

Datan visualisoinnin formaatit UNESCO Youth Newsroomille

Eemeli Martti



Tekijä(t) Eemeli Martti	
Koulutusohjelma Journalismin koulutusohjelma	
Opinnäytetyön otsikko Datan visualisoinnin formaatit UNESCO Youth Newsroomille	Sivu- ja liitesivumäärä 35 + 56
<p>Opinnäytetyöni on produktimuotoinen opinnäytetyö UNESCO Youth Newsroomille, joka tuottaa journalistista sisältöä vuosittain järjestettävästä World Press Freedom Day -sanavapaustapahtumasta. Opinnäytetyön tavoite oli luoda pohja ja edellytykset datan visualisoimiselle Youth Newsroomissa.</p> <p>Olin Youth Newsroomin jäsenenä toukokuussa 2015 Latvian Riikassa järjestetyssä tapahtumassa yhdessä seitsemän muun Haaga-Helian journalismin koulutusohjelman opiskelijan kanssa. Olimme osa toimitusta kun tapahtuma järjestettiin toukokuussa 2016 Helsingissä. Johdimme omia tiimejämme, jotka vastasivat sisällön eri muotojen tuottamisesta.</p> <p>Opinnäytetyön produktiosa on datan visualisoinnin tyylikirja ja opas Youth Newsroomille. Produkti määrittelee, miten Youth Newsroom visualisoi dataa. Produkti on tämän opinnäytetyön liitteenä.</p> <p>Produkti toteutettiin yhden vuoden aikana. Toimeksianto annettiin toukokuussa 2015. Produktin suunnittelu alkoi syksyllä 2015. Produkti työstettiin alkuvuodesta 2016 ja toteutettiin toukokuussa 2016. Raportti on kirjoitettu huhti-toukokuussa 2016.</p> <p>Produkti on syntynyt opinnäytetyön tietoperustassa käsiteltävän lähdemateriaalin ja oman empiirisen tiedon perusteella.</p> <p>Opinnäytetyö etenee niin, että johdannossa kerrotaan toimeksiantaja ja tavoitteet sekä määritellään keskeisimmät käsitteet. Tietoperustassa käsitellään ensin yleisesti datavisualisointeja journalismissa, infografiikan formaatteja, datan erilaisia tyyppisiä ja lopuksi datan esittämistä. Empiiriosassa osassa kerron aluksi produktin suunnitelma- ja toteutuskuvauksen, minkä jälkeen käsitelen tarkemmin sitä, millaista sisältöä Youth Newsroomin dataryhmä tuotti tapahtumassa. Empiiristä osaa seuraa pohdintaosuus, jossa pohdin, miten lopullinen produkti täytti sille asetetut tavoitteensa, miten tulevat toimitukset voivat kehittää sitä, ja mitä opin. Pohdinta-osuuden jälkeen on opinnäytetyön lähteenä käytetty materiaali ja liitteenä Youth Newsroomille tehty produkti.</p> <p>Produktin toteutuksessa ilmeni, että datajournalismia ja datavisualisointeja voidaan tehdä nopealla aikataululla ja tiimillä, joka ryhmäytyy erittäin lyhyessä ajassa. Lisäksi huomattiin, että sanan- ja ilmaisunvapautteen liittyvää dataa on hankalasti saatavasti avoimessa muodossa eikä sitä välttämättä saa edes organisaatioilta pyytämällä.</p>	
Asiasanat Datavisualisointi, datajournalismi, UNESCO, visualisointi,	

Sisällys

1 Johdanto.....	1
1.1 Opinnäytetyön tavoitteet.....	1
1.2 Henkilökohtaiset tavoitteet.....	1
1.3 Opinnäytetyön rakenne.....	2
1.4 Keskeisimmät käsitteet.....	2
2 Datavisualisoinnit ja journalismi.....	5
2.1 Infografiikan formaatit.....	6
2.1.1 Staattinen infografiikka.....	6
2.1.2 Liikkuva infografiikka.....	7
2.1.3 Interaktiivinen infografiikka.....	8
2.2 Visualisoinnin työvaiheet.....	8
2.3 Datan eri muodot.....	10
2.4 Tiedon esittäminen.....	11
2.4.1 Edellytykset onnistuneelle visualisoinnille.....	13
2.4.2 Kuvaajatyypit.....	14
2.4.3 Värien käyttö kuvaajissa.....	18
2.4.4 Typografian rooli visualisoinneissa.....	18
3 Riikasta Helsinkiin: Näin Youth Newsroom visualisoi dataa.....	20
3.1 Produktin tekemiseen käytetyt lähteet.....	20
3.2 Ei pelkästään datan visualisointia.....	21
3.3 Pikemminkin opas kuin tyylikirja.....	23
3.4 Toimeksiannosta tapahtumaan.....	24
3.4.1 Datatiimi aloittaa työskentelynsä.....	25
3.4.2 Mistä data kerättiin?.....	27
3.4.3 Miten dataryhmän artikkelit visualisoitiin?.....	28
3.5 Vuoden kestänyt produktio.....	29
4 Pohdintaa produktin onnistumisesta.....	31
4.1 Produktille asetetut tavoitteet ja miten niissä onnistuttiin.....	31
4.2 Itse prosessi ja sen arviointi.....	31
4.3 Miten produktia voisi kehittää?.....	32
4.4 Produktin toteutuksessa onnistuttiin, mutta produkti olisi voinut parempi.....	32
Lähteet.....	34
Liitteet.....	35

1 Johdanto

Opinnäytetyöni toimeksiantaja on UNESCO, joka on Yhdistyneiden kansakuntien kasvatust-, tiede- ja kulttuurijärjestö. Opinnäytetyöni produkti World Press Freedom Day -nimisestä sananvapaustapahtumasta uutisoivalle Youth Newsroomille, joka toimii UNESCON alaisuudessa. Youth Newsroom koostuu ympäri maailmaa tulevista nuorista toimittajista.

Produktini määrittelee miten, Youth Newsroom visualisoi dataa ja mitä työkaluja se käyttää datavisualisointien tekemiseen. Produktini on käytännössä opas datan visualisoimiseen ja myös visualisointityökalun käyttöön.

Produkti määrittelee, millä tavoin Youth Newsroomin tuottamassa sisällössä visualisoidaan dataa ja mitä työkaluja datan visualisoimiseen käytetään. Produkti määrittelee käytettävät kuvaajatyytit, värit ja työkalut.

1.1 Opinnäytetyön tavoitteet

Olin seitsemän muun Haaga-Helian journalismin koulutusohjelman tavoin Youth Newsroomin jäsen kun World Press Freedom Day järjestettiin Latviassa toukokuussa 2015. Kansainvälinen nuorten toimitus ei silloin tuottanut datavisualisointeja tai tehnyt datajournalismia, vaan pitkälti referoivia artikkeleja, videoita ja kuvia tapahtuman paneelikeskusteluista tai niiden käsittelemistä aiheista. Datajournalismilla ja datavisualisoinneilla toimituksen tuottamasta sisällöstä saadaan monipuolisempaa.

Opinnäytetyön ensimmäinen ja tärkein tavoite on luoda Youth Newsroomille pohja datavisualisointien tekemiselle. Tyylikirjan ja formaattien luomisella tulevaisuuden nuorten toimituksilla on valmiit ohjeet datavisualisointien tekemiseen ja jota joku voi halutessaan lähteä kehittämään.

Koska ei voida tietää, mikä nykyisen tai tulevien nuorten toimituksien osaamistaso on datapohjaisen sisällön saralla, on produktin tarkoitus olla selkeä ja palvella erityisesti niitä, joilla vain vähän osaamista datavisualisointien tekemisestä.

1.2 Henkilökohtaiset tavoitteet

Oppimistavoitteeni opinnäytetyöhön liittyen ovat oman osaamisen ja tietouden syventäminen varsinkin datavisualisointien ja datajournalismin osalta.

sekä toimivan tyylikirjan tekeminen toimitukselle, varsinkin sellaiselle, jolle asia voi olla uusin uusi. Lisäksi tavoitteenani on yhteistyön tekeminen muiden toimittajaopiskelijoiden

kanssa ja toiminta näin isossa produktissa sekä tiimin johtaminen ja toiminnan koordinoiminen. Datapohjaisen sisällön tuottaminen nopealla aikataululla sekä tiedon etsiminen.

1.3 Opinnäytetyön rakenne

Opinnäytetyö etenee niin, että johdannossa kerrotaan toimeksiantaja ja tavoitteet sekä määritellään keskeisimmät käsitteet. Tietoperustassa käsitellään ensin yleisesti datavisuaalisointeja journalismissa, infografiikan formaatteja, datan erilaisia tyyppejä ja lopuksi datan esittämistä. Empiirisessä osassa kerron aluksi produktin suunnitelma- ja toteutuskuvauksen, minkä jälkeen käsittelen tarkemmin sitä, millaista sisältöä Youth Newsroomin dataryhmä tuotti tapahtumassa. Empiiristä osaa seuraa pohdintaosuus, jossa pohdin, miten lopullinen produkti täytti sille asetetut tavoitteensa, miten tulevat toimitukset voivat kehittää sitä, ja mitä opin. Pohdinta-osuuden jälkeen on opinnäytetyön lähteenä käytetty materiaali ja Youth Newsroomille tehty produkti.

1.4 Keskeisimmät käsitteet

Datajournalismi

Datajournalismi on journalismia, jota tehdään datan kanssa. Digitaalisessa maailmassa lähes kaikki voi olla tai on kuvailtavissa numeroin. Datajournalismi eroaa muusta journalismista siinä, että se avaa uusia mahdollisuuksia kun perinteinen ”uutisvainu” ja kyky kertoa tarinoita yhdistetään digitaalisen informaation, joka nyt on saatavilla. (Gray, Bounegru & Chambers, 2.)

Datajournalismi voi auttaa journalisteja kertomaan monimutkaisia tarinoita vetoavilla infografiikoilla. Data voi olla lähde datajournalismille, se voi olla työkalu, jonka kanssa tarina kerrotaan – tai se voi olla molempia. Kuten mihin tahansa lähteeseen, dataan pitäisi suhtautua kriittisesti, ja kuten mihin tahansa työkaluun, meidän pitäisi olla tietoinen siitä kuinka data voi muokata ja rajoittaa tarinoita, jotka niillä luodaan. (Gray, Bounegru & Chambers, 2–3.)

Datavisualisointi

Datavisualisointi on visuaalinen esitys datasta tai tapa visualisoida dataa. Sen yleisimmät muodot käsittävät piirakka-, pylvä- ja viivadiagrammit, mutta dataa voidaan myös visualisoida ainutlaatuisemmilla tavoilla. (Lankow, Ritchie & Crooks, 20.)

Datavisualisointien yleisimpiä käyttötapoja on näyttää, kuinka numeeriset arvot ovat muuttuneet ajan kuluessa. Datavisualisoinnit kuitenkin auttavat lukijoita myös vertailemaan kahta tai useampaa erillistä arvoa ja näyttämään yhteyksiä, virtauksia ja hierarkioita. (Gray, Bounegru & Chambers, 193–196.)

Infografiikka

Infografiikka käyttää visuaalisia merkkejä informaation viestimiseen. Termiä käytetään kuvitetuista pitkistä pystysuuntaisista faktajoukoista. Infografiikka ei kuitenkaan vaadi tiettyä määrää dataa, sen ei tarvitse olla monimutkainen tai sen ei tarvitse esittää tietyn tasoista analyysia. Infografiikka voi olla yksinkertaisimmillaan liikennemerkki tai monimutkainen visuaalinen analyysi. Painetuissa julkaisuissa ja verkkojulkaisuissa käytettäviin infografiikoihin voidaan viitata toimituksellisinä infografiikoina. (Lankow, Ritchie & Crooks, 21.)

Infografiikka esittää informaatiota tilastotieteellisten kuvaajien ja karttojen avulla kun taas informaation visualisointi tarjoaa visuaalisia välineitä, joilla yleisö voi tutkia ja analysoida dataa. Infografiikat kertovat tarinoita ja informaatiovisualisoinnit auttavat lukijaa löytämään tarinoita itse. (Cairo 2013a, xvi.)

Infografiikat ja visualisoinnit voidaan kuitenkin nähdä jatkumona. Jokaisella infografiikalla ja visualisoinnilla on esittävä ja tutkiva osa. Ne esittävät, mutta myös helpottavat sen analysointia, mitä ne näyttävät, eri määrissä. (Cairo 2013a, xvi.)

Infografiikan ja datavisualisointien raja on hämärä, sillä molemmat perustuvat pitkälti samoihin muotoiluperaatteisiin. Infografiikka voidaan nähdä visuaalisena esityksenä, jonka tarkoitus on tehdä havainto, siinä missä datavisualisointi on työkalu datan interaktiiviseen tutkimiseen. (Cairo 2014b)

Infografiikka kertoo tarinan, jonka suunnittelija haluaa selittää, mutta datavisualisointi antaa ihmisten rakentaa omat havainnot. (Cairo 2014b)

Infografiikoiden ja datavisualisointien välillä ei ole kuitenkaan selvää rajaa. Nykyään monet interaktiiviset grafiikat korostavat ensin relevantteja faktoja ja sen jälkeen antavat lukijan tutkia dataa infografiikan takaa. Tällaiset grafiikat ovat samaan aikaan infografiikoita ja datavisualisointeja. Niillä on vähintään kaksi tasoa: esittävä ja tutkiva. (Cairo 2014b)

Visualisointityökalut

Visualisointityökalut ovat ohjelmistoja tai ohjelmointikieliä, joiden avulla data voidaan visualisoida.

Visualisointityökalut voidaan jakaa kahteen ryhmään: niihin, joiden käyttämiseen ei tarvitse ohjelmointitaitoja ja niihin, joiden käyttämiseen tarvitsee. Dataresurssien kasvu on lisännyt helppokäyttöisempien visualisointityökalujen määrää. (Yau 2013a, 278.)

Yksi tällaisista on Infogr.am, joka on selaimella käytettävä data visualisointien tekemiseen tarkoitettu visualisointipalvelu. Infogr.amilla pystytään tekemään visualisointeja ja infografiikoita helposti ja nopeasti. (Infogr.am.)

2 Datavisualisoinnit ja journalismi

Visualisoinnit ovat kehittyneet erinomaiseksi tavaksi kertoa asioita. Uutisorganisaatiot ovat oppimassa tekemään niitä hyvin datajournalismin saralla. (Yau 2013a, 61.)

Toisin kuin muu visuaalinen media, kuten kuvat tai videot, datavisualisoinnin juuret ovat myös mitattavissa faktoissa. Datavisualisoinnit ovat visuaalisesti viehättäviä, mutta ne tulkitaan enemmän järjellä kuin tunteella. Datavisualisoinnit voivat olla tehokkaita niin nopeassa uutisoinnissa kuin pidemmissä artikkeleissa. (Gray, Bounegru & Chambers 2012, 191–192.)

Journalismissa datan esittämiseen on monta tapaa. Raaka data voidaan julkaista jutun yhteydessä tai se voidaan visualisoida. Huono datavisualisointi on kuitenkin huonompi vaihtoehto kuin sen puuttuminen. (Gray ym 2012, 178.)

Visualisoinnit eivät pelkästään taivu hyvin kerrontaan vaan myös kommunikointiin ja ideoiden selventämiseen. Ne auttavat tiivistämään suuren määrän informaatiota kerralla. (Yau, 63.)

