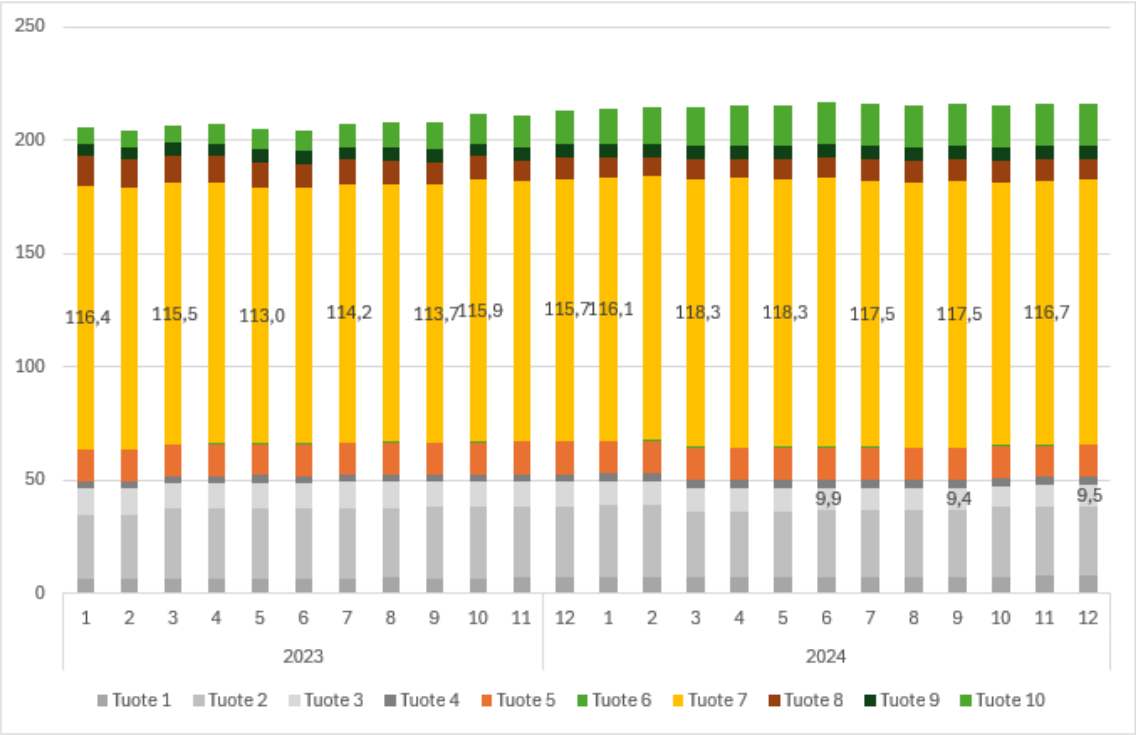
Tuottajien havainnot: Datanäkökulmasta kuvat ovat yksinkertaisia tiedon tullessa yhdeltä tuotealueelta, joskin omistajuuskategoriat eroavat kahden eri raportointilähteen välillä aiheuttaen hetkittäin haasteita. Tuottajat tunnistivat laajan joukon erilaisia tarpeita pilkkoa dataa, mitä haastaa omistus- ja omistajatiedon sijainti eri järjestelmissä hieman eri tavalla. Tietokantojen 1 ja 2 välillä on eroavaisuutta omistusten käsittelyssä, jolloin eri lähteestä katsottaessa samasta asiasta saadaan kahdet eroavat tiedot. Omistuksista saadaan tietyn ajanhetken Excel-tiedostoja, jolloin tietyn ajanhetken tilanne usein eroaa raportoinnissa olevasta ajantasaisesta tiedosta aiheuttaen kysymyksiä.

Teoriaan perustuva analyysi: Pylväskuvassa ihmisen silmä hahmottaa yleensä tasot tarkemmin, mutta pinotussa pylväskuvassa ja tässä tapauksessa kolmen eri kategorian sisäisten muutosten arviointi on haasteellista. Pinottuun pylväsgraafiin olisi mahdollista laittaa totaalitasot eri kategorioille, mutta se tekisi taas kuvasta heikosti luettavan liian informaation vuoksi. Voisi olla järkevää näyttää totaalitasot joko vuosien lopuilta tai vaihtoehtoisesti aina huipuilta ja pohjilta. Eri kategoriat näyttävät jokseenkin eri kokoisilta ja keskimmäisessä tummemman harmaassa tapahtuva prosentuaalinen isokin muutos voi hukkua oranssin

kategorian prosentuaalisesti pienen muutoksen alle. Huomattavaa on kahden vierekkäisen graafin eri y-akselisto, joista on vaikea hahmottaa oikealla olevan esityksen suhdetta vasempaan. Oikealla oleva kuva kuvaa siis vasemman kuvan pohjimmaisen pylvään kehitystä graafilla, johon on lisätty keskiarvot eri vuosilta näkyviin graafin päälle. Positiivista on saman värin käyttö molemmissa kuvissa kuvaamaan alinta kategoriaa.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmasta raportointilähteiden, käytännössä tietokantojen 4 ja 5, sekä tietyn ajanhetken Exceleiden (tarkennus lisätty), ero aiheuttaa ajoittain ihmetystä. Omistusta on mahdollista raportoida käytännössä kolmella tasolla, jotka eroavat toisistaan suhteellisen paljon, jonka lisäksi tietokannat tuottavat eriparista dataa. Yksi pyyntö olikin yrittää selkeyttää näkökulmia, jotta ymmärrys syvenisi. Yllättäen teoriasta noussut eroavan akseliston näkökulma ei herättänyt huomioita.

Yhteenveto: Tiedon eheys on suuri haaste visualisoinnin osalta. Kahden vierekkäisen graafin y-akselit olisi ensisijaisen tärkeää olla samat, jotta ei muodostu turhia väärinymmärryksiä. Muutosten arviointi pinotussa pylväskaaviossa on haastavaa, olisikin hyvä miettiä, halutaanko korostaa tiettyjä ajanhetkiä näyttämällä niiden arvot tai vaihtoehtoisesti käyttää pinoamatonta viivagraafia, jossa yhteensä olisi omana viivanaan.



Kuva 20. Palkkiotuotot eri kategorioista liukuvalta 12 kuukaudelta.

Kuva 20 esittää pinotussa pylväskuvaajassa liukuvan kahdentoista kuukauden palkkioita tuoteryhmistä A-J, yhteensä 10 tuoteryhmästä.

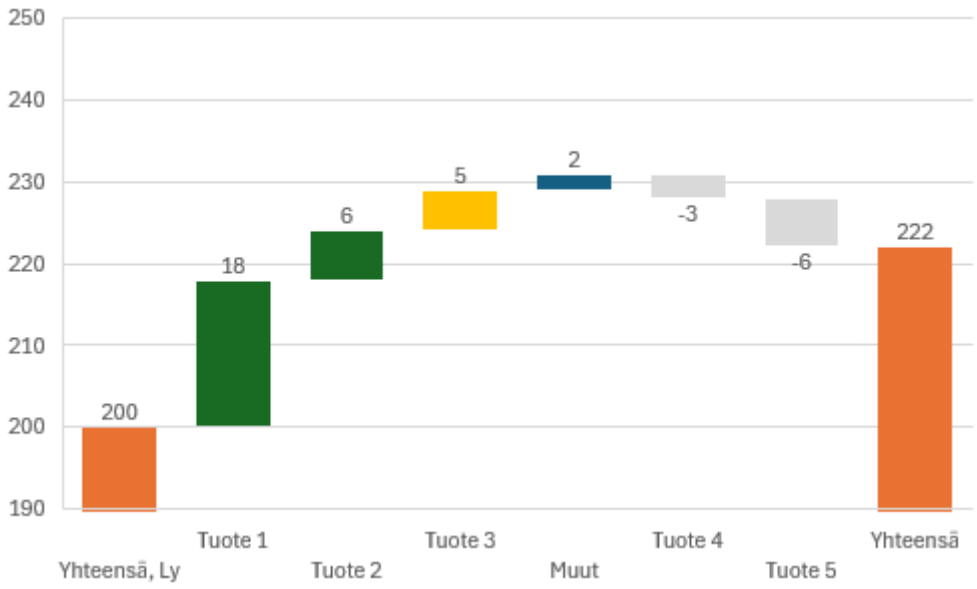
Tuottajien havainnot: Datan tuottamisen näkökulmasta graafi on monimutkainen, sillä kategorioiden data tulee monesta erillisestä järjestelmästä ja osa tuodaan käsin työkirjaan. Osa järjestelmistä käyttää eri desimaalierottimia ja kokonaisuutena kuvan tuottaminen nähtiin virheherkkänä manuaaliprosessina. Koettiin hankalaksi, että eri järjestelmissä valmistuu eri aikaan palkkiodata, jolloin kuvan kokoamisesta vastaava asiantuntija C joutuu odottelemaan asiantuntijoiden A, B, D ja E kuittauksia tietojen valmistumisesta ja syötöstä taustalla olevaan Exceliin.

