

## Organisaation datalukutaidon arvioiminen

Jussi Sandholm



<b>Tekijä</b> Jussi Sandholm	
<b>Koulutusohjelma</b> Liiketoiminnan teknologiat	
<b>Opinnäytetyön nimi</b> Organisaation datalukutaidon arvioiminen	<b>Sivu- ja liitesivumäärä</b> 55 + 7
<p>Datan vaikutusvalta lisääntyy jatkuvasti. Varsinkin liike-elämässä sen merkitys on jo huomattava. Moni meistä kuitenkin ymmärtää dataa huonosti tai ei ollenkaan. Mitä data on ja miten sitä voidaan käyttää hyväksi päätöksenteossa? Vasta viime aikoina organisaatiot ovat alkaneet kiinnittää huomiota tähän datalukutaidoksi kutsuttuun osaamisalueeseen. Nykypäivänä organisaatioissa ei ole montaakaan työtehtävää, joihin ei kuulu datan kanssa työskentelyä. Tai ainakin tällaisten työtehtävien voidaan olettaa katoavan lyhyen ajan sisällä. Organisaation tulee pystyä arviomaan ja kehittämään kyvykkyyksiään dataan liittyen, että kasvu ja kannattavuus säilyvät huomennakin. Ei kuitenkaan riitä, että organisaatioon palkataan parhaat tekijät, vaan organisaation tulee tukea datalukutaitoa omien prosessiansa ja IT-infrastruktuurinsa kautta.</p> <p>Opinnäytetyössä on tavoitteena luoda liike-elämässä tarvittavien datalukutaitojen perustaso organisaation tasolla arvioiva verkkokysely. Ensimmäiset toteutukset verkkokyselystä toteutettiin Google Forms- kyselytyökalulla, mutta kysely voidaan toteuttaa jatkossa myös muilla tavoilla. Verkkokyselyn kautta saatavat tulokset antavat arvokasta tietoa sekä työn toimeksiantajalle Cubiq Analytics Oy:lle sekä sen asiakkaille eli arvioinnin kohdeorganisaatioille. Toimeksiantaja pystyy vastausten ja palautteiden perusteella kehittämään datalukutaidon arvioinnin toteutustapaa, sekä parantamaan sen tehtävien laatua. Tulosten perusteella voidaan paremmin ymmärtää datalukutaidon tasoa, mikä auttaa kehittämään myös konsulttien osaamista. Tulosten perusteella kohdeorganisaatio saa arvion datalukutaidon nykytilasta ja sen on helppo löytää mahdolliset kehityskohteet sekä seurata kehitystoimenpiteiden vaikutuksia väliajoin toteutettavien uusinta-arvioiden avulla.</p>	
<b>Asiasanat</b> Data, monilukutaito, informaatio, visualisointi	

## Sisällys

1	Johdanto .....	1
1.1	Keskeiset käsitteet .....	1
1.2	Taustaa .....	1
1.3	Työn tavoite, tarkoitus ja rajaus .....	3
1.4	Tutkimuksellinen lähestymistapa ja tutkimusmenetelmät .....	5
2	Data .....	8
2.1	Tietämyksenhallinta .....	11
2.2	Tiedonkulun rajoitteita .....	16
2.3	Datastrategia .....	16
2.4	Korrelaatio ja kausaalisuus .....	18
3	Visualisoinnit .....	20
3.1	Visualisointi ja Infograafi .....	22
3.2	Visualisointi prosessina .....	24
4	Ihmismielen toiminta: kuinka teemme päätöksiä .....	32
5	Datalukutaidon selvittäminen .....	38
5.1	Datalukutaitoa selvittävä kysely .....	38
5.2	Testin analyysi ja tulokset .....	40
6	Pohdinta .....	48
	Lähteet .....	51
	Liitteet .....	56
	Liite 1. Kyselylomake .....	56

# 1 Johdanto

## 1.1 Keskeiset käsitteet

Tässä opinnäytetyössä esiintyvät usein seuraavat tärkeät käsitteet:

### Data

Tässä yhteydessä datalla tarkoitetaan informaatiota digitaalisessa muodossa, joka on jalostettavissa ja siirrettävissä. (Merriam-Webster 2020a).

sekä

Informaatiota, erityisesti faktatietoa tai numeroita, jota on kerätty tutkittavaksi, harkittavaksi ja käytettäväksi auttamaan päätöksenteossa. (Cambridge Dictionary 2020).

### Datalukutaito (eng. Data Literacy)

Kyky tulkita, ymmärtää, luoda ja kommunikoida dataa informaationa. Kuten lukutaito yleisenä käsitteenä, datalukutaito keskittyy tiedon käsittelyyn liittyviin kompetensseihin. Se ei kuitenkaan ole samanlainen taito kuin normaali lukutaito, koska se edellyttää tiettyjä taitoja, kuten datan ymmärtäminen ja lukeminen. (Baykoucheva 2015, 80).

## 1.2 Taustaa

Kaikki yritykset toimialasta riippumatta keräävät dataa enemmän kuin koskaan ennen. Vähittäiskaupat seuraavat jokaista ostotapahtumaamme, lentoyhtiöt seuraavat kuinka hyvin lennot ovat aikataulussa, ja melkein jokaista asiakaspalvelupuhelua, sähköpostia tai muuta verkon tapahtumaa seurataan. Mikä on kaiken tämän datan keräämisen tarkoitus? Datan avulla voidaan muun muassa paremmin ymmärtää asiakkaan tarpeita, kehittää palveluja ja toimintaa sekä pyrkiä parempiin rekrytointipäätöksiin. Olemme siirtyneet datan kultaiselle aikakaudelle. Teknologinen kehitys pilvipalveluiden muodossa on mahdollistanut aiemmin vain suurten yritysten saatavilla olevat ratkaisut kaikkien organisaatioiden saataville ja vielä kohtuullisin kustannuksin. (Bersin & Zao-Sanders 2020).

Vaikka yrityksillä on hallussaan valtavasti heidän liiketoimintaansa liittyvää dataa, ja siitä on tulossa yhä vaikutusvaltaisempaa, kohtaavat ne silti yhden kriittisen haasteen: suurin osa meistä ei ole kovin hyviä tulkitsemaan tai ymmärtämään dataa.

Muutama vuosi sitten The National Center for Education Statistics (NCES 2012a.) arvioi aikuisten datan tulkinta- ja ongelmanratkaisutaitoja 23 maassa ja havaitsi, että suurella osalla työikäisistä on puutteita datalukutaidoissaan. (Bersin & Zao-Sanders 2020).

Miksi sitten suurella osalla meistä on niin huono datalukutaito? Koulut ovat Bersin ja Zao-Sandersin mukaan usein huonosti valmistautuneita ratkaisemaan datalukutaitoon liittyviä haasteita. Matematiikan opetussuunnitelma priorisoi kriittisiä ydinkonsepteja (calculus ja algebra) muiden liike-elämään paremmin soveltuvien aiheiden kuten tilastotieteiden ja todennäköisyyslaskennan kustannuksella. Pragmatismien puolesta puhuneet kampanjat on suurelta osin jätetty huomiotta. Tänä päivänä vastuu on siirtynyt työnantajille ja erilaisten digitaalisten koulutusalojen suosio on kasvussa. (Bersin & Zao-Sanders 2020).

Viimeisten 5-10 vuoden aikana tarvittavat datataidot ovat muuttuneet. Aiemmin yritykset halusivat teknistä osaamista niin datan käsittelyyn, tietokantoihin kuin ohjelmointiin liittyen. Kuten IBM:n tutkimus (The enterprise guide to closing skills gap 2019) osoittaa, monet aiemmin halutuista taidoista ovat nyt sisäänrakennettuina data-alustoissa ja yritykset etsivät toisenlaisia taitoja. Nykyään liike-elämässä ei enää riitä, että ymmärtää, kuinka järjestelmät toimivat, vaan tulee myös ymmärtää, kuinka luodaan vakaa, turvallinen ja luotettava hallintotapa datalle. Tekoälyn kehittyessä yhä tärkeämmäksi osaksi liiketoimintaa tämän päivän ammattilaisten tulee pystyä arviomaan algoritmien tuotoksia kriittisesti, eikä vain olettaa järjestelmän olevan aina oikeassa. (Bersin & Zao-Sanders 2020).

Bersin ja Zao-Sanders toteuttivat kyselyn valitulle 20 yrityksen kohderyhmälle. Näiltä yrityksiltä kysyttiin, minkälaisia dataan liittyviä taitoja heidän organisaatiostaan puuttuu. Kävi ilmi, että puutteet eivät liittyneet niinkään teknisiin taitoihin, vaan datalähtöiseen ongelmanratkaisuun. Erityisesti mainittiin ihmisten olevan taidoiltaan vajavaisia:

- Kysymään oikeita kysymyksiä.
- Ymmärtämään, mikä on merkityksellistä dataa ja kuinka se todetaan toimivaksi.
- Tulkitsemaan dataa niin hyvin, että tulokset olisivat käyttökelpoisia ja merkityksellisiä.
- AB-testaamaan hypoteesit nähdäkseen miltä tulokset näyttävät.
- Luomaan helposti luettavia visualisointeja, joista johtajat ymmärtävät lopputulokset.
- Kertomaan tarinaa ja analyysia, jonka avulla päätöksentekijät näkevät kokonaisuuden ja jonka perusteella pystytään toimimaan.

Tarvitaan tietenkin myös muita edistyneempiä taitoja, mutta kyselyn mukaan juuri näiden ”pehmeiden” taitojen puuttuminen tekee taitovajeesta vieläkin huomattavamman. Lisäksi on tärkeää ymmärtää data oikeassa kontekstissa. (Bersin & Zao-Sanders 2020).

Dataan liittyvät taidot ovat nyt välttämättömiä melkein jokaisessa roolissa kaikenlaisissa organisaatioissa. Mutta koska nämä taidot ovat niin puutteellisia niin monissa yrityksissä, on vaikea tietää, mistä valmiuksien kehittäminen tulisi aloittaa. Bersin ja Zao-Sanders kannustavat ihmisiä keskittymään perusasioihin. Ensiksi on varmistettava, että henkilöstö osaa käyttää organisaation käytössä olevia työkaluja. Olivatpa ne sitten Excel tai jokin analytiikkaohjelmisto. Toiseksi, on järjestettävä koulutuksia ja perustettava sisäinen työryhmä viemään datakyvykkyyksiä eteenpäin organisaatioissa. Tämän lisäksi voidaan tarinoiden ja esimerkkien avulla opettaa ja inspiroida muita siitä, kuinka dataa voidaan käyttää hyväksi. Esimerkiksi Netflix käyttää valtavaa datamäärää hyväksi sekä alustansa personointiin että luomaan houkuttelevaa sisältöä. (Bersin & Zao-Sanders 2020).

Datalukutaidosta on tullut tärkeää melkein kaikille. Organisaatiot tarvitsevat enemmän ihmisiä, jotka kykenevät tulkitsemaan dataa, tekemään oivalluksia ja esittämään oikeita kysymyksiä oikeassa paikassa. Nämä ovat taitoja, joita kuka tahansa voi kehittää. (Bersin & Zao-Sanders 2020). Dataan perustuvan päätöksenteon omaksuneet yritykset ovat tutkitusti parantaneet liiketoimintansa suorituskykyä ja niissä nähdään huomattavaa tuottavuuden kasvua. (Brynjolfsson ym. 2011).

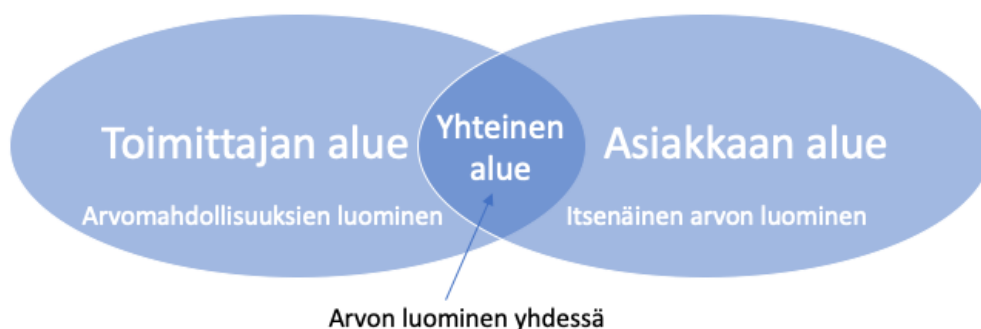
Tässä opinnäytetyössä pyritään keskittymään liike-elämässä tarvittaviin datalukutaidon perustarpeisiin. Työ tehdään Cubiq Analytics Oy:n toimeksiannosta. Cubiq Analytics Oy on vuonna 2016 perustettu suomalainen analytiikka-alan konsultointiyhtiö, joka palvelee sekä yksityisen että julkisen sektorin asiakkaita kaikissa dataan ja analytiikkaan liittyvissä asioissa. Valtaosa yhtiön liikevaihdosta tulee analytiikkakonsultoinnista, jonka lopputuloksena on yleensä yrityksen ylimmän johdon tai keskijohdon käyttöön tuleva kyseisen organisaation omasta datasta toteutettu interaktiivinen visualisointi eli liiketoimintasovellus. Myöhemmin yrityksestä käytetään nimeä Cubiq. Työn kirjoittaja on toiminut samassa yrityksessä analytiikkakonsulttina tätä työtä kirjoitettaessa keväällä ja kesällä 2020. Kirjoittajalla on usean vuoden kokemus liiketoiminnan analytiikasta, datan yhdistelystä sekä visualisointien toteuttamisesta.

### **1.3 Työn tavoite, tarkoitus ja rajaus**

Tämän opinnäytetyön tutkimustehtävä on luoda datalukutaitoa arvioiva kysely Cubiqin käyttöön. Tavoitteena on luoda uusi palvelu, joka voidaan toteuttaa verkkokyselynä asiakkaan toiveiden mukaisesti teknologiasta riippumatta. Datalukutaito ja sen mittaaminen valikoitui opinnäytetyöni aiheeksi sen uutuusarvon, liiketoimintapotentiaalin ja

hyödynnettävyyden ansiosta. Tämä on myös ollut osa-alue, jota ei ole vielä voitu arvioida kvantitatiivisesti. Usein liike-elämässä kuultavan viisauden mukaan: mitä mittaat, sitä voit kehittää. Datalukutaidon todettiin jo aiemmin olevan kriittistä organisaation tulevaisuuden kannalta eli saat parhaat hyödyt datasta ja analytiikkasovelluksista, kun organisaatio on datalukutaitoinen. Myöhemmin nähdään datalukutaitoon vaikuttavat eri osa-alueet, mutta tässä työssä käsitellään suurelta osin vain työntekijän henkilökohtaisiin kyvykkyyksiin liittyviä ominaisuuksia. Aiheen laajuuden vuoksi useaa datalukutaidolle tärkeää aluetta, kuten tilastotiedettä tai peruslaskutoimituksia ei käsitellä tässä työssä. Myöskään itse datan käsittelyyn, yhdistelyyn tai muuhun analyysiin valmistavaan työhön ei puututa tässä yhteydessä.

Toimittajana Cubiq pyrkii jatkuvaan palveluidensa kehittämiseen sekä uusien arvomahdollisuuksien luomiseen. Grönroosin ja Voiman 2011 esittelemän näkemyksen mukaan arvoa luodaan asiakkaan ja toimittajan omien alueiden lisäksi vuorovaikutuksessa yhteisellä alueella. Kehitystyötä onkin tehty yhteistyössä Cubiqin asiakkaiden kanssa. Ajatuksenamme on, että datalukutaitoa voitaisiin mitata ja sitten tarkastella sen mahdollista kehittymistä erilaisten toimenpiteiden, kuten koulutusten, jälkeen. Palvelulla tarkoitetaan asiakkaiden toimintojen ja prosessien tukemista. (Grönroos 2006, 313-337). Asiakkaat eivät osta vain palvelua, vaan sen tuottamaa hyötyä. (Levitt 1980, 83-91).



Kuvio 1. Arvonluonnin alueet. (Grönroos & Voima 2011, 31).

Kun on selvitetty organisaation datalukutaidon nykytaso, sitä voidaan kehittää ja sen kehitystä seurata teettämällä uusia arviointeja tai selvittämällä vain tiettyjä asiakkaan tarpeita datalukutaidon osa-alueilta, esimerkiksi uuden hankintasovelluksen ymmärtämistä. Datalukutaidon arvioinnin avulla sekä toimittaja (Cubiq) että asiakas (kohdeorganisaatio) saavat arvokasta tietoa analytiikkapalvelujen tilasta ja mahdollisten koulutusten tarpeesta. Kysely voi myös sellaisenaan toimia palautteena yksittäisen analytiikkaprojektin onnistumisesta. Esimerkiksi, jos uusi analytiikan kehitystyö sisältää visualisointeja, joita ei aikaisemmin ole käytetty, saadaan arvioinnin kautta palautetta niiden toimivuudesta ja ymmärrettävyydestä. Konsepti on siis hyvin monikäyttöinen ja sen tuottamat vastaukset palautteineen ovat arvokkaita sekä asiakkaalle että toimittajalle.

Opinnäytetyö on jaettu kuuteen päälukuun. Johdannossa esitellään aihe, sekä kerrotaan työn tavoitteista ja tarkoituksista sekä rajauksista. Lisäksi johdannossa esitellään käytetyt tutkimusmenetelmät sekä tutkimuksellinen lähestymistapa. Luku kaksi esittelee käsitteitä ja teoriaa datalukutaidon taustalla datan ja tietämyksenhallinnan osalta. Kolmas luku käsittelee visualisointeja ja näiden käyttötarkoituksia. Neljännessä luvussa kuvataan ihmismielen toimintaa ja sitä, miten tietyt asiat vaikuttavat päätöksentekoomme. Viidennessä luvussa esitellään ensimmäinen toteutettu verkkokysely henkilökohtaisten ominaisuuksien osalta, käydään läpi kyselyn kehitysvaiheita sekä analysoidaan tuloksia. Kuudes luku on pohdintaa ja yhteenvetoa työn tuloksista sekä jatkotoimenpiteistä.

#### **1.4 Tutkimuksellinen lähestymistapa ja tutkimusmenetelmät**

Kehitystyömme ajatuksena on ollut tuottaa datalukutaidon kyselystä ohjelmistokehityksestä tuttu MVP (minimum viable product) ja sen jälkeen koekäyttää kyselyä sekä sisäisesti että ulkoisesti aina mahdollisuuksien mukaan, esimerkiksi Cubiqin järjestämissä korkeakouluopiskelijoiden analytiikkatyöpajoissa tai ekskursioilla. Tässä prosessissa on tarkoitus kehittää tuotetta palautteiden perusteella, joten niiden kerääminen on kriittistä. (Lenarduzzi & Taibi 2016). Kehitystyön lopputuloksena on Cubiqin asiakasorganisaatioille tarjottava palvelu: kvantitatiivinen eli määrällinen verkkokysely, jolla voidaan arvioida organisaation datalukutaidon nykytilaa, ja jonka tulokset ovat analysoitavissa sekä esitettävissä digitaalisessa muodossa heti toteutuksen jälkeen. Jokaisen teetetyn arvioinnin jälkeen vastaajilta on kerätty palautetta ja jokainen palaute on analysoitu. Niiden perusteella on parannettu muun muassa taulukoiden ja muiden visualisointien selkeyttä sekä kysymysten asettelua.

Cubiqilla työskentelee noin 50 data-analytiikan ammattilaista, joiden päivittäiset työtehtävät sisältävät asiakasorganisaatioiden haasteiden ratkomista datan avulla. Hyvin monen projektin lopullisena tavoitteena on tuottaa sovelluksen kautta näkymä, jonkin liiketoimintaan liittyvän haasteen ratkaisemiseksi. Monesti haasteet liittyvät puutteelliseen tai kokonaan puuttuvaan näkymään jostain tietystä liiketoiminnan osa-alueesta, esimerkiksi varaston arvoa ei nähdä päivätasolla tuotekohtaisesti.

Kirjoittaja on toiminut vuosia asiakasrajapinnassa, jossa hänelle on kertynyt käytännön tietoa asiakasorganisaation kohtaamista haasteista sekä organisaatioiden käyttämistä sovelluksista ja visualisoinneista. Näitä kaikkia käytännöstä kertyneitä tietoja on käytetty hyväksi, määriteltäessä osaltamme datalukutaidon perustasoa ja tärkeänä pidettyjä aihealueita. Cubiqin asiantuntijat kohtaavat päivittäin asiakasorganisaation edustajia ja toimivat

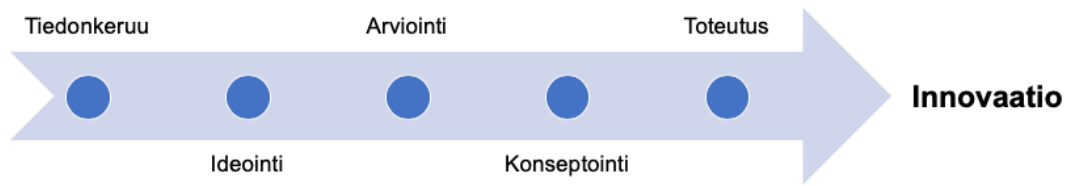
vuorovaikutuksessa heidän kanssaan. Näin on saatu jatkuvasti uutta tietoa mahdollisista visualisointeihin tai datan ymmärtämiseen liittyvistä haasteista.

Tärkeimpänä menetelmänä kehitystyössä ovat olleet perinteisten aivoriilien lisäksi erilainen ajatustentörmäyttäminen Slack- pikaviestimessä. Tätä projektia varten on perustettu oma datalukutaito-niminen kanava, jossa kaikilla halukkailla on mahdollisuus osallistua keskusteluun ja vaikuttaa itse arviointikyselyn syntyyn sekä kysymysten asetteluun. Aivoriilihissä ajatuksia on kerätty yhteen ja niistä on päätetty jatkokäsittelyyn valikoituneet tehtävät. Graafiset tehtävät on testattu värien ja selkeyden osalta erityisen tarkasti, esimerkiksi kuinka hyvin ne toimivat punavihervärisokeuden kanssa.

Itse kehitystyön lähestymistapaa voidaan pitää Ojasalo, Moilanen ja Ritalahti mukaan konstruktivisena, sillä kehitystehtävänä on luoda konkreettinen tuotos. Kyseissä tavassa pyritään ratkaisemaan ongelmia käytännönläheisesti uuden rakenteen luomisen kautta. Tämän uuden rakenteen luomiseksi tarvitaan sekä olemassa olevaa teoreettista tietoa että uutta empiiristä eli käytännöstä kerättävää tietoa. (Ojasalo ym. 2014, 65). Ojasalo ym. (2014, 83) tekevät kuitenkin eron konstruktivisen kehittämisen ja innovatiivisen kehittämisen välille. Heidän mukaansa kyseessä on innovaatioiden tuottaminen, mikäli lopputulos on täysin uusi. Kuitenkaan esimerkiksi uusi koulutusmateriaali ei ole innovaatio, vaan kehitystyön tuloksena syntynyt rakenne. (Ojasalo ym. 2014, 65-86).