Visualisoinneilla on journalismissa monta roolia. Raportointivaiheessa ne auttavat teemojen, kysymysten, poikkeavien havaintojen, tyypillisten esimerkkien ja aukkojen tunnistamisessa. Julkaisuvaiheessa ne voivat havainnollistaa jutussa nostettua huomiota houkuttelevammalla tavalla. Varsinkin interaktiiviset visualisoinnit mahdollistavat datan tutkimisen ja tuottavat läpinäkyvyyttä raportointiin. (Gray ym 2012, 186–187.)

Hyvin suunniteltu datavisualisointi voi antaa katsojalle välittömän ja syvällisen vaikutelman ja menee monimutkaisen tarinan sijaan suoraan asiaan. (Gray ym 2012, 191.)

Tehokas datavisualisointi on riippuvainen hyvästä, puhtaasta, tarkasta ja merkityksellisistä informaatiosta. Datavisualisointi on vain niin hyvä kuin data sen takana. Joskus data yksin ei kerro tarinaa kaikkein vangitsevimalla tavalla. Vaikka yksinkertainen kuvaaja, joka osoittaa trendiä tai tilastokatsausta voi olla hyödyllinen, kerronta, joka liittyy todelliseen elämään voi olla lukijalle välittömämpää ja vaikutuksellisempaa. Visualisointia ei myöskään tule käyttää, kun datapisteitä on vähän, datassa on vain pientä vaihtelua tai siinä ei ole lainkaan selkeää trendiä tai päätelmää. Myöskään spatiaaliset, eli sijaintia koskevat elementit eivät välttämättä aina ole merkityksellisiä tai mukaansatempaavia tai ne harhauttavat huomiota olennaisemmista numeerisista trendeistä, kuten muutos ajan suhteen. Jos data pisteitä on vähän, mutta informaatio voi silti olla hyödyllistä lukijoille, data kannattaa taulukoida, koska se on siisti, helppolukuinen eikä luo epärealistia odotuksia. (Gray ym 2012, 200–201.)

Journalismissa datan pitää olla tarkkaa ja luotettavaa. Data pitää esittää selkeästi, tiiviisti, mutta myös hienosti. Kuvaajan ja grafiikan suunnittelu ei ole pelkästään tilastollista visualisointia, vaan myös sen selittämistä, mitä visualisointi näyttää. Katsojan pitää ymmärtää grafiikka ja sen kautta data. Tärkeät kohdat tai alueet on selitettävä: symbolit ja värit on tarkasti selitettävä. (Yau 2011b, 3-4.)

Sen sijaan, että grafiikka vain sovitettaisiin sen mukaan, miten se näyttää hyvältä, datasta pitäisi puhua sen mukaan, miten se sopii jutun kontekstiin. Lukijat voidaan viedä läpi kirjoittajan analyysin isosta kuvasta yksityiskohtiin ja huomattaviin datapisteisiin tai se voidaan tehdä toisin päin. (Yau 2013a, 263.)

Visualisointien ja tekstien välille pitää luoda jatkuvuutta. Pitää huomioida se, miten yleisö lukee grafiikoita ja kuinka yleisön halutaan lukevan niitä. (Yau 2013a, 263.)

Tekstiä ja grafiikkaa pitää ajatella yhtenä yksikkönä sen sijaan, että ne olisivat erillisiä osia. (Yau 2013a, 264.)

Joskus jatkuvuus on niinkin yksinkertaista kuin sen huomioiminen, mihin grafiikka sijoitetaan leipätekstissä. On hyödyllistä suunnitella kuvaaja niin että se kiteyttää itsensä siltä varalta, että viittaus grafiikkaan on eri sivulla. Lukijat voivat silloin katsoa grafiikkaa suoraan sen sijaan, että he katsoisivat toiselle sivulle, mitä graafi tarkoittaa. Jos kuvaajat asettellaan tekstin sekaan, on kenties järkevämpää selittää, mistä kuvaajassa on kyse leipätekstissä kuin kuvaajassa. Tämä voi lisätä luettavuutta eikä se myöskään riko kokonaisuutta. (Yau 2013a, 263.)

Hyvä tapa lukijoiden ja visualisoinnin yhdistämiseen on lukijoiden päästäminen datan näkemiseen kun se kuuluu heille. Ihmiset käyttävät aikaa datan kanssa kanssakäymiseen kun he jotenkin näkevät itsensä siinä tai voivat yhdistää itsensä tai ympäristönsä siihen. (Yau 2013a, 264.)

Yleisöstä huolimatta, tekijän pitää mennä datan ulkopuolelle ja harkita, mitä se esittää todellisessa maailmassa. (Yau 2013a, 273.)

2.1 Infografiikan formaatit

Infografiikka voidaan formaateiltaan jakaa staattiseen kuvaan, interaktiiviseen käyttöliittymään ja liikkuva sisältöön. Paras formaatti määrittyy sen perusteella, kuinka tehokkaasti se sisältää ja välittää informaatiota. (Lankow ym 2012, 59.)

2.1.1 Staattinen infografiikka

Staattinen infografiikka on informaatiomuotoilun kaikkein hallitsevin muoto. Staattista infografiikka käytetään yleisimmin kuvana printissä ja websivuilla. Staattinen infografiikka on

erittäin tehokas rikkaan datan esittämisessä yhdessä kuvassa. Sen etuja on, että se on verrattain helppo ja halpa tehdä verrattuna interaktiiviseen ja liikkuvaan infografiikkaan. Staattisia infografiikkoja on myös helppo jakaa. Niitä on helppo liittää esimerkiksi blogeihin. (Lankow, Ritchie & Crooks ym 2012, 60.)

Staattisen infografiikan haasteita on, että informaatio vanhentuu. Tosin infografiikka, joka visualisoi säännöllisesti päivittyvää datasettiä, pitää kuitenkin arvonsa, koska se viittaa dataan tietyllä aikahetkellä. Infografiikka pitää kuitenkin päivittää niille, jotka etsivät kaikkein uusinta tietoa. (Lankow ym 2012, 60.)

Yksi staattisen infografiikan parhaista käytöistä on näyttää ikivihreää sisältöä, joka pysyy relevanttina ilman taajaa tarvetta päivittää informaation päivittämiseen. (Lankow ym 2012, 62.)

Staattisia kuvia voidaan käyttää laajasti. Ne ovat kolmesta formaatista yksinkertaisin ja ehkä myös halvin toteuttaa, ja ne myös yleensä ovat kaikkein monipuolisimpia. Staattisten infografiikoiden tekeminen on nopeampaa kuin interaktiivisten ja liikkuvien infografiikoiden, minkä takia staattiset kuvat ovat optimaalisia esimerkiksi nopeampaan uutistyyöhön. (Lankow ym 2012, 59.)

2.1.2 Liikkuva infografiikka

Liikkuva grafiikka tiiviillä käsikirjoituksella, hyvin ajoitetulla animoinnilla ja selkeillä selityksillä palvelee monimutkaisten lukujen tai ideoiden herättämistä eloon. Liikkuvalla grafiikalla ja animoiduilla pätkillä voidaan vahvistaa sitä, mitä yleisö kuulee. Kun taustaselostusta käytetään selittävien graafikoiden kanssa, saadaan aikaan voimakas ja muistettava tapa kertoa tarinoita. (Gray ym 2012, 178.)

Liikkuvalla infografiikalla on kyky sitouttaa ihmisiä erilaisella tavalla kuin staattisella tai interaktiivisella infografiikalla. Taustaselostuksen kanssa asiat voidaan esittää lineaarisella tyylillä. Tällöin ihmisten ei tarvitse aktiivisesti valita sitoutuvatko he informaation kanssa hetki hetkeltä, kuten he tekisivät staattisen infografiikan kanssa. Liikkuvalla infografiikalla on kyky vedota katsojaan tunteellisesti esimerkiksi musiikin kautta sekä taustaselostuksella ja liikkuvilla kuvilla informoimalla tarjoaa mahdollisuuden viestiä voimakkaalla tavalla. (Lankow ym 2012, 74.)

Toinen ero liikkuvan ja staattisen sisällön välillä on aika. Staattinen sisältö on olemassa pysyvä ajassa ja visuaalinen sisältö ei muutu. Liikkuva sisältö on olemassa yli ajan. Siinä missä staattinen sisältö on tilannekatsauksia, liikkuvan sisällön palasilla on oma elämänsä ja liikkeet. Liikkuva grafiikka on perinteisesti palvellut kertovaa funktiota rajoitetulla ja enimmäkseen passiivisella käyttäjäinteraktiolla. (Lankow ym 2012, 75.)

Liikkuva grafiikka on parhaimmillaan kun tehtävänä on kommunikoida yhdellä lineaarisella tarinalla ja luoda visuaalinen ja tunteellinen vetovoima, joka sitouttaa katsojan monella tasolla. (Lankow ym 2012, 75.)

2.1.3 Interaktiivinen infografiikka

Jotkut kertomukset ovat parempia, kun ne esitetään interaktiivisessa muodossa, joka mahdollistaa vaihtoehdon pitkälle, vertikaalisesti orientoituneelle infografiikalle tai tarinalle, jossa datalla on monta tasoa: esimerkiksi kartta, jonka takana olevalla datalla on monta muuttujaa. (Lankow ym 2012, 82–83.)

Interaktiivisen infografiikan käyttäjäinteraktio koostuu klikkaamisesta, datan etsimisestä, aktiivisesta näytettävän sisällön muotoilusta ja sen valitsemisesta, mihin informaatioon on pääsy ja mitä visualisoidaan. Interaktiivinen infografiikka voi olla kertovaa, tutkivaa tai molempia. (Lankow ym 2012, 59.)

Formaatti on erityisen hyödyllinen, jos visualisoitavana on suuri määrä dataa ja käyttäjä halutaan rohkaista tutkimaan sitä. (Lankow ym 2012, 82).

2.2 Visualisoinnin työvaiheet

Visualisoinnin työvaiheet voidaan jakaa neljään osaan: tutkimiseen, muokkaamiseen, visualisoinnin tekemiseen ja arviointiin. Tutkimisvaiheessa data kerätään ajan tasalla olevista lähteistä, esimerkiksi viranomaislähteistä. Jos datan käyttämiseen tarvitaan lupa, se pitää hankkia. Muokausvaiheessa pitää tunnistaa datan avainviesti. Sen jälkeen valitaan paras datasarja havaintojen visualisoimiseen. Dataa voidaan yksinkertaistaa ja suodattaa, jotta yleisölle saadaan välitettyä tärkein viesti. Tekemisvaiheessa valitaan oikea kuvaaja, jolle annetaan esimerkiksi otsikko ja kuvaus. Värejä ja typografiaa voidaan käyttää avainviestin korostamiseen. Arviointivaiheessa visualisoitu data tarkistetaan lähteistä. Kuvaajan järjestyttä tulee arvioida harkiten ja sitä katsotaan myös lukijan näkökulmasta. Kyseenalainen sisältö ja poikkeamat voidaan tarkistaa muista lähteistä ja tai asiantuntijoilta (Wong 2013, 20–21.)

Data on vain yhtä hyvä kuin sen lähde. On tärkeää saada data hyvämaineisesta ja puolueettomasta lähteestä. Dataan pitää aina suhtautua kriittisesti, koska yksi väärä datapiste voi tuhota kuvaajan luotettavuuden. Jokainen datasetti tarvitsee itsenäisen analyysin ja tulkinnan. Koska saman datan visualisoimiseen on monta tapaa, lopullinen visualisointitapa riippuu siitä, mikä on aiottu viesti. (Wong 2013, 26).

Tutkimusvaiheessa iso datasetti mahdollistaa suuremman syvällisen analyysin. Editointivaiheessa on tärkeää arvioida, hautaako kaikki lisäinformaatio tarinan keskeisimmän huomion vai laajentaako se sitä tehden jutusta vakuuttavamman. Rikas data on laadukasta ja

tarkkaa data, joka on kerätty luotettavista lähteistä yhdistettynä tehokkaaseen datan siivoamiseen yleisölle. Esityksessä joskus vähemmän on enemmän (Wong 2013, 28).

Nathan Yau jakaa (Yau 2013a, 137) visualisointiprosessin neljään kysymykseen:

1. Millaista dataa sinulla on?
2. Mitä haluat tietää datastasi?
3. Mitä visualisointitapaa sinun pitäisi käyttää?
4. Mitä näet ja käykö se järkeen?

Vastaus jokaiseen kysymykseen riippuu siitä, miten aikaisempaan kysymykseen on vastattu. Esimerkiksi, jos data-aineisto käsittää vain kourallisen havaintoja, rajoittaa se sitä, mitä datasta voidaan löytää ja mitkä visualisointitavat ovat hyödyllisiä. Toisaalta, jos käsissä on paljon dataa, visualisoinnin jälkeen voidaan havaita jokin uusi kiinnostava ulottuvuus, mikä johtaa uuteen grafiikkaan. (Yau 2013a, 137).

Millaista dataa sinulla on?

Ihmiset yleensä muodostavat kuvan siitä, miltä visualisoinnin pitäisi näyttää tai löytävät esimerkin, jota haluavat jäljitellä. Tällainen menettelytapa saattaa johtaa siihen, että käytettävä data ei ole riittävä tai se ei vain toimi haluttavan kuvaajatyypin kanssa. Sen sijaan pitäisi toimia päinvastoin: data ensin ja sen jälkeen visualisointi. Datan kerääminen on vaikein ja aikaa vievin osuus visualisointiprosessissa. (Yau 2013a, 137–138).

Mitä haluat tietää datasta?

Datan kanssa työskenteleminen on helppoa, jos sitä on vähän. Kun kyseessä on aineisto, jossa on tuhansia tai miljoonia havaintoja, voi olla haastavaa päättää, mistä aloittaa. Kysymyksien esittäminen datalle, auttaa tekijää löytämään aloituspaikan työlleen. (Yau 2013a, 138–139).

Mitä visualisointitapaa sinun pitäisi käyttää?

Datan visualisoimiseen voi olla monia eri vaihtoehtoja. Tutkimisen aikaisessa vaiheessa on tärkeää, että dataa katsotaan monesta eri näkökulmasta erilaisilla kuvaajilla. Tämä vaihe käsittää myös visuaaliset kokeilut, kuten erilaisten skaalojen, värien, muotojen, kokojen ja geometrioiden käytön. (Yau 2013a, 139–141).

Mitä näet ja käykö se järkeen?

Datan visualisoinnin jälkeen on tiettyjä asioita, joita pitää katsoa: kasvut, laskut ja poikkeamat. Havainnoissa voi olla eroja ihmisen tai koneen tekemän virheen vuoksi, koska tietyt arvot ovat epävarmasti arvioituja tai koska aineistossa on tapaus, joka eroaa muista.

Yaun mielestä on tärkeää, että kun jotain kiinnostavaa löytyy, pitää kysyä, käykö se järkeen ja miksi? Virheellisesti voidaan automaattisesti olettaa, että data on faktaa, koska numerot eivät voi mitenkään olla väärässä. (Yau 2013a, 142).

2.3 Datan eri muodot

Dataa voidaan luokitella eri tavoin. Nathan Yau jakaa datan viiteen eri tyyppiin: kategoriiseen dataan, aikasarjoihin, spatiaaliseen dataan, dataan, jossa on monta muuttujaa, ja jakautumiin. (Yau 2013a, 143–199) .

Lankow, Ritchie ja Crooks (Lankow ym 2012, 207) taas jakavat kvantitatiivisen eli määrällisen datan kuuteen eri tyyppiin, joita ovat diskreetti-, kategorinen, nominaalinen, ordinaalinen, aikavälillinen ja jatkuva data.