Teoriaan perustuva analyysi: Visualisoinnissa huomataan samoja elementtejä kuin edellisessä kuvassa 19, joskin tässä on tuotu graafeihin arvopisteet eri sarjoille. Kuvasta ei hahmota eri kategorioiden muutoksia ja arvopisteiden tihrustaminenkin on hieman haasteellista, käytetty mallinnustyökalu tuottaa arvopisteet yleensä joka toiseen pylvääseen, mutta välillä satunnaisesti. Kuvasta voidaan huomata keltaisen palkin edustavan suurehkoa osaa kokonaisuudesta ja pienehköt muutokset siinä voivat peittää alleen pienempien kokonaisuuksien muutokset tai jopa toisen kategorian katoamisen. Teoriaan perustuen suurimman ja vähiten muuttuvan palkin sijoittaminen alimmaksi olisi myös järkevää. Huomattavaa on kategorioiden massiivinen määrä (10) samassa kuvaajassa ja voisi olla hyvä miettiä jokin selkeämpi tapa viestiä samaa asiaa, mahdollisesti ryhmitellä pienempiä kategorioita loogisiin kokonaisuuksiin, jotta esitettävät ryhmät olisivat tasapainoisempia? Teoriassa maksimimäärä on seitsemän (Nguyen et al. 2021), mutta huomioiden yhden edustavan jo reilusti yli puolta kokonaisuudesta voisi vähempikin riittää. Harmaan sävyt luovat haasteita erottaa alempia kategorioita toisistaan ja graafi on värien ilotulitusta.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmat hieman hajosivat kuvan osalta ja käytiin laajaa pohdintaa, mitä kuvalla tarkalleen ottaen halutaan kertoa. Sinällään positiiviseksi nähtiin kattava yleiskuva kaikista tuotteista ja tunnistettiin teoriaan pohjautuva haaste eri kategorioiden kokoeroista. Intressitahoja on sen verran useita, että kuvaan ei oikein löydetty parannusehdotuksia.

Yhteenveto: Tiedon tuottamisessa ja keräämisessä on merkittäviä haasteita ja riskejä. Eri intressitahot tuottavat haasteen koostetumpien kategorioiden luomisesta, joka todennäköisesti merkittävästi selkeyttäisi kuvaa ja nostaisi sen lisäarvoa. Pieniä parannuksia olisivat

suurimman palkin siirto alimmaiseksi ja mahdollisesti manuaalisesti näytettävät arvopisteet, joka on työlästä.



Kuva 21. Palkkiotuottojen muutos vuoden aikana.

Kuva 21 esittää vuoden muutosta eri palkkiotuottokategorioissa. Kuva toimii parina edellisen kuvan kahdentoista kuukauden liukuville tuotoille ja käytännössä siltaa nykyhetken vuoden takaiseen hetkeen kertoen kategoriat, joissa tuotot ovat nousseet ja laskeneet.

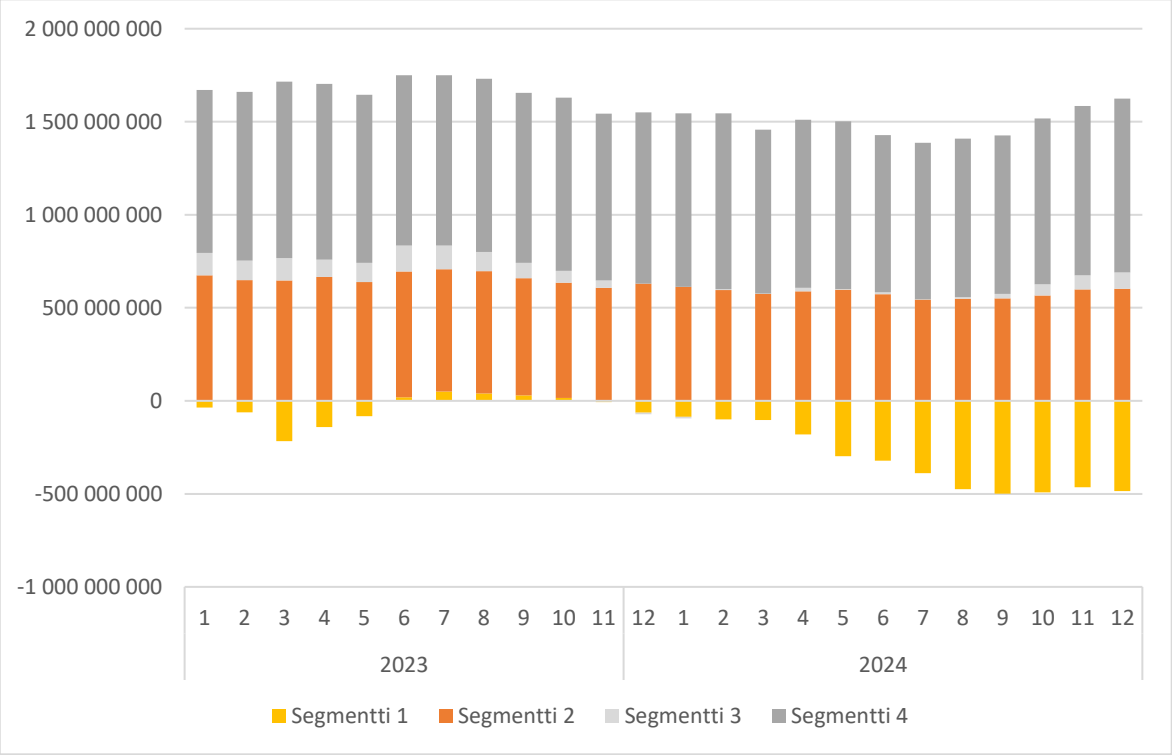
Tuottajien havainnot: Datassa näkyvät edellisen kuvan yhteydessä esitellyt haasteet. Koettiin haasteelliseksi, että tekniset muutokset tuotteiden välisessä tuotonjaossa saattavat joskus tehdä kuvasta harhaanjohtavan. Tuottajienkin keskuudessa kuvan informaatioarvo koettiin korkeaksi edelliseen kuvaan verrattuna, joka johti laajaan keskusteluun mahdollisuuksista parantaa ryhmittelyä. Tunnistettiin ongelma liiketoiminnan pirstaloitumisesta ja eri intressiryhmistä, joka vaikeuttaa kategorioiden ryhmittelyä.