Innovaatiolla tarkoitetaan uutta tuotetta, palvelua, prosessia, toimintamallia tai vastaavaa, jolla tuotetaan taloudellista tai muuta hyötyä. Uusi idea, keksintö tai tutkimustulos ei siis sellaisenaan vielä ole innovaatio, vaan kehitystyön tulokset pitää pystyä kaupallistamaan tai ottamaan käyttöön muulla tavoin. (Ojasalo ym. 2014, 86.)

Aikaisempien rakenteiden puuttuessa kehitystyön seurauksena syntyvää palvelua voidaan siis pitää innovaationa. Ojasalo ym. mukaan innovaatioprosessin vaiheet jaotellaan tiedonkeruuksi, ideoinniksi, arvioinniksi, konseptoinniksi ja toteutukseksi (Kuvio 2). Vertaan kehitystyötä näihin innovaatioprosessin vaiheisiin: tiedonkeruu, ideointi ja arviointi. Ideointia tapahtui työpajoissa, Slack- kanavalla. Arviointia suoritettiin sekä sisäisessä ajatustentörmäyttämisessä että ideointityöpajoissa aivoriilien muodossa, mutta myös kysymällä mielipiteitä datalukutaitoa arvioivista tehtävistä yksittäisiltä asiantuntijoilta.



Kuvio 2. Innovaatioprosessi. (Ojasalo ym. 2014).

Innovaatioihin liittyvä piirre on myös ainutlaatuisuus. Ei ole olemassa tiettyä kaavaa, miten innovaatioita tulisi tuottaa. Innovaatioihin liitetään lisäksi, että ne ovat usein yrittäjämäisiä yhden tai muutaman avainhenkilön henkilökohtaisen panoksen ja sitoutumisen tuotoksia. Innovaation ei myöskään tarvitse markkinoille tullessaan olla täysin viimeistelty. (Rilla & Saarinen 2007). Innovaatioprosessia mukaillen työtä on tehty ensin tiedonkeruun vaiheessa niin, että on etsitty datalukutaitoa koskevaa tutkimustietoa sekä kirjallisuutta ja näiden tietojen pohjalta on lähdetty ideoimaan, minkälaisilla kysymyksillä näitä datalukutaidon perustaitoja selvitettäisiin. Useista eri kysymysvaihtoehdoista on valittu arvioinnin kautta konseptointiin ja toteutukseen päätyneet tehtävät eri vastausvaihtoehdoineen.

## 2 Data

Dataa on sanottu uudeksi öljyksi (Economist 2017), tai kuten David McCandless kutsui dataa vuonna 2010 pitämässään TED- puheessa: uudeksi maaperäksi. (CJR 2010). Dataa pidetään maailman arvokkaimpana raaka-aineena. Se on joka tapauksessa yksi suurimmista trendeistä niin liiketoiminnassa kuin yleisestikin arjessamme.

Nykyään jokaisen yrityksen sanotaan olevan datayritys. Data muuttaa maailmaa ja sitä, kuinka elämme sekä työskentelemme, täysin ennustamattomalla tavalla. Eri näkökulmista katsoen voimme joko olla mukana jossain jännittävässä, tai sitten siirtymässä kammottavaan isoveljen valvovaan maailmaan, jossa jokainen liikkeesi voidaan jäljittää - tai jopa ennustaa. Liikkeenjohdon kannalta tällaiseen skeptisyyteen datan suhteen ei kuitenkaan ole aikaa. (Marr 2017, 1-2). Yritykset, jotka näkevät datansa strategisena voimavarana, tulevat kasvamaan ja menestymään. Kuitenkin tutkimusyhtiö Forresterin selvityksen mukaan 60 – 73 prosenttia kaikesta yritysdatasta jää analysoimatta. (Forrester 2016). Tähän ilmiöön monet yritykset etsivät nyt parrannusta.

Datasta itsessään on tullut varallisuutta, ja kyky kaupallistaa oma datansa onnistuneesti voi vaikuttaa liiketoiminnan tulokseen merkittävästi. Esimerkiksi Alphabet (Googlen emoyhtiö) tai Facebook ovat niitä yrityksiä, joiden koko liiketoimintamalli tai ainakin suurin osa siitä perustuu datan kaupallistamiseen. Informaation avulla voidaan siis ensin parantaa oman yrityksen toimintaa prosessien, tuotekehityksen ja asiakasymmärryksen kautta. Ja sitten tätä informaatiota voidaan lisäksi jakaa tai myydä valituille kohderyhmille. (Marr 2017, 73).

Informaation ja siitä mahdollisesti saatavan ymmärryksen sekä datan rahallisen arvon merkityksen lisääntyessä korostuu datalukutaidon merkitys. Organisaatioiden ja ihmisten kykyä käsitellä dataa on vasta viime aikoina selvitetty. Teknologiayhtiö Accenture selvitti omaan datalukutaitoprojektiinsa liittyen työntekijöiden datavalmiuksia ympäri maailmaa. Kyseisen tutkimuksen toteutti markkinatutkimusyhtiö Opinium. Kysely suoritettiin 9000 kokoaikaiselle työntekijälle yli 50 hengen yrityksissä 10.-24. syyskuuta 2019. Kyselyyn osallistui työntekijöitä yhdeksästä eri maasta: Iso-Britannia, Yhdysvallat, Saksa, Ranska, Singapore, Ruotsi, Japani, Australia ja Intia.

Kyselyn vastaukset kokoavan raportin mukaan yritykset pyrkivät poistamaan aukon datan tuottamien mahdollisuuksien ja nykyisten rakenteidensa väliltä. Yritysjohtajista vain 32 % vastasi pystyvänsä luomaan datasta mitattavaa arvoa, ja 27 % heistä vastasi data-analytiikkaprojektiensa tuottavan toimintakelpoisia oivalluksia. Kyselyssä selvisi myös, että monilla työntekijöillä ei ole taitoja, jotka auttaisivat heitä työskentelemään ja tulkitsemaan

dataa luottavaisesti, sillä vain 21 % vastaajista oli täysin luottavaisia omiin kykyihinsä työskenneltäessä datan kanssa. (Accenture 2020).

Harvard Business Review:n artikkelissaan Bowne-Anderson sanoo datalukutaidon riippuvan siitä, kuinka ymmärtää erilaisia datatyyppejä ja datan tallennustapoja. Kyky ymmärtää ja kommunikoida datan avulla on yhä tärkeämpää kolmesta syystä: ensinnäkin, datatiede ja tekoäly vaikuttavat globaalisti moniin toimialoihin. Toiseksi, monet uutiset kerrotaan datan ja ennustavien mallien kautta. Ja kolmanneksi, paljon henkilökohtaista dataamme käytetään määrittelymään, miten olemme vuorovaikutuksessa maailman kanssa. Kun päätöksentekoon käytetään niin paljon dataa eri toimialoilla, vaaditaan arkisista asioista keskustelemiseen ainakin perusymmärrys siitä, kuinka data vaikuttaa ekosysteemissä. Toimiala todennäköisemmin kohtaa data-analytiikan vaikutukset kuin säästy niiltä. Datalta ei voi välttyä. (Bowne-Anderson 2018).

Asiaa havainnollistavaksi esimerkiksi voidaan nostaa keskustelu todennäköisyyksistä Yhdysvaltojen presidentinvaaleista vuodelta 2016. FiveThirtyEight-journalistipalvelu antoi Clintonin voitolle 71,4 % ja Trumpin voitolle vain 28,6 % todennäköisyyden. Varmasti monelta järkytykseltä olisi välttytty esittämällä Trumpin voiton todennäköisyys hiukan todennäköisempänä, kuin kahden perättäisen kruunan saaminen kolikonheitossa. (Bowne-Anderson 2018).

Bowne-Anderson luettelee dataan liittyviä käsitteitä, joita myös ei-teknisten ihmisten on hyvä ymmärtää: ensiksi, datan tuottaminen, kerääminen ja varastointi. Ja toiseksi, miltä data näyttää ja tuntuu datatieteilijän tai analyytikon näkökulmasta. Aina internetiä käytettäessä kaikki toiminta havaitaan ja monet niistä myös tallennetaan. Tavallisen internet-selaimen keräämää tietoa havainnollistetaan projektissa, jonka tarkoituksena avata verkossa tapahtuvan tiedonkeruun laajuutta. Kaiken tekemisen lisäksi myös passiivisuus havaitaan. Testin voi tehdä osoitteessa: <https://clickclickclick.click>. (Bowne-Anderson 2018).

Tietojen kerääminen ei kuitenkaan rajoitu pelkästään tietokoneisiin tai mobiililaitteisiin, vaan paljon laajempaan kokonaisuuteen nimeltään esineiden internet (IoT). Tämä tarkoittaa, että aiemmin ilman älyä toimineet laitteet, kuten valot, jääkaappi tai televisio voidaan tehdä niin sanotusti älykkäämmiksi yhdistämällä ne internettiin. Aivan kuten mitä tahansa muu dataa keräävä laite: älykello, älykaiutin ja itse ajava auto. (Bowne-Anderson 2018).

Kuten johdannosta kävi ilmi, teknologinen kehitys ja pilvipalvelut ovat kaiken tämän datan keräämisen takana. Kaikki tämä data tallennetaan niin kutsuttuun pilveen. Itse pilveä ei kuitenkaan ole olemassa sanan varsinaisessa tarkoituksessa, vaan data tallennetaan kiintolevylle joka, sijaitsee jossain fyysisessä tilassa. Ainoa ero normaalin tietokoneen tai

älypuhelimien tallennukseen on se, että pilvipalvelua käytetään internetin kautta. Tiedot ovat todennäköisesti jonkun kolmen suuren palveluntarjoajan pilvessä. Nämä kolme suurinta ovat Amazon, Microsoft ja Google. Data siis sijaitsee fyysisessä paikassa, ja siihen sovelletaan sen maan lakeja, jossa tiedot sijaitsevat. Dataa suojaavat erilaiset säädökset kuten EU:n General Data Protection (GDPR), joka vaikuttaa käyttäjän tietosuojaan ja vaatii hänen suostumuksensa henkilökohtaisen datan käyttöön. (Bowne-Anderson 2018).

Analyttikot ja datatieteilijät kohtaavat datansa yleensä yhdessä kolmesta eri muodosta: taulukkomainen, kuvadata tai ei-strukturoitu (strukturoimaton), kuten luonnollisena tekstinä tai verkkosivujen html-koodina, jotka muodostavat suurimman osan maailman datasta. Kuviossa 3. on esimerkki taulukkomaisesta datasta (tabular data). Data, joka on taulukkomuodossa (tietokannassa), on datan yleisin olomuoto. (Bowne-Anderson 2018).

Nimi	Ikä	Sukupuoli	Tulot (\$)
John Smith	19	M	36000
Joe Bloggs	23	M	42000
Sally Jones	21	N	50000
Linda Lee	22	N	30000
Harry Jameson	18	M	28000

Nämä ovat mitattavan tai laskettavan tietoyksikön ominaisuuksia, kuten pituus, syntymämaa tai tulot. Ominaisuuksiin viitataan myös muuttujana, koska ominaisuus voi vaihdella tietoyksiköiden välillä ja voi vaihdella ajan myötä.

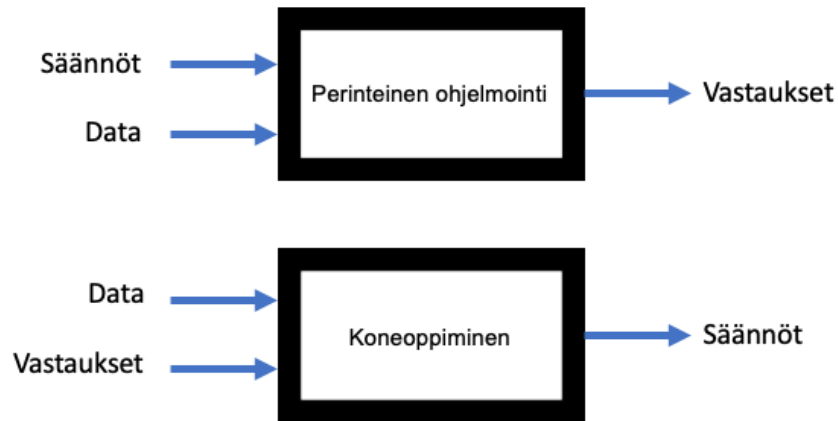
Tietoyksikkö, josta tietoa on kerätty. Tässä tapauksessa henkilö.

Numeerinen muuttuja.

Kategorinen (ei-numeerinen) muuttuja.

Kuvio 3. Dataa taulukossa. (ABS 2020).

Kuvadata on nimensä mukaisesti kuvia. Monet syväoppimisen onnistumiset ovat tapahtuneet juuri kuvien luokittelussa. Esimerkiksi kyky diagnosoida sairauksia kuvien perusteella, kuten ihosyöpää. (Bowne-Anderson 2018). Tässä yhteydessä on hyvä käydä läpi sitä, miten koneoppiminen eroaa normaalista ohjelmoinnista tai analytiikasta, jossa tietyillä ehdoilla datasta etsitään vastauksia. Kuviosta 4. nähdään, että koneoppimisessa ihminen syöttää datan lisäksi odotettavat vastaukset ja saa tuloksena alkuperäisiin vastauksiin sovellettavissa olevia sääntöjä tai malleja.



Kuvio 4. Koneoppiminen: ohjelmoinnin uusi malli. (Chollet 2017, 2)

Koneoppimisella tarkoitetaan yleisesti tietokoneen kykyä oppia kerätyn datan pohjalta asiayhteyksiä. (Müller & Guido 2016). Syväoppiminen taas on yksi koneoppimisen muoto, jossa yksinkertaisista prosessiyksiköistä koostuvia kerroksia yhdistetään verkoksi, jotka pystyvät oppimaan monimutkaisemman datan pohjalta ja toimimaan ihmisen hermoston tapaan. (Chollet 2017, 2)

Ja lopuksi on vielä strukturoimaton data, joka on nimensä mukaisesti dataa, jota ei ole mitenkään järjestettyä tai jäsentynyttä. Osa datatieteilijän tai analyttikon työtä on järjestää data sellaiseen muotoon, että sitä voidaan analysoida ja hyödyntää. (Bowne-Anderson 2018).

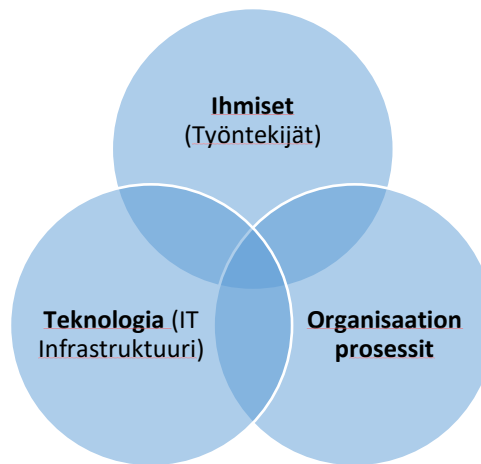
## 2.1 Tietämyksenhallinta

Liiketoiminta muuttuu teknologisen kehityksen myötä nopeasti. Kilpailu ei ole enää pitkään aikaan ole ollut suoraviivaista tai ennustettavaa. Selviytyminen ja menestys riippuvat täysin organisaation kyvystä sopeutua liiketoimintaympäristöön ja sen muutoksiin. Kuinka tietoa sitten tulisi soveltaa saavuttaakseen lisäarvoa ja kilpailuetua? Kuinka informaatio muutetaan tietämykseksi? Tietämyksenhallinnan (Knowledge Management) avulla saadaan organisaation tekniset kyvykkyudet tuottamaan ymmärrystä, joka vie organisaatiota eteenpäin. (Awad & Ghaziri 2003, 26-27).

Tietämyksenhallinta (Knowledge Management) on monitieteinen ala, jonka tavoitteena on organisaation mukauttaminen jatkuviin muutoksiin hallitsemalla systemaattisesti sen toiminnassa tarvittavaa tietoa. Se on saanut vaikutteita monelta eri tieteenalalta, mukaan lukien liiketoiminta, taloustiede, psykologia ja tiedonhallinta (Information Management). Sen perimmäinen tarkoitus on luoda yritykselle kilpailuetua. Tietoa luodaan vuorovaikutuksessa, mutta myös yksittäiset henkilöt oppivat uutta. Muiden ihmisten sekä

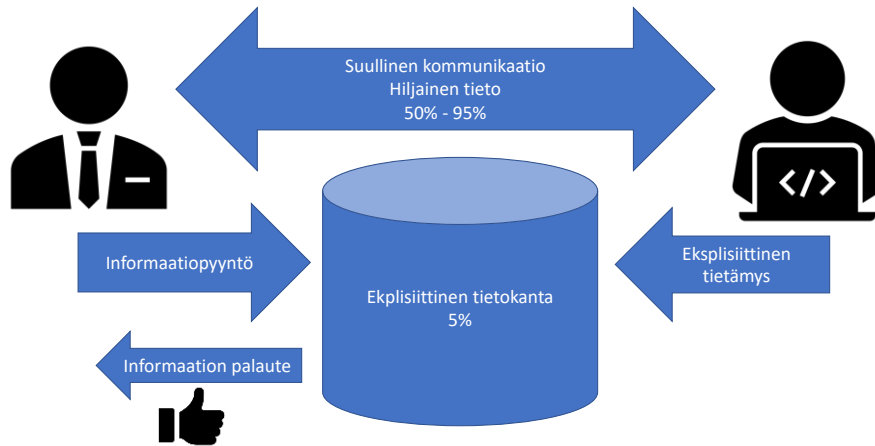
tietokannoissa olevan tiedon yhdistämisen seurauksena syntyy oivalluksia. (Awad & Ghaziri 2003 26-27).

Tietämyksenhallinta sisältää kolme eri osa-aluetta: ihmiset, teknologian ja prosessit. (Awad & Ghaziri 2003, 26-27). Tämä on syytä ottaa huomioon selvittäessä organisaation datalukutaitoa, sillä kokonaisuus on tekijöiden summa, eli arvioinnin kohteeksi on syytä ottaa kaikki kolme osa-aluetta. Tietämyksen syntyyn tarvitaan siis kaikkia alueita ja niiden tulee osaltaan tukea tiedon vaihtoa.



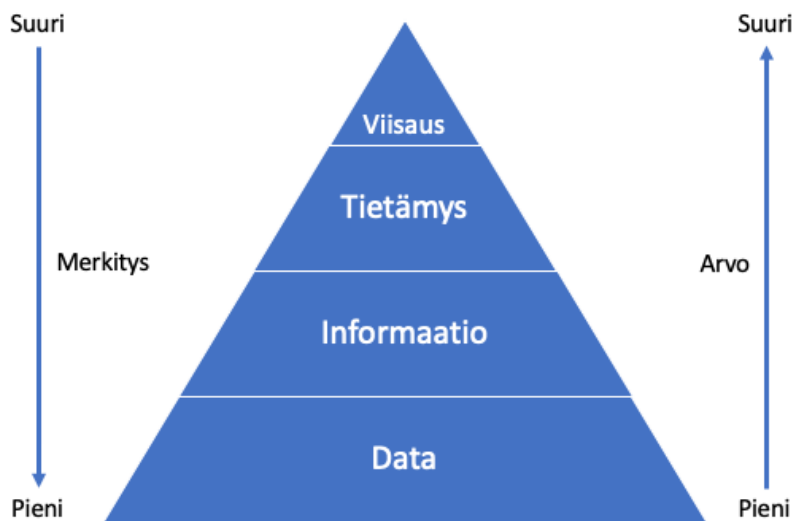
Kuvio 5. Tietämyksenhallinnan osatekijät. (Awad & Ghaziri 2003)

Tietämyksenhallinnassa kaikki organisaation prosessit nähdään tietoa tuottavina. Ja itse tieto hallittavana sekä johdettavissa olevana asiana. Tietämyksenhallinta on vahvasti sidoksissa tiedonhallintaan, kuten myöhemmin nähdään. Ihanteellisena organisaationa kuvaillaan sellainen, jossa ihmiset vaihtavat tietoa poikkifunktionaalisesti käyttämällä hyväksi teknologiaa ja vakiintuneita prosesseja. Tietämyksenhallinnan fokus on oikeiden asioiden tekemisessä, eikä niinkään asioiden tekemisessä oikein. Niin, että tämän päivän ydinkompetensseista ei tule huomisen rasitteita. (Malhotra 2000).



Kuvio 6. Tallentamaton hiljainen tieto. (Bair 2001).

Organisaation tietoja katsotaan koostuvan sekä hiljaisesta että ekspliittisestä tietokantoihin tallennetusta tiedosta. Yllä oleva kuvio 6. kertoo, että suuri osa tiedosta on hiljaista tietoa, eikä sitä ole tallennettu organisaation tietokantoihin. Varsinkin innovaatiot syntyvät tästä työntekijöiden välisestä kommunikaatiosta. (Awad & Ghaziri 2003, 27).

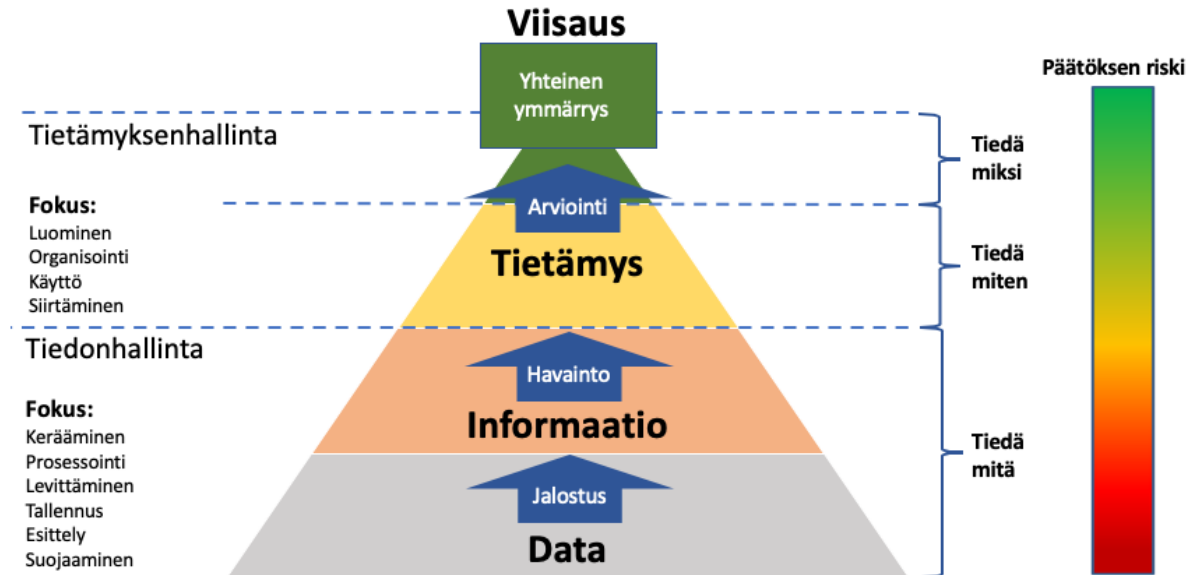


Kuvio 7. DIKW- pyramidi. (Rowley 2007, 163)

Alkuperäinen DIKW-pyramidi, jossa datasta tulee sen merkityksen kasvun myötä informaatiota, josta jälleen arvon ja merkityksen lisääntymisen kautta tietämystä. Lopuksi tietämys muuttuu viisaudeksi. (Rowley 2007,163).