Luokittelu ja kategorisointi tuovat dataan rakennetta, joka muuten saattaisi olla muodoton. (Yau 2013a, 143.)

Kun kategoriat laitetaan yhteen, niiden summa on yhtä suuri kuin kokonaisuus. Kategorioiden näkeminen yhtenä yksikkönä voi olla eduksi, jos halutaan nähdä esimerkiksi jakauma. (Yau 2013a, 147).

Kategorisoidulla datalla voi olla myös alakategorioita eli kategorioita kategorioiden sisällä, jotka ovat yleensä kertovat enemmän kuin pääasialliset kategoriat. (Yau 2013a, 150.)

Kategorisoidusta datasta etsitään yleensä ensin maksimi- ja minimiarvot, joka antaa käsityksen datan vaihteluvälistä, ja joka myös löytyy arvojen nopealla lajittelulla. Tämän jälkeen voidaan etsiä osien jakaumat, ovatko arvot enimmäkseen korkeita, matalia vai jotain siitä väliltä. Viimeiseksi etsitään muotoja rakenteesta. Jos kategorioilla voi olla sama arvo tai ne eroavat suuresti muista, on syytä kysyä, miksi ja mikä tekee kategorioista samanlaisia tai erilaisia. (Yau 2013a, 152–153).

Aikasarjoja visualisoidessa, tavoitteena on nähdä, mitä on kulunut, mikä on muuttunut ja mikä on pysynyt samana, ja kuinka paljon. Mitkä syyt vaikuttavat kasvuun, laskuun, tai muuttumattomuuteen? Onko toistuvuutta, ja onko se hyvää vai huonoa, odotettua vai odottamatonta? Yleensä muutoksen suuruus aikasegmenttien välillä on kiinnostavampaa kuin arvo tietyssä pisteessä. (Yau 2013a, 154).

Spatiaalinen data muistuttaa paljon kategorisoitua dataa, mutta siinä on geograafinen komponentti. Globaali data on yleensä kategorisoitu maittain ja kansallinen data esimerkiksi osavaltioittain. Spatiaalisesta datasta pitäisi etsiä sitä, keskittyvätkö korkeat ja matalat arvot tietyille alueille. (Yau 2013a, 176).

Data on yleensä taulukoidussa muodossa, jossa on useampi sarake, jotka kukin esittävät eri muuttujaa. Jotkin visualisointitavat pystyvät visualisoimaan tällaisen datan yhdessä näkymässä, jolloin muuttujien väliset suhteet, sekä yksittäisten muuttujien trendit näkyvät. Yleensä muuttujien väliset suhteet eivät kuitenkaan ole suoraviivaisia. Esimerkiksi selvää trendiä ei löydy. (Yau 2013a, 176–177).

Kun datassa on enemmän kuin yksi muuttuja, tarkastellaan sitä, miten muuttujat muuttuvat suhteessa toisiinsa. (Yau 2013a, 177).

Tilastollisia suhteita muuttujien välillä kutsutaan korrelaatioksi. Kun yhden muuttujan arvo nousee, toisen muuttujan arvo nousee myös. (Yau 2013a, 177.)

Korrelaatiota ei tule sekoittaa kausaatioon, eli syy-seuraussuhteeseen. (Yau 2013a, 79).

Jakautumilla tarkoitetaan sitä, miten eri datassa olevat havainnot jakautuvat ja miten ne poikkeavat esimerkiksi havaintojen keskiarvosta tai mediaanista. Keskiarvoa ja mediaani käytetään kuvailemaan, mikä on normaalia jollekin joukolle. Kaikki mikä poikkeaa joukon keskiarvosta tai mediaanista on poikkeavaa ja alle tai yli keskiarvon. Tietääkseen, mikä on suuri tai pieni poikkeama keskiarvosta tai mediaanista, on tiedettävä datan laajuus. (Yau 2013a, 189.)

Kun tutkitaan jakaumia, tulee etsiä maksimi- ja minimipisteitä, vaihteluväliä ja laajuutta, mitkä kertovat enemmän kuin mediaani tai keskiarvo. (Yau 2013a, 199).

Vajavainen data

Jos data on vajavaista, ja tehtävänä on näyttää laaja trendi ja datasta puuttuu vain muutama datapiste, datan visualisointi voi silti olla hyödyllistä. Pitää kuitenkin arvioida, esittääkö kuvaaja silloin kokonaisuudessaan vääristymättömän tarinan. On hyväksyttävää täydentää dataa yhdistelemällä sitä eri lähteistä, jos lähteet keräävät dataa samalla tavalla. (Wong 2013, 124.)

Useimmissa tapauksissa kuvaaja, josta puuttuu yksi tai kaksi datapistettä, on arvokas paitsi, jos puuttuvat datapisteet ovat välttämättömiä esitettävälle viestille. Pylväs- ja viivakuviota käytettäessä datan voi visualisoida, jos siitä puuttuu vähemmän kuin viidesosa datapisteistä. Piirakkakuviota ei voida koskaan käyttää, jos yksi sen segmenteistä ei ole saatavilla, koska piirakkakuviokuva kuvastaa aina 100 prosenttia. (Wong 2013, 124–125.)

2.4 Tiedon esittäminen

Kerätyn tilastotiedon hyödyllisyys riippuu esitystavasta. Tilastotieto voidaan esittää kirjoitettuna tekstinä, taulukkona tai kuvaajana. (Kuusela 2000, 8–9.)

Kokonaisuudet sidotaan yhteen tekstillä, jolla myös valotetaan taustoja sekä arvioidaan tiedon merkitystä. Taulukko soveltuu massatiedon esittämiseen ja kuvion rooli on havainnollistaa, tutkia ja samalla painottaa keskeisiä ja merkittäviä tuloksia, mutta ensisijaisesti välittää tietoa. (Kuusela 2000, 14–15.)

Tilastokuvion on ominaista, että se välittää tietoa nopeasti ja havainnollisesti, mutta epä-tarkasti. Verrattuna tekstiin, se kuitenkin kertoo enemmän ja tieto säilyy muistissa pidempään. (Kuusela 2000, 10.)

Kuvaaja ilmaisee informaatiota lukijalle visuaalisesti ja narratiivisesti. Kuvaajassa numerot näyttävät konkreettisimmilta ja tarkemmilta kuin jos ne olisi kirjoitettu tekstiin, otsikkoon tai kuvatekstiin. Kuvaaja on muistettavampi kuin pätkä numeroita tekstin joukossa. Se myös mahdollistaa sen, että lukija pystyy tekemään arvion vilkaisulla. (Wong 2013, 22.)

Paras kuvaaja on sellainen, joka on vapaa kaikista häiriötekijöistä ja antaa lukijan verrata tai rinnastaa dataa ja tehdä päätelmiä. Puhdas ja terävä kuvaaja antaa lukijan keskittyä dataan. (Wong 2013, 22.)

Koristeet eivät ole vaihtoehto datan esittämiseksi ja järjestämiseksi oikealla tavalla. Jos data on riittämätön tarinan kertomiseen, koristeellisia elementtejä ei tulisi lisätä, vaan sen sijaan dataa pitäisi kerätä muista lähteistä. (Wong 2013, 23.)

Yksi datasarja voidaan visualisoida monella tapaa. Visualisoinnissa tulisi käyttää oikeaa kontekstia oikean viestin kommunikointiin. Kuvaajan viestin pitäisi olla johdonmukainen kaikkine saatavine faktoineen ja todisteineen. (Wong 2013, 25.)

Taulukko on tekstiä ja tilastokuvion objektiivisempi, koska taulukon laatimiseen on vähemmän vaihtoehtoja. Tekstin ja kuvion laatijalla on enemmän mahdollisuuksia esityksen muuttamiseen ilman, että totuus muuttuu. (Kuusela 2000, 11.)

Taulukon tehtävä on antaa objektiivinen tieto asioista ja materiaalia jatkoanalyysiin. Se antaa perustiedot ilmiöstä ja sen seurauksena lukijalle perusteet johtopäätöksille ja mahdollisuuden muiden tekemien johtopäätöksien arvioinnille. Suurta tietomassaa ei tule esittää kuin taulukossa, jossa data voidaan esittää järjestetyssä ja hallittavassa muodossa toisin kuin tekstissä tai kuviossa. (Kuusela 2000, 11–12.)

Tekstiä käyttämällä kirjoittajalla on vapaus tilastollisen tiedon esittämiseen ja laajojen yhteyksien esittämiseen. Teksti eroaa kuviosta ja taulukosta siinä, että lukijalla on vähemmän mahdollisuuksia arvioida tilannetta, mikä on myös tekstin ongelma, sillä kuviosta ja taulukosta lukija voi itse todeta asian. Tekstin avulla kirjoittaja voi käyttää johdattelevia ja tunnepitoisia ilmaisuja sekä adjektiiveja. (Kuusela 2000, 13.)

Luotettavuus rakentuu, kun faktat esitetään reilusti. Prosenttilukujen esittäminen ilman kantalukuja on merkityksetöntä. Isoja muutoksien esittämistä kannattaa välttää, jos muutos perustuu pieniin lukuihin. (Wong 2013, 27.)

On hyväksyttävää eriyttää muutama luku sarjasta, jos nämä datapisteet kertovat tarinan ilman lukijan harhaanjohtamista väriin oletuksiin menneestä ja tulevasta. Datan paljous ei tarkoita datan rikkautta. Suuren datapistemäärän visualisointi ei ole välttämättä hyvä. Sarja datapisteitä on tarkoituksellista ja merkittävää, jos se indikoi muutosta kantalinjan muotoon. (Wong 2013, 29.)

2.4.1 Edellytykset onnistuneelle visualisoinnille

Tilastokuvioita voidaan arvioida erilaisilla kriteereillä. Näitä ovat esimerkiksi tehokkuus, tietotiheys, kuvioroinan määrä ja visuaalinen vertailtavuus. Tehokkuutta mitataan siten, kuinka paljon silmänliikkeitä tarvitaan kuvion sanoman hahmottamiseen. Tietotiheyttä mitataan indeksillä, joka saadaan laskemalla kuviossa olevien erillisten lukujen määrä ja jakamalla se kuvion pinta-alalla. Kuvioroinaksi lasketaan sellaiset elementit kuvaajassa, jotka eivät välitä tietoa, kuten koristeet. Visuaalisella vertailtavuudella tarkoitetaan kuvioalkioiden kooka suhteutettuna toisiinsa samoin kuin luvut, joita ne esittävät. (Kuusela 2000, 23–24)

Edward Tufte on esittänyt yhdeksän kriteeriä, joita pitäisi vaatia hyvältä tilastokuvioilta. Kaikkia kriteerejä ei tosin välttämättä pystytä soveltamaan kaikissa kuvioissa. (Kuusela 2000, 24.)

Tuften kriteerit onnistuneelle tilastokuvioille ovat:

- Tiedon esittäminen
- Katsojan houkuttelemine ajattelemaan esitettävää asiaa
- Tietoihin sisältyvää sanomaa ei saada vääristöä
- Lukuja pitää esittää paljon pienessä tilassa
- Suureen tietojen joukkoon liittyvän sanoman tiivistäminen
- Vertailujen tekemisen rohkaisu
- Tietoihin liittyvän asian esiintuominen monella eri tasolla
- Tietyn päämäärän kohtuullisen selvä palvelu
- Kiinteän kokonaisuuden muodostaminen muun taulukoissa ja tekstissä olevan aineiston kuvauksen kanssa.

(Kuusela 2000, 24–25.)

Kolme hyvän infografiikan elementtiä ovat rikas sisältö, joka tuo tarkoitusta grafiikkaan, kutsuva visualisointi, joka tulkitsee sisältöä ja korostaa informaation tärkeyden lukijalle, ja hienostunut toteutus, joka herättää sisällön ja grafiikan eloon. (Wong 2013, 14.)

Kirjaisimet, värit, design ja kuinka kriittisen analyysin syvyys näytetään, tekevät kuvaaja-
ta tehokkaan. (Wong 2013, 19).

2.4.2 Kuvaajatyypit

Viiva-, pystypylväs-, vaakapylväs- ja piirakkakuviot ja niiden muunnelmät ovat yleisimmin
käytettyjä kuviotyyppejä. (Kuusela 2000, 49).

Parvikuvio

Parvi- eli pistekuvio on yksinkertainen tilastokuvio, jolla on helppo todeta riippuvuuksia ja
niiden voimakkuuksia sekä niiden laadun. (Kuusela 2000, 70). Parvikuviota voidaan käyt-
tää osoittamaan, miten ilmiöt kytkeytyvät toisiinsa. (Kuusela 2000, 71).

Viivakuvio

Viivadiagrammi voi näyttää suuren määrän dataa erittäin pienessä tilassa. Viivadiagrammi
on parhaimmillaan kun visualisoidaan aikasarjoja ja se soveltuu hyvin trendien, kasvun,
laskun tai volatiliteetin visualisoimiseen. (Wong 2013, 50.)

Viivakuvio kuvaa muutosta tai kehitystä tai niiden puutetta. Se korostaa kehityssuuntaa ja
vaihtelua sen suhteen. Jos kuviossa on useampi viiva, nähdään siitä myös kehityssuun-
tien erot. (Kuusela 2000, 78.)

Viivakuvion perusvaatimus on, että molempien asteikkojen on oltava tasavälisiä eikä niitä
saa typistää keskeltä. (Kuusela 2000, 81). Viivakuvioon vaikuttaa eniten x- ja y-akselin vä-
linen suhde, eli aspektisuhde. Sitä muuttamalla voidaan muuntaa viivan kehitystä. (Kuuse-
la 2000, 90.) Perusohje oikealla aspektisuhteelle on, että tasaista muutosta kuvataan 45
asteen kulmassa kulkevalla viivalla. (Kuusela 2000, 91).

Yleisimpiä virheitä viivakuvion tekemisessä on, että x-akselilla kuvattava muuttuja on epä-
jatkuva, samassa kuviossa on liian monta viivaa, jolloin yksittäisiä viivoja ei erota, eri vii-
voilla on eri skaalat, käytetty asteikko ei ole tasavälinen tai x- ja y-akselin suhde vääristää
kuvaajaa. (Kuusela 2000, 101–102).

Viivadiagrammiin ei saisi lisätä enempää kuin neljää viivaa. Muuten kuvaaja on hämmen-
tävä, eikä enää palvele tarkoitusta. (Wong 2013, 54–55.)

Pysty- ja vaakapylväskuvio

Pylväskuviot soveltuvat määrän esittämiseen, mutta toisin kuin viivakuviossa, pylväsku-
viossa trendi tai vaihtelu ei nouse niin voimakkaasti esille. Vaaka- ja pystypylväskuviot
ovat kaksi eri kuviotyyppiä, vaikka ne rakenteellisesti ja alatyypeiltään ovatkin samanlai-

sia. Yksinkertaiset pylväät yksin, ryhmissä tai pinottuina soveltuvat hyvin määrän kuvaamiseen. (Kuusela 2000, 108–109.)

Vaakasuuntaisessa pylväskuviossa pylväitä ei tule järjestää satunnaiseen järjestykseen vaan ne tulee järjestää esimerkiksi pienimmästä suurimpaan tai päinvastoin. Poikkeus sääntöön on, jos pylväät halutaan tai pitää järjestää esimerkiksi aakkosjärjestykseen, jos se on välttämätöntä lukemisen helpottamiseksi. Jos vaakapylväskuviota käytetään aikasarjojen kuvaamiseen, tulee tuorein datapiste olla ylimpänä. (Wong 2013, 70–71.)