Teoriaan perustuva analyysi: Tässä huomiona on pienempi määrä kategorioita verrattuna pylväskuvaan, osa pienemmistä kategorioista ja niiden muutoksista on ryhmitelty kategoriaan muut. Tämä on periaatteessa hyvä asia, mutta osa tiedosta katoaa matkalla ryhmittelyssä, joka ei ole läpinäkyvä. Värityksen kanssa voi huomata useita ongelmia, punaviherhaasteet tekevät kuvasta haasteellisen lukea, kategorioiden värit eivät linkity edelliseen kuvaan. Kuva

20 ja samaan suuntaan kehittyvät kategoriat eivät ole aina saman värisiä keskenään. Kasvavissa on vihreää ja summatasot oranssia. Akseli on katkaistu, jossa hyötynä on erojen hahmottaminen paremmin, mutta kun vieressä on 0:sta alkavalla akselilla oleva kuva, on nopealla silmäyksellä haastava hahmottaa muutosten kokoluokkaa. On ensin katsottava selitteestä, mitä kategoriaa kuvataan, jonka jälkeen on vertailtava kokonaissummaa toisesta graafista ja muutosta toisesta.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmasta esitystapa koettiin lisäarvoa luovaksi – kuvasta saa nopeasti käsityksen suurimmista muutoksista ja niille voi pyytää selityksen, jos sitä ei ole jo valmiiksi tarjottu. Hyödyntäjät eivät täysin tunnistaneet tuoteryhmien tuotonjaosta johtuvien muutosten haastetta, eikä keskusteluun noussut eroavat värit edelliseen kuvaan.

Yhteenveto: Tuottajien ja hyödyntäjien näkökulmissa on jonkun verran eritahtisuutta ja keskustelua onkin hyvä käydä. Kuva olisi hyvä ryhmitellä maksimissaan viiteen eri tuotekategoriaan, jotta kuvasta saisi paremmin selvää.



Kuva 22. Myynti eri asiakassegmenteissä.

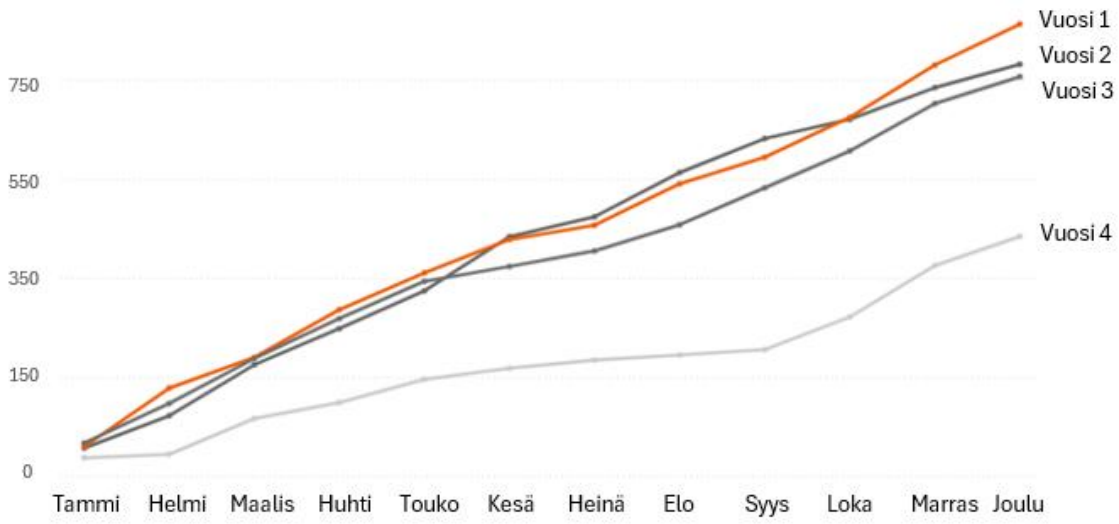
Kuva 22 esittää pinottuna pylväskaaviona liukuvaa kahdentoista kuukauden myyntiä eri asiakassegmenteissä kahdelta vuodelta kuukausittain.

Tuottajien havainnot: Datan tuottamisen näkökulmasta haasteena nähdään myyntidatan erilainen käsittely tuotealueilla. On huomattu, että tuotealueella B käsitellään ainoastaan bruttomyyntiä, kun taas useimmat tuotealueet käsittelevät oikeellisesti nettomyyntiä, jossa huomioidaan myynnin lisäksi lunastukset tai erääntymiset. Lisäksi myyntiä lasketaan hieman eroavasti tietokannoissa 4 ja 5, joka aiheuttaa eri asiakassegmenteissä mittausvirhettä. Tietokantojen eroista johtuva mittausvirhe tosin on pienempi kuin tuotteen B myynnin aiheuttama mittausvirhe.

Teoriaan perustuva analyysi: Kuva sisältää samoja haasteita kuin edellisissä kuvissa, pinottu pylväskuvaaja kertoo selkeästi totaalitason, mutta eri kategorioiden muutoksia on hankala tunnistaa. Kategoriat näyttävät hieman tasapainoisemmilta kuin aiemmissa kuvissa. Toki asiakassegmenttien ollessa eri kokoisia ei usein olekaan tarkoituksenmukaista pakottaa niitä samankokoisiksi, vaikka sekin olisi mahdollista rajauksia muuttamalla. Kuvasta näkee silmämääräisesti vaaleanharmaan palkin pienenemisen ja taas kasvun sekä keltaisen sahaamisen nollan molemmin puolin, joskin pääsääntöisesti se näyttää olevan negatiivinen. Kuvasta on hankala saada kiinni nopeampia muutoksia, koska aikajakso on pitkä. Jos myynti laskee, voi se johtua huonosta viimeisimmästä kuukaudesta tai vaihtoehtoisesti tarkastelusta pois ”putoavasta” erinomaisesta kuukaudesta. Olisi järkevää näyttää kuukausitason tai lyhyemmän jakson kehitystä kuvan lisäksi.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmasta on haasteellista, jos tieto ei ole yhteismitallista. Joinain vuosina virheellisesti käsitellyn myynnin osuus on saattanut olla neljänneskin myynnistä, tällöin kuva ei tuota neutraalia ja oikeellista kuvaa tilanteesta. Hyödyntäjät keskustelivat aikajaksosta, mutta parempaa ei oikein löydetty huomioiden lomakaudet ja tarve niiden tasoittamiseen. Budjettinäkökulmasta seuraava kuva tuo esille kumulatiivisen myynnin vuoden alusta, joten tässä sitä ei ole tarpeen näyttää.

Yhteenveto: Kuvassa on ilmeisiä datahaasteita myynnin ollessa heikosti vertailtavissa eri tuotealueiden kesken, joka olisi ensisijaisen tärkeää korjata. Pitkä tarkastelujakso näyttää trendinomaisia muutoksia, mutta nopeammat muutokset jäävät pimentoon.



Kuva 23. Myynti segmentissä 1.

Kuvassa 23 on vertailtu eri vuosien myyntiä yhdessä segmentissä kumulatiivisesti vuoden aikana.

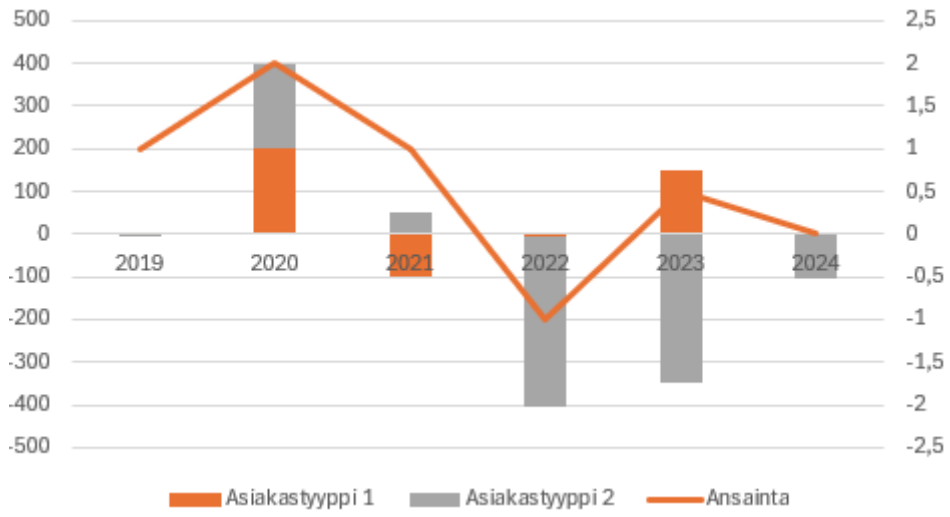
Tuottajien havainnot: Datan rajoitteet osuvat kuvan osalta syvälle tietokantakerroksiin, joista eri tuotekategoriat yhdistellään segmentin myynniksi. Datojen päivittymisessä on merkittäviä haasteita, historian säilyttämistä joudutaan laajentamaan ja tuotealueilta tulevien tietojen päällekkäisyys on ongelmallista. Tuotealueella B mittaustapa eroaa muista tuotteista.