Perinteisestä DIKW-pyramidista on tehty useita muunnoksia. Tämä tiedonhallinnan Yhdysvaltojen armeijan päälliköiden käyttämä malli kuviossa 8. esittelee tarkemmalla tasolla, miten päätöksenteon riskiä vähennetään ja hallitaan yhteisen ymmärryksen

kautta. Tämä kuvio tekee myös selkeän eron tiedonhallinnan ja tietämyksenhallinnan välillä. Tiedonhallinnan fokuksen ollessa tiedon keräämisessä, prosessoinnissa, levittämisessä, tallennuksessa, esittelyssä ja suojaamisessa. Kun taas tietämyksen hallinta keskittyy tiedon luomiseen, organisointiin, käyttöön ja siirtämiseen. Tässä mallissa ylin taso eli viisaus saavutetaan arvioinnin ja yhteisen ymmärryksen kautta.



Kuvio 8. Tiedonhallinnan kognitiivinen pyramidi. (mukaillen ARMY 2015).

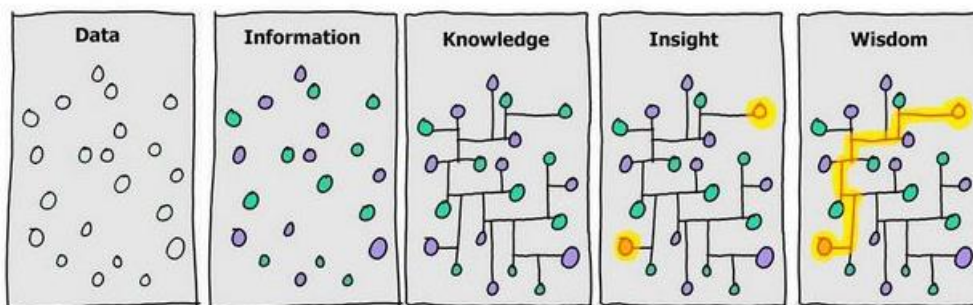
Tässä yhteydessä olevat ihmismielen tasot voidaan systeemiteoreetikko Russel Ackroffin mukaan jakaa viiteen eri kategoriaan:

1. Data: merkkejä, tieto on raakaa. Se on olemassa, mutta sillä ei itsessään ole merkitystä. Se voi olla missä tahansa muodossa, käyttökelpoisessa tai ei.
2. Informaatio: dataa, joka on prosessoitu hyödylliseksi ja tarjoaa vastauksia kysymyksiin: kuka, mitä, missä ja milloin. Informaatio on siis dataa, jolle on annettu merkitys. Merkitys voi olla hyödyllinen, mutta sen ei tarvitse olla.
3. Tietämys: datan ja informaation soveltaminen vastaamaan kysymykseen: miten. Tietojen tarkoituksenmukaista käyttöä tietyn tietämyksen saavuttamiseksi.
4. Ymmärtäminen: kun tiedät miksi. Se on kognitiivinen ja analyttinen prosessi. Ero ymmärryksen ja tietämyksen välillä on kuin ero oppimisen ja muistamisen välillä. Ymmärrystä omaavat ihmiset voivat ryhtyä hyödyllisiin toimiin, koska he voivat syntetisoida uutta tietoa tai joissain tapauksissa ainakin luoda uutta tietoa aiemmin tiedetystä sekä ymmärrästä.
5. Viisaus: arvioitua ymmärrystä, toisin kuin neljältä edelliseltä tasolta, siinä kysytään kysymyksiä, joihin ei ole helposti saavutettavissa olevaa vastausta. Se vetoaa kaikkiin aikaisempiin tasoihin ja ottaa keskusteluun mukaan moraaliset ja eettiset

säännöt. Viisaus on myös prosessi, jolla erotamme tai tuomitsemme oikean ja väärän, hyvän sekä pahan.

Ackoffin mukaan neljä ensimmäistä kategoriasta liittyvät menneisyyteen, sillä ne käsittelevät sitä, mikä on ollut tai mennyttä. Viisaus sen sijaan käsittelee tulevaisuutta, koska siihen sisältyy visio ja suunnittelu. Viisauden avulla ihmiset voivat luoda tulevaisuuden sen sijaan, että pitäisivät kiinni nykyisestä ja menneestä. Viisauden saavuttaminen ei kuitenkaan ole helppoa, vaan ihmisten on ennen sen saavuttamistaan liikuttava onnistuneesti muiden kategorioiden läpi. (Ackoff 1989, 3-9).

Kuviossa 9. on vielä DIKW-pyramidia mukailleen esitetty, miten raakadatasta muodostetaan oivalluksen kautta viisautta. Kirjoittajan kokemuksen mukaan tämä kuvaus sopii myös perinteisen, suoraviivaisen liiketoimintasovelluksen toteutuksen jälkeisen analyysin kuvaukseksi. Oivallukset syntyvät (tai jäävät syntymättä) interaktiivisen käyttöliittymän kautta, tekemällä valintoja ja hakemalla vastauksia esitettyihin kysymyksiin.



Kuva 9. Erilainen kuvaus DIKW-pyramidista lisättyä oivalluksen tasolla. (Figuroa 2019).

DIKW-pyramidia on kritisoitu siitä, että tietämys ei synny vain informaatiota suodattamalla, vaan se on paljon esitettyä monimutkaisempi prosessi. On totta, että faktatietoa saadaan tietokannoista, ja algoritmien avulla voidaan löytää korrelaatioita. Tietämys on kuitenkin paljon laajempi ja monimutkaisempi prosessi, joka on sosiaalinen, tavoitteellinen ja kulttuuriin sekä asiayhteyteen sidottu. Tämän kuvauksen todellinen ongelma on siinä, että se on pyramidi. Mielikuva, jonka mukaan tieto on vain seurausta erilaisten suodattimien käytöstä kullakin tasolla, vääristää todellisuutta. Tämä näkemys on kuitenkin luontainen informaation aikakaudelle, joka on keskittynyt epäolennaisen kohinan suodattamiseen ja siitä hiljaa juoksuttamalla puhtaaseen, selkeään ja hallittavissa olevaan tietoon. Todellisuudessa tieto on luovaa, sotkuista, vaikea saavuttaa ja paljon epäjatkevampaa kuin edellä on kuvattu. (Weinberger 2019).

## 2.2 Tiedonkulun rajoitteita

Tiedon kulkua organisaatiossa voi haitata myös muulla tavalla. Nykyään on paljon keskustelua siilojen purkamisesta tiedon ympäriltä. Tätä kutsutaan datan demokratisoinniksi. Tämän periaatteen mukaan organisaation dataan tulisi olla kaikilla vapaa pääsy, eikä sen tulisi olla vain tiettyjen asiantuntijoiden tai johtajien saatavilla, kuten aiemmin on saatettu menetellä. Yritykset ovat ymmärtäneet, että sallimalla yhä useammalle pääsyn dataan ja analyyseihin, työtapoihin on saatu monimuotoisuutta, ja sitä kautta on päästy parempiin tuloksiin. Tiedon jakamisessa on tietenkin otettava huomioon tietoturva ja yksityisyys, mutta kuten muissakin toiminnoissa, kaikenlaisten siilojen purkaminen tehostaa toimintaa ja auttaa näkemään asioita uudella tavalla. (technopedia 2020).

Päätöksentekoon ja datalukutaitoon organisaatiossa vaikuttavat myös kognitiiviset vääristymät, joista mainitsen yhden yleisimmistä, mutta menestyksen kannalta myös yhden vaarallisimmista. Auktoriteetin päätöksentekoon mahdollisesti aiheuttama vääristymä todistettiin oikeaksi kuuluisassa Milgarmin kokeessa 1960-luvun alussa. Tässä kokeessa koehenkilöt olivat valmiita antamaan toisille koehenkilöille hengenvaarallisia sähköshokkeja vain, koska kokeen järjestäjät eli auktoriteetit niin pyysivät. Koehenkilöt eivät millään tavalla kyseenalaistaneet auktoriteetteja, vaikka uskoivat aiheuttavansa toisille koehenkilöille suunnatonta kipua. Sähköshokkeja vastaanottaneet henkilöt olivat kokeessa mukana olleita näyttelijöitä, eikä heihin todellisuudessa kohdistunut minkäänlaista vaaraa. (Milgram 1963, 371-378). Liike-elämässä auktoriteetin aiheuttamaa viinoutta tai vääristymää on kutsuttu nimellä hippos (highest paid person's opinion). Eniten tienavaan henkilön ollessa auktoriteetti eli johtaja. On kuitenkin osoitettu, että tällainen ajattelu on liiketoiminnalle erittäin epäsuotuisaa. Tutkimuksen mukaan alemman tason projektipäällikköjen projektit onnistuvat todennäköisemmin kuin korkeampiarvoisten vanhempien ja kokeneempien projektipäällikköiden. Tämä johtuu siitä, että auktoriteetin puuttuessa tai ollessa heikompi, projektiryhmän jäsenet ovat halukkaampia jakamaan mielipiteitä, haastamaan oletuksia sekä antamaan rehellistä palautetta. (Szatmari 2016).

## 2.3 Datastrategia

Organisaation tietohallinnon tärkein omaisuus on data. Monet tietohallintojohtajat kuitenkin joutuvat vastaamaan heiltä kysyttäessä, että tätä varallisuuserää ei käytetä suunnitelmallisesti hyväksi. Heidän on tunnustettava, että dataa ei ole inventoitu, eikä ole täysin selvää, mitä dataa missäkin tietokannassa on. Heillä ei myöskään ole aavistustakaan sen laadusta. Ei siis ole mitään suunnitelmaa, miten tämä omaisuus saataisiin tuottamaan. (Adelman ym. 2005).

Vain harvoilla organisaatiolla on hyvin määritelty selkeä datastrategia. Useimmissa organisaatioissa datan arvoa ei ymmärretä. Dataa pidetään vain sitä luovan osaston omaisuutena. Ja usein tämä osasto suojaa näitä tietoja liiankin tehokkaasti, eivätkä kaikki pääse tarvitsemaansa tietoon käsiksi. Vielä nykyäänkin joillakin organisaatioilla on omaa varjoanalytiikkaa. Tämä data saattaa olla Exceleissä tai jopa omissa tietokannoissa, joista organisaation virallinen IT-osasto ei tiedä. Tällainen toimintatapa on syntynyt, sillä on haluttu toimia nopeasti ja ketterästi ohittaen organisaation viralliset käytännöt. Nämä sovellukset ja tietovarastot ovat joko itse tehtyjä, tai sitten ne on tilattu ulkopuoliselta konsultilta. Sovellukset eivät noudata organisaation standardeja, eikä niitä ole suunniteltu integroitavaksi organisaation virallisiin järjestelmiin. Usein tällaiset ratkaisut ovat puutteellisesti dokumentoituja tai sitten dokumentaatio puuttuu kokonaan. (Adelman ym. 2005).

Työskentely ilman datastrategiaa on verrattavissa organisaatioon, jossa jokainen henkilö saa luoda oman kirjanpidon tilihierarkiansa. Mikäli jokainen saa mahdollisuuden valita käytettävät tilinsä muista riippumatta, on hyvin suuri todennäköisyys, että tapahtuu virheellisiä kirjauksia. Strategian puuttuminen antaa organisaatiossa tilaa ajaa omia agendoja liiketoiminnan kustannuksella. Datastrategia vähentää järjestelmäkehityksen riskiä ja johtaa paljon todennäköisemmin muihinkin onnistumisiin. Se myös johtaa laadukkaampiin järjestelmiin ja tarjoaa organisaation johdolle perusteen torjua kaikki dataan liittyvät asiat, mitkä eivät ole nykyisen strategian mukaisia. (Adelman ym. 2005).

Yhdeksi datastrategian osaksi mainitaan liiketoiminta-arvo ja investoidun pääoman tuottoaste. Organisaation datalla on yhtä suuri ellei suurempi arvo kuin muilla varallisuuserillä: rakennukset, varasto, myyntisaatavat ja tarvikkeet. Datastrategian kehittäminen ja toteuttaminen vievät kuitenkin rahaa ja muita arvokkaita resursseja. Sitä tulisikin mitata kuten mitä tahansa investointia ja mahdollisesti sitoa yrityksen johdon bonukset projektin onnistumiseen. (Adelman ym. 2005).

Data vaikuttaa liiketoimintaan kolmella eri tavalla: päätöksentekoon, operaatioihin ja kaupallisena hyödykkeenä. Ilman tarkkaa suunnitelmaa datasta voi tulla kallista ja hyödyttöntä. Järjestelmistä maksetaan joka tapauksessa kuukausittaiset ylläpitokustannukset. Datastrategian tulee olla koko organisaation laajuinen, eikä vain esimerkiksi markkinoinnin oma. Ja lisäksi on unohdettava perinteinen käsitys, että data ja analytiikka kuuluisivat vain IT-osastolle. (Marr 2016, 17-33).

Dataa on saatavilla nykyään niin paljon, että ilman fokusta on vain ympäriinsä harhailua. Päätöksenteossa korostuu juuri oikea data, eikä vain sen määrä. Aivan kuten

liiketoimintaakaan ei pitäisi toteuttaa ilman selkeää strategiaa. Suuryritykset kuten Google, Amazon ja Facebook ovat mestareita datan hyödyntämisessä. Nämä yritykset eivät vain kerää kaikkea mahdollista dataa, vaan ovat myös löytäneet uusia ja innovatiivisia tapoja käyttää sitä suunnitelmallisesti hyväkseen. Organisaatioiden on ensin määriteltävä liiketoiminnalle kriittiset kysymykset, joihin datasta odotetaan saatavan vastauksia, ja sen jälkeen kerätä sekä analysoida tätä dataa. (Marr 2016, 17-33).

Organisaation on ensin päätettävä strategiset tarpeet. Kaikkea mahdollista dataa ei ole kannattavaa kerätä tai analysoida, kuten jo aiempaan mainittiin niin nykyisestäkin datasta jää huomattava määrä käsittelemättä. Nämä tarpeet voivat olla esimerkiksi prosessien parantamista tai päätöksentekoon liittyviä. Esimerkiksi asiakkaiden tarpeiden parempaan ymmärtämiseen. Esimerkki prosessien parantamisesta on Amazonin suosittelujärjestelmä, joka osaa valtavan ostosdatamäärän perusteella suositella tuotteita, joita kuluttaja ei ole vielä itsekään tajunnut tarvitsevänsä. Amazon on kasvanut kirjojen verkkokaupasta nykyiseen kokoonsa pitkälti datan ansiosta. Amazon on yksi maailman parhaista yrityksistä käyttämään dataa hyväksi. Datan tulisi siis olla ennen kaikkea liiketoiminnan mahdollistaja. (Marr 2016, 17-33).

Marr toteaa, että IT-vetoiset datastrategiat keskittyvät liian usein datan tallentamiseen, omistajuuteen ja tietojen yhtenäisyyteen, eivätkä liiketoiminnan pitkäaikaisten tavoitteiden saavuttamiseen datan avulla. Suuremmat organisaatiot saattavat hyötyä oman CDO:n (Chief Data Officer / Chief Digital Officer) palkkaamisesta. Hän on se, joka vastaa organisaation kaikesta datasta ja sen tehokkaasta käytöstä. Tämän datajohtajan on tunnettava hyvin sekä liiketoimintaa että ymmärrettävä teknologiaa. Hänen vastuullaan on myös dataohjautuvan kulttuurin edistäminen organisaatiossa, mukaan lukien datalukutaidosta ja sen kehityksestä huolehtiminen. (Marr 2016, 17-18).

## **2.4 Korrelaatio ja kausaliteetti**

Kausaliteetti tarkoittaa periaatetta, että kaikelle tapahtuvalle on olemassa jokin syy. Jos tällaista yhdistävää tekijää ei ole, näiden kahden tapahtuman välillä ei ole syy-seuraussuhdetta. Jo aiemmin mainittu korrelaatio taas tarkoittaa olemassa olevaa suhdetta ilmiön, asian tai matemaattis-tilastollisen muuttujan välillä, joilla on taipumus vaihdella, liittyä toisiinsa tai tapahtua yhdessä tavalla, jota ei odoteta pelkästään sattuman perusteella. (Merriam-Webster 2020b & Merriam-Webster 2020c).

Ihmiset ovat hyvin nopeita tekemään johtopäätöksiä asioiden A ja B yhteyksistä. Korrelaatio on helpompi havainta kuin kausaliteetti. Kesäinen myyntipiikki jäätelön osalta ei aiheuta lämpöhalvauksia, eikä lisääntyneet lämpöhalvaukset ole syy kasvaneelle

jäätelön myynnille, vaan syy molempien taustalla on lämmin sää. Aina syyt eivät ole näin selkeitä (Reinhart, 2015 s. 84-85).

Tämä luku on hyvä päättää taloustieteilijä Ronald Coasen toteamukseen, että data kyllä tunnustaa kaiken mahdollisen, kun sitä tarpeeksi kauan kidutetaan.

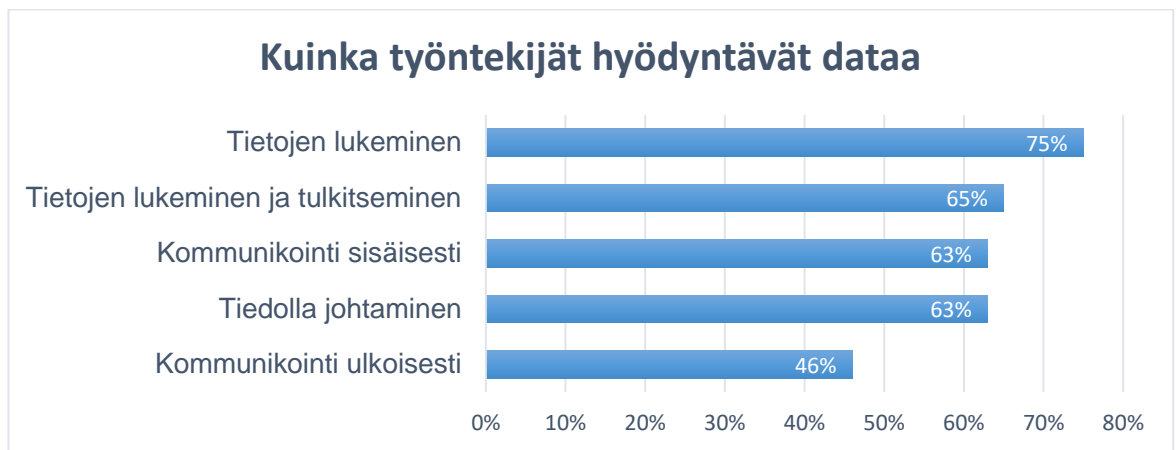
### 3 Visualisoinnit

Data voi olla melkein mitä tahansa. Se voi olla nimiä, numeroita, tilastollisia arvoja, päivämääriä, kommentteja tai paikkatietoja koordinaatteina. Voimme suhteellisen vaivattomasti etsiä yksittäisiä arvoja esimerkiksi Excel-taulukoista tai tietokannoista. Emme kuitenkaan pysty vertailemaan suhteellisia eroja tai näkemään yhteyksiä useiden eri datapisteiden välillä. Meidän mieleemme ei ole hyvä muuntamaan tekstiä ja numeerisia arvoja kvalitatiiviseksi ja kvantitatiiviseksi merkitykseksi. Ymmärtääksemme dataa paremmin se tulee muuttaa visuaaliseen muotoon. Visualisointien tarkoituksena on siis esitellä ja kuvantaa dataa ymmärryksen helpottamiseksi. (Kirk 2016, 19)

Näköemme on ylivertainen muihin aisteihimme verrattuna. Tutkijat ovat laskeneet silmän pystyvän siirtämään informaatiota noin 10 miljoonaa bittiä sekunnissa, kun taas korvat pystyvät noin sadan tuhannen bitin siirtonopeuteen sekunnissa. (Britannica 2020a). Graafinen esitystapa on siis kaikkein tehokkain tietoa tapa siirrettäessä dataa tajuntaamme. Koska graafinen kommunikaatio on visuaalista, se tulee ilmaista tavalla, jota silmämme voivat vastaanottaa ja aivomme ymmärtää. Arviolta noin 70 % kehomme aistireseptoreista on omistautunut näkökyvylle ja 30 % muille aisteille. (Few 2012, 61).

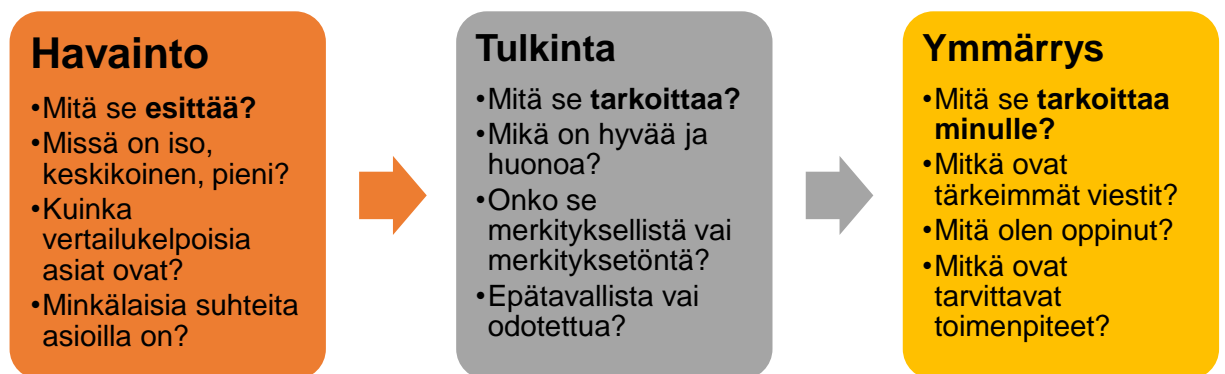
Datavisualisoijan luodessa visualisointia hän kuvaa dataa edustavalla tavalla. Kuinka data siis visuaalisesti esitetään merkkien ja ominaisuuksien kautta. Merkit ovat pisteitä, viivoja tai alueita. Ominaisuudet taas ovat näiden merkkien ilmentymiä, kuten koko, värit tai positio. Yhdessä nämä merkit ja niiden ominaisuudet, mukaan lukien muut komponentit kuten X-, ja Y- akselit tai ruudukoviivat sekä selitteet, muodostavat kaavion anatomian. (Kirk 2020, 20-21). Kvantitatiivisia taulukoita ja grafiikoita voidaan käyttää organisaation tietojen analysoimiseen, kommunikaation, suunnitteluun tai seurantaan. (Few 2012, 10).