Pystypylväskuvio soveltuu yhden asian määrän ja sen vaihtelun tai erityisesti aikasarjojen esittämiseen. (Kuusela 2000, 116). Vaakapylväskuviota voidaan taas käyttää yhden asian, eri luokkien tai osien määrän tai suuruuden vertailuun. (Kuusela 2000, 129).

Ryhmitetyssä pylväskuviossa voidaan esittää esimerkiksi useampia aikasarjoja samassa kuviossa. Ryhmitetyissä pystypylväskuvioissa pylväiden on oltava samassa – ja mieluiten suuruusjärjestyksessä. Pylväiden määrä on hyvä pitää enintään kolmessa, sillä muuten kuvioista voi tulla liian sekava (Kuusela 2000, 117–118.)

Ryhmitetyissä pylväsdiagrammeissa tulisi käyttää enintään vain neljää kategoriaa. Pylväiden varjostukset tulee määrittää vaaleimmasta tummimpaan, jotta lukija pystyy helposti vertailemaan dataa. (Wong 2013, 67.)

Summapylväskuviolla esitetään osasummia ja kokonaismäärän kehitystä samassa kuviossa. Pylväistä voidaan tehdä samanmittaisia, jolloin ne kuvaavat suhteellisia määriä. Kuvioita voidaan käyttää piirakkakuvion sijaan. (Kuusela 2000, 119–120.)

Tyypillisimpiä virheitä pylväskuvioiden tekemisessä ovat määräästeikon katkaiseminen, liian monen pylvään sijoittaminen pylväsryhmään, pylväiden sijoittaminen liian kauas tai lähelle toisiaan, pylväiden värittäminen esimerkiksi liian erilaisilla väreillä tai summapylväiden tekeminen niin, ettei pylväiden osien summa ole mielekäs. (Kuusela 2000, 143–144.)

Piirakkakuvio

Piirakkakuvio on yksi käytetyimmistä kuvaajatyypeistä. Piirakkakuvio kuvaa, kuinka monen osaan jokin asia on jakautunut, eikä voi esittää muuta kuin prosenttijakaumaa. (Kuusela 2000, 145–146.)

Sektorit voidaan sijoittaa ympyrään eri tavoin, mutta ne on sijoitettava aina suuruusjärjestykseen niin, että suurin sektori on ensimmäisenä, koska lopputulos on tällöin havainnollisempi. (Kuusela 2000, 146).

Kuvioon voidaan sijoittaa kokonaismäärää esittäviä lukuja. On suotavaa, että sektorien määrä on enimmillään kuusi, sillä muuten eri sektorien vertailu on vaikeaa. Jos esitettäviä luokkia on enemmän kuin kuusi, voi luokkia yhdistellä. (Kuusela 2000, 148.)

Piirakkakuviota ei tule käyttää monimutkaisten suhteiden esittämiseen. On helpompaa vertailla kahta pylvästä kuin kahta piirakkakuvion sektoria. On intuitiivista lukea piirakkakuviota myötäpäivään, siksi piirakkakuvion segmenttejä ei tulekaan koskaan sijoittaa myötäpäivään pienimmästä suurimpaan. Tällöin vähiten tärkeä segmentti on kaikkein merkittävimmässä paikassa. (Wong 2013, 74–75.)

Piirakkakuvion segmenttejä on vaikea vertailla, joten piirakkakuviossa ei tule olla enempää kuin 5-6 segmenttiä. Jos segmenttejä on enemmän kuin viisi, tulee pienimmät ja vähiten merkittävimmät segmentit yhdistää ja nimetä esimerkiksi ”muut”. Jos kaikki segmentit on esitettävä erillään, on pinottu pylväskuvio parempi vaihtoehto. (Wong 2013, 74–75.)

Teemakartat

Kartat ovat journalismissa yksi yleisimmin käytetyistä kuvaajatyypeistä. Varsinkin virtaus- ja koropleettikartat ovat hyödyllisiä työkaluja journalisteille. Aika, määrä ja maantieteellisyys ovat yleisiä kartoille. Ihmiset haluavat aina tietää, kuinka paljon yksi alue on verrattuna toiseen ja kuinka data virtaa alueelta toiselle. (Gray ym 2012, 205–206.)

Käytännössä, mikä tahansa, joka voidaan liittää paikkaan, on esitettävissä kartalla. Teemakartat ovat väline alueellisen tiedon ja yhteyksien kuvaamiseen. Teemakartalla voidaan esittää maantieteellisiä riippuvuuksia, joiden ilmaiseminen muuten voi olla vaikeaa tai epähavainnollisempaa. (Kuusela 2000, 154–155.)

Karttaa ei tule kuitenkaan käyttää, ellei maantieteellisyys ole relevanttia. Karttaa ei tule käyttää paljouden tai määrän vertailuun, ellei maantieteellisyys ole relevantti viestille. (Wong 2013, 90.)

Koropleettikartta on tilastollisen teemakartan perustyyppi, jossa luokitellulle tiedolle määritellään tummuusaste tai väri, jolla kukin alue kokonaisuudessaan väritetään. Alueet ovat yleensä hallinnollisia alueita. Koropleettikartalla voidaan esittää vain yhtä asiaa kerrallaan ja vain sen suhteellinen vaihtelu voidaan tuoda esiin. (Kuusela 2000, 158.)

Tehokkuuden vuoksi koropleettikartassa on hyvä käyttää vain yhtä väriä. Poikkeus on, jos esitettävä asia voidaan jakaa kahteen vastakkaiseen luokkaan. (Kuusela 2000, 160.)

Koropleettikartan käytössä keskeistä on datan luokitus: niiden määrä ja välit. Erilaisella luokituksella ja sitä kautta värityksellä saadaan aikaan erilaisia karttoja samasta asiasta. Luokat voidaan laskea eri tavalla, mutta hyvä sääntö niiden laskemiseen on, että luokkien

määrä on yhtä kuin kuutiojuuri esitettävien havaintojen määrästä. (Kuusela 2000, 159–160.)

Koropleettikartan ongelmia ovat mahdolliset pinta-ala harhat, rajaamisharha ja se, että alueiden sisäinen vaihtelu ei tule esille. (Kuusela 2000, 161).

Yksi teemakartan alalajeista on pistesymbolikartta, jolla esitetään asian esiintymistä tietyssä pisteessä. Kartta voi esittää myös absoluuttisia lukuarvoja tai esimerkiksi jotain paikkaa pistemäisillä symboleilla. Pisteet voivat olla samankokoisia tai niiden koko voi kuvata esitettävää määrää. (Kuusela 2000, 161.)

Pistetiheyskartassa pisteet ovat samankokoisia ja pisteiden tiheys esittää alueellista jakaumaa. Erilaisia värejä ja symboleja käyttämällä voidaan pistetiheyskartassa esittää useampaa ilmiötä samassa kartassa. (Kuusela 2000, 162.)

Suhteellisessa symbolikartassa pisteiden koko määrittyy esitettävän määrän perusteella: suurempi ympyrä esittää suurempaa tihentymää. (Kuusela 2000, 163).

Yksikkösymbolikuviot

Yksikkösymbolikuviot on kuva-aiheinen tilastokuva, jossa kaikki symbolit ovat samankokoisia ja ne esittävät määrää. (Kuusela 2000, 184).

Toisin kuin hahmokuviossa, jossa hahmon koko määrittyy lukuarvon perusteella, yksikkösymbolikuviossa symbolien koko pysyy samana ja ne edustavat laskentayksikköjä, jotka ovat esimerkiksi peräkkäin tai päällekkäin. Ne ovat käytännössä pylväskuvioita, jotka on rakennettu kuvalliseen muotoon. (Kuusela 2000, 186.)

Piktogrammi on visuaalisesti viehättävämpi kuin pylväsdiagrammi, mutta se on tehottomampi suuren datamäärän esittämisessä. Piktogrammia tulisi käyttää vain yksinkertaisten datasettien kanssa. (Wong 2013, 86–87.)

Yksikkösymbolien käytössä on otettava huomioon seuraavat asiat:

- Symbolien tulee olla yksikäsitteisiä, helposti tulkittavia ja ne liittyvät esitettävään asiaan.
- Jokainen symboli esittää tiettyä määrää, ne ovat samankokoisia ja yhdessä ne tuovat esiin määrän vaihtelun.
- Kuvio on niin yksinkertainen kuin mahdollista.
- Kuviolla annetaan ylimalkainen kuva, koska yksinkertaisten asioiden esittämiseen se ei sovi.
- Kuvio tuo esiin yleisen vertailun.

(Kuusela 2000, 187.)

Piktogrammin tarkoitus on antaa datasta tiivistelmä. Osittaisten ikonien käyttö on hämmentävää. Ainoa poikkeus on neliö, koska se toimii, vaikka siitä käytettäisiin vain pientä osaa. (Wong 2013, 86–87.)

Muuttujien esittämiseen ei tule käyttää eri versiota symbolista. Yhdistelmät voivat olla häiritseviä, jolloin lukijoille on vaikeaa vertailla alla olevaa dataa. Fokus pitäisi olla informaatiossa, ei piirroksessa. Piktogrammit voivat antaa nopean tiivistelmän määrästä ja volyyminä, mutta ne eivät sovi suuren datamäärän esittämiseen. Pylväskuvio on paljon tehokkaampi kuin piktogrammi, kun vertaillaan useiden monimutkaisten datasettien diskreettejä määriä. Muuttujia tulisi esittää yhdellä symbolilla ja eri sävyillä. Näin lukijat pystyvät keskittymään datan vertailuun eikä ikonien eri tyyliin. Hyvä ikoni on yksinkertainen, symmetrinen, selkeä vaikka se olisi pieni, mahtuu neliöön. (Wong 2013, 86–87.)

2.4.3 Värien käyttö kuvaajissa

Värejä tulee käyttää kuvaajissa sulavasti. Kaikkia väripaletin värejä ei tule käyttää samaan aikaan, sillä liian monen värin käyttö samaan aikaan tekee kuvaajasta hämmentävän ja räikeän. Värejä määriteltäessä tulisi valita harmonisia yhdistelmiä, kuten yhden värin eri sävyjä tai värejä, jotka ovat väriympyrässä samalla puolella. Aina kun väri vaihtuu, se merkitsee muutosta informaatiossa tai lisätasoa datassa. Esitettävän informaation tulisi määrittää jokainen väri, joka valitaan kuvaajaan. (Wong 2013, 40–41.)

Samankaltaisen datan esittämisessä ei tule käyttää montaa väriä, vaan samaa väriä tulisi käyttää saman muuttujan esittämisessä, jotta lukija voi keskittyä datan vertailuun. (Wong 2013, 40–41.)

Jos kuvaajassa käytetään enemmän kuin yhtä väriä, värien sävyt eivät saa olla yhtä tummia, koska lukijan on helpompi erottaa sävyjen ero. Riittävä sävyero tekee kuvaajasta luettavamman myös värisokeille lukijoille. (Wong 2013, 44–45.)

2.4.4 Typografian rooli visualisoinneissa

Kirjasimien rooli kuvaajissa on kuvailla informaatiota eikä koristella sitä. Typografia pitäisi valita luettavuuden perusteella. (Wong 2013, 30.)

Typografian ei tule olla keskeinen osa kuvaajia vaan datan. Tekstin tulee kuvailla kuvaajaa selkeästi eikä herättää tunteita. Huono typografia herättää kohtuutonta huomiota datan sijaan, joka kantaa pääasiallista viestiä. (Wong 2013, 32.)

Kirjasimien ei tulisi olla liian pieniä tai tiiviitä. Tyylin pitää olla yksinkertainen. Lihavointia tai kursivaa voidaan käyttää vain huomion korostamiseen, mutta niitä ei tule käyttää samaan

aikaan. Tekstiä ei tule kirjoittaa kapitaaleilla, koska se on vaikealukuisempi. Tavutusta tulee välttää. (Wong 2013, 31.)

3 Riikasta Helsinkiin: Näin Youth Newsroom visualisoi dataa

Produktin kohde oli UNESCON alla toimiva Youth Newsroom, joka hoitaa kansainvälisen sananvapaustapahtuman World Press Freedom Dayn viestintää ja tuottaa tapahtumasta journalistista sisältöä liittyen sanan- ja lehdistönvapauteen.

Toimeksiantaja määritteli produktin tavoitteiksi datan visualisoinnin formaattien määrittelyn ja sen, että produkti on sovellettavissa myös tulevina vuosina.

Produktin tavoite oli yksinkertaistettuna vastaus kysymykseen, miten Youth Newsroom visualisoi dataa? Datavisualisoinneille ei ollut toimituksessa aiempaa pohjaa, sillä niitä ei oltu tehty esimerkiksi vuoden 2015 tapahtumassa, joten produktin tekemiselle ei ollut toimituksen puolesta mitään mallia. Se oli sinänsä hyvä, koska produktin pääsi tekemään puhtaalta pöydältä, mutta toisaalta suuri haaste, koska ei ollut tiedossa, mitä erityisesti pitäisi ottaa huomioon tai kehittää, ja mahdolliset ongelmat piti osata ennakoida etukäteen. Tämän vuoksi joitain asioita jouduttiin tekemään lennosta.

Produktin tavoitteena oli myös luoda pohja datavisualisoinneille Youth Newsroomissa ja edellytykset sille, että datavisualisointeja voivat tehdä myös tulevat toimituksessa. Formaattit ja työkalut piti valita niin, että myös tulevat toimitukset voivat käyttää niitä ja tahtoaan myös kehittää niitä. Ohjeet kirjoitettiin siinä mielessä, että lukijan osaaminen ja tietämys datavisualisoinneista ja datajournalismista on matala.

3.1 Produktin tekemiseen käytetyt lähteet

Produktin suunnitteluvaiheessa määrittelin, että tarvitsen tietoa käsitteiden tarkkaan määrittelyyn datan visualisointien visuaaliseen puoleen ja formaattivalintojen perusteluun. Visuaaliseen puolella minua askarrutti erityisesti, millainen kuvaaja on paras minkäkin datan visualisoimiseen, ja mitä värejä ja sävyjä on hyvä käyttää tai olla käyttämättä keskenään.

Kuuselan Tilastografiikan perusteet käsitteli on julkaistu vuonna 2000 ja käsitteli aihetta lähinnä tilastotieteellisestä näkökulmasta, mutta tarjosi tietoa erityisesti eri kuvaajatyypeistä, ja niiden käytöstä sekä onnistuneen esityksen edellytyksistä. The Data Journalism Handbook antoi tietoa journalistisesta näkökulmasta, ja Infographics: The Power of Visual Storytelling infografiikan formaateista. Yaun teoksista sai parhaiten tietoa datan eri tyypeistä ja visualisointiprosessista, kuten myös Wongin kirja, joka oli erityisen hyödyllinen värien ja typografian suhteen.

Infogr.amin käyttöohjeet rakentuivat pitkälti omasta kokemuksepäisestä tiedosta. Epäselvät asiat tarkistutin Infogr.amin tekniseltä tuelta.

3.2 Ei pelkästään datan visualisointia

Alkuperäisenä suunnitelmana oli, että Youth Newsroom tekee datavisualisointeja ja datajournalismia World Press Freedom Daysta tekemäni tyylikirjan mukaan. Pyrin siihen, että tyylikirjan avulla Youth Newsroomin toimittajat osaavat visualisoida dataa tarkasti ja tiiviisti, mutta visuaalisesti ilman liiallista kuvioroinaa. Halusin tehdä produktista sellaisen, että se ei rajaa tulevia Youth Newsroomeja liikaa, joten tyylikirjassa pyrin universaaliuteen ja yleisiin visualisoinnin sääntöihin: mitä voi tehdä ja mitä ei kannata tehdä.