Teoriaan perustuva analyysi: Myynnin mittaaminen YTD-näkökulmasta pohjautuu vahvasti budjetointikauteen eikä sinällään aina ole täysin järkevää, vaan voisi olla usein järkevämpää katsoa 12 kuukauden jaksoja, kuten yllä on katsottu. Yhdessä kuvat muodostavat tarkoituksenmukaisen kokonaisuuden. Kumulatiivisesti esitettynä toki kuukausien vaihtelun näkee paremmin. Oranssi viiva näyttää vahvaa myyntiä helmikuulle, kun taas alimpana oleva vaalean harmaa viiva näyttää vahvemman myynnin maaliskuulle. Toki graafista ei näe, miten oranssilla olevan vuoden heikko maaliskuun myynti suhtautuu kokonaisuutena heikomman vuoden maaliskuun hyvään myyntiin. Kokonaisuuden tarkasteluun esitystapa toimii, kun vertaillaan myyntien kehittymistä eri vuosina. Kuvassa esitellyt värit ovat sinällään loogisia, kun korostetaan oranssilla viimeisin vuosi ja edelliset vuodet ovat harmaalla.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmasta budjettivuosi näkökulmana saa aikaan luonnollisesti nyökkäilyä, koska se on normaali seurantajakso johdon palkitsemissa.

Lisäarvo tulee useammasta aiemmasta vuodesta, mutta ehkä tieto markkinan kehityksestä voisi tukea kuvan analysointia. Kuten on aiemmin todettu, myynnillä on vahva korrelaatio markkinan kehitykseen. Seuraavassa kuvassa esitellään toisen segmentin myyntiä, josta inspiraationa tähän kuvaan tai omaan kuvaansa toivottiin tietoa myynnin ansaintavaikutuksesta. Kun hyödyntäjiltä saatua toivetta käsiteltiin jälkikäteen datan rajoitteiden näkökulmasta, todettiin sen olevan varsinainen miinakenttä. Tästä päädyttiin jatkokehityksenä tekemään koeversio.

Yhteenveto: Kuvaaja täydentää eri näkökulmia myynnin mittaamiseen, mutta datahaasteet vievät siltä uskottavuutta. Olisi tärkeää korjata datapohja, jotta kuvaajien luotettavuus paranee.



Kuva 24. Myynti segmentissä 2.

Kuva 24 esittää vuositason myynnin segmentissä 2 hieman aiemmista poikkeavalla tavalla ja mukaan on lisätty ansaintavaikutusta.

Tuottajien havainnot: Datan tuottamisen näkökulmasta segmentin 2 data on hyvässä kunnossa segmenttiin 1 verrattuna. Dataa voidaan tarkastella eri kulmista ja merkittävän lisäarvon tuovaa ansaintavaikutusta pystytään laskemaan eri vuosilta vuosittain tai kumulatiivisesti. Segmentissä 2 tuotealueen B mittausongelmat saadaan oikaistua oikeellisella laskennalla tietokannassa 1.

Teoriaan perustuva analyysi: Kuva 24 esittää vuositason myynnin hieman eri tavalla ja mukaan on lisätty ansaintavaikutusta. Voisi olla tarkoituksenmukaista, että kahden kategorian myynnit esitettäisiin samalla tavalla vertailukelpoisuuden saavuttamiseksi, kun nyt peräkkäisillä dioilla on erilaiset esitystavat. Akselistolle voisi antaa otsikot, kun ne ovat eripariset, mutta muuten alkaminen nollasta on hyvä asia. Silmämääräisesti näyttäisi, että segmentin 2 myynnin ansaintavaikutus ei täysin korreloi myynnin määrän kanssa. Korreloimattomuus toki voi johtua myynnissä olevien tuotteiden koostumuksesta, ostavatko asiakkaat korkeakatteista ja lunastavat matalakatteisia tuotteita. Värien näkökulmasta oli outo valinta käyttää samaa väriä viivassa kuin pylväissä, olisi tarkoituksenmukaista valita näihin eri värit.

Hyödyntäjien näkökulma: Hyödyntäjissä kuvan osalta löydettiin merkittävä kahtiajako, jossa korostui henkilöiden tausta ja jopa tottuminen ratkaisemattomiin datahaasteisiin. Johtajat, joilla taustaa segmentin 1 puolelta, eivät nähneet vastaavaa lisäarvoa ansaintavaikutuksessa ja segmentin 2 puolella toimineille ansaintavaikutus oli jopa absoluuttista myyntiä tärkeämpi mittari. Keskustelun pohjalta toivottiin koeversiota ansaintavaikutuksesta kaikelle myynnille.

Yhteenveto: Eroava esitystapa muista myyntikuvaajista voi tehdä asennoitumisen kuvaan hieman hankalaksi. Värit tulisi valita uudelleen niin, ettei pylväissä ja viivassa ole samaa väriä kuvaamassa eri asioita.

Asiakaskategoria 1: top tuotteet

Tuotetaso 5	Netto yhteensä
Tuote 1	135 200 K€
Tuote 2	133 709 K€
Tuote 3	92 072 K€
Tuote 4	79 893 K€
Tuote 5	40 162 K€
Tuote 6	28 250 K€
Tuote 7	25 177 K€
Tuote 8	15 770 K€
Tuote 9	8 412 K€
Tuote 10	5 381 K€

Asiakaskategoria 2: top tuotteet

Tuotetaso 5	Netto yhteensä
Tuote 3	90 000 K€
Tuote 4	49 725 K€
Tuote 1	43 027 K€
Tuote 6	39 829 K€
Tuote 2	23 900 K€
Tuote 11	16 328 K€
Tuote 12	8 544 K€
Tuote 9	4 569 K€
Tuote 13	3 603 K€
Tuote 14	2 523 K€

Kuva 25. Tuotemyynti kahdessa eri asiakaskategoriassa, kymmenen myydyintä tuotetta.

Kuva 25 esittää kahden eri asiakaskategorian tuotemyynnin parhaat kymmenen tuotetta.

Tuottajien havainnot: Datanäkökulmasta tuotteiden myynnissä asiakaskategorioille on kaksi haastetta. Asiakkaiden tunnistaminen eri kategorioihin ei ole yksiselitteistä ja asiakkailla on käytännössä kaksi tapaa omistaa tuotteita, joista tunnistetaan nyt vain toinen. Tunnistettiin, että kuvissa tarkastellaan rajattua osatotuutta ja on tärkeä viestiä kuvan yhteydessä väärinymmärrysten välttämiseksi.

Teoriaan perustuva analyysi: Silmiinpistävää on vaikeus vetää yhteys esitetyistä luvuista mihinkään kokonaisuuteen. Onko myyty määrä paljon vai vähän? Asiakaskategoriat ovat segmentin 1 osajoukko, mutta niiden koko ei käy ilmi ja siten suhteuttaminen on haasteellista. Kokonaismyynnin esittäminen asiakaskategorioille erikseen sekä kokonaisuudelle olisi erinomaisen hyödyllistä. Summa esitetyistä tuotteista antaisi yhdellä silmäyksellä kokonaisvaltaisemman kuvan. Mahdollisesti tasapainoisuuden vuoksi voisi näyttää huonoiten pärjänneet tuotteet, josko rahaa olisi siirtynyt tuotteesta toiseen ja sen näkisi suoraan. Kuvat olivat hieman viimeistelemättömän oloisia.