Accenture selvitti kyselyssään (Accenture 2020. kuvio 10.), kuinka työntekijät hyödyntävät dataa työssään. Suurin osa kyselyyn vastaajista ilmoitti tulkitsevansa dataa lukemalla tietoja sekä tulkitsemalla niitä erilaisten graafisten esitysten kautta esimerkiksi analytiikkasovelluksessa. Pääpaino tämän työn kehityksessä on ollut työntekijöiden datan käytössä, siis tietojen lukemisessa ja tulkitsemisessä. Tästä syystä keskityttiin vahvasti tähän osa-alueeseen, kun valikoitiin tehtäviä perusdatalukutaitoa arvioivaan kyselyyn.



Kuvio 10. Miten työntekijät hyödyntävät dataa. (Accenture 2020).

Visualisointia katsottaessa katsoja käy läpi ymmärryksen prosessin kolme vaihetta: havainnon, tulkinnan ja ymmärtämisen (Kuvio 11). Jokainen vaihe on riippuvainen edellisestä vaiheesta ja visualisoijana voit vaikuttaa näihin, mutta täyttä kontrollia sinulla ei ole. Olet enimmäkseen katsojan armoilla. Et tiedä mitä hän tietää ja mitä hän ei tiedä. Et myöskään aina tiedä mistä hän on kiinnostunut ja mikä on hänelle merkityksellistä. Näin ollen on monta muuttujaa, mihin itse visualisoinnin tekijä ei voi vaikuttaa. Vastuu ymmärryksen saavuttamisesta onkin näin ollen sekä visualisoijan että katsojan yhteistä. (Kirk 2016, 22)



Kuvio 11. Ymmärtämisen kolme vaihetta. (Kirk 2016, 22)

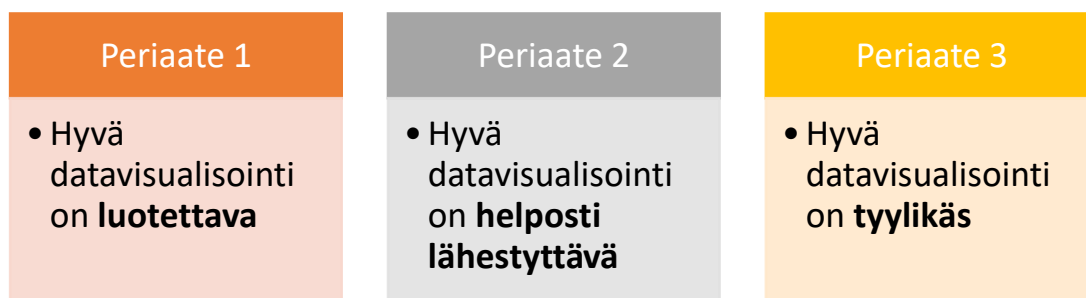
Ensimmäisessä vaiheessa on kyse havainnosta: osataanko esitystä lukea, mitä kaavio esittää, missä eri suuruiset arvot sijaitsevat sekä minkälaisia riippuvuuksia asioilla on. Toisessa eli tulkintavaiheessa katsoja luo ymmärrystä näkemästään asiasta. Kolmannessa, viimeisessä vaiheessa katsoja käsittelee asiaa, sekä pohtii sen vaikuttavuutta itselleen. (Kirk 2016, 22).

Hyvää ja kestävä suunnittelua pohdittaessa voidaan hyväksi koettuja tapoja etsiä myös perinteisten rajojen ulkopuolelta. Saksalainen teollinen muotoilija Dieter Rams on tullut tunnetuksi palkituista minimalistisista sekä ajattomista tuotteistaan. Hän on luonut oman

listansa kestävän kehityksen puolesta, kamppaillessaan huonoa käytettävyyttä ja sekavaa värien käyttöä vastaan 1970- ja 80- lukujen taitteessa. Ramsin listan jokaisen kohdan tarkoituksena on vastata kysymykseen: ”onko minun suunnitteluni hyvää suunnittelua?”. Soveltaen tätä listaa voidaan käyttää hyväksi kaikessa suunnittelussa, myös visualisoinneissa (Vitsø 2020):

1. Hyvä design on innovatiivista.
2. Hyvä design tekee tuotteesta hyödyllisen.
3. Hyvä design on esteettistä.
4. Hyvä design tekee tuotteesta ymmärrettävän.
5. Hyvä design on huomaamatonta / vaatimatonta
6. Hyvä design on rehellistä.
7. Hyvä design kestää aikaa.
8. Hyvä design perusteellista viimeistä yksityiskohtaa myöten.
9. Hyvä design on ympäristöystävällistä.
10. Hyvä design on niin vähän designia kuin mahdollista.

Andy Kirk on kirjassaan Data Visualisation luonut osittain Ramsin listan perusteella oman visuaalisen suunnittelun periaatteiden listansa. Jokainen visualisoija voi näiden kolmen periaatteiden avulla arvioida omien graafisten tuotosten käytettävyyttä:



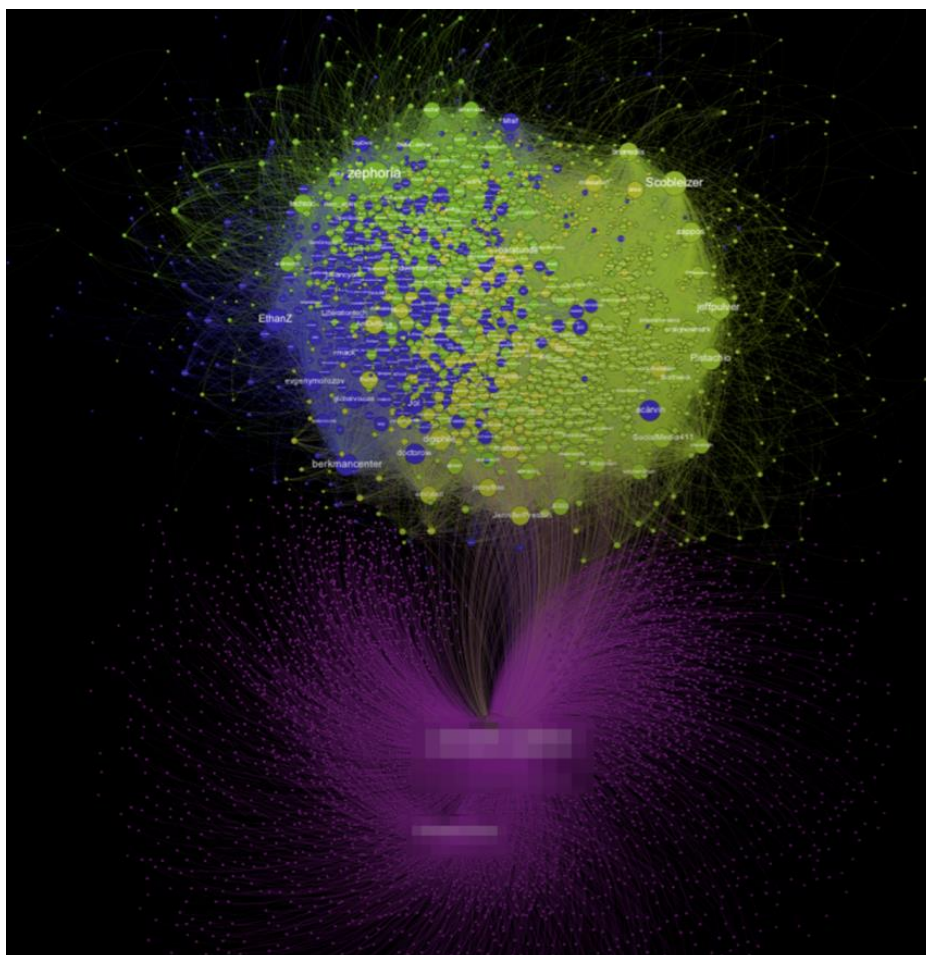
Kuvio 12. Hyvän visualisoinnin kolme periaatetta. (Kirk, 2016. s.30)

Jokainen päätös, visualisointiin tai taulukkoon tehty lisäys tai muutos voidaan arvioida vastaamalla kysymyksiin: tekeekö se visualisoinnista luotettavan, helpommin lähestyttävän tai tyylikkäämmän.

### 3.1 Visualisointi ja Infograafi

Tietokirjailija ja suunnittelija Alberto Cairo määrittelee hyvän visualisoinnin olevan tiedoiltaan luotettava ja visualisoitu niin, että merkityksellinen informaatio on havaittavissa. Sen tulee myös olla järjestetty niin, että se tarjoaa tutkittavaa, kun niin on

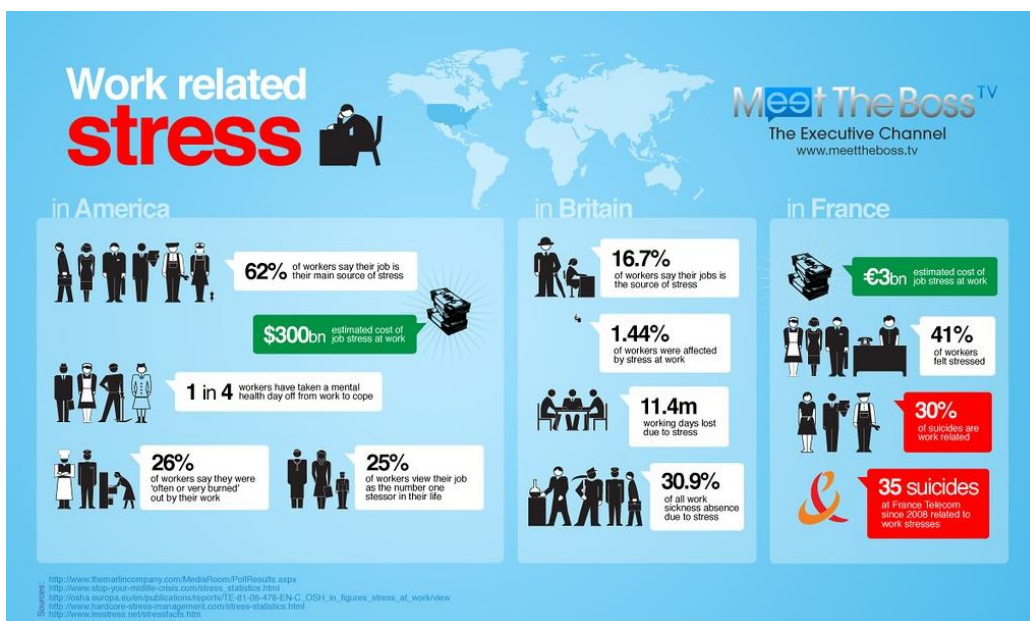
tarkoituksenmukaista. Ja sen tulee olla esitetty houkuttelevalla tavalla, muistaen aina kuitenkin, että rehellisyys, selkeys sekä syvyys ovat etusijalla. Visualisointien ja infograafien tarkoitus on ihmisten valistaminen – ei viihdyttäminen, ei minkäänlainen myyminen, vaan ainoastaan informaation siirtäminen. (Cairo 2016, 13). Visualisoinnit eroavat infogaafeista niiden käyttötavan lisäksi niiden toteutustavasta johtuen. Kuviossa 13. on kuvattu hyvin monen datapisteen visualisointi.



Kuvio 13. Gilad Lotanin Twitter- seuraajien suhdeverkko. (Lotan 2014).

Hieman eksoottiselta kasvilta näyttävä visualisointi kuviossa 13. on datatieteilijä Gilad Lotanin kuvaus hänen Twitter-kokeilustaan. Kokeensa aluksi hän muodosti Twitter-datan avulla oikeiden seuraajien verkko-kaavion. Kaaviosta on selkeästi nähtävissä yhteisöllisyyttä eli moni seuraa toisiaan Twitterissä. Mitä suurempi piste, sitä enemmän yhteyksiä (kuvassa ylempänä). Tämän jälkeen hän osti lisää niin kutsuttuja valseuraajia. Nämä eivät ole oikeita ihmisiä, vaan jonkun tähän tarkoitukseen luomia käyttäjätilejä. Tämän jälkeen hän visualisoi seuraajadatansa uudelleen. Valetilien näkee olevan selvästi yksittäisiä pisteitä ilman keskinäisiä suhteita toisiinsa. (Lotan 2014).

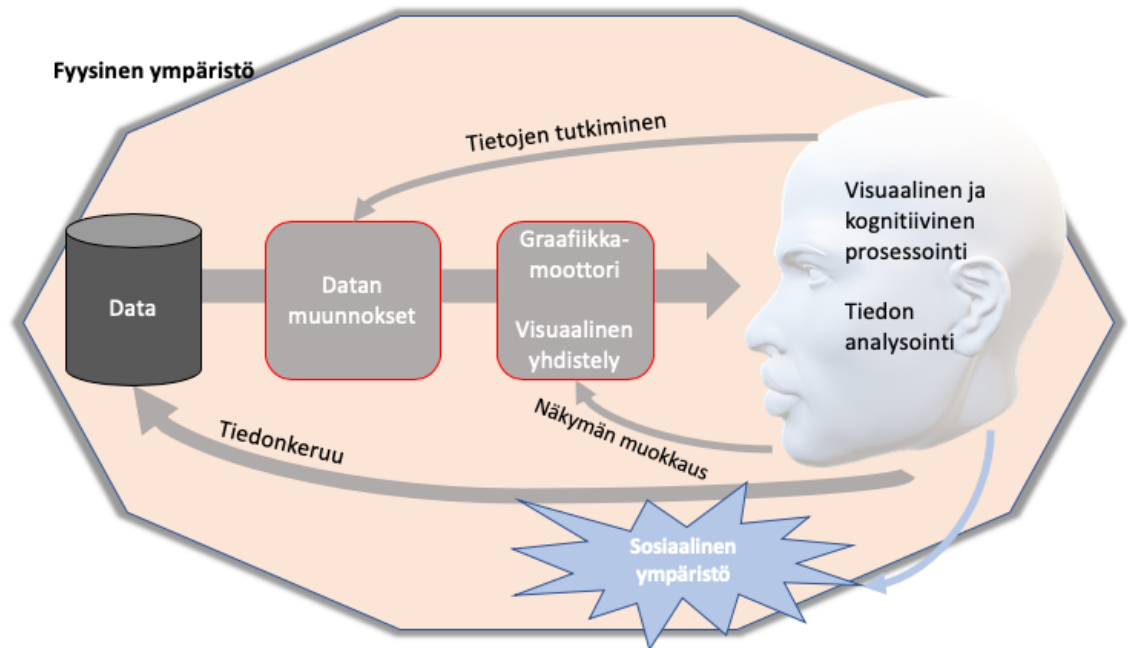
Visualisointi voi siis näyttää tuhansien yksittäisien eri pisteiden yhteyden tai niiden puuttumisen, kun taas infograafi voi kertoa hyvinkin yksityiskohtaisen tarinan. Visualisointi edustaa dataa ja mahdollistaa analyysin, tutkimisen ja löytöjen tekemisen. Infograafit ovat sekoitus erilaisia piirroksia, kuvioita ja tekstiä. Infograafin tekijä ei esitä kaikkia sen tekemiseen keräämäänsä tietoa, vaan vain sen, mikä on olennaista esitettävän asian kannalta. Infograafeja käytetään esimerkiksi uutisissa, verkkosivuilla ja painetussa mediassa. (Cairo 2016, 31).



Kuvio 14. Infograafi työhön liittyvästä stressistä. (meettheboss.tv 2020).

### 3.2 Visualisointi prosessina

Tässä opinnäytetyössä keskitytään tietokoneella toteutettaviin visualisointeihin. Datan saaminen graafiseen muotoon on monivaiheinen prosessi. Se on usean eri palautesilmukan tulos. Dataa tulee ensin tutkia tiedonkeruun ohella. Kaiken tämän on siis parannettava ymmärrystä esitettävään dataan.



Kuvio 15. Visualisointiprosessi. (Ware 2020, 4).

Visualisointiprosessi sisältää neljä perusvaihetta yhdistettynä useisiin palautesilmukoihin. Nämä vaiheet on visualisoitu kuviossa 15. Ja ne ovat: 1. Datan kerääminen ja tallentaminen. 2. Käsittelyvaiheessa data muunnetaan helpommin käsiteltäväksi, ellei se jo valmiiksi ole siinä muodossa. Tämä saattaa sisältää myös epäolennaisen tai puutteellisen datan poistamista valittujen näkökohtien paljastamiseksi. 3. Valittu data yhdistetään sopivaan visualisointiin, joka saavutetaan tietokoneen algoritmien avulla (kuva muodostetaan ruudulle). Käyttäjän syöte voi muuntaa datan yhdistelyä, korostaa tietyn tiedon tarpeellisuutta tai muuttaa näkymää. Yleensä tämä tapahtuu käyttäjän omalla tietokoneella. 4. Ihmisen vastaanotto: havainnointi ja kognitiivinen järjestelmä. (Ware 2020, 4-5).

Palautesilmukoista pisin liittyy tiedon (datan) keräämiseen. Datankerääjä: kun tutkija tai osakemarkkinoiden analyysoija saattaa päättää tarvitsevansa enemmän dataa seuratakseen mielenkiintoista löydöstä. Toinen palautesilmukka valvoo laskennallista tapahtumaa ennen visualisointia. Analyttikosta saattaa tuntua, että data voidaan alistaa tiettyyn muunnokseen, jonka jälkeen se menettää merkityksensä. Tämä vaihe saattaa sisältää valtavan tietomäärän läpikäymistä tärkeän tiedonjyvän löytämiseksi. Lopuksi tapahtuva visualisointivaihe saattaa olla vuorovaikutteinen, esimerkiksi tutkija voi liikkua eri näkökulmien välillä datassa esiintyvien rakenteiden ymmärtämiseksi. Vaihtoehtoisesti datasta voidaan valita vain tietty kiinnostava joukko tai alue tarkasteltavaksi. Ympäristöistä sekä fyysinen että sosiaalinen osallistuvat tiedonkeruusilmukkaan. Fyysinen ympäristö on datan lähde, kun taas sosiaalinen ympäristö määrittelee hienovaraisesti ja monitahoisesti mitä dataa kerätään ja kuinka sitä tulkitaan. (Ware 2020, 4-5).

Kuten aiemmin jo mainittiin, vaikka haluisit, et voi vaikuttaa kaikkeen datan visualisoijana. Yleisö tulkitsee kaavioita ja grafiikkaa oman maailmansa kautta. (Kirk 2016, 39-42).

Vain yleisö vaikuttaa näihin asioihin:

- Onko visualisoinnin aihe heistä kiinnostava.
- Tarvitseeko heidän sitoutua näkemäänsä tai onko siitä heille muuta hyötyä.
- Mitä yleisö tietää tai ei tiedä aiheesta.
- Onko visualisointi tai esitystapa heille vieras.
- Onko heillä aikaa tutkia visualisointia.
- Onko visualisointisi käytettävissä kuluttajan kannalta sopivassa formaatissa.
- Katsojan henkilökohtaiset mieltymykset, asenteet ja tunteet.

Sinä taas voit visualisoijana vaikuttaa näihin asioihin:

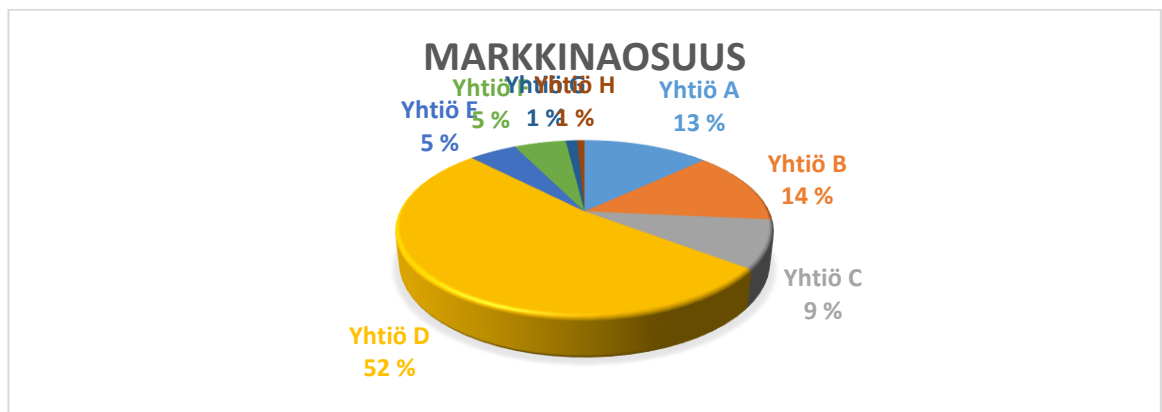
- Fokus olennaisessa sisällössä.
- Visualisointi on tehty yleisöä tarpeita varten.
- Monimutkaista aihetta ei ole yksinkertaistettu liikaa.
- Kokonaiskuvan hahmottamiseen ei tarvitse käyttää liikaa aikaa.
- Toteutus ei edusta dataa, tai erityspiirteitä, kuten värisokeutta ei ole huomioitu.
- Tuotos on luettavissa eri kokoisilta näytöiltä.
- Tarjoa interaktiiviseen visualisointiin vain tarvittavat toiminnot.
- Aihe tai analyysi ei ole monimutkainen, eikä käytössä ole vaikeaselkoisia sanoja tai lyhenteitä.
- Selitteitä on käytetty tarpeen mukaan.
- Visualisointia ei ole tehty monimutkaisella kaaviolla.
- Kaikki tarpeelliset selitteet ja merkinnät ovat mukana (asteikot, yksiköt, kuvaukset yms.)

Aina graafinen esitystapa ei kuitenkaan ole paras lähestymistapa, vaan toteutustapa tulisi valita aina tietotarpeen mukaan. Taulukossa 1. eriteltyinä käyttötapaukset kuvaajille ja taulukoille.

Käytä taulukkoa kun	Käytä kuvaajaa kun
<ul style="list-style-type: none"> <li>Halutaan tarkastella yksittäisiä arvoja.</li> </ul>	<ul style="list-style-type: none"> <li>Halutaan esittää trendiä, poikkeamaa tai muuta havaittavaa muotoa.</li> </ul>
<ul style="list-style-type: none"> <li>Halutaan vertailla yksittäisiä arvoja.</li> </ul>	<ul style="list-style-type: none"> <li>Halutaan tarkastella kuinka yksittäiset arvot liittyvät suurempaan kokonaisuuteen.</li> </ul>
<ul style="list-style-type: none"> <li>Vaaditaan tarkkoja arvoja.</li> </ul>	
<ul style="list-style-type: none"> <li>Määrälliset arvot sisältävät enemmän kuin yhden mittayksikön.</li> </ul>	
<ul style="list-style-type: none"> <li>Halutaan sekä yksityiskohtaiset että kokonaisarvot.</li> </ul>	

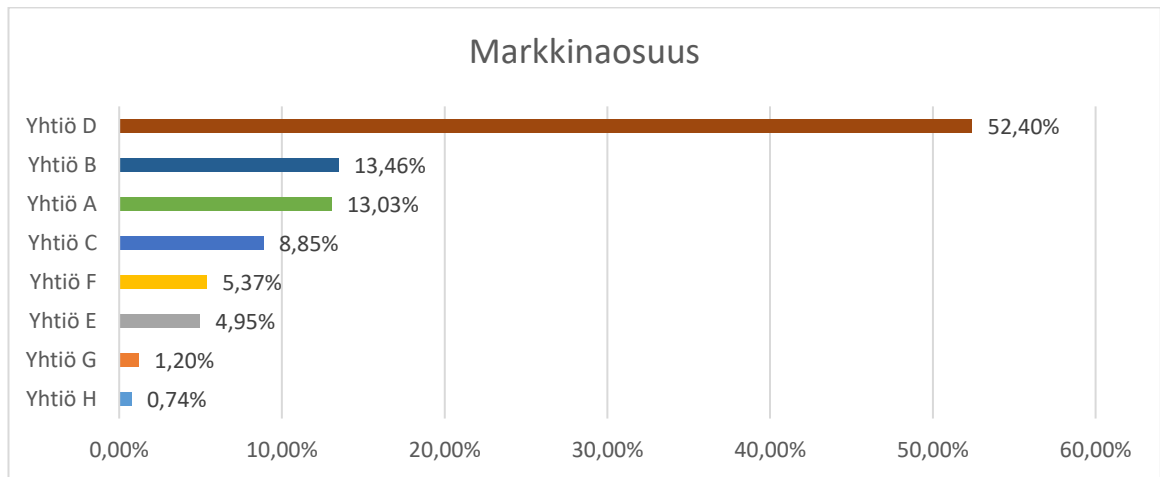
Taulukko 1. Taulukon ja visualisoinnin käyttötapaukset. (Few 2012, 51).