Suunnittelin työkirjan sisältävän selkeät ohjeet työvälineiden käyttämiseen. Niin selkeät ja seikkaperäiset ohjeet, että asiasta mitään tietämätön henkilökin oppisi visualisointien tekemisen tyylikirjan perusteella.

Alusta lähtien minulle oli selvää, että datatiimi visualisoi dataa yksinkertaisilla ja perinteisillä kuvaajilla. Produktin keskeinen vaihe olikin, kuvaajatyypien valitseminen ja niiden hyvien ja huonojen puolien sekä käytettävyyden erittely. Lisäksi piti valita, mitkä Infogr.amin kuvaajatyypit ovat hyödyllisiä ja mitä ei vastaavasti tarvittaisi, koska esimerkiksi joku toinen kuvaajatyypit pystyy visualisoimaan saman asian yhtä hyvin tai jopa paremmin.

Jo alkuvaiheessa oli selkeää, että Youth Newsroomin dataryhmä ei tee pelkästään datavisualisointeja vaan myös juttuja datan pohjalta, joita voitaisiin julkaista niin printissä kuin Youth Newsroomin verkkosivuilla. Määritin, että datatiimin jutut sisältäisivät datan analysointia, syiden erittelyä ja datasta paljastuvat ilmiöt konkretisoitaisiin tosielämään mielellään henkilöiden kautta, jos mahdollista, koska se konkretisoi datan tosielämään.

Määritin niin, että tiimi tekisi staattisia visualisointeja printtiin ja interaktiivisia nettiin, koska printti ei tue muita formaatteja kuin staattisia toisin kuin internet, jossa interaktiivisuudella voidaan lisätä visualisointeihin lisää dataa, ja sitouttaa paremmin lukijoita.

Syksyllä 2015 suunnitelmissa oli myös liikkuvien visualisointien tekeminen, joita oltaisiin voitu käyttää videossa ja julkaista nettisivujen lisäksi myös sosiaalisessa mediassa, mutta näitä suunnitelmia ei koskaan viety yhtä tai kahta palaveria pidemmälle.

Staattista infografiikkaa oltaisiin voitu julkaista myös nettisivuille mutta koska Infogr.amilla tehdään visualisointeja ensisijaisesti internetiin, oli jotenkin ilmiselvää, että teemme visualisoinnit tavallaan netti edellä, kuitenkin niin, että ne on myös nopeasti sovellettavissa printtiin. On kuitenkin huomioitava, että Infogramilla pystyi tekemään visualisointeja, joissa interaktiivisuus oli varsin pientä. Kun käyttäjä esimerkiksi vie hiiren pylvään päälle se kertoo pylvään tarkan numeerisen arvon. Kun käyttäjä klikkaa nimiöstä jotain muuttujaa. Visualisoinnissa korostetaan pelkästään valittu alue tai muuttuja.

Päätin, että visualisoinnit tehdään perinteisillä kuvaajilla: viiva-, pylväs-, piirakka-, ja parvi-kuvioilla. Tarvittaessa dataa voitaisiin visualisoida myös koropleettikartoilla sekä yksinker-taisilla piktogrammeilla, joita voitaisiin käyttää myös mielenkiintoisten havaintojen nostami- seen leipätekstiin. Jos datan visualisoimisessa ei olisi mieltä, kirjoitettaisiin se pelkästään tekstissä tai esitettäisiin taulukossa.

Visualisointeja varten tarvittiin 2–3 erilaista väriä ja niille eri sävyjä. Päätin, että visualisoin- tien päävärinä käytettäisiin samaa sinistä, mitä tapahtuman logossa, jonka perusteella määritettäisiin kaksi lisäväriä, jotka ovat samalla puolella väriympyrää kuin sininen, eli vih- reä ja purppura.

Tarvittavat tekstielementit tehtäisiin taittovaiheessa suoraan sivulle tai nettisivuilla, jotta kuvaajissa pystyttäisiin käyttämään varmasti samaa fonttia kuin printtijulkaisussa ja nettisi- vuilla.

Suunnitelmavaiheessa oli ehdolla kaksi työkalua, Infogr.am ja Piktochart. Jälkimmäinen siksi, että sillä pystyi tekemään visualisointeja ja infografiikkaa vapaamuotoisemmin julis- teita, joiden tekeminen oli pöydällä vielä syksyllä 2015, mutta pudotettiin myöhemmin pois.

Selkeyden vuoksi päätin, että tiimi käyttäisi vain Infogr.amia, koska sen käyttäminen on helppo ja nopea oppia. Infogr.am valikoitui työkaluksi myös siksi, että olin datavisualisoin- tiohjelmista käyttänyt sitä eniten, joten pystyin tarpeen vaatiessa opastamaan ja autta- maan muita sen käytössä. Lisäksi se oli helppokäyttöinen ja dynaaminen. Dataa pystyi, tuomaan suoraan Google Drivesta, muokkaamaan lennosta, jolloin myös visualisointi päi- vittyi automaattisesti. Se teki paljon käyttäjän puolesta, joten helpoimmillaan käyttäjän piti vain valita kuvaaja, tuoda data, valita värit ja säätää asetukset. Lisäksi Infogr.amilla tehdyt visualisoinnit skaalautuivat automaattisesti websivuilla automaattisesti käyttäjän näytön koon mukaan.

Suunnittelin, että dataryhmä käyttää Googlen Google Spreadsheetsia datan siivoamiseen ja analysointiin.

Spreadsheets on Googlen etäresurssipalvelupohjainen versio Microsoftin Excel-taulukko- laskentaohjelmasta, mutta toiminnoiltaan yksinkertaisempi. Spreadsheetsin avulla käyttäjä pääsee nopeasti käsiksi dataansa eri koneilta ja laitteilta. Reaaliaikainen editointi ja si- säänrakennettu keskustelupalvelu mahdollistavat yhteistyön. (Yau 2013a, 278).

Lisäksi datatiimi käyttäisi Google Drivea, jonne tallensimme alkuperäisen, muokkaamatto- man version datasta, ja myös työkopiot kustakin datasetistä. Drive on Googlen tarjoama etäresurssipalvelu, joka antaa käyttäjälleen 15 gigatavun verran tallennustilaa. Spreads- heetsin tapaan Driveen pääsee käsiksi eri koneilla ja laitteilla ja muita henkilöitä voi kutsua

yhteiskäyttämään ylläpitäjän valitsemia tiedostoja. (Google.) Googlen työkalut valikoituivat käyttöön siksi että niiden avulla yhteistyön tekeminen oli helppoa ja dataan pääsi käsiksi tarvittaessa niin omalta kuin toimituksenkin koneelta.

Alkuperäisen tavoitteena oli, että datatiimissä olisi minun lisäksi viisi jäsentä. Tällöin tavoittelin, että datatiimi pystyisi tekemään yhteensä seitsemän datajournalistista juttua. Ne oltaisiin julkaistu niin, että jokaisena tapahtumapäivänä yksi ja ennen tapahtumaa neljä, yksi jokaisena huhtikuun viikkona. Toisaalta olin sitä mieltä, että tällaisessa julkaisutavassa on mieltä vain, jos Youth Newsroom julkaisee muutakin kuin datajournalistista sisältöä. Lopulta tiimissä oli minun lisäksi neljä jäsentä ja ennen tapahtumaa asetin tavoitteeksi, että jokaiselta julkaistaan vähintään yksi juttu.

3.3 Pikemminkin opas kuin tyylikirja

Tyylikirjasta tuli lopulta kuvineen 40 sivuinen, joka jakautuu neljään pääluukuun: Johdanto, datatiimi, prosessi, datavisualisointien formaatit, ja kuinka Infogr.amia käytetään.

Tyylikirja kertoo datatiimin koon, jäsenien osaamisvaatimukset, visualisointiprosessin, millaista dataa voi olla, edellytykset onnistuneelle visualisoinnille sekä määrittelee käytettävät kuvaajat, värit ja työkalut. Lisäksi tyylikirja sisältää tiiviit ohjeet visualisointityökalu Infogr.amin käyttöön.

Tiimi ja vaatimukset -osuus määrittelee, mitä teknisiä taitoja ryhmän jäseniltä vaaditaan ja millä välineillä he työskentelevät. Tämän jälkeen produkti käsittelee sitä, minkälaista dataa voi olla, mitä siitä kannattaa etsiä ja miten sitä kannattaa visualisoida. Lisäksi käydään läpi, miten toimitaan, jos data on vajavaista, mitkä ovat edellytykset onnistuneelle visualisoinnille ja milloin visualisoinnin sijaan kannattaa avata data tekstillä tai esittää se esimerkiksi taulukossa. Formaatti-osuus käsittelee, millaisia visualisointeja Youth Newsroom tekee, millaisia visualisointielementtejä käytetään ja mikä on niiden paikka jutussa. Lopuksi guidelines käsittelee yhden visualisointityökalun, jota Youth Newsroom käyttää: Infogr.amin. Tyylikirja opastaa lukijalle tiiviisti, miten Infogr.am toimii.

Koska lopullisesta tyylikirjasta, tai pikemminkin oppaasta, tuli 40-sivuinen, tein siitä tiimille myös tiivistelmän, jossa oli käytettävien formaattien lisäksi käytettävät kuvaajat sekä ohjeistukset väreistä ja kirjasimista. Tiivistelmän tarkoitus oli palvella tiimiä niin, että jäsenet voisivat tarvittaessa tarkistaa siitä joitain pieniä asioita, esimerkiksi värikoodit.

Määritin värit yhdessä Riikka Kajanderin (valokuvatiimin johtaja) kanssa, joka vastasi valokuvaajien tyylikirjan tekemisestä. Päätimme, että pääväri tulisi olemaan sama vaalean sininen, jota käytettiin myös tapahtuman logossa. Lisäksi käytimme vihreää ja tummempaa sinistä, jossa oli vähän punaisen sävyä. Näille väreille määritettiin kaksi vaaleampaa ja

tummempaa sävyä visualisointeja varten. Määritimme värit Paletton-työkalulla ja tarkistimme, että värit ja niiden sävyt erottuvat toisistaan harmaasävyisinä, jotta mahdollisesti väri-sokeat ihmiset pystyvät lukemaan niitä.

Väreistä vaaleansininen oli se, jota ja jonka sävyjä käytettäisiin ensisijaisesti kuvaajissa. Vihreää käytettäisiin tarvittaessa ja tummempaa sinistä silloin kun tarvittaisiin ehdottomasti kolme eri väriä.

Saimme käyttöömmme lopulta Infogr.amin business-version, joka mahdollistaa infografiikoiden tallentamisen kuvatiedostoina, ja tilien yhteenkytkemisen, jolloin ryhmän jäsenet voivat tehdä visualisoinnit omilla käyttäjätunnuksillaan.

3.4 Toimeksiannosta tapahtumaan

Produkti eteni niin, että koulutusohjelman johtaja Anne Leppäjärvi antoi toimeksiannot minulle ja viidelle muulle opiskelijalle toukokuussa 2015. Aloitin oman produktin tekemisen elokuussa 2015 käymällä läpi Infogr.amin. Kävin alustavasti läpi, mitkä Infogr.amin tarjoamat kuvaajaelementit ovat käyttökelpoisia ja mitkä eivät. Loppusyky kului tapahtuman ja produktin suunnitteluun. Keräsin kirjallista lähdemateriaalia ja laadin tutkimussuunnitelman. Lisäksi aloitin lähdemateriaalin lukemisen.

Tammikuussa 2016 luin suurimman osan lähdemateriaalista ja työstin produktin viimeistelyä vaille valmiiksi. Viimeistelin produktin helmikuussa 2016 ja tein vielä joitain pieniä korjauksia siihen maaliskuussa 2016. Tällöin työstin produktista myös tiiviimmän paketin pelkästään tämän vuoden toimitusta varten. Maaliskuun lopussa sain tietooni, keitä kaikkia tiimiini tulisi kuulumaan.

Maaliskuun alussa sain tietooni, että datan visualisoinnin perusteet -opintojaksolle osallistuvat opiskelijat tuottaisivat visualisointeja tapahtumaan, jotka olisivat myös Youth Newsroomin käytettävissä.

Tästä kurssista muodostui itselle hieman hämmentävä tekijä. Minulle jäi epäselväksi, käytetäänkö tekemääni produktia kurssilla ja missä opiskelijoiden tuottamaa sisältöä julkaistaan, sillä tästä liikkui erilaista tietoa. Yleisradion Plusdeskin toimittaja ja koodari Ville Juutilainen, joka opetti jakson toisen osion, kertoi, että häntä oli ohjeistettu niin, että opiskelijat tekevät visualisointeja ensisijaisesti Finlandia talon näytöillä ja Youth Newsroom voi halutessamme julkaista niitä nettisivuillaan. Journalismin koulutusohjelman johtaja Anne Leppäjärvi taas ilmoitti, että opiskelijat tuottavat sisältöä ensisijaisesti nettisivuille ja mahdollisesti Finlandia-talon näytöillä.

Opiskelijoiden ideat menivät lopulta jossain määrin päällekkäin datatiimin ideoiden kanssa. Kaksi opiskelijoiden tekemään visualisointia meinattiin julkaista, mutta molemmat jou-

duttiin tekemään uudestaan. Ensimmäinen siksi, että kaikki visualisoinnissa esitetyt luvut eivät sopineet yhteen lähdeaineiston kanssa. Toista, joka esitti aika sarjoja piktogrammista, ei voitu julkaista alkuperäisessä muodossaan lehdessä, koska se luotti niin paljon interaktiivisuuteen. Tämä visualisointi tehtiin uudestaan, mutta se tiputettiin pois lehdestä viime metreillä.

3.4.1 Datatiimi aloittaa työskentelynsä

Dataryhmä aloitti työskentelynsä huhtikuun ensimmäisellä viikolla. Otin jäseniin yhteyttä huhtikuun 4. päivänä, jolloin ohjeistin ryhmääni kuuluneet henkilöt, lähetin heille tyylikirjan ja annoin heille viikon aikaa ideoimiseen. Sillä aikaa ideoin myös itse ja jatkoin datan keräämisen eri organisaatioilta ja avoimista lähteistä.

Dataryhmän jäsenien rekrytointi ei ollut ongelmaton. Alkuperäisenä toiveenani oli yhteensä viisi henkilöä. Alun perin oli suunniteltu, että jäsenet rekrytoitaisiin Jyväskylän yliopistosta, koska hankkeen apurahan saaminen edellytti yhteistyötä muiden journalismia opettavien korkeakoulujen kanssa. Ehdotin itse Jyväskylää, koska tiesin, että heillä järjestetään datajournalismin kurssi. Lopulta Jyväskylästä saatiin rekrytoitua yksi opiskelija. Olin myös toivonut, että rekrytointi hoidettaisiin niin, että dataryhmä pääsisi aloittamaan työskentelyn helmi-maaliskuussa, koska datavetoisten juttujen tekemiseen saattaa mennä aikaa, mutta lopulta pääsimme aloittamaan huhtikuussa. Tähän vaikutti se, että Youth Newsroomin osallistujat olivat selvillä maaliskuun lopulla.

Ryhmääni kuului lopulta minun lisäksi neljä henkilöä. Jyväskylän yliopistossa journalismia opiskeleva Ines Sirén, Haaga-Heliassa journalismia opiskeleva Malviina Hallamaa, perulainen toimittaja Aramis Castro ja israelilainen Yuval Finkel.