Hyödyntäjien näkökulma: Hyödyntäjien näkökulmasta datan aiheuttamat rajaukset ovat harmilliset, eikä kuva oikeastaan kerro kokonaisuudesta, vaan lisäarvo on enemmänkin yksittäisten tuotteiden menestyksessä – mikä kiinnostaa asiakkaita juuri nyt? Taustana hyödyntäjät käyttävät vahvaa oletusta asiakkaiden samanlaisesta käyttäytymisestä eri omistustavoilla. Eräs kommentti oli toive ansaintavaikutuksen lisäämisestä kuvaan kokonaisnäkemyksen muodostamiseksi ja kokonaismyynnin selvempi esiintuonti. Periaatteessahan kokonaismyynti on aiemmilla kalvoilla, mutta kalvojen selaamista ei nähty kovin järkevänä ajankäyttönä, kun nämä kaksi asiakaskategoriaa eivät yhdessä summaudu mihinkään kokonaisuuteen. Kuvan osalta jatkotoiveet olivat aluksi hieman epäselviä, mutta näkymä tuotemyyntiin koettiin kiinnostavaksi.

Yhteenveto: Kuva on viimeistelemätön ja sen linkki muuhun liiketoimintakatsaukseen on häilyvä, ellei jopa puuttuva. Olisi järkevää miettiä uudelleen kokonaan kuvan tarkoitus ja viesti.

Seuraavassa aliluvussa tarkastellaan analyysin tuloksia kokonaisuutena.

4.4 Työn keskeiset tulokset

Teorian läpikäynnissä tunnistettiin kolme keskeistä näkökulmaa tiedon vääristämiseen ja sudenkuoppiin, jotka toimivat selkänojana analyysille. Ensimmäinen on akselisto, joka voi olla manipuloitu katkaisemalla, venyttämällä, epäloogisessa järjestyksessä tai epälineaarinen, mikä voi johtaa virhetulkintaan. Toinen näkökulma liittyy mittasuhteisiin - epätarkkuuksia esitettyjen kokojen suhteessa tai pinta-alassa sekä asioiden summaamisessa harhaanjohtavasti. Kolmas kokonaisuus käsittelee värien käyttöä: punavihersokeille haasteelliset, turhan kirjavat (ns. joulukuusi) tai läpi raportin vaihtuvat värit.

Nämä kolme kategoriaa on nostettu taulukkoon 8, joka tiivistää havainnot ja antaa kokonaiskuvan nykytilanteesta yhdellä silmäyksellä. Taulukossa on mukana huomioita, joita nousi tiedon tuottajien tai hyödyntäjien kanssa tehdyistä läpikäynneistä.

Taulukko 8. Yhteenveto käsitellyistä visualisoinneista.

Käsiteltävä visualisointi	**Akselisto**	**Kokojen**	**Värit**	**Muut huomiot**
Taulukko 6. Liiketoiminnan KPI-mittarit.				Epätahtinen päivittyminen
Taulukko 7. Kvartaalisuunnittelun status.				Yhteismitallisuus
Kuva 18. Käsitellyt tapaukset eri ajanjaksoilla.				
Kuva 19 Omistusten kehitys tuotteessa C	x			
Kuva 20. Palkkiotuotot eri kategorioista liukuvalta 12 kuukaudelta			x	Liikaa dataa, datahaasteita
Kuva 21. Palkkiotuottojen muutos vuoden aikana	x		x	Datahaasteita
Kuva 22. Myynti eri asiakassegmenteissä				Datahaasteita
Kuva 23. Myynti segmentissä 1.				Datahaasteita
Kuva 24. Myynti segmentissä 2.	x			
Kuva 25. Tuotemyynti kahdessa eri asiakaskategoriassa.				Datahasteita, kokonaisuus?

Analyysissä huomattiin, että teoriassa löydettyjä tarkoituksellisia tai vahingossa tapahtuneita vääristymiä tulisi rikastaa asioiden suhteuttamisen näkökulmalla. Kokonaisuuteen

suhteuttamisen puute ei ole graafisessa esityksessä mitenkään virheellistä, mutta se voi heikentää datan ymmärrettävyyttä. Kuvassa *25* esitettiin myynti tuotteittain, mutta rinnalta puuttui tieto kokonaismyynnistä, jolloin suhteuttaminen oli yksittäisenä kuvana hankalaa.

Käytetyissä graafeissa on vältetty katkaistujen akseleiden käyttöä, joka vastasi teorian mukaisestikin hyvää käytäntöä. Joissakin tapauksissa huomattavat erot tarkasteltavien kohteiden välillä olisivat perustelleet poikkeuksellisen akselien katkaisun, mikäli se olisi ilmaistu selkeästi. Akselistot olivat johdonmukaisesti vasemmalta oikealle ja asteikot olivat loogiset kuvaa 19 lukuun ottamatta, jossa yhtenäinen asteikko olisi lisännyt vertailukelpoisuutta.

Kokojen suhteellisuuden osalta ei esityksessä havaittu teorian mukaisia virheitä, mutta useammassa kuvassa esitettyjen pylväiden suhteuttaminen toisiin oli hankalaa ja kehityksen tarkka seuranta oli haasteellista. Tätä voisi kehittää tuomalla mukaan arvopisteet sarjoille, puolivuosittain tai kvartaalien lopuille. Arvopisteiden kanssa tulisi välttää informaatioähkyä ja lisätä niitä harkiten.

Värien käyttö keskittyi perinteisesti brändiväreihin sekä liikennevaloihin. Jälkimmäinen ei sovi punaviherhaasteellisille ja parempi olisi käyttää muita symboleja, esim. nuolia tai vastaavia indikaatioita kehityksen suunnasta. Kuvien 19 ja 20 värit voisivat olla selkeämmät ja yhteneväiset toistensa kanssa, erityisesti kuvassa 20 esiintyneet lukuisat harmaan sävyt ja monet eri värit vaikeuttivat tulkintaa. Lisäksi huomattiin kuvan 21 katkaistu akseli ja kuvasta 24 puuttuneet akselien selitteet, joiden lisääminen olisi parantanut luettavuutta.

Yleisesti ottaen analysoiduissa esityksissä ei havaittu selkeitä virheellisyyksiä, mutta monien visualisointien selkeyttä voitaisiin parantaa pienillä muutoksilla. Teoriasta nousi esiin karttavisualisointien puute: markkinaosuuksia tai myyntiä voitaisiin tarkastella maantieteellisesti kuntatasolla tai asiakkaiden varallisuuteen perustuen. Näin voitaisiin tunnistaa alueita, joissa myynti on vahvaa tai heikkoa sekä vertailemaan tuotteiden alueellista menestystä. Samalla voitaisiin tunnistaa parhaat käytännöt myynnin onnistumiseen.

5 Johtopäätökset

Työssä lähestyttiin tiedon visualisointia eri näkökulmista ja pyrittiin löytämään keinoja tuoda lisäarvoa haastateltuun kohdeyritykseen. Usein yrityksissä visualisointeja työksi tekevät henkilöt ovat oppineet paljon kokeilemisen ja erehdyksen kautta. Käytännönläheinen tapa oppia ei ole väärin, mutta teoreettinen tarkastelu tuo uudenlaisia näkökulmia. Uudet tavat nähdä visualisointeja voivat joko vahvistaa jo tehtyjä visualisointipäätöksiä ja tuoda mahdollisuuksia kehittää teoreettiseen selkänojaan nojaten raportteja eteenpäin. Haastatellut henkilöt saivat arvokasta palautetta visualisointien ymmärrettävyydestä, kun visualisointeja arvioitiin ulkopuolisen silmin.

Tutkimuksen tulokset osoittavat, että suurin osa vääristymistä ei synny tahallisesta manipuloinnista, vaan datan rakenteellisista rajoitteista, yhteismitallisuuden puutteesta ja raportointityökalujen teknisistä ominaisuuksista. Seuraavassa esitetään vastaukset tutkimuskysymyksiin, työn merkitys ja jatkotutkimuskohteet.

5.1 Vastaukset tutkimuskysymyksiin

Tässä aliluvussa esitellään vastaukset kolmeen tutkimuskysymykseen tutkimukseen pohjautuen.