Visualisoinnin tulee olla selkeä ja sen tietojen hyvin luettavissa. Usein kuitenkin näkee houkuttelevalla tavalla toteutettuja graafeja, joiden tiedoista saa vain summittaisen käsityksen. Kuviossa 16. on esimerkki tehottomasta suunnittelusta. Kyseinen kuvio esittää kahdeksan eri yhtiön markkinaosuutta. Kuvioon on vielä lisätty kolmiulotteinen tehokeino graafisen vetovoiman lisäämiseksi. (Few 2012, 7-8).



Kuvio 16. Esimerkki tehottomasta visualisoinnista, piirakkakaavio. (Few 2012, 7)

Pylväsgraafi kuviossa 17. esittää täysin samat tiedot yhtiöiden markkinaosuuksista paljon tehokkaammalla tavalla. Jokaisen yhtiön markkinaosuus on selkeästi nähtävillä ja paremmin vertailtavissa, vaikka numeraalisen osuuden tasapuolisuuden nimissä poistaisikin tästä kuvioista. Tässä kuviossa ei ole häiriötekijöitä, kuten kolmiulotteisuutta tai liian samankaltaisia värejä, vaan se on selkeä ja helppolukuinen. Se antaa numeroiden tehdä tehtävänsä eli kertoa tarinansa, eikä korosta esitystapaa.



Kuvio 17. Esimerkki pylväsgraafina. (Few, 2012. s.8).

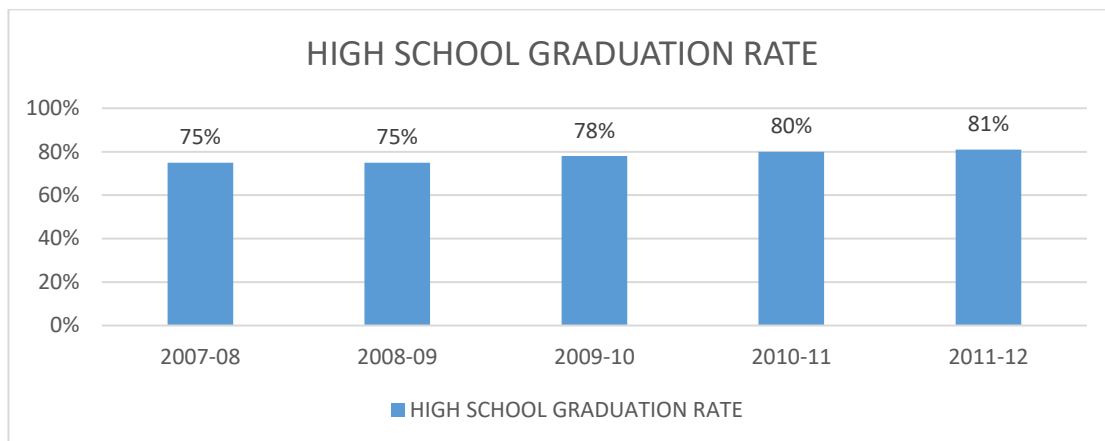
Esitystavan tulisi siis korostaa dataa ja kertoa sen tarinaa mahdollisimman selkeästi. Visualisointi voi myös olla hieno ja muuten tiedoiltaan selkeä, mutta vääristää asiaansa monella eri tavalla.



Kuvio18. Esimerkki asiaansa vääristävästä visualisoinnista. (White House Archived data 22.1.2015 & NCES 2012b).

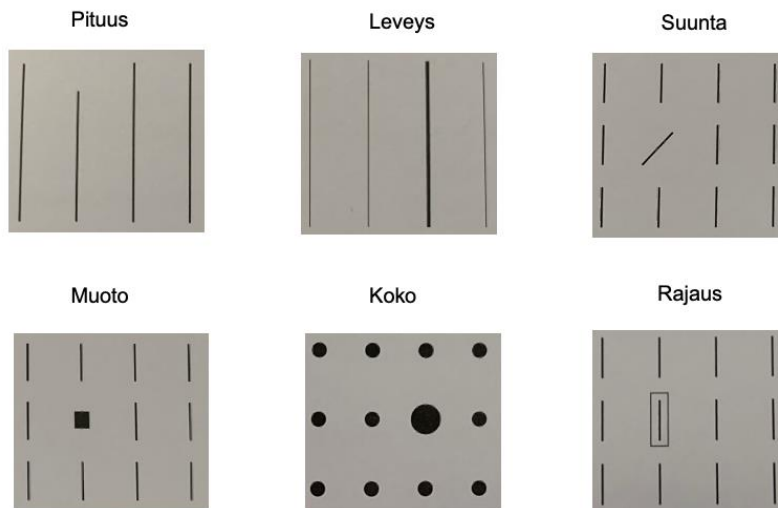
Valkoisen talon saavutuksista kertovan Twitterin tilin mukaan amerikkalaisista lukioista valmistui kyseisellä mittausjaksolla 2011-12 kaikkien aikojen suurin määrä opiskelijoita. Visualisointi on muuten hieno, pylviäitä kuvastavia kirjoja myöten, mutta asteikko on asetettu virheellisesti ja katkaistu. Kuvassa 75 % - valmistumisastetta kuvastaa 5 kirjaa, ja 81 % - valmistumisastetta 15 kirjaa. Myös 1 % kasvu välillä 2010-11 ja 2011-12 on ilmoitettu yhden kirjan lisäyksellä, vaikka oikeassa mittakaavassa yksi kirja vastaisi noin 5,7% kasvua. Tämä kirjapylväiden suhde tarkoittaisi, että valmistuneiden opiskelijoiden määrä olisi kaksi kertaa enemmän kuin mittausjaksolla 2007-08 tai 2008-09. Väärin skaalattuja tai katkaistuja asteikkoja näkee valitettavan usein, varsinkin kun tavoite on

suurennella tai vähätellä esitettyä asiaa. Alla olevassa kuvassa on käytetty samoja arvoja mitä kirjapylväätkin edustavat, jolloin nähdään asia oikeassa mittakaavassa. Tällöin valmistuneiden määrän lisääntyminen ei näytä enää niin merkittävältä, vaikka merkittävästä asiasta onkin kyse. Kuvio 19. huomaa myös kuinka paljon selitteiden ja asteikon puuttuminen vaikuttaa sen luettavuuteen.



Kuvio 19. Pylväsgraafi kuvion 18. datalla, kun asteikkoa ei ole katkaistu.

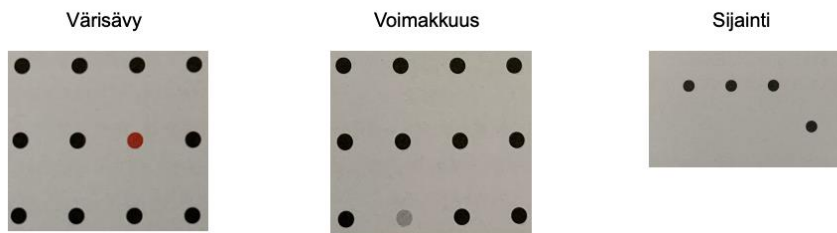
Kuviossa 20. on listattu muodon ominaisuuksia tärkeysjärjestyksessä, niin että ne on helppo havaita. Jos katsot vaikka kuvion 17. Pylväsgraafia, huomaat että pituus on helppo havaita. Kun taas esimerkiksi palkkien leveys taas ei ole niin helposti havaittavissa.



Kuvio 20. Muodon ominaisuudet. (Few 2012, 69).

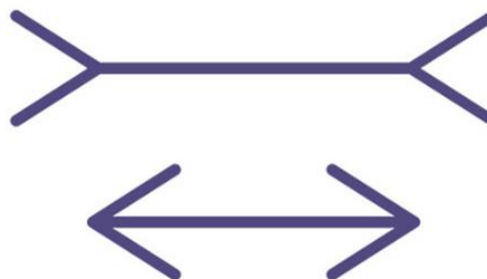
Kuten muodon eri ominaisuuksilla, voidaan väreillä tai niiden voimakkuuksilla vaikuttaa havaintoihimme, ja kiinnittää huomiomme haluttuihin kohtiin kaaviossa tai taulukossa (kuvio 21). Näitä ominaisuuksiavoidaan verrata esimerkiksi aiemmin esitettyyn piirakka-kaavioon, markkinaosuuden ollessa tarkasteltava asia. Osuuden leveyden tai sen koon ollessa ratkaiseva tekijä, emme huomaa eroa enää niin selkeästi, vaikka eroja on korostettu vielä värien kanssa. (Few 2012, 69). Havaintokykymme tilan suhteen on pääasiassa

kaksiulotteinen. Havaitsemme muutokset pystysuunnassa (ylös ja alas) ja horisontaalisesti (vasemmalle ja oikealle) selvästi ja tarkasti. (Few 2012, 71).



Kuvio 21. Värien ja tilan ominaisuudet. (Few 2012, 69).

Kuviossa 22. olevat viivat kuitenkin muistuttavat siitä, kuinka helposti tehokkain aistimme voi johtaa ensivilkaisulla harhaan. Viivat ovat yhtä pitkät, mutta ylempänä oleva viiva näyttää tehosteidensa ansiosta pidemmältä. Kyseessä on maineikas Müllerin-Lyerin harha, tiedät että viivat ovat yhtä pitkät, mutta silti sinun on vaikea uskoa siihen. Mikäli pystyt kyseenalaistamaan viivojen pituuden, kun niiden päässä on nuolet, pystyt jatkossa välttämään tämän harhan vain muistamalla ja tiedostamalla edellisen. (Kahneman 2012, 38). Muodon ominaisuuksilla voidaan siis joko helpottaa tai vaikeuttaa viestiä. Optiset harhakuvat kuten tämä osoittavat selkeästi kontekstin vaikutuksen näköaistin osalta. Vaikka tiedämme katsovamme optista harhaa, väärinkäsitys säilyy, koska näkökyky mittaa eroja ominaisuuksissa, ei absoluuttisissa arvoissa. On visualisoijan vastuulla käyttää tätä tietoa suunniteltaessa taulukoita ja kaavioita estääksemme vääristymiä viestissä, jota yritämme kommunikoida. (Few 2012, 74).



Kuvio 22. Müller-Lyer illuusio. (Few 2012, 74).

Visualisoinnit kattavat hyvin laajan kokonaisuuden kymmenistä erilaisista kaavioista erityyppisiin taulukoihin. Hyväksi datan visualisoijaksi oppii vain tekemällä ja harjoittelemalla, sekä tutkimalla muiden visualisointeja. On tiettyjä suuntaviivoja siihen, mitkä visualisoinnit sopivat mihinkin tarkoitukseen. Tärkeintä on kuitenkin helpottaa katsojan ymmärrystä kyseisestä datasta ja kiinnittää huomio olennaisiin kohtiin. Helppo tapa parantaa organisaation datalukutaitoa on yksinkertaisesti suunnitella käytössä olevat visualisoinnit ja infograafit tehokkaammiksi ja selkeämmiksi tulkita. Oli kyse sitten tiedon käyttämisestä

eli tiedon visualisointien lukemisesta tai niiden tekemisestä, voi jokainen organisaatiossa vaikuttaa niiden kehitykseen antamalla aktiivisesti palautetta. Tehottomien tai muuten sekavien kaavioiden ei pitäisi hidastaa tiedon kulkua tai estää liiketoimintaa edistävien oivallusten syntymistä. Kuten aiemmin jo todettiin, graafisten esitysten tarkoituksena on siirtää tietoa ja helpottaa ymmärryksen saavuttamista. Kaikki muu on toissijaista.

Omia visualisointitaitojaan voi harjoitella ja parantaa esimerkiksi Makeover Mondayn viikottain ilmestyvällä datalla osoitteessa: <https://www.makeovermonday.co.uk/>

## 4 Ihmismielen toiminta: kuinka teemme päätöksiä

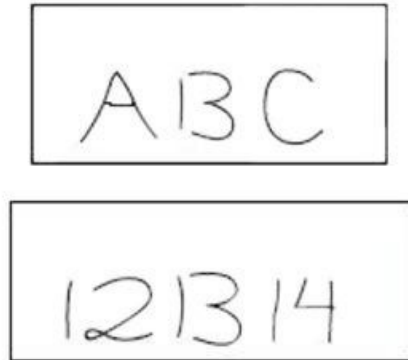
Daniel Kahneman on tutkinut ihmismielen toimintaa 1950-luvun alusta lähtien, jolloin hän aloitti psykologian opinnot. Vuonna 2002 hänet palkittiin talouden Nobel-palkinnolla elämäntyöstään, joka yhdisti psykologisen tutkimuksen taloustieteeseen. Hänen urauurtava tutkimuksensa selvitti arviointia ja päätöksentekoa epävarmuuden vallitessa. Kahneman aloitti palkitun tutkimuksensa 1960-luvun lopulla lisätäkseen ymmärrystä siitä, kuinka ihmiset tekevät taloudellisia päätöksiä. (Britannica 2020a).

Kahnemanin tutkimukset yhdessä Amos Tverskyn päätöksenteosta epävarmuuden vallitessa muodostivat uudet haaran taloustieteeseen: prospektiteoria. Aiemmin taloustieteilijät uskoivat, että ihminen on rationaalinen ja arvioi päätöksiä tulevaisuuden skenaarioiden todennäköisyyksien perusteella. Mutta jos ihmiset tekevät irrationaalisen päätöksen antamalla joillekin skenaarioille enemmän arvoa kuin toisille, heidän päätöksensä eroavat siitä, mitä perinteinen talusteoria ennustaa. Kahnemanin haastatteluihin ja kokeisiin perustuva tutkimus osoitti, että hänen koehenkilönsä eivät kyenneet analysoimaan monimutkaisia päätöksentekotilanteita, kun sen seuraukset olivat epävarmoja. Sen sijaan he tukeutuivat heuristisiin oikoteihin tai peukalosääntöihin ja vain harvat ihmiset arvioivat taustalla olevaa todennäköisyyttä. (Britannica 2020a).

Kahnemanin kirja ”Ajattelu, nopeasti ja hitaasti” kuvailee hänen tutkimuksensa tuloksia prospektiteoriasta ja kognitiivisista vääristymistä. Systemaattisia virheitä sanotaan biaaseiksi eli vinoutumiksi, joista voidaan myös käyttää sanaa vääristymä. (Kahneman 2012, 11). Kirjassa käsitellään paljon kahta ajattelun muotoa. Näitä muotoja voidaan kutsua nimillä järjestelmä 1 ja järjestelmä 2. Käytän jatkossa myös nimitystä ensimmäinen ja toinen järjestelmä.

Kahneman kuvailee, että järjestelmä 1 toimii automaattisesti ja nopeasti, hyvin vähäisin ponnisteluin tai ilman ponnisteluja tai tahdonalaisen säätelyn aistimusta. Automaattisten toimintojen esimerkkejä, jotka liitetään järjestelmään 1: mikä on vastaus:  $1+1 = ?$ , havaitsee esineiden etäisyyden tai täydentää sanonnan: ”hehän ovat kuin kaksi...”. Kun taas järjestelmä 2 on täysin päinvastainen toiminnaltaan. Se huomioi ponnisteluja vaativia mentaalisia toimintoja, esimerkiksi monimutkaisia laskutoimituksia. Se edellyttää tarkkaavaisuutta ja häiriintyy tarkkaavaisuuden kohdistuessa johonkin muuhun. Esimerkkejä tämän järjestelmän toiminnasta on veroilmoituksen täyttämisen, tietyn sanan toistumisen laskeminen kirjan sivulta tai tulon (ellet, satu tietämään vastausta jo ennestään)  $13 * 21$  laskeminen. (Kahneman 2012, 30-31).

Järjestelmä 2 vaatii perusteellista keskittymistä, etkä Kahnemanin mukaan voi käyttää sitä kääntyessäsi risteyksessä vasemmalle, jolloin sinun tulee keskittyä vastaan tulevaan liikenteeseen. Ensimmäisen järjestelmän toimintaa havainnollistaa kuvio 23. Siinä osaamme sijoittaa saman B- kirjaimen tai luvun 13 viitekehyksen mukaan luvuksi tai kirjaimeksi.



Kuvio 23. Kahneman ABC – 121314. (2012, 97).

Keskittyessään voimakkaasti johon tiettyyn tehtävään ihminen voi olla jopa sokea ärsykkeille, joihin hän normaalisti kiinnittäisi huomion. Tämä ilmiö on kuvattu esimerkiksi lyhytelokuvassa, jossa valkoisiin ja mustiin paitoihin pukeutuneet koripalloilijat heittelevät toisilleen palloa. Elokuvan katselijoita pyydetään laskemaan, kuinka monta heittoa valkoisen joukkueen pelaajat tekevät toisilleen ja jättämään kaiken muun huomiotta. Katsojan uppoutuessa vaikeaan tehtävään ja laskiessa syöttöjä, puolivälissä elokuvaa kentän keskelle ilmestyy gorillaksi pukeutunut musta hahmo. Tämän gorillan esiintyminen videolla jää kuitenkin huomaamatta monilta sen katsojilta, sillä he ovat keskittyneitä annettuun tehtävään. Vaikka gorilla esiintyy videolla 9 sekunnin ajan, tämä jää noin puolella katsojista huomaamatta. Tämä The Invisible Gorilla esimerkki on Daniel Simonsin ja Christopher Chabrisin psykologisesta kokeesta ja samannimisestä kirjasta. Voimme siis olla sokeita ilmeiselle ja sokeita myös sokeudellemme. (Kahneman 2012, 34). Videon voi katsoa myös Youtubesta hakusanoilla ”The Invisible Gorilla”.

Järjestelmän 1 toimiessa automaattisesti järjestelmä 2 on yleensä vain vähän ponnisteluja vaativassa miellyttävässä tilassa. Ensimmäinen järjestelmä tuottaa jatkuvasti ehdotuksia toiselle järjestelmälle. Nämä ehdotukset ovat vaikutelmia, intuitioita, intentioita ja tunteita. Toisen järjestelmän hyväksyessä nämä, tulee vaikutelmista ja intuitioista uskomuksia. Impulsseista taas tulee tahdonalaisia toimintoja. Kun asiat etenevät ongelmitta, toinen järjestelmä hyväksyy ensimmäisen ehdotukset niitä juurikaan muuttamatta. Yleensä ihmiset uskovat vaikutelmiinsa ja toimivat halujensa perusteella. Yleensä näin onkin hyvä, mutta ei kuitenkaan aina. Ensimmäisen järjestelmän joutuessa ongelmiin se pyytää tukea toiselta järjestelmältä. Tämä yksityiskohtainen ja erityinen prosessi saattaa ratkaista

senhetkisen ongelman. Toinen järjestelmä käynnistyy, kun ensimmäisellä järjestelmällä ei ole vastausta, kuten todennäköisesti tapahtui kertolaskun  $13 * 21$  kohdalla. Näiden kahden järjestelmän työnjako on todella tehokas: minimoi ponnistus ja optimoi suoritus. Tämä työnjako toimiikin yleensä hyvin, koska yleensä ensimmäinen järjestelmä on suurimman osan ajasta hyvä toiminnassaan. (Kahneman 2012, 34-36).

Sen tuttujen tilanteiden mallit ovat tarkkoja, ennusteet lyhyelle aikavälille normaalisti hyviä ja reaktiot ovat nopeita sekä asianmukaisia. Ensimmäinen järjestelmä toimii näin yleensä ja monet haasteet saadaan ratkaistua. Tällä järjestelmällä on kuitenkin vinoutumia, systemaattisia virheitä. Toisinaan ensimmäinen järjestelmä vastaa asioihin kysytyä helpommin, eikä sen vahvuuksiin kuulu logiikka tai tilastotiede. Sitä ei myöskään saa pois päältä, vaan se on joko apunamme tai hidasteenamme. (Kahneman 2012, 34-36).

Kuten Müller-Lyerin illuusiosta kuvioista 22. huomasimme, on olemassa optisia harhoja. Kaikki harhat eivät kuitenkaan ole visuaalisia, vaan on olemassa myös ajattelun harhoja, joita sanotaan kognitiivisiksi harhoiksi. Kognitiivisia harhoja on vaikea välttää ensimmäisen järjestelmän toimiessa automaattisesti. Ja koska järjestelmää ei myöskään saada pois päältä, näitä virheitä intuitiivisessa ajattelussa on vaikea estää. Vääristymiä ei aina voida myöskään välttää, sillä toisella niitä mahdollisesti havaitsevalla järjestelmällä ei välttämättä ole niistä aavistustakaan. Ja vaikka viitteitä vääristymistä olisikin, ne voidaan estää vain seurantaa lisäämällä, sekä toisen järjestelmän ponnistelujen avulla. Tällainen jatkuva valppaus on kuitenkin elämäntapana erittäin epäkäytännöllinen ja kuluttava toisen järjestelmän ollessa tehoton ja liian hidas korvaamaan ensimmäistä rutiininomaisissa päätöksissä. (Kahneman 2012, 37-39).

Mentaalinen ponnistelu on raskasta ja toinen järjestelmä hyväksyy usein ensimmäisen järjestelmän ehdotukset itseään vaivaamatta. Pohdi seuraavaa arvoitusta, pyrkimättä kuitenkaan ratkaisemaan sitä, kuuntele vain intuitiotasi:

Maila ja pallo maksavat 1,10 euroa.

Maila maksaa euron enemmän kuin pallo.

Kuinka paljon pallo maksaa?