Ohjeistin ryhmän niin, että ensimmäisten ideoiden jälkeen he aloittavat juttujen kanssa työskentelyn: datan analysointi, siivoaminen sekä mahdolliset taustahaastattelut asiantuntijoiden kanssa olisi hyvä tehdä ennen tapahtumaa. Rohkaisin myös juttujen kirjoittamisen ja visualisointien suunnittelun aloittamiseen.

Kun jäsenet olivat vastanneet, arvioin ideat toimituspäällikkö Taika Dalhbomin kanssa. Kaikki ideat olivat mielestämme toteuttamiskelpoisia, joten jäsenet saivat aloittaa ne.

Aramis Castro halusi työskennellä Unescon datan kanssa, mutta ehdotti myös, että keräisimme dataa henkilöistä, jotka ovat osallistuneet World Press Freedom Day -tapahtumiin ja analysoisimme esimerkiksi heidän ikä- ja sukupuolijakaumaa. Ines Sirén halusi tehdä jutun Jyväskylän yliopiston hankkeesta, jossa suomalaisille kunnille on lähetetty tietopyyntöjä, ja mitattu, kuinka nopeasti kunnat vastaavat tietopyyntöihin. Malviina Hallamaalla ei ollut ideoita, joten annoin hänelle aiheeksi journalisteja ja mediaa vastaan tehdyt terroristi-

iskut. Yuval Finkelin ensimmäinen idea ei ollut datapohjainen, vaan hän kirjoitti israelilaisesta dokumentista, joka on herättänyt Israelissa keskustelua sananvapaudesta.

Huhtikuun 14. päivä Taika Dalhbom toivoi ryhmän juttuja valmiiksi huhtikuun 25. päiväksi. Päivämäärä tuli yllätyksenä, sillä sitovasta päivästä ei oltu puhuttu. Keskustelimme asiaa Dahlbomin kanssa ja sen pohjalta ohjeistin ryhmän nopeasti uudelleen: On toivottavaa, että artikkelin ensimmäinen versio on valmiina huhtikuun 25. päivään mennessä tai viimeistään ennen tapahtumaa, mutta jos kumpikaan ei toteudu, artikkelit viimeistellään tapahtumassa.

Tilanne ennen tapahtumaa oli, että Yuval Finkel ehti palauttaa artikkelinsa ensimmäisen version noin viikko ennen tapahtumaa. Lopulta artikkelin editoimiseen kului sen verran aikaa, että se saatiin valmiiksi vasta tapahtuman viimeisenä päivänä, keskiviikkona 4.5. Malviina Hallamaa sai ensimmäisen versionsa valmiiksi myös viikko ennen tapahtumaa ja se saatiin viimeistelyä tapahtuman toisen päivän aamuna tiistaina 3.5. Aramis Castro ja Ines Sirén eivät olleet ehtineet työskennellä juttujensa parissa ennen tapahtumaa, joten he aloittivat työskentelyn vasta maanantaina 2.5.

Tapahtuma alkoi maanantaina toukokuun 2. päivä. Ryhmän työt olivat siinä vaiheessa, että Malviina Hallamaa viimeisteli juttuaan terrorismista, Ines Sirén aloitti työn tietopyyntöjen kanssa. Aramis Castro aloitti Unescon datan läpikäymisen ja Yuval Finkel oman artikkelinsa viimeistelyn.

Artikkelit ja niiden visualisoinnit olivat valmiita seuraavana päivänä. Osaan tehtiin viimeiset korjaukset aamulla, Aramis Castron artikkeli oli valmis noin kaksi tuntia ennen lehden deadlinea ja Yuval Finkelin artikkeli viimeisteltiin vasta viimeisen päivän aamuna, koska lehteen tulevien artikkelien editointi oli etusijalla. Sillä aikaa Yuval Finkel työskenteli World Press Freedom Dayn osallistujalistoihin keräämäni datan parissa, josta yritimme kartoittaa muun muassa sitä, millainen sukupuolijakauma eri paneelikeskusteluissa ja tapahtumissa ylipäätään on ollut. Juttua ei kuitenkaan saatu viimeistelyä tapahtuman aikana. Koska datatiimi sai juttunsa valmiiksi tiistaina, siirrettiin jäsenet muihin tiimeihin, koska uusien projekteja ei oltaisi enää saatu valmiiksi, koska toimitus lopetti toukokuun 4. päivä kello 14.

Datatiimi teki tapahtumassa lopulta yhteensä kolme datapohjaista juttua. Nämä olivat Malviina Hallamaan artikkeli terrorismi-iskuista journalisteja ja mediaa kohtaan 2000-luvulla, Ines Sirénin artikkeli Suomen kunnille lähetetyistä tietopyynnöistä ja Aramis Castron artikkeli tapetuista toimittajista Committee to Protect Journalist -järjestön datan pohjalta. Lisäksi Yuval Finkel teki artikkelin israelilaisesta dokumentista, joka on herättänyt keskustelua sananvapaudesta Israelissa ja työskenteli World Press Freedom Dayn puhujadatan kanssa. Tätä juttua ei kuitenkaan saatu valmiiksi tapahtuman aikana.

Dataryhmän jutut julkaistiin Youth Newsroomin verkkosivuilla, ja jokainen datapohjainen artikkeli julkaistiin lopulta myös Youth Newsroomin tekemässä Freezine-lehdessä. Lisäksi lehdessä piti julkaista kaksi datavisualisoinnin perusteet -kurssin opiskelija visualisointia, mutta koska ne eivät sellaisenaan soveltuneet lehteen, tein minä ne uusiksi erilaisella visualisointitavalla.

Datatiimi ei selvinnyt toimitustyöstä ongelmitta. Ainoa tekninen ongelma oli, että Infogr.am ei jostain syystä toiminut kunnolla Mozilla Firefox -selaimissa, jotka oli asennettu toimituksen koneille. Tämän vuoksi visualisointeja jouduttiin visualisoimaan muilla tietokoneilla ja koneille jouduttiin asentamaan toinen verkkoselain.

Joitain ongelmia ja haasteita oli myös tiimin työskentelyssä. Jäsenillä oli ensinnäkin vaihtelevasti aikaa työskennellä juttujen parissa ennen tapahtumaa. Toisaalta, oletimme ehkä liian optimistisesti, että kaikilla rekrytoitavilla on aikaa työskentelyyn ennen tapahtumaa. Jotkut jäsenet eivät pysyneet aina lupaamissaan aikamääreissä ja siitä ilmoitettiin vasta jälkikäteen, jos ilmoitettiin. Yksittäiset henkilöt taas saattoivat kadota useiksi tunneiksi ja ilmoittivat sijaintinsa ja syyn poissaololle vasta myöhemmin.

3.4.2 Mistä data kerättiin?

Youth Newsroomin datatiimi käytti pääasiassa avointa dataa. Käytimme eniten Committee to Protect Journalists -järjestön avointa tietokantaa, johon on tilastoitu tapetut toimittajat vuodesta 1992 alkaen. Tätä käytettiin lähteenä Aramis Castron artikkelissa ja myös yhdessä lehteen tehdyssä visualisoinnissa.

Aluksi Aramis Castron artikkelissa oli tarkoitus käyttää UNESCON keräämää dataa, mutta CPJ:n data yksityiskohtaisempaa ja oli helpommin kerättävissä, sillä Unescon tilastot olivat pdf-tiedostoissa.

Malviina Hallamaan artikkeliin käytettiin datana Marylandin yliopiston ylläpitämää Global Terrorism Database -nimistä tietokantaa, johon on tilastoitu tehdyt terrorismi-iskut 1970-luvulta alkaen.

Ines Sirénin artikkelissa käytettiin datana Jyväskylän yliopiston tutkijoiden keräämää dataa, johon on merkitty paikkakunnittain, milloin he ovat vastanneet tutkijoiden lähettämään tietopyyntöön ja kuinka kauan vastaamiseen on mennyt aikaa päivissä.

Tämän lisäksi keräsin datatiimin käyttöön itsenäisen Freedom House -järjestön keräämää dataa, johon on kerätty dataa siitä, kuinka vapaa media on ympäri maailmaa. Data käsittelee erikseen esimerkiksi digitaalisen median ja broadcast-median. Dataan on laskettu numeeriset pisteet ja arvioitu medioiden laillista ympäristöä: Kuinka paljon poliittinen paine ja

taloudelliset tekijät vaikuttavat medioiden toimintaan ja uutisten ja informaation saatavuuteen.

Hyödyllisen datan saatavuus ja käytettävyys oli ongelma. Dataa kyllä löytyi, mutta se ei ollut avointa, vaan esimerkiksi PDF-tiedostoissa, joista datan kerääminen oli haastavaa ja aikaavievää. Löydetty data ei myöskään välttämättä ollut ajantasalla. Dataa pyydettiin myös esimerkiksi Toimittajat ilman rajoja -järjestöltä, Amnestyilta, PENiltä, ja OSCElta, mutta yhdeltäkään organisaatiolta ei saatu sitä. Syiksi luettiin resurssipula tai neuvottiin datan löytyvän jo paikasta x. Amnesty ei orastavan sähköpostikeskustelun jälkeen enää vastannut viesteihin.

Toinen ongelma datan suhteen oli se, että eri lähteistä kerätty data ei ollut vertailukelpoista. Esimerkiksi Unescon ja Committee to Protect Journalists -järjestöjen luvut tapetuista toimittajista erosivat toisistaan.

3.4.3 Miten dataryhmän artikkelit visualisoitiin?

Dataryhmän visualisoima data oli pääasiassa erilaisia aikasarjoja tai ne lopulta siivottiin erilaisiksi aikasarjoiksi, kuten terrorismidata ja Committee to Protect Journalists -järjestöltä kerätty data, jotka alkuperäisessä muodossaan sisälsivät monta muuttujaa. Myös kategorisoitua dataa visualisoitiin.

Malviina Hallamaan jutun nettiversiossa data visualisoitiin kahdella pinotulla pylväskuvioilla. Ensimmäisessä oli mediaa vastaan tehdyt terroristi-iskut vuosina 2000-2014, niin että pylväissä oli eroteltuna onnistuneet ja epäonnistuneet iskut. Toisessa oli ne kymmenen maata, joissa terroristi-iskuja mediaa vastaan oli tehty eniten ja pylväissä oli eroteltu myös iskutyypit. Freezinessä käytetyssä versiossa käytettiin vain ensimmäistä visualisointia tilanpuutteen vuoksi.

Ines Sirénin artikkelissa datan visualisoimiseen käytettiin karttaa ja yksinkertaista viivadiagrammia. Viivadiagrammilla visualisoitiin, kuinka moni kunta vastaa tietyn päivän kuluessa. Kartalla taas visualisoitiin jokainen kunta niin, että lukija pystyi valitsemaan kunnan ja katsomaan, kuinka kauan kunnalla kesti vastata. Mitä tummempi kunnan värisävy oli, sitä kauemmin kunnalla kesti vastata. Jos kunta oli oranssi, se ei ollut vastannut tietopyyntöön. Teknisesti karttavisualisointia ei tehty Infogr.amilla vaan Google Fusion Tablesilla, koska sillä pystyttiin tekemään tarkempi Suomen kuntia käsittelevä visualisointi, toisin kuin Infogr.amilla, jolla pystyy tekemään vain maakuntatason visualisointeja. Lisäksi Ineksen jutussa käytettiin yksinkertaista piktogrammia, korostamaan sitä, kuinka moni kunta ei ole vastannut, ja kuinka moni kunta vastasi lain määräämässä ajassa. Lehdessä visualisointeja ei käytetty, osaksi sen takia, että juttu nostettiin lehteen viime hetkellä, koska lehteen alunperin suunniteltu artikkeli ei valmistunut ajoissa.

Aramiksen jutussa data visualisointiin yksinkertaisella vaakasuuntaisella pylväskuviolla, jossa oli ne kymmenen maata, joissa journalisteja oli tapettu eniten vuosina 2000-2016.

Dataryhmän tuottaman sisällön lisäksi lehdessä julkaistiin visualisointi tapetuista ja vangituista toimittajista vuosina 2005-2015, joka julkaistiin palestiinalaisen Lubna Sahebin jutun yhteydessä. Visualisointina oltaisiin voitu käyttää Haaga-Helian opiskelijoiden visualisointia, mutta alkuperäisen visualisoinnin kaikki luvut eivät täsmänneet lähteeseen, joten visualisointi piti tehdä kokonaan uudestaan.

3.5 Vuoden kestänyt produktio

World Press Freedom Day 2016 oli vuoden kestänyt projekti. Se alkoi oikeastaan jo vuoden 2015 tapahtumassa, jossa minä ja seitsemän muuta opiskelijaa osallistuimme Youth Newsroomin toimintaan. Suomeen palattuamme aloitimme seuraavan tapahtuman suunnittelun ja saimme toimeksiannot tuotteille. Valtaosa suunnittelutyöstä tehtiin syksyllä 2015 ja itse työstin produktin alkuvuodesta 2016. Huhti-toukokuussa produkti toteutettiin käytännössä ja sen toimivuus testattiin.

Produktin tavoitteena pääasiallisena tavoitteena oli luoda pohja datavisualisoinneille Youth Newsroomissa, ja edellytykset datavisualisointien tekemiselle tulevaisuudessa.

Produkti määritteli, miten Youth Newsroomin tuottamassa sisällössä visualisoidaan dataa ja mitä työkaluja niiden tekemiseen käytetään. Tavoitteena oli tehdä opas, jonka avulla asiasta tietämätönkin pystyi tekemään datavisualisointeja. Datan visualisoimiseen Youth Newsroomin datatiimi käytti perinteisiä tilastollisia kuvaajia. Visualisointien tekemiseen käytettiin Infogr.amia, joka on selaimella käytettävä datavisualisointipalvelu.

Produktin lähteinä käytin erilaisia datavisualisointeja, infografiikkaa tai datajournalismia käsitteleviä teoksia. Lähdemateriaalista kerättiin tietoa erityisesti kuvaajien käyttöön, käsitteiden määrittelyyn sekä värien ja typografian käyttöön.

Produktista tuli lopulta enemmän opas kuin pelkästään tyylikirja. Se oli pituudeltaan mitta-va, jonka takia teinkin siitä vielä tämän vuoden toimitukselle tiivistetyn version. Lopulta produkti määritteli dataryhmän kokoonpanon, datan eri tyyppejä, visualisointiprosessin, onnistuneen visualisoinnin edellytykset, eri kuvaajatyypit ja formaatit, värien ja typografian käytön sekä Infogr.amin käyttöohjeet.

Produkti toteutettiin huhti-toukokuussa 2016. Dataryhmä aloitti juttujen ideoinnin ja työstämisen noin kuukausi ennen tapahtumaa ja viimeisteli artikkelinsa ja visualisoinnit World Press Freedom Day -tapahtuman toisena päivänä. Jokainen artikkeli julkaistiin Youth Newsroomin sivuilla ja lisäksi jokainen datajournalistinen artikkeli julkaistiin myös Youth Newsroomin toimittamassa Freezine-lehdessä. Visualisoinnit tehtiin yhtä poikkeusta lukuun ottamatta produktin mukaan.

Datatiimi käytti pääasiallisena lähteenään avointa dataa. Ongelma oli, että hyödyllistä dataa ei ollut juurikaan avoimena, eikä sitä myöskään ollut kauhean helposti saatavilla tai se ei ollut ajantasalla.