Mitä asioita yrityksen tulisi huomioida datan visualisoinnissa?

Tutkimus osoitti, että yrityksen tulisi datan visualisoinnissa huomioida visualisoinnin tarkoitus ja kuvattavan asian luonne, jotta osataan valita tilanteeseen sopiva visualisointikeino. Visualisoinnit tulisi valita kohdeyleisö ja käyttötarkoitus huomioiden. Datan visualisoinnissa tulisi keskittyä avaintietojen esittämiseen, eikä visualisointiin pitäisi sisällyttää ylimääräisiä tietoja. Datan visualisointi tulisi toteuttaa mahdollisimman yksinkertaisella tavalla. Datan ymmärrettävyys paranee yrityksen visualisoidessa sen mahdollisimman yksinkertaiseen muotoon ja sisältäen ainoastaan tarvittavat osat. Visualisointityökalujen suunnitteluominaisuuksia käyttämällä voidaan edesauttaa visualisointien oikeanlaista tulkintaa, ja visualisointeihin tulisi sisällyttää todellisia lukuja ymmärrettävyyden lisäämiseksi. Värien käyttö on tärkeässä roolissa visualisoinnissa, sillä asianmukaisesti valitut sävyt tukevat datan

järjestelmällistä vertailua. Analyysissä havaittiin (aliluku 4.4 Työn keskeiset tulokset) akselien ja värien johdonmukaisuuden tukevan ymmärrettävyyttä ehkäisten tahattomia vääristymiä.

Miten datan visualisoinnin avulla voidaan vääristää vahingossa tai tarkoituksellisesti visualisoinnin antamaa informaatiota?

Tutkimuksen perusteella voidaan todeta, että monet tekijät voivat vaikuttaa kaavioiden ja visualisointien oikeellisuuteen. Tekijä voi joko tahallaan vai vahingossa luoda sellaisen visualisoinnin, mikä ei vastaa todellista kuvaa. Kaavioiden avulla voidaan johtaa kohdeyleisöä harhaan muuttamalla akseleiden suhteita tai kuvasuhdetta. Katkaistun akselin tai kahden kaavion vertaileminen eri asteikoilla saattaa aiheuttaa sen, että kaavio johtaa harhaan. Kaavioita ja visualisointeja luodessaan tulisikin aina ottaa huomioon, että kaavio kuvaa oikeaa ja todellista asiaa.

Kuvaavatko yrityksen raportoinnissa käytetyt visualisoinnit esitettävää dataa tarkoituksenmukaisesti ja neutraalisti tarjoten pohjan oikeaan tietoon perustuvalle päätöksenteolle?

Osan haasteista loi jo etukäteen ongelmalliseksi tunnistettu brändivärien käyttö ja niiden sopivuus asioiden esittämiseen. Tiedon hyödyntäjien haastatteluista saatiin yritykselle arvokasta lisätietoa visualisointien ja kokonaiskuvan yhteydestä – monesti ongelmat kokonaisuuden esittämisen haasteista kumpusivat tiedon rajoitteista.

5.2 Tutkimuksen merkitys, rajoitteet ja jatkotutkimus

Tutkimuksen haaste ja lisäarvo tulee staattisesta ajanhetkestä, jolloin tutkimuskohdetta päästään tutkimaan ja kuvailemaan sekä tekemään parannusehdotuksia. Tietojen visualisoinnin teoria ei itsessään muutu jatkuvasti, mutta samaa ei voida sanoa visualisointiin käytetyistä työkaluista ja tekniikasta niiden taustalla. Tulevaisuuden tutkimussuuntana tulee varmasti vahvasti olemaan AI:n luomien kuvien tutkimus. Jo nyt on tunnistettu AI:n osalta vakavaa hallusinointia. Siirrettäessä aina enemmän rutiinitöitä sekä datan murskaamista pois ihmisiltä tulee lopputuotosten varmentamisesta entistä tärkeämpää. Ehkä meillä on lopulta useampi AI-työkalu tai -agentti, jotka tarkastavat ristiin toistensa tuotoksia toimien itsenäisinä arvioijina. Voi hyvin olla, että visualisointien luojille ja datan ymmärtäjille on jatkossakin kysyntää.

Kohdeyrityksessä pikavoittoja ei ole saatavissa. Nykyisellä datapohjalla nykyaikaisen business intelligenceä tukevan raportoinnin luominen ja ylläpitäminen on haastavaa, manuaalista ja sitä kautta virheherkkää. Harmillisesti tietokantojen uudistaminen ja dataosaamisen vahvistaminen eivät ole mitenkään helppoja tai nopeita harjoituksia – yrityksestä riippuen vuoden tai viiden vuoden projekteja. Kohdeyrityksen monimutkainen organisaatiorakenne tuo omat haasteensa datakehittämiseen ja vaatii erityisen vahvaa johtajuutta saavuttaakseen tason, jolla sen olisi hyvä olla.

Yhteenvetona voidaan todeta, että eettinen ja informatiivinen datan visualisointi perustuu tasapainoon kolmen tekijän välillä: datan luotettavuus, visuaalinen selkeys ja konteksti. Tutkimus osoitti, että vaikka datan tekniset rajoitteet ja organisaation rakenteet voivat rajoittaa täydellistä toteutusta, pienilläkin parannuksilla voidaan merkittävästi lisätä ymmärrettävyyttä ja päätöksenteon laatua. Datan visualisointi ei ole vain tekninen tehtävä, vaan olennainen osa tiedon tulkintaa ja vastuullista johtamista.

Lähteet

6sense (2024). "Data Visualization Software Market Share." Viitattu 18.8.2024. Saatavissa: https://6sense.com/tech/data-visualization

Adolfsson, C., Arvidson, G., Gidlund, P., Nordberg, A. & Nordberg, L. (2010). A General Methdology for Selective Data Editing. Saatavissa: http://gauss.stat.su.se/master/statdatabaser/HT10/Literature/SwedishEditingMethods.pdf

Alhadad, S. (2018). Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science. Journal of Learning Analytics. 5. 60-85. Saatavissa DOI: 10.18608/jla.2018.52.5.

Ahmed, S.K., Mohammed, R.A., Nashwan, A.J., Ibrahim, R.H., Abdalla, A.Q., Ameen, B.A.M. & Khdhir, R.M. (2025). Using thematic analysis in qualitative research, Journal of Medicine, Surgery, and Public Health, Volume 6, 2025, 100198, ISSN 2949-916X. Saatavissa DOI: 10.1016/j.glmedi.2025.100198.

Armitage, A. (1966). Edmond Halley. London: Thomas Nelson and Sons, 1966. A biography that focuses on Halley's many scientific achievements.

Best, J. (2005) 'Lies, Calculations and Constructions: Beyond "How to Lie with Statistics"', Statistical Science, vol. 20, no. 3, pp. 210-214. Saatavissa DOI: 10.1214/088342305000000232.

Cook, A. (1998). Edmond Halley: Charting the Heavens and the Seas. Oxford, England: Clarendon Press, 1998. A comprehensive account of Halley's life and science.

Coherent Market Insights (2024). "Data Visualization Tools Market - Price, Size, Share & Growth." Available at: Coherent Market Insights

D3.js 2024. Viitattu 18.8.2024. Saatavissa: D3 by Observable | The JavaScript library for bespoke data visualization (d3js.org)

Diehl, J., 2004. Preprocessing and visualization. Aachen, Germany: RWTH Aachen University.

Dougherty, J. & Ilyankou, I. (2025) Hands-On Data Visualization – Interactive Storytelling from Spreadsheets to Code. O'Reilly Media Inc. Saatavissa: https://handsondataviz.org/

Elgendi, M., Howard, N. & Hussain, A. (2020). From ancient times to modern: realizing the power of data visualization in healthcare and medicine. Big Data Analytics 5, 4 (2020). Saatavissa DOI: 10.1186/s41044-020-00047-z.