Mieleesi tuli todennäköisesti luku 10 senttiä. Tämän esimerkin tarkoituksena on tuottaa helppo, intuitiivinen, miellyttävä ja väärä vastaus. Oikea vastaus on 5 senttiä. Tämän saisi selville pohtimalla vastausta hiukan pidempään, eikä vain hyväksymällä ensimmäistä mieleen tullutta lukua. 10 senttiä on saattanut tulla mieleen ensimmäiseksi myös tähän arvoitukseen oikein vastanneille, he ovat kuitenkin tarkistaneet tulokset vielä toistamiseen tai pystyneet vastustamaan intuitiotaan. Väärin vastanneet siis hyväksyivät ensimmäisen

järjestelmän ehdotuksen, jonka toinen järjestelmä olisi voinut pienen ponnistelun jälkeen hylätä. Vastaaajien olisi pitänyt myös päätellä että, onko kyselylomakkeessa tarkoitus olla näin ilmeinen arvoitus. (Kahneman 2012, 56-57).

Tähän arvoitukseen on vastannut tuhansia yliopisto-opiskelijoita Harvardista, MIT:stä ja Princetonista. Ja yli puolet vastaajista on antanut helpon ja intuitiivisen eli väärän vastauksen. Kahnemanin mukaan tulokset kertovat motivaation puutteesta ajattelussa, eikä niinkään älykkyydestä, vaan siitä että on valmis hyväksymään ensimmäisenä mieleen tulevan luvun. Vielä huonommat tulokset arvoituksesta ovat olleet, kun testiä on tehty muissa vähemmän valikoivissa yliopistoissa. Näiden vastaajista yli 80% jätti tuloksen tarkastamatta. Maila ja pallo -testi osoittaa, että monet ihmiset luottavat liikaa intuitioihinsa ja ovat näin liian itsevarmoja. (Kahneman 2012, 56-57).

Toinen samaa asiaa valaiseva esimerkki on seuraavanlainen looginen väittämä:

Kaikki ruusut ovat kukkia.

Jotkin kukat kuihtuvat nopeasti.

Niinpä jotkin ruusut kuihtuvat nopeasti.

Suurin osa korkeakouluopiskelijoista pitää tätä loogista väitettä totena, vaikka on mahdollista, että ruusut eivät kuulu nopeasti kuihtuviin kukkiin. Maila ja pallo -esimerkin mukaisesti helppo ja hyvä vastaus tuli heti mieleen. Tämä viittaa siihen, että ihmisen uskoessa päätelmään hän todennäköisesti uskoo myös perusteluja. Hän näyttää uskovan siis silloinkin, kun perustelut eivät ole päteviä. (Kahneman 2012, 58-59).

Ensimmäisen järjestelmän toimiessa päätelmät tulevat ennen perusteluja. Kahnemanin mukaan päättelykyvyn lisäksi älykkyys on kykyä paikantaa muistista olennaista aineistoa ja kohdistaa tarkkaavaisuuttaan tarpeen mukaan. Ensimmäisen järjestelmän ominaisuuksiin kuuluu muistitoiminto. Toisen järjestelmän ominaisuuksia taas ovat tietoinen tarkistaminen ja tutkimisen määrä. Ja se taas vaihtelee yksilöiden välillä. Osa meistä kuitenkin suhtautuu intuitioonsa skeptisemmin, eivätkä kaikki ole valmiita hyväksymään pintapuolisesti miellyttäviä vastauksia. Kahnemanin mukaan ainakin kaikki korkeakouluopiskelijat pystyisivät kyllä päättelemään oikeat vastaukset molempiin edellä mainittuihin tehtäviin. (Kahneman 2012, 58-59).

Ensimmäisen järjestelmän hyvyttä mitataan sen tarinankerronnan johdonmukaisuudesta. Tarinan perustana olevien tietojen määrällä tai laadulla ei kuitenkaan ole suurestikaan merkitystä. Tämä ensimmäinen järjestelmä on erittäin taitava tunnistamaan

automaattisesti ja vaivattomasti näitä jo käsiteltyjä syy-yhteyksiä eri tapahtumien välillä, vaikka yhteys olisikin väärä tai puuttuva. (Kahnemann 104-105).

Kahneman kirjoittaa laajalti ilmiöstä, kuinka ihmiset kiirehtivät johtopäätöksiin rajoittuneen informaation perusteella. Hän käyttää tästä ilmiöstä nimitystä WYSIATI (what you see is all there is) eli vapaasti suomennettuna: ei ole muuta kuin nähty. Tämä ilmiö saa ihmiset keskittymään olemassa olevaan tietoon ja hylkäämään puuttuvan aineiston. Ja tämän tuloksena ensimmäinen järjestelmä kehittää pikaisesti yhtenevän ja uskottavan tarinan, joka perustuu rajalliseen tietoon. Toinen järjestelmä saattaa sitten hyväksyä näitä virheellisiä vaikutelmia ja intuitiota niin, että niistä tulee perimmäisiä arvoja ja uskomuksia. WYSIATI voi siis aiheuttaa ensimmäisen järjestelmän keksimään ja päättelemään syitä sekä aikomuksia riippumatta, ovatko nämä aikomukset totta vai keksittyjä. (Kahneman 2012, 104-105).

Ihmiset siis hyppäävät pikaisesti päätelmiinsä, olettavat huonot aikomukset, antautuvat ennakkoluuloille tai puolueellisiksi ja uskovat salaliittoihin. He keskittyvät vain saatavilla olevaan rajalliseen informaatioon, eivätkä huomioi ollenkaan puuttuvaa tietoa. Intuitioita ja vaikutelmia synnyttävästä tiedon laadusta ensimmäinen järjestelmä on täysin välinpitämätön. (Kahnemann 2012, 104-105). Mikäli esimerkiksi liikkeenjohdolle toimitettava analyysi vaatisi kaikkiaan kymmen elementtiä ollakseen kokonainen, mutta ensimmäisellä järjestelmällä on käytössään vain seitsemän komponentin tiedot, saattaa se joko keksiä loput tai jättää puuttuvat huomiotta.

Myös kehystämisaikatuksiksi kutsuttu ilmiö kuuluu samaan WYSIATI- kategoriaan. Saman tiedon esittäminen eri tavalla herättää usein eri tunteita. Väite ”myynti on 90% edellisen kuun myynnistä” on rauhoittavampi kuin ”myynti on laskenut 10% viime kuusta”. Vaikka näiden muotoilujen sisältö on sama. (Kahneman 2012, 106-107).

Tämän lisäksi päätöksiin ja valintoihin vaikuttavat niin kutsutut ankkurit. Ei voida jättää huomioimatta meille aikaisemmin esitettyjä lukuja arvioitaessa jotain suuretta. Ankkurointivaikutuksesta on kyse silloin, kun joku altistaa meidät tietyille luvuille, juuri ennen kuin meidän on arvioitava suuretta. Kysyttäessä kuoliko Gandhi yli 144-vuotiaana hänen arvioidaan menehtyneen paljon vanhempana, kuin kysymyksen viitatessa hänen menehtyneen vain 35-vuotiaana. Kahnemanin mukaan ankkuroitumisvaikutus on hyvin yleinen ja yksi kokeellisen psykologian luotettavimpia ja vahvimpia tuloksia. Sama vaikutus kohdataan myös asuntokaupoilla. Pyyntihinta vaikuttaa tarjottuun hintaan. Ilmiö kannattaa siis tiedostaa, sillä ilmeisen merkityksetönkin luku vaikuttaa ihmisten arvioihin. Ihmiset uskovat intuitiivisesti, että äskettäin hankitut tiedot ovat merkityksellisiä päätöksenteossa – jopa silloin, kun ne eivät ole. (Kahneman 2012, 141-143).

Miksi sitten kannattaa arvioida, osaavatko ihmiset hidastaa ajatteluaan ja käyttää kriittistä ajatteluaan? Ajattelemme intuitiivisesti, että jos päätös tuntuu hyvältä, se on oikea päätös. Meillä on tapana arvioida todennäköisyyksiä väärin, sillä WYSIATI muistaa vain tapahtuneen ja jättää puuttuvan datan huomioita. Samaan tapaan ankkuroidumme epäolennaiseen tai kehystämme ajattelua sen perusteella, miten tiedot esitetään. Kaikkea tätä tietoa voidaan ja on käytetty hyväksi datalukutaitoa arvioivien tehtävien tekemisessä.

Kaikki tämä edellä esitetty voidaan tiivistää muutamaankin merkittävään havaintoon päätöksenteon kannalta. On välttämätöntä saada kunnon dataa ja käyttää sovellettua analytiikkaa päätöksenteon yhteydessä. Vaikka ensimmäinen järjestelmämme onkin tehokas omalla toimialueellaan ja sen rajoissa toimiessaan, se ei kuitenkaan ole hyvä arvioissaan tai tekemään päätöksiä. Kun on kyse kriittisistä liiketoimintapäätöksistä (tai datalukutaitoa mittaavista tehtävistä), tarvitsemme hitaampaa, analyttistä-ajattelua. Ja kuten olemme edellä jo nähneet, jopa hyvän datan kanssa tarvitaan ylimääräistä kurinalaisuutta perinteisten päätöksenteon prosesseihin liittyen. Omien tunnistettujen harhojen tai vääristymien välttämiseksi pitää meidän tietoisesti saada ajattelumme käyttämään toista hitaampaa ja kriittiseen ajatteluun pystyvää järjestelmää.

Pyri hidastamaan ajattelusi ainakin tehdessäsi arviointia tai analyysia ensimmäistä kertaa. Käy läpi viiallisen ajattelun tarkistuslistaa, esimerkiksi:

- Onko tämä paras vaihtoehto, vai vaihtoehto, jolle vain olen usein altistunut?
- Onko tämä paras vaihtoehto vai kärsinkö tunnelinäöstä (WYSIATI) ja onko minun tutkittava muitakin mahdollisuuksia?
- Mikäli kysymys vaikuttaa olevan helppo ja yksinkertainen, pohdi sitä vielä hetki ja anna hitaan ajattelusi tehdä tehtävänsä ennen vastausta.

Voimme siis itse tietoisesti hidastaa ajatteluamme (tai ainakin yrittää sitä), ja olla hyväksymättä ensimmäisenä mieleen tullutta helppoa vastausta. Esimerkiksi epäselvällä fontilla kirjoitettuja kysymyksiin tai hyvin sekaviin visualisointeihin voidaan saada enemmän oikeita vastauksia, sillä ihmiset joutuvat näin keskittymään tehtävään paremmin. Tämä on kuitenkin ristiriidassa eräiden tavoitteidemme kanssa, jotka ovat verkkokyselyn mobiilikäytettävyys ja helppolukuisuus. Lopuksi voidaan vielä selventää, että aivoissamme ei kuitenkaan ole kahta erillistä järjestelmää, vaan järjestelmät ovat havainnollistava kuvaus mieleemme toiminnasta.

## 5 Datalukutaidon selvittäminen

Verkkokyselyn kaupallisen luonteen vuoksi esittelen siitä vain osan tässä työssä. Esimerkkitehtävät ovat verkkokyselyn ensimmäisen version osia.

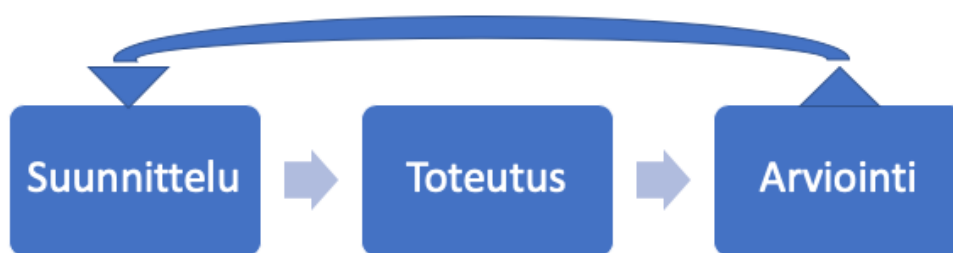
### 5.1 Datalukutaitoa selvittävä kysely

Verkkokysely on kvantitatiivinen eli määrällinen tutkimus, joka sisältää myös kvalitatiivisia eli laadullista tutkimusta. Kvalitatiiviset kysymykset perustuivat vastaajien taustatietoihin. Jokainen yksittäinen kysymys on käynyt läpi useamman muutoksen ja arvioinnin ennen päätymistään lopulliseen muotoonsa. Kysymyksiä pohdittaessa on kysytty mielipiteitä esimerkiksi kaavioiden värityksiin, muotoiluihin ja kysymysten kirjalliseen muotoon. Kysymykset ovat muodostuneet pala palalta sitä mukaa, kun uusia ideoita on syntynyt. Taustatietoja kerättiin tämän työn lisäksi kohdeorganisaatioiden sisäisiin tarpeisiin, että tulosten perusteella tehtäviä toimenpiteitä voitaisiin kohdentaa oikein osastokohtaisesti.

Toteutukseen asti selviytyneet kysymykset on valittu tärkeinä pitämiemme asioiden perusteella. Olemme käyttäneet aluksi vain hyvin yleisiä visualisointeja ja tehtäviä, joita liikelämässä kohdataan, sekä katsomme datalukutaidon perustaitojen osalta tarpeellisiksi. Lisäksi olemme keskittyneet pääosin tulkintaa ja lukemista arvioiviin tehtäviin. Hyvin paljon tärkeitä ja tarpeellisia asioita esimerkiksi tilastotieteeseen ja todennäköisyyksiin on jouduttu jättämään alkuvaiheessa toteutuksen ulkopuolelle. Olemme tietoisia, että organisaatioiden datalukutaidon taso saattaa vaihdella suurestikin, eikä yksi ja sama kysely välttämättä toimi jokaisessa organisaatiossa.

Työn tekeminen aloitettiin luomalla eräänlaiset kysymysaihiot, joita sitten on iteratiivisesti kehitetty lopulliseen muotoonsa MVP:n hengessä. Ojasalo ym. mukaan kaikenlainen kehittämistyö voidaan jäsentää yksinkertaiseksi muutostyön prosessiksi. Ensimmäiseksi selvitetään kehittämishaasteet ja niitä koskevat tavoitteet, sekä suunnitellaan, miten tavoitteet voitaisiin saavuttaa. Tätä vaihetta kutsutaan suunnitteluvaiheeksi. Toisena vaiheena muutosprosessissa on suunnitelman toteutusvaihe. Viimeisenä vaiheena arvioidaan, miten prosessissa on onnistuttu. Tämän vaiheen pohdinnan seurauksena onkin jo saatettu aloittaa uuden kehitystarpeen suunnittelu. (Ojasalo, Moilanen ja Ritalahti 2014, 22-23). Myös tämä opinnäytetyö on seurannut tätä kaavaa yksittäisten tehtävien osalta, sillä yleisesti hyvin toimivien kysymysten ja selkeiden visualisointien tekeminen on vaatinut iterointia useamman toteutuskierroksen verran sekä paljon sisäistä keskustelua tätä aihetta varten perustetulla Slack-kanavalla (#datalukutaito).

Verkkokyselyn yksittäisten tehtävien iteratiivinen kehitysprosessi:



Kuvio 24. Muutostyön prosessi (Ojasalo, Moilanen ja Ritalahti, 2018, 23)

Esimerkiksi organisaation suorituskykyä voidaan kehittää tehokkaasti vain, jos se on mittavissa ja arvioitavissa. Sama pätee muihinkin asioihin organisaatiossa.

Verkkokyselyssä meidän piti myös vastata asiakkaiden toiveisiin eli selvittää heidän datalukutaitoaan mahdollisimman laaja-alaisesti arvioimalla IT-infrastruktuuria, datan saatavuutta ja analytiikka-alustan kyvykkyyksiä, mutta myös organisaatioiden päätöksentekoa tukevia prosesseja ja selvittää mahdollisia esteitä tehokkaalle datan hyväksikäytölle (siilot yms.)

Tehtävien muodostumiseen ovat vaikuttaneet vahvasti Kahnemanin teoriat ihmismielen toiminnasta. Esimerkiksi siitä, kuinka kyselyyn vastaaja saattaa valita ensimmäisenä mieleen tulevan helpon vaihtoehdon, ja kuinka me voimme tarjota hänelle tehtävän ratkaisun kannalta epäolennaista tietoa ankkurivaikutuksen muodostamiseksi. Lisäksi tietomme visualisointien parhaista käytännöistä sekä teoriakirjojen perusteet vaikuttivat graafisten tehtävien syntyyn. Kyselyn kieleksi valikoitui englanti, sillä ensimmäisten vastaajien joukossa oli useita eri kansallisuuksia. Kielivalinta on saattanut osalla hankaloittaa ja hidastaa kysymyksen ymmärrystä sekä vastaamista.

Laskentaa edellyttävissä tehtävissä helpotimme vastaamista vaihtoehdoilla niin, että laskinta ei tarvita yhtenkään tehtävän ratkaisemiseksi. Tärkeintä on hahmottaa suhteet eli mitä luvulle tapahtuu: kasvaako se, vai väheneekö se. Verkkokyselyn muissa osioissa kaikki tehtävät olivat kirjallisia kysymyksiä. Nämä kaksi muuta osiota arvioivat siis organisaation IT-kyvykkyyksiä sekä prosesseja datalukutaitoon liittyen. Organisaation kypsyytensä on selvitetty muun muassa seuraavilla kysymyksillä: ”Organisaatiomme päätöksenteko perustuu dataan?” ja ”Kannustetaanko organisaatiossa hyödyntämään dataa?”. Vastausvaihtoehdot näissä kysymyksissä ovat olleet: 1 = Completely disagree, 2 = Slightly disagree, 3 = Neither agree nor disagree, 4 = Slightly agree ja 5 = Completely agree. Pisteitä on annettu vastauksen mukaan eli vastausvaihtoehdosta 5 on annettu 5 pistettä.

Hirsjärvi, Remes ja Sajavaara ovat koostaneet seuraavan suosituslistan, jota olemme noudattaneet verkkokyselyn suunnittelussa (Hirsjärvi ym. 2004):

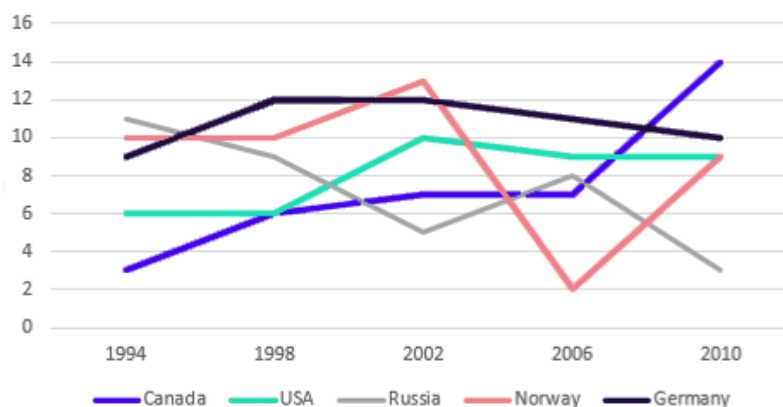
1. Selvyys on tärkeintä. Laadi yksiselitteisiä kysymyksiä.
2. Tarkat kysymykset ovat parempia kuin yleisluontoiset kysymykset.
3. Lyhyet kysymykset ovat parempia kuin pitkät.
4. Vältä kysymyksiä, joihin liittyy kaksoismerkityksiä. Kysy vain yhtä asiaa kerrallaan.
5. Tarjoa valittavaksi "ei mielipidettä" -vaihtoehto (en tiedä, tässä yhteydessä).
6. Käytä monivalintavaihtoehtoja mieluummin kuin "samaa" mieltä tai "eri mieltä"-väittämiä.
7. Harkitse kysymysten määrää ja järjestystä lomakkeessa.
8. Tarkista sanojen valinta ja käyttö. Käytä kieltä, jota vastaaja ymmärtää.

Mittauksen eli havaintoyksikön kohteita ovat olleet organisaatiot ja ryhmät, jolle verkkokysely on toteutettu. Tätäkin kyselyä suunniteltaessa etsittiin aiheesta aikaisemmin tehtyjä tutkimuksia ja kyselyitä. Verkkokyselyn viemää vastausaikaa kelloitettiin Cubiqin omien työntekijöiden vastausten perusteella. Ja vastausajaksi saatiin suositellun 15-20 minuutin sijaan alle 15 minuuttia. (Ojasalo ym. 2014, 131).

## 5.2 Testin analyysi ja tulokset

Ensimmäiseen asiakkailta ja Cubiqin ulkopuolisilla henkilöillä toteutettuun versioon vastasi 161 henkilöä kolmen eri toteutuskerran aikana. Seuraavaksi käydään läpi nämä vastaukset henkilökohtaisia ominaisuuksia arvioivan osuuden osalta.

23. According to line chart beneath, which country has the greatest total (1994 – 2010) amount of sales? Sales are in millions.



Kuva 1. Verkkokyselyn kysymys numero 23.

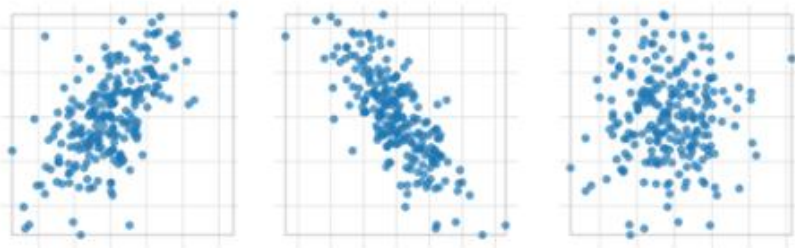
A) Canada, B) USA, C) Germany, D) Norway, E) I don't know

Kyseessä on yleisesti käytössä oleva viivagraafi, joka soveltuu hyvin trendin kuvaamisen. Tehtävän ratkaisemiseksi tulee osata tulkita viivagraafia sekä ymmärtää, että kysymyksessä haetaan suurinta kokonaismyyntiä ajalta 1994 – 2010. Oikea vastaus on C - Saksa, sen viivan alle jäävän alueen ollessa suurin. Kysymykseen on vastannut oikein 84 % vastaajista. Kysymyksen tiedettiin olevan yksi helpoimmista ja siksi se valikoitui sarjan ensimmäiseksi.

Taulukko 2. Verkkokyselyn kysymyksen 23. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä
A		5	11	16
B		1		1
C		66	68	136
D		1	1	2
E		1	5	6
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>161</b>
<b>Oikein-%</b>		<b>89 %</b>	<b>80 %</b>	<b>100 %</b>
			<b>100 %</b>	<b>84 %</b>

24. What is the best option to describe these scenarios from left to right?



Kuva 2. Verkkokyselyn kysymys numero 24.

A) Positive correlation, negative correlation, correlation, B) Positive correlation, negative correlation, no correlation, C) Negative correlation, positive correlation, no correlation, D) No correlation, no correlation, correlation, E) I don't know

Kuvan kolmessa eri hajontakaaviossa on kuvattu heikko positiivinen korrelaatio, heikko negatiivinen korrelaatio sekä tapaus, jossa minkäänlaista korrelaatiota ei ole havaittavissa. Parhaiten kuvan tapahtumasarjaa kuvaa vaihtoehto B.

Taulukko 3. Verkkokyselyn kysymyksen 24. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä
A	7	12		19
B	59	57	2	118
C	2	3		5
D		1		1
E	6	12		18
<b>Kaikki yhteensä</b>	<b>74</b>	<b>85</b>	<b>2</b>	<b>161</b>
<b>Oikein-%</b>	<b>80 %</b>	<b>67 %</b>	<b>100 %</b>	<b>73 %</b>

25. Here is a list of ages of group of people. Find the mean, median, mode and range for the following values: 7, 7, 7, 10, 11, 12, 16

Kuva 3. Verkkokyselyn kysymys 25.