Muita tuotteeseen liittyneitä olivat aikataulujen muuttumiset ja se, että miten tuotteeseen löyhästi liittyneet asiat/sidosryhmät toimivat jäi hieman epäselviksi

Tapahtuma meni datatiimin osalta hyvin. Ongelmia ja haasteita toki oli, mutta ne eivät lopulta vaikuttaneet ratkaisevasti lopputulokseen. Tiimin jäsenien taidot ja valmiudet olivat enimmäkseen riittävät. Yhtä luukun ottamatta jäsenet eivät tarvinneet apua kuin näkemysasioihin ja muutamaan tekniseen asiaan.

4 Pohdintaa produktin onnistumisesta

Produktin keskeinen tavoite oli määritellä Youth Newsroomille formaatit datan visualisoimiseen ja luoda sellainen pohja niiden tekemiselle, että myös tulevat toimitukset pystyvät tekemään niitä. Produkti toteutettiin toukokuun alussa Helsingissä järjestetyssä World Press Freedom Dayssa.

4.1 Produktille asetetut tavoitteet ja miten niissä onnistuttiin

Yleinen produktille asetettu tavoite oli visualisointiformaattien luomien niin, että myös tulevat toimitukset pystyvät tekemään visualisointeja. Tavoitteenani oli tehdä selkeät ohjeet, jotta myös henkilöt, joilla on vähäiset taidot visualisointien tekemiseen, pystyvät produktin avulla tekemään niitä. Tämän vuoksi helppokäyttöisyys oli keskeistä visualisointitapojen määrittelyssä ja työkalujen valitsemisessa.

Tässä tavoitteessa mielestäni onnistuttiin. Näen, että tyylikirja on tarpeeksi seikkaperäinen vasta-alkajillekin, mutta produktin toteutuksessa tätä ei pystytty testaamaan, sillä tiimin jäsenillä oli jo ennestään kokemusta visualisointien tekemisestä ja toimituksen käyttämän visualisointityökalun käytöstä.

Henkilökohtaisia tavoitteitani olivat oman osaamisen ja tietouden syventäminen, toimivan ja hyödyllisen tyylikirjan tekeminen toimitukselle, jolle asia saattoi olla aivan uusi, yhteistyö muiden toimittajaopiskelijoiden sekä ulkomaalaisten toimittajien kanssa ja tiimin sekä sen toiminnan ohjaaminen varsinkin nopealla aikataululla.

Opinnäytetyöprosessin aikana oma osaaminen ja tietous datan visualisoimisesta kasvoi lähinnä visuaalisesta näkökulmasta. Opin onnistuneen visualisoinnin edellytykset, värien ja kuvaajien oikeaoppisen käytön. Syvensin osaamistani tilastollisen tiedon esittämisessä. Ehkä keskeisin oppi oli, että datan ja tiedon visualisoiminen ei ole visuaalisesta näkökulmasta niin yksinkertaista, mitä etukäteen luulin, vaan on monia asioita, joita voi tehdä huomaamattaan väärin. Lisäksi sain kokemusta visualisointien ja datajournalismin tekemisestä sekä tiimin johtamisesta hehtisessä toimitustyössä. Yhteistyö ulkomaalaisten toimittajien ja suomalaisten toimittajaopiskelijoiden kanssa sujui hyvin.

4.2 Itse prosessi ja sen arviointi

Produktin kaari toimeksiannosta itse toteutukseen kesti yhteensä noin vuoden. Produktin suunnittelu aloitettiin syksyllä 2015, työstettiin alkuvuodesta 2016 ja toteutettiin keväällä 2016.

Produktin ensimmäinen haaste ja ongelma oli, että toisin kuin esimerkiksi videoiden tekemisessä, datavisualisoinneissa ei ollut valmiiksi mitään pohjaa tai toimintamallia. Tyylikir-

jaa tehdessä ei ollut olemassa olevaa mallia esimerkiksi edeltävältä vuodelta, jonka avulla olisi pystynyt paremmin esimerkiksi ennakoimaan mahdollisia ongelmia tai haasteita.

Produktin tekemiseen oli aikaa, mutta jälkikäteen ajateltuna sen tekeminen olisi pitänyt jaksottaa tasaisemmin. Henkilökohtaisesti intensiivisimmät vaiheet olivat tammikuussa ja huhtikuussa tapahtumaviikko mukaan lukien. Tämä olisi vähentänyt omaa stressiä ja todennäköisesti myös tuottanut entistä paremman lopputuloksen. Produktista olisi saanut ehkä paremman, jos sitä olisi käynyt läpi enemmän toimeksiantajan kanssa. Produktin työstövaiheessa toimeksiantaja ei juurikaan antanut palautetta, onko produkti hyvä vai huono.

Tein produktin käytännössä kokonaan tammikuun kolmen ensimmäisen viikon aikana. Tämä käsitti myös jäljellä olleen lähdemateriaalin läpikäymisen. Lähdemateriaaliin olisi pitänyt tutustua aikaisemmin, sillä uutta tietoa tulikin enemmän mitä olin ennakoanut. Se toisaalta kehitti omaa oppimista ja osaamista, mutta näin jälkikäteen ajateltuna vaikutti siihen, että produkti piti osittain suunnitella lennosta uudelleen. Paremmalla ajalla uutta tietoa olisi pystynyt arvioimaan paremmin.

Produktin työstämisen aikataulukseen vaikutti se, että pääasiassa muu opiskelu vei enemmän aikaa, mitä olin olettanut.

Aikatauluongelmista huolimatta produktin ja opinnäytetyön prosessi meni mielestäni hyvin. Produkti ei esimerkiksi myöhästynyt yhdestäkään sille asetetusta aikataulusta.

4.3 Miten produktia voisi kehittää?

Yksi selkeä osa-alue, jota voi ja pitääkin kehittää, on toimituksen käytettävän visualisointityökalun arviointi. Infogr.am oli mielestäni sopiva Youth Newsroomille nopeutensa ja helpoutensa vuoksi. Tulevaisuudessa tai jo nyt voi olla työkalu, jolla pystytään tekemään parempia tai hienompia visualisointeja, tai niiden tekeminen on helpompaa tai edullisempää.

Toinen kehitettävä osa-alue olisi produktin laajentaminen niin, että se käsittelee datajournalismia. Nyt produkti käsittelee enimmäkseen datan visualisointeja, joten yksi kehittämisalue voisi olla datajournalistiset alueet. Tiedon kerääminen avoimista lähteistä tai viranomaisilta, datan analysoiminen ja hyvien datajournalististen juttujen kirjoittaminen.

4.4 Produktin toteutuksessa onnistuttiin, mutta produkti olisi voinut parempi

Olin ennen tapahtumaa epäluuloinen siitä, kuinka paljon järkeä on datavisualisointien ja datajournalististen juttujen tekemisessä on toimituksessa, joka käytännössä työskentelee kaksi ja puoli päivää.

Lopulta kaikki epäluulot osoittautuivat turhiksi. Oli palkitsevaa nähdä, että ryhmä pystyi tuottamaan tehokkaasti sisältöä, vaikka ryhmäytyminen tehtiin periaatteessa alle kahdeksassa tunnissa lukuun ottamatta sitä, että tiimi oli käynyt ennen tapahtumaa sähköposti- ja Facebook-keskusteluja noin kuukausi ennen tapahtumaa. Jokaiselta jäseneltä julkaistiin vähintään yksi juttu, ja tiimin sisältöä julkaistiin Freezinessa enemmän, mitä oli aluksi suunniteltu.

Olen tyytyväinen siihen, miten produkti toteutettiin, mutta itse tyylikirja eli produkti olisi voinut olla parempi. Siitä olisi voinut tehdä rohkeasti tiiviimmän poistamalla asioita tai vain tiivistämällä sitä, mitä tyylikirjassa kerrotaan. Asioita ei olisi avattu niin paljon, vaan kerrottu jämäkästi, että toimitus visualisoi näin ja näitä työkaluja ja asioita käyttäen. Tällainen menettelytapa olisi toisaalta saattanut rajata ehkä liikaa tulevien toimituksien tilaa toimia.

Lähteet

- Lankow, J, Ritchie, J & Croocks, R. 2012. Infographics: The Power of Visual Storytelling. Wiley. Luettu: 6.1.2016
- Kuusela, V. 2000. Tilastografiikan perusteet. Edita. Luettu: 28.12.2015
- Wong, D. 2013. The Wall Street Journal – Guide to Information Graphics: The Dos and Dont's of Presenting Data, Facts, and Figures. W. W. Norton & Co., cop. Luettu: 7.11.2015
- Gray, J, Bounegru, L. & Chambers, L. 2012. The Data Journalism Handbook: How Journalists Can Use Data to Improve the News. O'Reilly. Luettu: 7.1.2016
- Nathan Yau (A), P. 2013. Data Points: Visualization That Means Something. Wiley. Luettu 8.1.2016
- Nathan Yau (B) P. 2011. Visualize This: The FlowingData Guide to Design, Visualization and Statistics. Wiley. Luettu 9.1.2016
- Cairo, A. 2013a. The Functional Art: An Introduction to Information Graphics and Visualization. New Riders. Luettu 10.1.2016
- Infogr.am. About Us. Luettavissa: <https://about.infogr.am/> Luettu: 2.4.2015
- Google. Google Drive. Luettavissa: <https://www.google.com/intl/fi/drive/> Luettu: 2.4.2015
- Cairo, A. 16.3.2014b. Infographics to explain, data visualizations to explore. The Functional Art: Alberto Cairo's weblog on visualization, infographics, and data journalism. Luettavissa: <http://www.thefunctionalart.com/2014/03/infographics-to-reveal-visualizations.html> Luettu 1.6.2016

Liitteet

Liite 1. Handbook for UNESCO WPFD: Guideline for Data Visualizations

Liite 2. Guideline for Data Visualizations: Synopsis for Guideline for Data Visualizations



Handbook for UNESCO WPFDF

Guideline for Data Visualizations

Eemeli Martti

Contents

Contents	2
1. Introduction	4
2. Youth Newsroom Organization – Team Data-Driven Journalism & Online Production in 2016	4
2.1 Team	4
2.2. Technical requirements.....	5
2.3. Workspace	5
2.3.1. Tools: Google Drive and Spreadsheets.....	5
3. Process	6
2.4. Different Types of Data.....	6
2.4.1. Categorical data	7
2.4.2. Time Series Data.....	7
2.4.3. Spatial Data.....	7
2.4.4. Multiple Variables.....	7
2.4.5. Distributions.....	8
2.5. Data Narrative.....	8
2.6. Visualize the Data.....	9
2.6.1. Three Ways to Present Data: Visualization, Table and Text.....	9
2.6.2. A Successful Visualization.....	10
2.6.3. Missing Data Points	12
3. Data Visualization Formats	13
3.2. Different Chart Types	13
3.2.1. Common Glossary.....	13
3.2.2. Column and Bar Chart	14
3.2.3. Regular Column Bar Chart.....	15
3.2.4. Stacked Column Bar Chart.....	16
3.2.5. Grouped Column Chart.....	17
3.2.6. Regular Bar Chart	18
3.2.7. Grouped Bar Chart	19
3.2.8. Stacked Bar Chart.....	20
.....	20
3.2.9. Pie Chart.....	21
.....	21
3.2.10. Line Chart.....	22
.....	22

3.2.11.	Multiple Axes.....	23
3.2.12.	Scatter Chart	25
3.2.13.	Bubble chart.....	26
3.2.14.	Pictorial charts / Pictograms.....	27
3.2.15.	Thematic Map.....	29
3.3.	Colours.....	31
3.4.	Fonts.....	32
4.	How to Visualize Data with Infogr.am.....	33
4.2.	The First View.....	33
4.3.	Upgrades	34
4.4.	Let's Visualize.....	36
	Settings.....	36
4.4.1.	Making Things Visual.....	37
4.4.1.1.	Text Elements.....	38
4.4.1.2.	Pictures and Videos.....	38
4.4.1.3.	Facts & Figures.....	38
4.4.1.4.	Editor	39
4.4.1.5.	Exceptions in editors and how the values are plotted in table editor...	40
4.5.	Time to Publish	41

1. Introduction

Data visualizations have a relative long history and different charts are common in paper. In digital age data visualizations are strongly become related to data-driven journalism.

This guideline defines the formats of data visualizations made by Youth Newsroom and the methods used to visualize data. It is also meant to be a tutorial for persons who have never visualized anything. The principles are easiness, simplicity, informative, but still stylish.

I have made guidelines to usage of basic charts and how to make good visualizations with them. At the end of the guideline there is also a guide to use Infogr.am which is a very simple and easy visualization tool to make quite appealing charts.

But before you can visualize, you have to analyse your data. Therefore I have gathered some of basic things to look from data sets and also written down the team configuration and workspace we used during World Press Freedom Day 2016 held in Finland.

UNESCO has published some infographics about the freedom of speech, but there are no guidelines. Therefore this guideline will be a somewhat foundation for that. I hope that in future this guideline will be seen as useful and a ground work to somebody for developed.

It should be remembered, that in the end the visualization is as good as the data behind it. The data defines what kind of stories and visualizations can be made of.

2. Youth Newsroom Organization – Team Data-Driven Journalism & Online Production in 2016

2.1 Team

The data team consist altogether seven members: one team leader, one team subleader and five other person who are working with data-related content. The required skills for the team members are:

1. The person can use spreadsheets program like Google Spreadsheets.
2. The person can write data driven stories and make visualizations.

3. The person can use data visualization tools like Infogr.am to make visualizations.

Members work in pairs or in the groups of three. Work pairs and/or groups plan and write the data driven stories and make visualizations. The leaders gets an approval to the stories from the editor-in-chief.

The team leader is also in charge of the event's website. Team's subleader is working as a substitute website admin, in case the leader is unable. The admin is responsible for that the website supports different content like interactive data visualizations, videos, and the possible problems are fixed.

The amount of stories for the event is at least three stories but more is preferred. Three stories would be published on each day of event and if possible before the event for example one in week. However this depends on how much there are other kind of stories published beforehand. It would look silly if the Youth-Newsroom would publish only data driven stories before the event.

Copyright contracts will be signed at the event.

2.2. Technical requirements

The Youth Newsroom makes interactive visualizations that are published mainly on web page but also in print. Chart elements are embed to the web page with script or iframe tag. It is necessary that the site supports scripts or it is possible to embed charts with iframe.

2.3. Workspace

2.3.1.Tools: Google Drive and Spreadsheets

Data analysis can be made with spreadsheet program like Google Spreadsheets, which is simple and used in online. The data can be accessed quickly and in any place that with a computer and an internet connection.

Google Spreadsheets is part of Google's cloud service, Google Drive which can be used to collaborative work. For workgroup that works with great distances Google Drive and Spreadsheets are good choices.

3. Process

Visualizations should not be considered as separate part of the whole process. Instead it should be started early on. A common mistake is to form visualization first and get the data later. It should be done otherwise: data first and visualization follows.

At first phase data should be gathered and analysed. If the use of data needs permission, it must be obtained.

The data is only as good as its source. It is crucial to gather data from unbiased source. Still, data should be evaluated critically. Even one wrong data point can make the whole chart unreliable.

At edit phase key message should be recognized and the best data series to illustrate the point chosen. Data has to be simplified and filtered to deliver the essence of the data to audience. Numerical adjustment can be made – such as to change absolute values to percentage change – to enhance the point.

After editing phase, the right chart should be chosen to present the data. This phase also includes adjustment of the scale, writing of the title, description, legends and source.

The last phase is the review phase. Plotted data should be checked against the sources, and the chart's sensibility should be evaluated as well. Data can be verified from additional sources, or it should be consulted with experts for questionable content and outliers, if needed. It is also important to try watching the final chart from the reader's perspective.