Eriksson, P & Koistinen, K. (2014). Monenlainen tapaustutkimus. Kuluttajatutkimuskeskus, tutkimuksia ja selvityksiä 11/2014. Saatavissa: https://helda.helsinki.fi/items/db985ead-f6d1-4537-a432-4267f321a5c5

Famili, A., Shen, WM., Weber, R. & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. Intelligent Data Analysis, Volume 1, Issues 1–4, 1997, Pages 3-23, Saatavissa DOI: 10.1016/S1088-467X(98)00007-9.

Ferrell, O.C., Harrison, D.E., Ferrell, L.K., Ajjan, H. & Hochstein, B.W. (2024). A theoretical framework to guide AI ethical decision making. AMS Review, vol. 14, no. 1-2, pp. 53-67. Saatavissa DOI: 10.1007/s13162-024-00275-9.

Friendly, M. (1999). Re-Visions of Minard. Saatavissa: /users/faculty/friendly/Library/Documents/tex/minard/minard.dvi (yorku.ca)

Friendly, M. & Denis, D. (2000). Discussion and comments. Approche graphique en analyse des données. The roots and branches of modern statistical graphics. Journal de la société française de statistique 141.4 (2000): 51-60. Saatavissa: http://eudml.org/doc/199813

Friendly, M. (2005). Milestones in the History of Data Visualization: A Case Study in Statistical Historiography. In: Weihs, C., Gaul, W. (eds) Classification — the Ubiquitous Challenge. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. Saatavissa DOI: 10.1007/3-540-28084-7_4.

Gebco (2024). World map of bathymetric data. Viitattu 5.11.2025. Saatavissa: https://commons.wikimedia.org/wiki/Category:Bathymetric_maps_of_the_world#/media/File:World_map_of_bathymetric_data_-_GEBCO_2014.jpg

Giering, M. (2008). Retail sales prediction and item recommendations using customer demographics at store level. SIGKDD Explor. Newsl. 10, 2 (December 2008), 84–89. Saatavissa DOI: 10.1145/1540276.1540301.

GMI (2024). Data Visualization Tool Market Size - By Component (Software, Services), By Application (Finance, Human Resources (HR), Sales & Marketing, Supply Chain), By Organization Size (SMEs, Large enterprise), By Deployment Mode, By End Users & Forecast, 2024 – 2032. Saatavissa: Data Visualization Tool Market Size & Share, Growth Report 2032 (gminsights.com)

Google Data Studio 2024. Viitattu 18.8.2024. Saatavissa: data_studio_product_overview.pdf (google.com)

Grandview Research (2024). Data Visualization Tools Market Size, Share & Trends Analysis Report By Component (Software, Service), By Application, By Organization Size, By Deployment, By End-users, By Region, And Segment Forecasts, 2023 – 2030. Viitattu 18.8.2024. Saatavissa: Data Visualization Tools Market Size & Trends Report, 2030 (grandviewresearch.com)

Graunt, J. (1662). Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality. London: Martin, Allestry, and Dicas.

Guerry, A.-M. (1833). Essai Sur La Statistique Morale de la France. Paris: Crochard. English translation: Hugh P. Whitt and Victor W. Reinking, Lewiston, N.Y. : Edwin Mellen Press, 2002. 10, 39.

Guerry, A.-M. (1864). Statistique morale de l'Angleterre comparee avec la statistique morale ´ de la France, d'apres les comptes de l'administration de la justice criminelle en Angleterre et ´ en France, etc. Paris: J.-B. Bailliere et fils. ` BNF: GR FOL-N-319; SG D/4330; BL: Maps 32.e.34; SBB: Fe 8586; LC: 11005911. 11

Halley, E. (1686). An historical account of the trade winds, and monsoons, observable in the seas between and near the Tropicks, with an attempt to assign the physical cause of the said winds. Philosophical Transactions Society 16: 153–168 Saatavissa DOI: 10.1098/rstl.1686.0026.

Heer, J, & Robertson G.G. (2007). Animated Transitions in Statistical Data Graphics. IEEE transactions on visualization and computer graphics, vol. 13, no. 6, pp. 1240-1247, Nov.-Dec. 2007. Saatavissa DOI: 10.1109/TVCG.2007.70539.

Hill, A. (2024). Are pie charts evil? An assessment of the value of pie and donut charts compared to bar charts. Information Visualization, 24(1), 3-23. Saatavissa DOI: 10.1177/14738716241259432.

Hoogland, J., van der Loo, M., Pannekoek, J. & Schoitus, S. (2011). Data editing: Detection and correction of errors. Statistics Netherlands.

Ipsen, D.C. (2004). Edmond Halley: More than a Man with a Comet. Philadelphia: Xlibris, 2004.

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nat Mach Intell 1, 389–399. Saatavissa DOI: 10.1038/s42256-019-0088-2.

Joseph, G.P. (1992). Modern Visual Evidence. New York.

Kittivorawong, C. (2024). Legible Label Layout for Data Visualization, Algorithm and Integration into Vega-Lite, Ithaca. Saatavissa DOI: 10.48550/arXiv.2405.10953.

Krok, Ewa. (2021). Visualization on charts – manipulations and distortions. Procedia Computer Science. 192. 3932-3944. Saatavissa DOI: 10.1016/j.procs.2021.09.168.

Kosslyn, S. M. (2006). Graph design for the eye and mind. 1st ed. New York: Oxford University Press. Saatavissa DOI: 10.1093/acprof:oso/9780195311846.001.0001.

Lisnic, M., Polychronis, C., Lex, A. & Kogan, M. (2023). Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 817, 1–21. Saatavissa DOI: 10.1145/3544548.3580910.

Lisnic, M., Lex, A., & Kogan, M. (2024). "Yeah, this graph doesn't show that": Analysis of Online Engagement with Misleading Data Visualizations. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 199, 1–14. Saatavissa DOI: 10.1145/3613904.3642448.

Mansoor, H. & Harrison, L. (2018). Data visualization literacy and visualization biases: Cases for merging parallel threads, Cognitive Biases in Visualizations, pp. 87-96. Saatavissa DOI: 10.1007/978-3-319-95831-6_7.

Muskan, Singh, G., Singh, J. & Prabha, C. (2022). Data Visualization and its Key Fundamentals: A Comprehensive Survey, 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1710-1714. Saatavissa DOI: 10.1109/ICCES54183.2022.9835803.

Mordor Intelligence (2024). "Data Visualization Tools Market - Companies, Trends & Share."

Nguyen, V.T., Jung, K. & Gupta, V. (2021). Examining data visualization pitfalls in scientific publications. Vis. Comput. Ind. Biomed. Art 4, 27. Saatavissa DOI: 10.1186/s42492-021-00092-y.

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The Persuasive Power of Data Visualization. IEEE Transactions on Visualization and Computer Graphics, 20(12), 2211-2220. Saatavissa DOI: 10.1109/TVCG.2014.2346419.

Pandey, A. V., Rall, K., Satterthwaite, M., Nov, O. & Bertini, E. (2015). How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques. CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1469 – 1478. Saatavissa DOI: 10.1145/2702123.2702608.

Petty, W. (1665). The economic writings of Sir William Petty: together with the observations upon the bills of mortality. Cambridge: The University Press. C. H. Hall (ed.) (more probably by Captain John Graunt).

Power BI 2024. Viitattu 18.8.2024. Saatavissa: Power BI - Data Visualization | Microsoft Power Platform

Qlik 2024. Viitattu 18.8.2024. Saatavissa: Qlik Analytics | Qlik Sense

Runkler, T. (2020). Data Analytics: Models and Algorithms for Intelligent Data Analysis. Saatavissa DOI: 10.1007/978-3-658-29779-4.

SAS (2025). Viitattu 1.11.2025. Saatavissa: https://www.sas.com/fi_fi/software/viya.html#operational-decisioning

Siirtola, H., Räihä, K.-J., Istance, H. & Špakov, O. (2019). Dissecting Pie Charts. 17th IFIP Conference on Human-Computer Interaction (INTERACT), Sep 2019, Paphos, Cyprus. pp.688-698, ff10.1007/978-3-030-29384-0_41. hal-02544607.