A) 8, 10, 70, 16, B) 10, 10, 10, 7, C) 7, 10, 10, 10, D) 10, 10, 7, 9, E) I don't know

Tehtävässä kysytään keski- ja hajontalukuja. Tehtävän voi ratkaista monella eri tavalla. Yksinkertaisin tapa on ratkaista vaihteluväli (range), joka on suurimman ja pienimmän arvon väli eli tässä tapauksessa 9 (16 – 7). Toinen vaihtoehto on etsiä moodi eli yleisin luku 7, sillä vain yhdessä vastausvaihtoehdossa moodi on 7.

Taulukko 4. Verkkokyselyn kysymyksen 25. Vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä
A	2	2	1	5
B	13	11		24
C	13	7		20
D	34	31	1	66
E	12	34		46
<b>Kaikki yhteensä</b>	<b>74</b>	<b>85</b>	<b>2</b>	<b>161</b>
<b>Oikein-%</b>	<b>46 %</b>	<b>36 %</b>	<b>50 %</b>	<b>41 %</b>

26. Correlation always implies causality

Kuva 4. Verkkokyselyn kysymys 26.

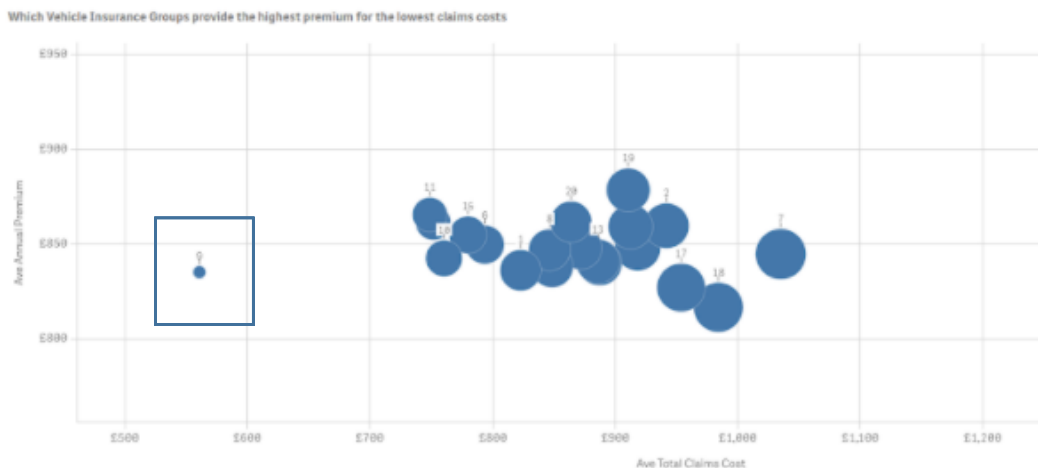
A) Yes, B) No, C) I don't know

Oikea vastaus on ei, kuten aiemmin jo mainittiin. Korrelaatio on helpommin havaittavissa. Kausaliteetti taas on vaikeampi todeta.

Taulukko 4. Verkkokyselyn kysymyksen 26. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä
A		22	16	38
<b>B</b>		<b>38</b>	<b>37</b>	<b>76</b>
C		14	32	47
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>161</b>
<b>Oikein-%</b>		<b>51 %</b>	<b>44 %</b>	<b>50 %</b>

27. Which outlying Vehicle Insurance Group provides the highest premium for the lowest Claims costs?



Kuva. 5 Verkkokyselyn kysymys 27.

A) 18, B) 9, C) 7, D) 19, E) I don't know

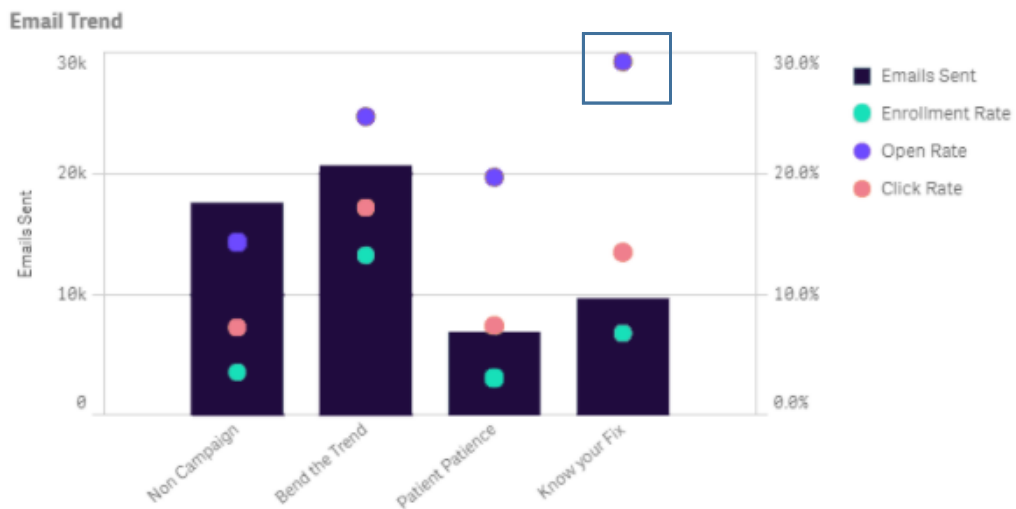
Kuvassa on hajontakaavio, jossa kaksi mittaria. Tiettyjen ajoneuvoryhmien vuosittainen keskimääräinen vuosittainen palkkio ja vakuutuskorvaukset. Kysyttäessä mikä ryhmä tuottaa parhaiten alimmilla korvauksilla. Vastaus on B eli ryhmä- 9. Tämän ryhmän vakuutuskorvaukset ovat olleet keskimäärin huomattavasti matalammat kuin muilla ryhmillä, vakuutusmaksujen ollessa hyvällä tasolla. Huomattavasti vaikeampi tehtävä edelliseen verrattuna, kuten vastauksistakin voi päätellä.

Taulukko 5. Verkkokyselyn kysymyksen 27. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä
A		9	11	20
<b>B</b>		<b>23</b>	<b>16</b>	<b>39</b>
C		2	4	6
D		18	25	44
E		22	29	52
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>161</b>
<b>Oikein-%</b>		<b>31 %</b>	<b>19 %</b>	<b>0 %</b>

28. Beneath is the Email campaign performance chart. Which campaign has the highest open rate?

### Email Campaign Performance



Kuva 6. Verkkokyselyn kysymys 28.

A) Non Campaign, B) Bend the Trend, C) Patient Patience, D) Know your Fix, E) I don't know

Kuvassa on niin kutsuttu yhdistelmä kaavio (combo chart). Vastatakseen kysymykseen on tulkittava kuvassa oikealla näkyviä selitteitä ja sen jälkeen etsittävä kampanjassa, missä kyseinen selite on suurin. Suhteellisen helppo tehtävä. Vastaavanlaisia kaaviota esiintyy paljon mediassa. Samalla kaaviolla voidaan esittää esimerkiksi ennusteen ja toteuman suhdetta.

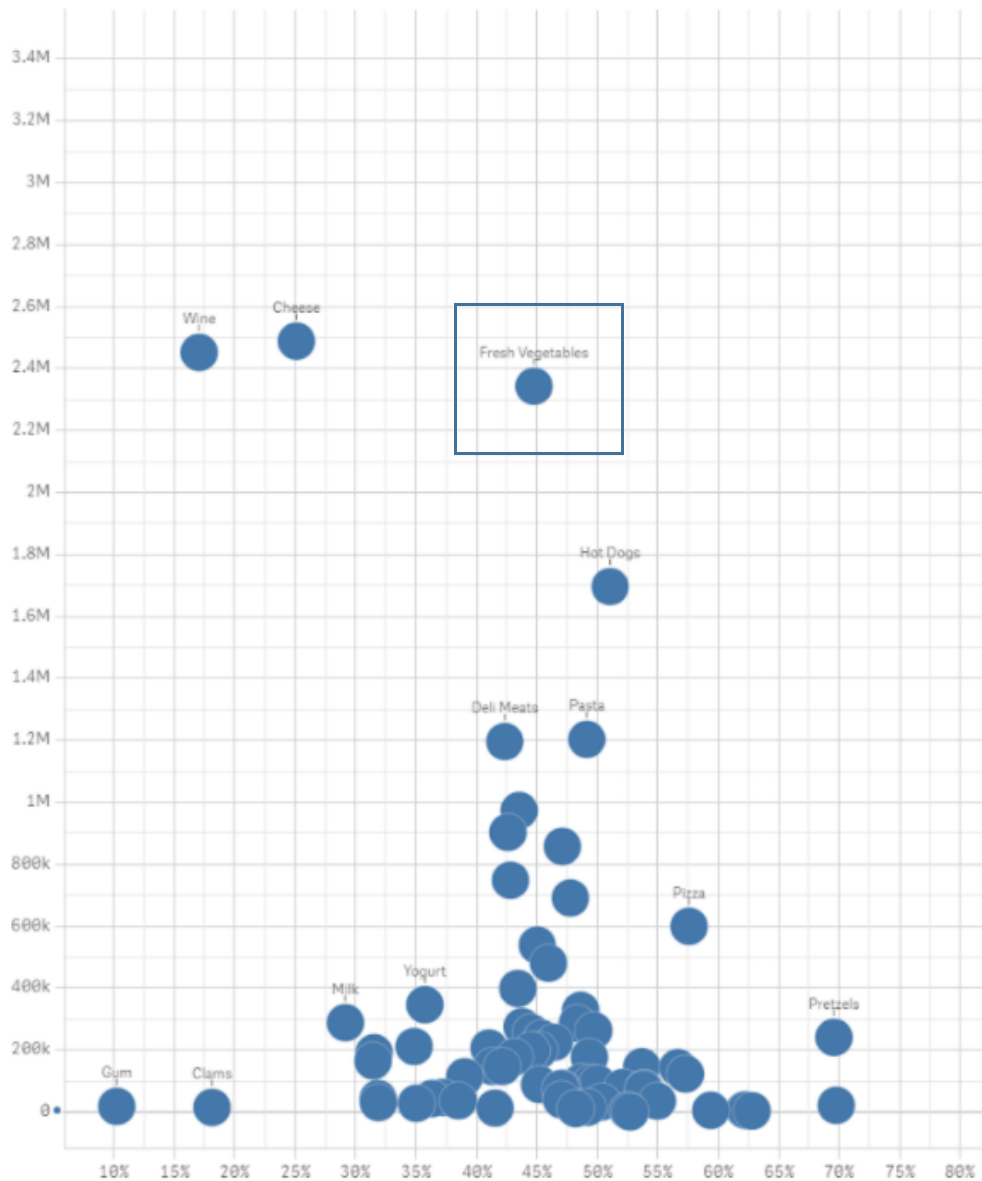
Taulukko 6. Verkkokyselyn kysymyksen 28. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä	
A			3	3	
B		8	5	13	
C		3	1	4	
D		53	60	2	115
E		10	16		26
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>2</b>	<b>161</b>
<b>Oikein-%</b>		<b>72 %</b>	<b>71 %</b>	<b>100 %</b>	<b>71 %</b>

29. In the Sales & Margin Analysis scatter plot above, the most profitable (in margin \$) product group is

### Sales & Margin Analysis

Western Sales vs Margin (Sales \$23,889,222)



Kuva 7. Verkkokyselyn kysymys 29.

A) Cheese, B) Fresh Vegetable, C) Pizza, D) Pretzels, E) I don't know

Kyseessä siis hajontakaavio, jossa etsitään katteen perusteella tuottoisinta tuoteryhmää. Mikäli kysyttäisiin myynnin perusteella suurinta tuoteryhmää, se olisi juustot, mutta vastatakseen tulee hahmottaa mitä tuoteryhmää on myyty eniten hyvällä katteella. Oikea vastaus on B – tuoret vihannekset. Kateprosentin ollessa noin 45 ja myynnin 2.35 miljoonaa dollaria, katteeksi jää noin 1.05 miljoonaa dollaria. Juuston kohdalla katteeksi jää noin  $(2.5 * 0,25)$  625 tuhatta dollaria. Pretzeleitä toki myydään kovalla katteella, mutta vain noin 250 tuhannella dollarilla, joten katetta jää noin 175 tuhatta dollaria.

Taulukko 7. Verkkokyselyn kysymyksen 29. vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä	
A		16	27	43	
B		16	16	2	34
C		2	2	4	
D		26	19	45	
E		14	21	35	
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>2</b>	161
<b>Oikein-%</b>		<b>22 %</b>	<b>19 %</b>	<b>100 %</b>	<b>21 %</b>

30. A number is increased by 30 % and then decreased by 30 %. Find the decrease per cent.

Kuva 8. Verkkokyselyn kysymys 30.

A) 3 %, B) 9 %, C) 15 %, D) It does not decrease, E) I don't know

Havainnollistetaan esimerkkiä luvulla 1. Lisätään lukuun ensin 30%,  $1 * 1,3 = 1,3$ . Tämän jälkeen vähennetään luvusta 30%.  $1,3 * 0,7 = 0,91$ . Alkuperäinen luku pienenee  $1 - 0,91 = 9$  %. Tässä esimerkissä Kahnemanin kuvailema järjestelmä 1 on saattanut tuottaa helpon ja intuitiivisen sekä väärän vastauksen. Tämän tehtävän perusteella seuraavaan versioon tehtiin muutama kysymys, joissa ensimmäisenä mieleen tuleva vaihtoehto on helppo ja intuitiivinen, mutta kuitenkin vääränä vastauksena. Toisin sanoen palkkaasi voidaan leikata 1000 rahasta 50 %, jolloin sinulle jää 500 rahaa. Tähän kun lisätään 50 %, niin palkkasi on enää 750 rahaa.

Taulukko 8. Verkkokyselyn kysymyksen 30. Vastaukset.

Vastaus	Mies	Nainen	N/A	Yhteensä	
A		15	6	21	
B		39	17	2	58
C		9	10	19	
D		3	22	25	
E		8	30	38	
<b>Kaikki yhteensä</b>		<b>74</b>	<b>85</b>	<b>2</b>	161
<b>Oikein-%</b>		<b>53 %</b>	<b>20 %</b>	<b>100 %</b>	<b>36 %</b>

Kokonaisuutena kysymysten haasteellisuus yllätti meidät, sillä tästä joukosta vain yksi henkilö onnistui saamaan täydet pisteet. Kysely toteutettiin aina suuremmalle joukolle kerrallaan ja vastauksia ei ole voinut miettiä pitkään, sillä aikaa vastata kaikkiin 30. kysymykseen on ollut vain 15 minuuttia. Kaikki vastaukset on annettu mobiilipäätteillä, mikä on

myös osaltaan tuonut lisähaastetta vastaamiseen. Osaa tehtävistä, kun saattanut joutua liikuttelemaan pienellä ruudulla paljonkin ymmärryksen muodostamiseksi.

Taulukko 9. Verkkokyselyn henkilökohtaisen osuuden pistejaukauma.

Pisteet	Mies	N/A	Nainen	Yhteensä
0			3	3
1	2		10	12
2	5		9	14
3	13		19	32
4	18		20	38
5	17	1	11	29
6	14		9	23
7	4	1	4	9
8	1			1
<b>Yhteensä</b>	<b>74</b>	<b>2</b>	<b>85</b>	<b>161</b>
<b>N=161</b>				

Tämän osion datalukutaidon indeksiksi muodostuisi näin 0,49. Olemme laskeneet indeksin kerättyjen sekä jaossa olevien pisteiden suhteeksi. Näissä tapauksissa kerätyt pisteet olivat 642, ja jaossa olleet pisteet olivat 1288 ( $n=161 * 8$ ). Samaa kaavaa olemme toteuttaneet muidenkin kyselyn osioiden kanssa eli IT-infrastruktuuria sekä organisaation prosesseja arvioivissa kysymyksissä. Nämä kysymykset olivat vaihtoehdoiltaan: 1 = Completely disagree 2 = Slightly disagree, 3 = Neither agree nor disagree, 4 = Slightly agree ja 5 = Completely agree. Pisteitä annettiin vaihtoehdon perusteella eli Completely agree tuotti 5 pistettä.

## 6 Pohdinta

Opinnäytetyön tutkimustehtävänä oli luoda organisaation datalukutaidon perustasoa arvioiva kvantitatiivinen verkkokysely. Verkkokyselyssä arvioitiin kolmea eri osa-aluetta: organisaation prosessit, IT-infrastruktuuri sekä henkilökohtaiset kyvykkyydet. Lisäksi verkkokyselyn kysymykset sekä tulokset esiteltiin henkilökohtaisten kyvykkyyksien osalta. Ainakaan liike-elämässä lukutaitoa ei enää voida määritellä pelkkään kykyyn lukea ja kirjoittaa. On olemassa useita muita taitoja, jotka ihmisten on hallittava ongelmien ratkaisemiseksi ja tiedon hankkimiseksi. Kyvykkyydet datan kanssa eli kyky lukea, tulkita ja kommunikoida sitä ovat nykyään organisaatioiden tärkeimpiä kykyjä.

“A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it” theoretical physicist Max Planck (1858-1947)

Organisaation datalukutaitoon vaikuttaa moni asia. Se koostuu edellä mainittujen kolmen eri osa-alueen yhteen sovittamisesta. Yhdenkin osa-alueen laiminlyönti kostautuu kokonaisuuden kustannuksella. Et siis voi saada parasta mahdollista lopputulosta vain rekrytoimalla, vaan organisaation tulee tukea yksilöitä hyvällä IT-infrastruktuurilla, sekä parantaa datalukutaitoa sitä edistävillä prosesseilla. Tiedon ympärillä olevista silloista tulisi päästä mahdollisuuksien mukaan eroon ja varoa hippoa neuvotteluhuoneessa. Tarkimman arvion organisaation datalukutaidon tilasta saa räätälöimällä verkkokyselyn kysymykset organisaation tarpeiden mukaisesti. Datalukutaidon arviointi tulisikin aina tehdä organisaation omien tavoitteiden ja käytössä olevien visualisointien tai taulukoiden pohjalta, eikä yleinen verkkokysely ole tähän tarkoitukseen aina sopiva. Datalukutaitoa tukevien prosessien ja IT-infrastruktuurin arvio on paremmin yleistettävissä, ja samat kysymykset sopivat näiden osalta paremmin kaikkiin organisaatioihin.

Itse kehitystehtävää voidaan pitää onnistuneena, niin työn tilaajan, kuin asiakkaidenkin näkökulmasta. Asiakkaiden osalta verkkokyselyn onnistumista arvioitiin testiin vastanneilta ja asiakasorganisaatioiden avainhenkilöitä saatujen palautteiden perusteella. Tässä ote eräästä asiakaspalautteesta:

”Mobiilisti kick off -tilaisuuksien osana toteutettu datalukutaitotesti osoittautui erinomaiseksi työkaluksi. Sen avulla saimme innostavassa ja interaktiivisessa muodossa osoitettua datalukutaidon tärkeyden henkilöstöllemme ja samalla hahmotettua organisaatioidemme nykytilanteen ja kehityskohteet

tarkasti ja nopeasti. Henkilöstöltä saatu palaute oli erittäin positiivista, ja aiomme mitata jatkossa säännöllisesti datalukutaidon kehitystä.”

Kriittisimmät palautteet ovat koskeneet lähinnä kyselyn pituutta sekä kaavioiden luettaavuutta mobiilipäätteiltä. Tämä on saattanut näkyä vastauksissa siten, että on valittu vain jokin vaihtoehto kysymystä sen enempiä pohtimatta, tai sitten on suoraan valittu vaihtoehto: 'en tiedä'. Ihmiset eivät välttämättä ole kovin innokkaita vastaamaan vastaaviin kyselyihin, ainakaan epävarmuuden vallitessa, mikäli kyselyn perusteella voidaan tarkasti yksilöidä huonoiten pärjänneet henkilöt ja mahdollisesti hankkiutua heistä eroon heti seuraavan tilaisuuden koittaessa. Verkkokyselyn tarkoituksena ei ole kuitenkaan selvittää parhaiten tai huonoiten suoriutuvia yksilöitä, vaan arvioida, kuinka hyvin organisaatiossa osataan käsitellä ja ymmärtää dataa yleisesti. Organisaation pitää osata käyttää dataa hyväksi päätöksenteossa ja analyyseissä.

Haasteita jatkoon kannalta tuottaa se, miten onnistumme tuottamaan kysymys- ja tehtäväpankin, jossa olisi eri vaikeustasoisten kysymysten lisäksi useampi tasoltaan vastaava kysymys. Täysin samaa kyselyä ei ole mielekästä toteuttaa samassa organisaatiossa kehitystä seuratakseen, sillä ihmiset tietävät oikeat vastaukset ensimmäisen vastauskerran palauteosion perusteella.

Mikäli itse datalukutaidon määritelmä on epäselvä, se voidaan käsittää samankaltaiseksi kyvyksi kuin se, että pystyy lukemaan kirjoitusta: katsot jotain ja ymmärrät, mistä siinä on kyse. Datalukutaito ei ole sen monimutkaisempaa, eikä termeistä kannata tehdä sen suurempaa asiaa. Datalukutaito on kuitenkin hyväksytty omaksi termiksi ja sitä käytetty tutkimuskirjallisuudessa ainakin vuodesta 2012 lähtien.

Tietoperustaan tutustuminen on lisännyt ymmärrystäni monimutkaisen kokonaisuuden moninaisuudesta: kuinka datasta syntyy mahdollisesti viisautta ja miten organisaation olisi tuettava tätä prosessia. Teoriaan perehtyminen antoi minulle käsityksen siitä, miten moni asia vaikuttaa lopulta ymmärryksen muodostumiseen, sekä siitä, kuinka haastava tehtävä on sen arvioivan verkkokyselyn luominen. Aiheena datalukutaito on antanut tukea analytiikkakonsulttina kehittymiselle, koska aiheesta on tulossa yhä tärkeämpää ja työme aiheen ympärillä varmasti lisääntyy lähiaikoina. Myös ajatusten kirkastaminen visualisointiprosessista ja kaikesta siihen liittyvästä on ollut antoisaa.

Kahnemanin mukaan ihmisillä on taipumus vastata kysymyksiin ensimmäisellä mieleen tulevalla ajatuksella tarkistamatta sitä, varsinkin kun aikaa vastauksen muodostamiseen on vähän. Olemme itsekkin huomanneet tämän ilmiön. Vertaamalla Cubiqin yleisen ja kai-

kille avoimen datalukutaitotestin vastausdataa tässä työssä esiteltyyn dataan. Havaintomme mukaan tulokset ovat olleet parempia, kun aikaa on käytössä enemmän tai sitä ei valvota. Arviomme mukaan yleisen testin kysymykset ovat jopa edellä esitettyjä kysymyksiä haastavampia, ja silti kerätyt pisteet yleisestä testistä ovat olleet parempia. Tämän henkilökohtaista datalukutaitoasi arvioivan testin voit tehdä osoitteessa: <https://datalukutaitotesti.cubiq.fi/>

Kehitystyön tekeminen on ajoittunut aikaan, jolloin koronavirus on tuonut arkeemme ennen näkemätöntä epävarmuutta. Eikä tämä aika ole vähentänyt datalukutaidon tärkeyttä, vaan päinvastoin. Mediassa on ollut esillä vaikka minkälaista taulukkoa ja kaaviota siitä, kuinka korona mahdollisesti leviää, ja miten se vaikuttaa talouteen. Ja sitten on vielä se tarttuvuusluku.

“It ain’t what you don’t know that gets you into trouble. It’s what you know for sure that just ain’t so.” – tuntematon ajattelija.

## Lähteet

- Accenture 2020. The Human Impact of Data Literacy. Luettavissa: [https://www.accenture.com/\\_acnmedia/PDF-118/Accenture-The-Human-Impact-Data-Literacy.pdf#zoom=50](https://www.accenture.com/_acnmedia/PDF-118/Accenture-The-Human-Impact-Data-Literacy.pdf#zoom=50). Luettu: 10.7.2020.
- Adelman, S. & Moss, L. & Abai, M. 2005. Data Strategy. Addison-Wesley Professional.
- Australian Bureau of Statistics 2020. What are data?. Luettavissa: <https://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+data>. Luettu: 10.5.2020
- Awad, E. M. & Ghaziri, H. 2003. Knowledge Management. Pearson Education.
- Bair, J. 2001. Knowledge management technology: beyond document management: Awad, E. M., Ghaziri, H. 2003. Knowledge Management, Pearson Education.
- Baykoucheva, S. 2015. Managing Scientific Information and Research Data. Waltham, MA.
- Bersin, J. & Zao-Sanders, M. 2020. Boost Your Team's Data Literacy. Harvard Business Review. February 2020. Luettavissa: <https://hbr.org/2020/02/boost-your-teams-data-literacy>. Luettu: 20.4.2020
- Bowne-Anderson, H. 2018. Your Data Literacy Depends on Understanding the Types of Data and How They're Captured. Harvard Business Review. October 2018. Luettavissa: <https://hbr.org/2018/10/your-data-literacy-depends-on-understanding-the-types-of-data-and-how-theyre-captured>. Luettu: 16.4.2020
- Britannica 2020a. Physiology. Luettavissa: <https://www.britannica.com/science/information-theory/Physiology>. Luettu: 16.6.2020.
- Britannica 2020b. Daniel Kahneman. Luettavissa: <https://www.britannica.com/biography/Daniel-Kahneman>. Luettu: 25.7.2020.
- Brynjolfsson, E. & Lorin M, H. & Heekyung H, K. 2011. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?. Luettavissa [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1819486](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486). Luettu: 12.5.2020

- Cairo, A. 2016. *the truthful art – data, charts, and maps for communication*. New Riders.
- Cambridge Dictionary 2020. Data. Luettavissa: <https://dictionary.cambridge.org/dictionary/english/data>. Luettu: 10.6.2020.
- Chollet, F. 2017. *Deep Learning with Python*. Manning Publications.
- Few, S. 2012. *Show Me the Numbers – Designing Tables and Graphs to Enlighten*. Analytics press. California.
- Figuroa, A. 2019. Data Demystified — DIKW model. Luettavissa: <https://towardsdatascience.com/rootstrap-dikw-model-32cef9ae6dfb>. Luettu: 30.4.2020.
- Forrester 2016. The Forrester Wave™: Big Data Hadoop Distributions, Q1. Luettavissa: 2016<https://www.forrester.com/report/The+Forrester+Wave+Big+Data+Hadoop+Distributions+Q1+2016/-/E-RES121574>. Luettu: 10.5.2020.
- Grönroos, C. 2006. Adopting a service logic for marketing. *Marketing Theory*. 6. Luettavissa: [https://www.researchgate.net/publication/215915799\\_Adopting\\_a\\_service\\_logic\\_for\\_marketing](https://www.researchgate.net/publication/215915799_Adopting_a_service_logic_for_marketing). Luettu: 29.3.2020.
- Grönroos, C. & Voima, P. 2011. Making sense of value and value co-creation in service logic. *Hanken school of Economics Working Papers 559* : 32. Luettavissa: <https://helda.helsinki.fi/handle/10138/29218>. Luettu 7.7.2020.
- Hirsjärvi, S. & Remes, P. & Sajavaara, P. 2004. *Tutki ja kirjoita*. Tammi. Helsinki.
- IBM 2019. The enterprise guide to closing skills gap. Luettavissa: <https://www.ibm.com/thought-leadership/institute-business-value/report/closing-skills-gap#>. Luettu: 10.5.2020.
- Kahneman, D. 2012. *Ajattelu – nopeasti ja hitaasti*. Terra Cognita.
- Kircher, L. 2010. Data Is the New Soil. *Columbia Journalism Review*. Luettavissa: [https://archives.cjr.org/the\\_news\\_frontier/data\\_is\\_the\\_new\\_soil.php](https://archives.cjr.org/the_news_frontier/data_is_the_new_soil.php). Luettu: 13.6.2020.
- Kirk, A. 2016. *Data Visualisation – A Handbook for Data Driven Design*. Sage.

Lenarduzzi, V. & Taibi, D. 2016. MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product. Luettavissa: [https://www.academia.edu/35534315/MVP\\_Explained\\_A\\_Systematic\\_Mapping\\_Study\\_on\\_the\\_Definitions\\_of\\_Minimal\\_Viable\\_Product](https://www.academia.edu/35534315/MVP_Explained_A_Systematic_Mapping_Study_on_the_Definitions_of_Minimal_Viable_Product). Luettu: 8.4.2020.

Levitt, T., Marketing Success Through Differentiation – Of Anything. Harvard Business Review. January-February 1980. Luettavissa: <https://hbr.org/1980/01/marketing-success-through-differentiation-of-anything>. Luettu: 7.7.2020.

Lotan, G. 2020. (Fake) friends with (Real) benefits. Medium.com. Luettavissa: <https://medium.com/i-data/fake-friends-with-real-benefits-eec8c4693bd3>. Luettu: 10.7.2020.

Malhotra, Y. 2000. Knowledge Management for the New World of Business. Luettavissa: <https://www.brint.com/km/whatis.htm>. Luettu: 22.7.2020.

Marr, B. 2017. Data Strategy. Kogan Page. UK and US.

Ackoff, R. L. 1989. From Data to Wisdom, Journal of Applied Systems Analysis, Volume 16.

Meettheboss.tv 2020. Work related stress. Luettavissa: <https://meettheboss.tv/blog/work-related-stress-statistics>. Luettu: 15.7.2020.

Merriam-Webster 2020a. Data. Luettavissa: <https://www.merriam-webster.com/dictionary/data>. Luettu: 10.6.2020.

Merriam-Webster 2020b. Causation. Luettavissa: <https://www.merriam-webster.com/dictionary/causality>. Luettu: 11.6.2020.

Merriam-Webster 2020c. Correlation. Luettavissa: <https://www.merriam-webster.com/dictionary/correlation>. Luettu: 11.6.2020.

Milgram, S. 1963. Behavioral Study of obedience. The Journal of Abnormal and Social Psychology. Luettavissa: <https://content.apa.org/record/1964-03472-001>. Luettu: 10.7.2020.

Szatmari, B. 2016. We are (all) the champions: The effect of status in the implementation of innovations. Luettavissa: <http://hdl.handle.net/1765/94633>. Luettu: 20.7.2020.

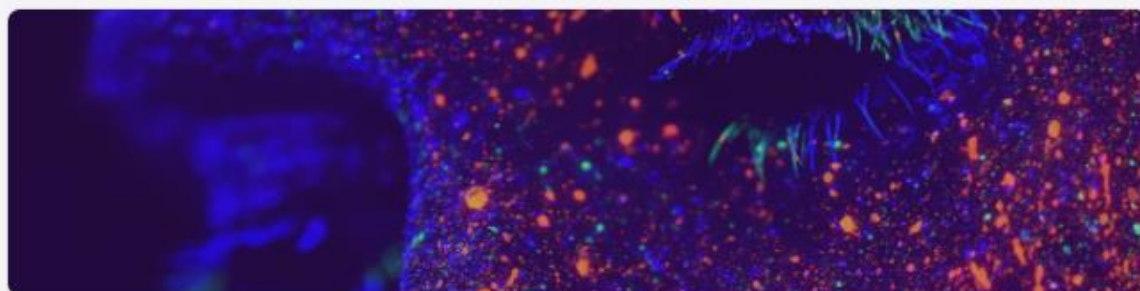
- Müller, A. C. & Guido, S. 2016. Introduction to Machine Learning with Python. O’reilly Media, Inc.
- Ojasalo, K. & Moilanen, T. & Ritalahti, J. 2014. Kehittämistyön menetelmät. Sanoma Pro Oy. Helsinki.
- Reinhardt, A. 2015. Statistics Done Wrong – the woefully complete guide. No Stress Press.
- Rilla, N. & Saarinen, J. 2007. Tutkimusmatka innovaatioihin. TEKES. Teknologia katsaus 197/2007. Luettavissa: <https://docplayer.fi/4608517-Tutkimusmatka-innovaatioihin.html>. Luettu: 15.3.2020
- Rowley, J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information and Communication Science. Vol. 33, Issue 2. Luettavissa: <https://journals.sagepub.com/doi/pdf/10.1177/0165551506070706>. Luettu 30.4.2020
- Technopedia 2020. Data Democratization. Luettavissa: <https://www.techopedia.com/definition/32637/data-democratization>. Luettu: 4.4.2020.
- The Economist 2017. The world’s most valuable resource is no longer oil, but data. May 6th, 2017 edition. Luettavissa: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Luettu: 12.6.2020
- The National Center for Education Statistics 2012a. Luettavissa: [https://nces.ed.gov/ccd/tables/ACGR\\_2010-11\\_to\\_2012-13.asp](https://nces.ed.gov/ccd/tables/ACGR_2010-11_to_2012-13.asp). Luettu: 22.7.2020.
- The National Center for Education Statistics 2012b. Literacy, Numeracy, and Problem Solving in Technology-Rich Environments Among U.S. Adults. Luettavissa: <https://nces.ed.gov/pubs2014/2014008.pdf>. Luettu: 5.7.2020.
- Vitsoe 2020. The power of good design. Dieter Rams’s ideology, engrained within Vitsoe. Luettavissa: <https://www.vitsoe.com/eu/about/good-design>. Luettu: 11.7.2020.
- Ware C. 2020. Information Visualization, perception for design. Elsevier.
- Weinberger, D. 2010. The Problem with the Data-Information-Knowledge-Wisdom Hierarchy Harvard Business Review. February 2010. Luettavissa: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>. Luettu: 16.4.2020

White House Archived 22.1.2015. "Our high school graduation rate has hit an all-time high." —President Obama #LeadOnEducation. Twitter-viesti @ObamaWhiteHouse.  
Luettavissa: <https://twitter.com/ObamaWhiteHouse/status/558318033639198720>. Luettu: 12.4.2020.

## Liitteet

### Liite 1. Kyselylomake

Opiskelijoille tarkoitettu kyselylomake toteutettuna Google Forms- verkkokyselyllä. Rajattu vain opinnäytetyössä käsiteltävään osioon yksilön henkilökohtaisista kyvykkyyksistä. Opiskelijoiden organisaation (IT-infrastruktuuri) ja organisaation prosessien kysymyksiä ei tässä yhteydessä käsitellä, sillä opiskelijat eivät pysty vastaamaan kaikkiin organisaation kyvykkyyksiä selvittäviin kysymyksiin objektiivisesti.



## Data Literacy Assessment

Welcome to Cubiq Analytics Data Literacy Assessment.

This form consists of three parts: student's background information, student's experiences, and student's competence. The purpose of this assessment is to find out about data literacy among individuals. The assessment is completely anonymous and we are not collecting any personal information.

**Data =** In this context data is understood as quantitative information distributed in digital form.

**Data literacy =** Is the ability to derive meaningful information from data, just as literacy in general is the ability to derive information from the written word. "Data literacy is the ability to read, work with, analyze and argue with data regardless of your role, skill level, or the BI tools you use." - Qlik.

The answer options are:

- 1 = Completely disagree
- 2 = Slightly disagree
- 3 = Neither agree or disagree
- 4 = Slightly agree
- 5 = Completely agree

**Background information**

Gender

- Female
- Male
- Prefer not to say

Age

- 18-20
- 21-25
- 26-30
- 31-35
- 36-40
- 41 +

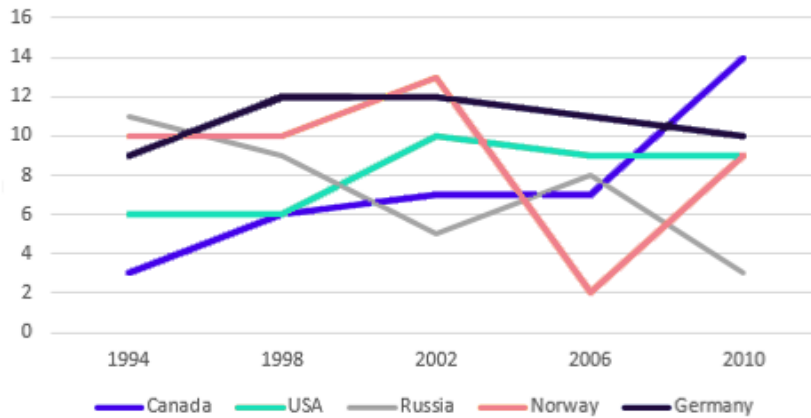
Years at university

- 0-1
- 1-2
- 2-3
- 3-4
- 4-5 or more

Home town (not your current studying city)

Oma vastauksesi \_\_\_\_\_

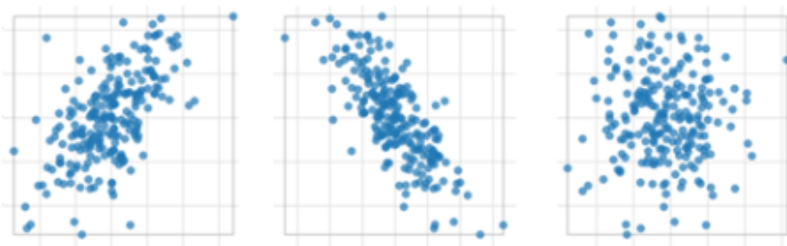
23. According to line chart beneath, which country has the greatest total (1994 – 2010) amount of sales? Sales are in millions.



- A) Canada
- B) USA
- C) Germany
- D) Norway
- E) I don't know

Tyhjennä valinta

24. What is the best option to describe these scenarios from left to right?



- A) Positive correlation, negative correlation, correlation
- B) Positive correlation, negative correlation, no correlation
- C) Negative correlation, positive correlation, no correlation
- D) No correlation, no correlation, correlation
- E) I don't know

25. Here is a list of ages of group of people. Find the mean, median, mode and range for the following values: 7, 7, 7, 10, 11, 12, 16

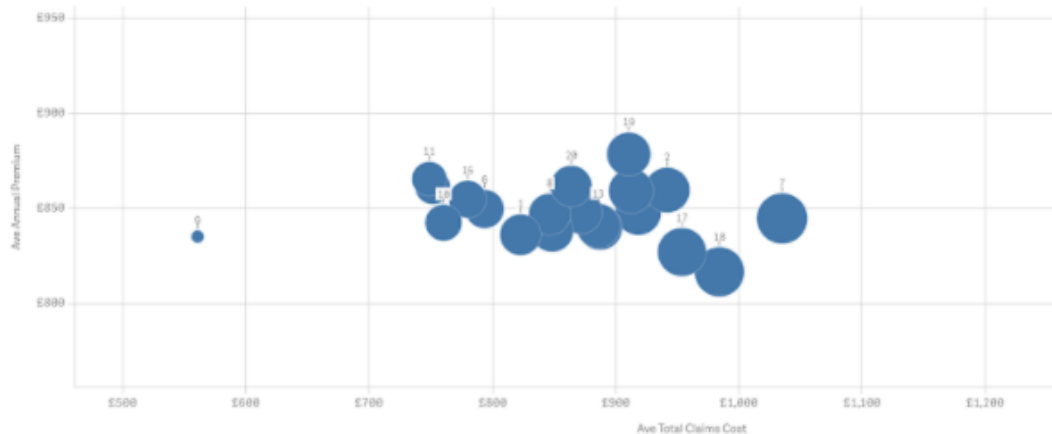
- A) 8, 10, 70, 16
- B) 10, 10, 10, 7
- C) 7, 10, 10, 10
- D) 10, 10, 7, 9
- E) I don't know

26. Correlation always implies causality

- A) Yes
- B) No
- C) I don't know

27. Which outlying Vehicle Insurance Group provides the highest premium for the lowest Claims costs?

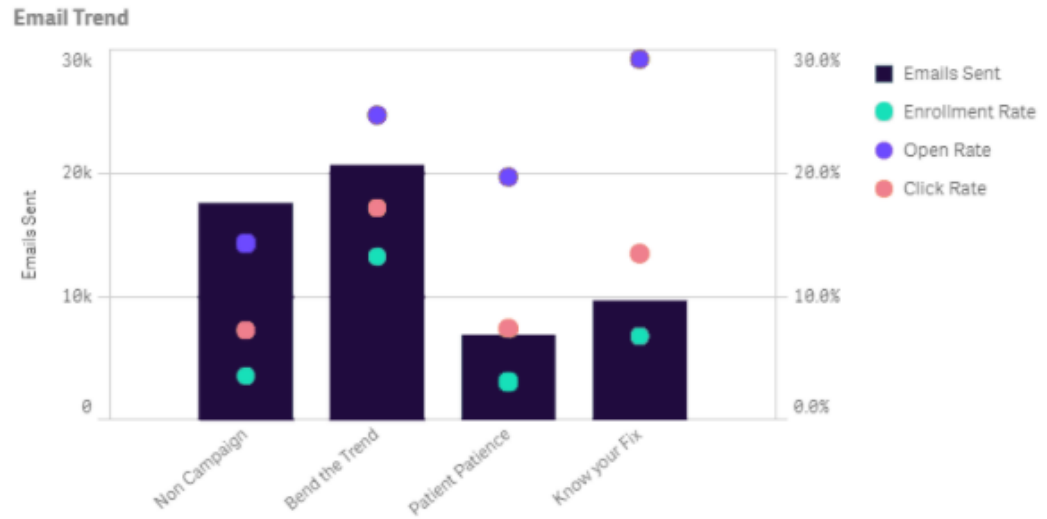
Which Vehicle Insurance Groups provide the highest premium for the lowest claims costs



- A) 18
- B) 9
- C) 7
- D) 19
- E) I don't know

28. Beneath is the Email campaign performance chart. Which campaign has the highest open rate?

### Email Campaign Performance

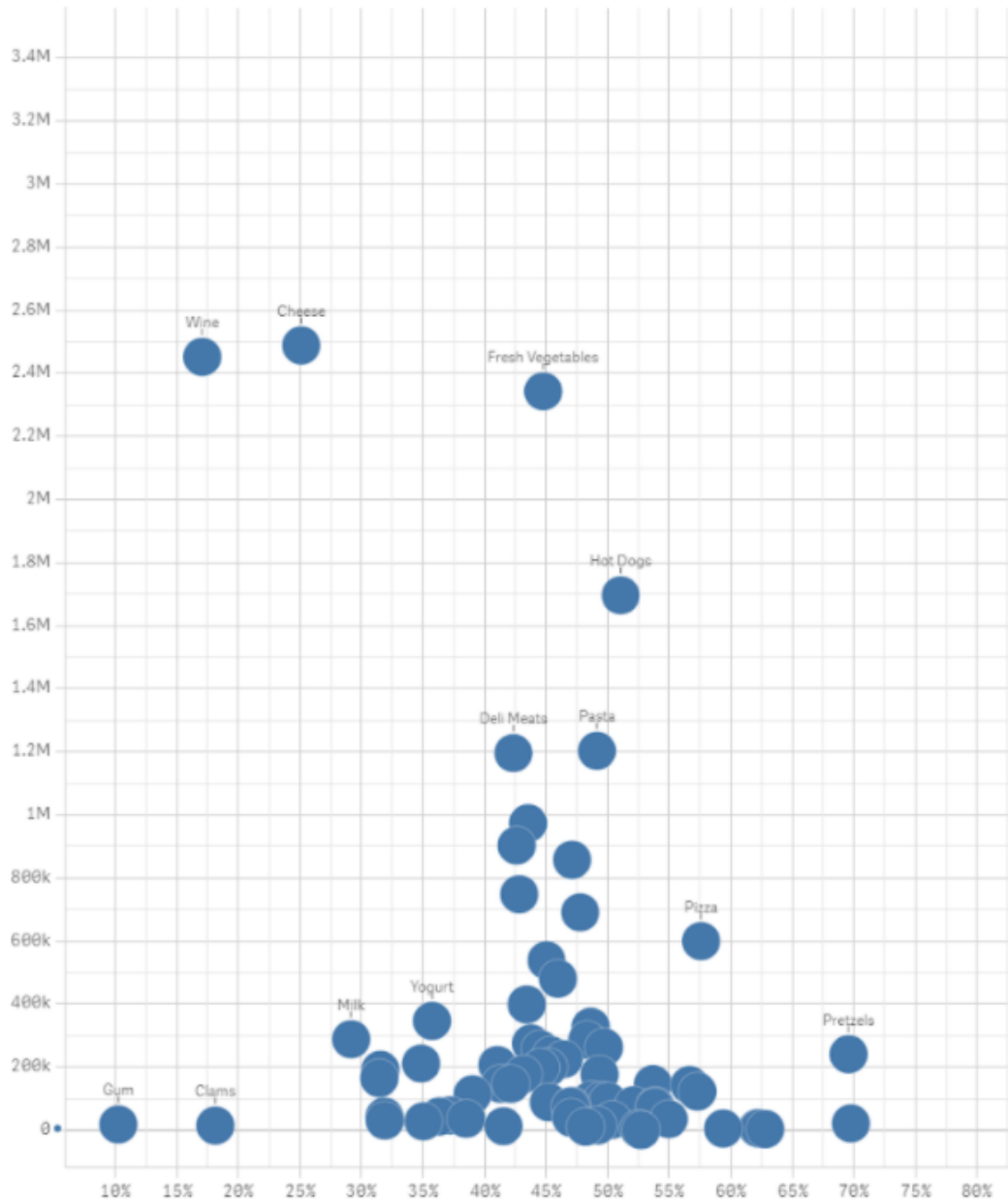


- A) Non Campaign
- B) Bend the Trend
- C) Patient Patience
- D) Know your Fix
- E) I don't know

29. In the Sales & Margin Analysis scatter plot above, the most profitable (in margin \$) product group is

### Sales & Margin Analysis

Western Sales vs Margin (Sales \$23,889,222)



- A) Cheese
- B) Fresh Vegetable
- C) Pizza
- D) Pretzels
- E) I don't know

30. A number is increased by 30 % and then decreased by 30 %. Find the decrease per cent.

- A) 3 %
- B) 9 %
- C) 15 %
- D) It does not decrease
- E) I don't know