SNS Insider (2024). Data Visualization Tools Market Size, Share & Segmentation By Tool (Standalone, Integrated), By Organization Size (Large Enterprises And Small And Medium-Sized Enterprises), By Deployment Mode (On-Premises And Cloud), By Business Function (Marketing And Sales, Human Resources, Operations, And Finance), By Vertical (BFSI, Government, Healthcare And Life Sciences, Retail And ECommerce, Manufacturing, Transportation And Logistics, Telecommunications And IT, Other Verticals), By

Regions And Global Market Forecast 2024-2032. Saatavissa: Data Visualization Tools Market Size, Share, & Forecast 2032 (snsinsider.com)

Tableau 2024. Viitattu 18.8.2024. Saatavissa: Business Intelligence and Analytics Software | Tableau

Tufte, E. R. (1983). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press. Saatavissa DOI: 10.1109/MPER.1988.587534.

Tufte, E. R. (1997). Visual Explanations. Cheshire, CT: Graphics Press.

Vantage (2025). Stocks and USD crater after liberation day shock. Viitattu 28.7.2025. Saatavissa: https://www.vantagemarkets.com/market-analysis/stocks-and-usd-crater-after-liberation-day-shock/#:~:text=A%20combined%20%242.4%20trillion%20in%20stock%20market%20value,the%20Nasdaq%20since%20the%20pandemic%20in%20March%202020

Wainer, H., Friendly, M. (2022). On the Origins of Data Visualization. In: Carriquiry, A.L., Tanur, J.M., Eddy, W.F. (eds) Statistics in the Public Interest. Springer Series in the Data Sciences. Springer, Cham. Saatavissa DOI: 10.1007/978-3-030-75460-0_27.

Weissgerber, T. L., Milic, N.M., Winham, S.J. & Garovic, V.D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS biology. 13 (4), e1002128. Saatavissa DOI: 10.1371/journal.pbio.1002128.

Welhausen, C.A. (2022). Precarious Data: Crack, Opioids, and Enacting a Social Justice Ethic in Data Visualization Practice. IEEE Transactions on Professional Communication, vol. 65, no. 1, pp. 50-69, March 2022. Saatavissa DOI: 10.1109/TPC.2022.3144826.

Wikicommons (2025). Viitattu 5.11.2025. Saatavissa: https://commons.wikimedia.org/wiki/File:Charles_Minard_-_1858_-_Carte_figurative_et_approximative_des_quantit%C3%A9s_de_viandes_de_boucherie_envoy%C3%A9es_sur_pied_par_les_d%C3%A9partements_et_consommateurs_%C3%A0_Paris.jpg

Yin R.K. (2014). Case study research: Design and methods. Sage. Thousand Oaks, CA. Saatavissa DOI: 10.3138/cjpe.30.1.108.

Liite 1. Datan tuottajien työpajamuotoiset teemahaastattelut

Työpajat järjestettiin 15.8.2024, 22.8.2024 ja 28.8.2024. Tavoitteena oli tarkastella yrityksen nykyistä raportointia ja sen kehittämismahdollisuuksia eheämmäksi kokonaisuudeksi erityisesti datan rakenteellisten rajoitteiden näkökulmasta. Työpajoissa keskityttiin uuden liiketoimintakokonaisuuden mukaisiin tarpeisiin ja erityisesti eri tuotteisiin jakautuneiden omistusten sekä niiden myynnin tarkasteluun.

Raportoinnin keskeisenä haasteena tunnistettiin datan eheyden varmistaminen —kaiken olennaisen tiedon saaminen mukaan raportteihin ja duplikaattien ehkäisy.

Työpajojen osallistujat: päällikkö A sekä asiantuntijat A, B, C, D ja E.

Työpajojen tavoitteet:

- Tunnistaa nykyisen raportoinnin keskeiset rajoitteet ja tiedon katvealueet.
- Selvittää, miten eri järjestelmistä tuleva data yhdistyy ja missä kohtaa syntyy päällekkäisyyksiä.
- Arvioida nykyisten visualisointien ja mittarien riittävyyttä uuden liiketoimintakokonaisuuden seurannassa.
- Löytää konkreettisia kehitysehdotuksia raporttien sisällön, eheyden ja käytettävyyden parantamiseksi.

Keskustelun teemat ja kysymykset

- Datan lähteet ja laatu
    - Mistä lähteistä raportoitava data on peräisin?
    - Kuinka paljon dataa on muokattu ennen visualisointia?
    - Mitkä ovat suurimmat datan eheys- tai yhteismitallisuusongelmat?
    - Missä vaiheessa prosessia datan virheet yleensä havaitaan?
- Visualisointien tekeminen ja työkalut
    - Mitä työkaluja visualisointien tuottamiseen käytetään?
    - Onko yrityksen mallivärit käytössä kaikissa visualisoinneissa?
    - Mitkä asiat rajoittavat visualisointien muokkaamista tai tuottamista?
    - Mitä valintaperusteita on käytetty visualisointityyppien valinnassa?
    - Miten varmistetaan, että raportit ovat visuaalisesti johdonmukaisia ja sisällöllisesti tarkkoja?

Haastattelut toteutettiin puolistrukturoituna teemahaastatteluna työpajamuotoisesti.

Liite 2. Tiedon hyödyntäjien näkökulma

Tiedon hyödyntäjien palaute saatiin perustuen yhtiö A:n hallituksen kokouksiin 25.10.2024 ja 13.12.2024, liiketoiminnan kvartaalikokouksiin 24.10.2024 ja 6.2.2025, sekä erillisissä läpikäynneissä johtaja C:n kanssa 24.1.2025. Haastatteluaineisto koostui käytännössä kahdesta eri kanavasta.

1. Epäsuora palaute. Ensimmäinen kanava oli hallituksen kokouksista ja niiden ulkopuolisista keskusteluista johtajien A ja B kautta joko suoraan tai päällikkö B:n tai johtaja C:n kautta. Toinen kanava oli kvartaalikokouksien jälkeiset keskustelut johtajien C, D ja E kanssa.
2. Suora palaute. Tätä saatiin erityisesti johtaja C:n kanssa läpikäynnissä 24.1.2025, jossa käytiin rauhassa kuva kuvalta liiketoimintakatsausta läpi. Hallituksen kokouksissa saatiin lisäksi muutamia suoria mainintoja.

Ensimmäinen kanava, hallituksen ja liiketoiminnan kvartaalikokoukset, heijasti käytännön päätöksentekotilanteita, joissa toimitusjohtajan katsauksena toiminut liiketoimintakatsaus käsitellään tiiviisti tarttuen kiinnostaviin asioihin. Toisessa kanavassa käytiin syvällisempi ja rauhallinen läpikäynti visualisointien tarkoituksenmukaisuudesta.

Käsittelyiden tavoitteet

- Selvittää kuinka nykyisiä raportteja ja visualisointeja hyödynnetään päätöksenteossa.
- Tunnistaa visualisointien tulkintaan, selkeyteen ja luotettavuuteen liittyviä haasteita.
- Koota mahdollisia kehitysehdotuksia visualisointien ymmärrettävyyden ja viestinnällisyyden parantamiseksi liiketoimintakatsauksen seuraavaan versioon.

Käsitellyt teemat ja kysymykset

- Miten toimitusjohtajan katsaus (=liiketoimintakatsaus) tukee hallituksen päätöksentekoa?
- Mitä osia katsauksesta hallituksen jäsenet pitävät tärkeimpänä ja miksi?
- Mitkä kaaviot tai kuvat herättävät eniten kysymyksiä tai vaativat lisäselvitystä?
- Onko esitystavoissa tai väreissä piirteitä, jotka hämmentävät tai tuntuvat harhaanjohtavilta?
- Mitä tietoja tai visuaalisia esityksiä toivoisitte lisää liiketoimintakatsaukseen?
- Millä tavoin visualisointien kehittämistä voitaisiin jatkossa tukea yhteistyössä tuottajien ja hyödyntäjien välillä?