2.4. Different Types of Data

The data specifies what kind of stories can be told: How much there is to be found and what visualization methods are useful. Data can be divided to different types and every data set requires individual analysis.

During the exploration phase, it can be beneficial to see an overview of the data first and then decide where to go from there.

2.4.1. Categorical data

Categorical data is data that has been categorized to a different kind of groups. When the data is categorized, the minimum and maximum should always be looked for as they give a sense of range of the dataset and is easily found with a quick sorting. Distribution – are the most values high, low, or somewhere between – should be also looked for. Finally there is structure and patterns – why and what makes the categories similar or different?

Horizontal bar graph and symbol plot are suitable for visualizing categorical data. When the parts of a whole are wanted to be represented, pie chart and stacked bar chart are a good choice.

2.4.2. Time Series Data

When visualizing time series data, the goal is to show what has passed, what is different, and what is the same. Usually the magnitude of changes between time segments is more interesting than the value at each point. When dealing with the time series data, changes over time and the nature of the changes should be looked for. Column bar chart, line chart and dot plot charts are useful ways to visualize time series data.

2.4.3. Spatial Data

Spatial data is data that can be categorized by locations. The most obvious way to explore spatial data is maps. However a map is not always the most informative way to visualize spatial data. The regional patterns should be looked from spatial data.

2.4.4. Multiple Variables

The data is often in table form, in multiple columns which each represent a variable.

Though the data might fit onto screen and the relationships between variables and trends can be interpreted, the relationships between variables are not often straightforward. There is no clear increasing or decreasing trend. In such cases, multiple views using more straightforward charts can help a lot.

For example, with time series data should be looked for how a variable changes when another variable, time, does.

Relationships can be visualized similarly to how you look for them with time series data.

Causal and correlative relationships should not be mixed. They look the same when visualized, but correlation is more difficult to prove than causality.

Assumption should not be made when exploring relationships. When it comes to correlation and causation, all the context is needed to take in before can be assigned to causality.

2.4.5. Distributions

Means and medians are used to describe a group of people, places, or things, and these measurements typically imply what is normal or average. Anything that is far away from these measurements is abnormal or above or below the average.

Different plot charts can be used to analyse distribution summary as histogram and density plot can be used for analyse the distribution of one variable. The distribution of multiple variables can be examined for example with a heat map.

2.5. Data Narrative

Visualization is often framed as a medium for storytelling. The numbers are source material, and the graphs are how the source is described.

When having a data story, there should be talk about data in a way that fits with the context, instead of just arranging graphics in a way that looks good. Readers can be taken through analysis, from big picture down to the details, and notable data points or vice versa from case study to overall summary.

Readers can be connected with visualization by letting them see the data as it pertains to them. People spend more time interacting with data that they somehow see themselves in or can relate to their surroundings.

If the work is more like a reference than a story, the data can be scattered into groups or categories. It is all about how it is thought that the audience will read graphics, coupled with how it is wanted that audience will read them. The same goes for charts embedded within text. There should be continuity between the visualization and the words which sometimes is as simple as paying attention to where a graphic is placed in text.

The one-column layout is common online. It is easier to place figures within the flow of the text like at the bottom, halfway or $\frac{3}{4}$ of the page. Text and graphics should be thought as single unit instead of separate parts.

In Youth Newsroom, visualizations are part of story. They pass on information and can demonstrate the phenomenon better than text, but and also act like pictures, as elements that complete and variate the whole story.

The data driven stories can contain analysis and opening of the data, but more important is that data has "faces". The phenomenon behind the data is concretized with examples from world and with real persons. This method makes story more meaningful for the reader.

2.6. Visualize the Data

2.6.1. Three Ways to Present Data: Visualization, Table and Text

Data can be presented in three ways: visualizing it with charts or presenting it with table or just writing it down in text. The usefulness of gathered data depends on how you present it.

Visualizations can transmit information quickly and illustratively. When compared to plain text, well designed visualization can tell a lot more. Readers can also remember the information better from visualization than from the text.

Tables are good choice if large mass of data is wanted to be represented in ordered and controlled form. Also, if having very few data points, or if the data doesn't have much variability, clear trend or conclusion, but still the information might be useful for readers, it is better publish the data in table. It can be an efficient layout for simple information.

A table should not be used unless a huge amount of data has to be included and the space is limited. A table does not have any visual impact and it requires a lot of work from the reader to compare and contrast the data.

Whenever there is space in table, it is always helpful to chart the column of data that is the main message. Expressing quantitative and descriptive information in a table is often the simplest method of presenting copious amounts of data. However, table should be used sensibly and in most cases as a last resort. A chart is more memorable than a table.

The data visualization is only as good as the data behind it. The data alone doesn't always tell the story in the most compelling way. A simple chart can be useful, but narrative can be more immediate and impactful to a reader.

Charting is a powerful tool that puts a series of numbers in a close proximity to each other. The numbers in a chart convey information to the reader both visually and narratively. The same set of numbers looks more concrete and precise when charted than when presented in a story or a caption. A chart is more memorable than a string of numbers held together by words. It also allows reader to make a judgement at a glance.

2.6.2. A Successful Visualization

A good visualization is visually inviting, has a rich content, and is executed with sophistication. The rich content brings meaning to visualization, the invitingness highlights the essence of the information, and the sophisticated execution brings the content and the graphics to life.

A high-quality chart has not necessarily be a complex presentation which is artistically made with decorations. Also a colourful chart doesn't make visualization successful. It may ruin it.

When making visualization, tools, situation and the audience has to be considered. In this case we are using making visualizations from freedom of speech -related data with Infogr.am to web page and paper. How does the visualization look on both web page and paper and who will view the charts must be considered.

The benefit of a successful visualization is naturally data. The visualization is only as good as the data. A good visualization can't be made from unclear mass of data and vice versa. A missing chart is better than a failed chart.

The main goal of the chart is to pass on information so that even with a quick view the reader can't misunderstand it.

Because the chart communicates with symbols and visual comparison, a successful chart doesn't always need additional numbers (for example in a bar chart on the top of the bars). One of the meanings is to give reader a quick notion about the nature, interrelationship, correlation and order of presented things.

A high-quality chart is a truthful, well composed presentation from interesting subject. The chart forms from complex things which are visualized clearly, precise and efficiently. The visualization gives a mass amount of insights in a short time using a small amount of colour as possible.

Truthfulness is acquired with detailed and thorough labelling which attaches chart's symbolic message to reality. There are no information loose from content or context. The chart presents the information and the change in information, but no presentation technic or the change of presentation technic.

Chart should be presented in the right context. Viewer can interpret the message of the chart correctly and place it to right context and relation with other information. This requires that axes are marked clearly and precise as are the variables. Title and the description are made so that they make attention to time and place.

The demand of the right context means usually that the comparison is expanded to greater compilations so that it is possible to give meaning to the represented information. Another good aspect in a good chart is that it answers to the questions it has made. To be truthful, chart has to give answer to question: compared to what?

Visual comparison is the most important thing in a chart, because the information related to quantities is born only from the comparison of different parts of the chart.

Additionally visual comparative concerns charts that are side by side, one below another or are made with the same construct. For this reason the parts of the charts has to be comparable. If not, a so called visual labyrinth is born. A gaze wanders in charts and tries to understand the message but does not succeed. In worst case visual labyrinth is not born, but the assumption is false.

A chart that passes on information efficiently and fast has one important feature: simplicity. A good chart can tell complex information, but can also have a simple structure. A chart can have artistic value, but it must never be used as decoration.

Visual simplicity is composed from different things. Scales are used right, data density index is high and the amount of chart junks is small: colours and pictorial icons are used with consideration.

Diagonalization is an important part of visual simplicity. The perfect equivalence of things appears best if points are placed on diagonal. The most important phase is to assemble chart elements to decreasing or increasing order whenever it is possible.

2.6.3. Missing Data Points

In case that the data is incomplete, but a trend must be shown, it is still worthwhile to plot the data, if only a few data points are missing. It has to be assessed, that does the chart presents an unbiased story though it does not have all the data points. It is acceptable to complete the data set with combining data from different sources, as long as the sources have the same methodology for data collection.

Usually a chart with one or two missing data points still has value. The exception is when the missing data point is crucial to message.

With bar chart, a space should be left for the missing data point and point it out in a footnote. If more than 20 % of the data points are missing, bar chart is a bad choice.

If a line chart is used to visual a long data series with a few missing data points, the gap should be spanned and continue the line. Since the line chart represents trend, a few missing data points don't have a big effect to the final result. However, with a short data series, a line chart should not be plotted if more than 20 % of data points are missing.

If there a several missing data points in data series, a sample of directly relevant data points that convey the message should be selected. Additional data to substantiate the chart should be researched.

The most recent value should be plotted and historical data points selected that are directly relevant to the message. In addition, new information to make the chart more demonstrative should be incorporated.

A pie chart should not be made if the data for any segment is unavailable, as the pie chart always represents a whole.

3. Data Visualization Formats

Infographics and data visualizations can be divided to three formats: static, motion and interactive. In Youth Newsroom, static visualizations are used in print if needed, and the interactive visualizations on the web page.

Static visualizations are effective at representing rich data in a single image. The user interaction consists of viewing and reading. Static visualization can be used in wide range of applications, and are also simple to execute. The context and purpose of static visualization is determined by the layout and size of the image itself. Static visualization is usually used in print and web page. The purpose of the content may vary, but the overall size and shape of the visualization is mostly determined by the demands of publishing context.

Interactive visualizations: User interactions can consists clicking, searching for specific data, actively shaping the content displayed, and choosing which information is accessed and visualized. Interactive visualizations can be narrative, explorative, or both. With simple interactive methods like clicking an element in chart, there can be presented more information.

Interactive visualizations are particularly useful when there is a vast amount of data and interactive content that draws the user in to encourage further exploration is wanted to be created.

Interactivity allows you and your audience to engage with the information in deeper way and to unearth new insights.

3.2. Different Chart Types

3.2.1. Common Glossary

The charts base on a simple structure. Some parts are common to all charts, but some of the charts also differ from each other, so they also have their special structure.

Gridlines are the dim lines that run across the chart from axis to axis. In line and scatter plot chart gridlines can run in vertical and horizontal direction and in bar chart only across the bars. The necessity of the grid lines depends on the size of the chart. If chart is small, the grid lines are not necessary. However in large charts – especially when comparing details – gridlines are great help.

Both **axes** are marked with used scale and unit of measure. Axes are divided in equal parts. In line and column bar chart the y-axis is usually called quantity scale and x-axis time scale.

Scale units must be chosen so that the related numbers are easy to interpret. For this reason there should be used numbers that are dividable with five, ten, hundred or thousand with quantity scale.

A legend gives the reader the key to the information. It is confusing to list the elements of the legend in a different order than they are in the chart.

Text elements in charts are as central part of the presentation as is the chart. Without them the chart is usually impossible to interpret correctly.

By common advice, the **title** of the chart should answer shortly but exhaustively to questions, what, where and when. Another option is to use title that tells something about the phenomenon which is represented in the chart.

The **description** can be placed either on top or below the chart. It is usually broader than and longer than title. It describes the chart and usually contains something that interprets the chart.

Different labels are used to point out which part of the chart does mean. It is usual to place tooltips to separate legend below the chart.

3.2.2.Column and Bar Chart

In this guide a bar chart that is placed on horizontal axis (x-axis) is called column bar chart and bar chart placed on vertical axis (y-axis) is called simply bar chart.

The bar chart creates association to quantity and answers to question, how much? However, trend or variation regarding to the trend does not rise as strongly as in line chart.

Although column and bar charts are different charts for different utilization, they share same constructional elements.

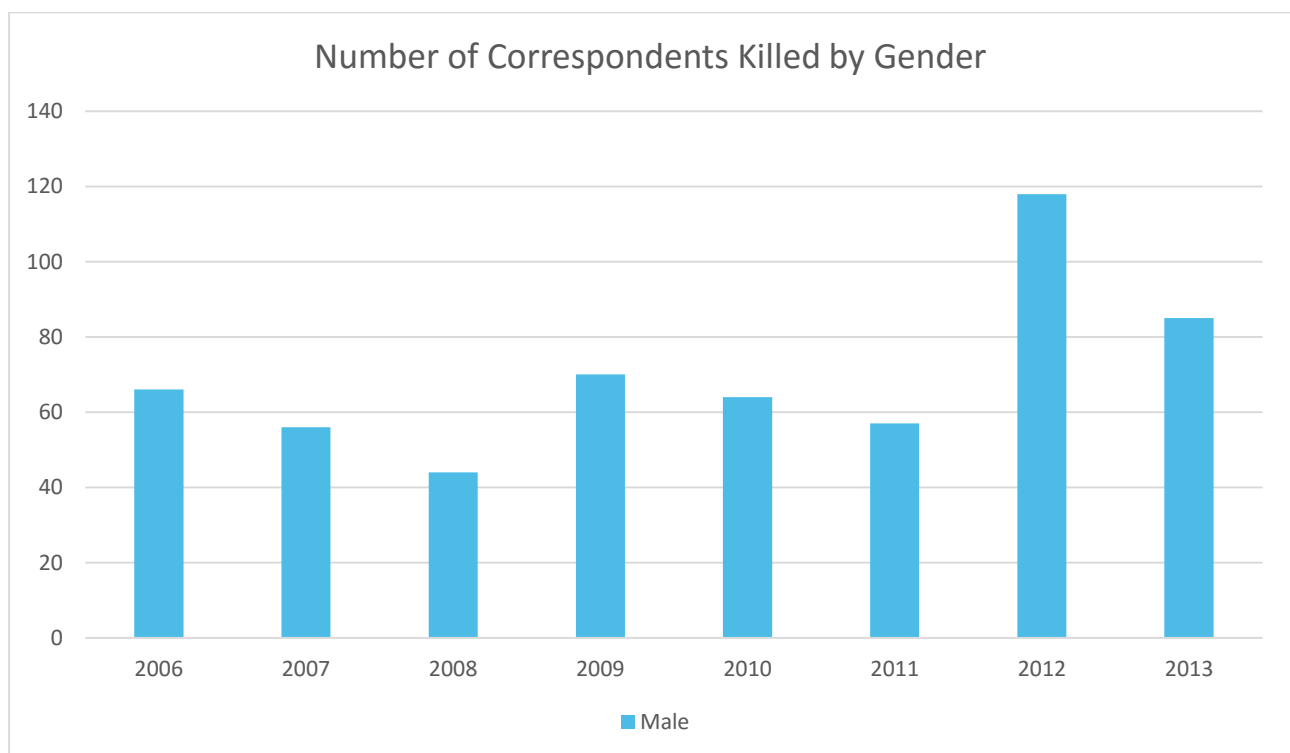
There are approximately same amount of different bar and column charts which mainly work on the same way. The difference is that column charts are bound to order of information, for example time, whereas with bar charts you can't often make so called

natural order. Because of that, the chart has to be composed by bars. Usually bars are placed in length order, which is balanced and more informative as you can see immediately the quantity order. Although, if the visualization has any particular factor that can be used with ordering, there is no reason to break it.

In regular column and bar chart all the bars should be the same colour and shade since they measure the same variable. However, a darker hue can be used to highlight a particular bar. With grouped and stacked chart the shading of the bars should move from lightest to darkest for easy comparison.

One of the most common mistakes made with bar and column bar chart is to cut of the scale. This is done because it is thought that the line on top of the bar is the part that transmits the information. Cutting the scales vanishes the visual comparison from the chart.

3.2.3.Regular Column Bar Chart

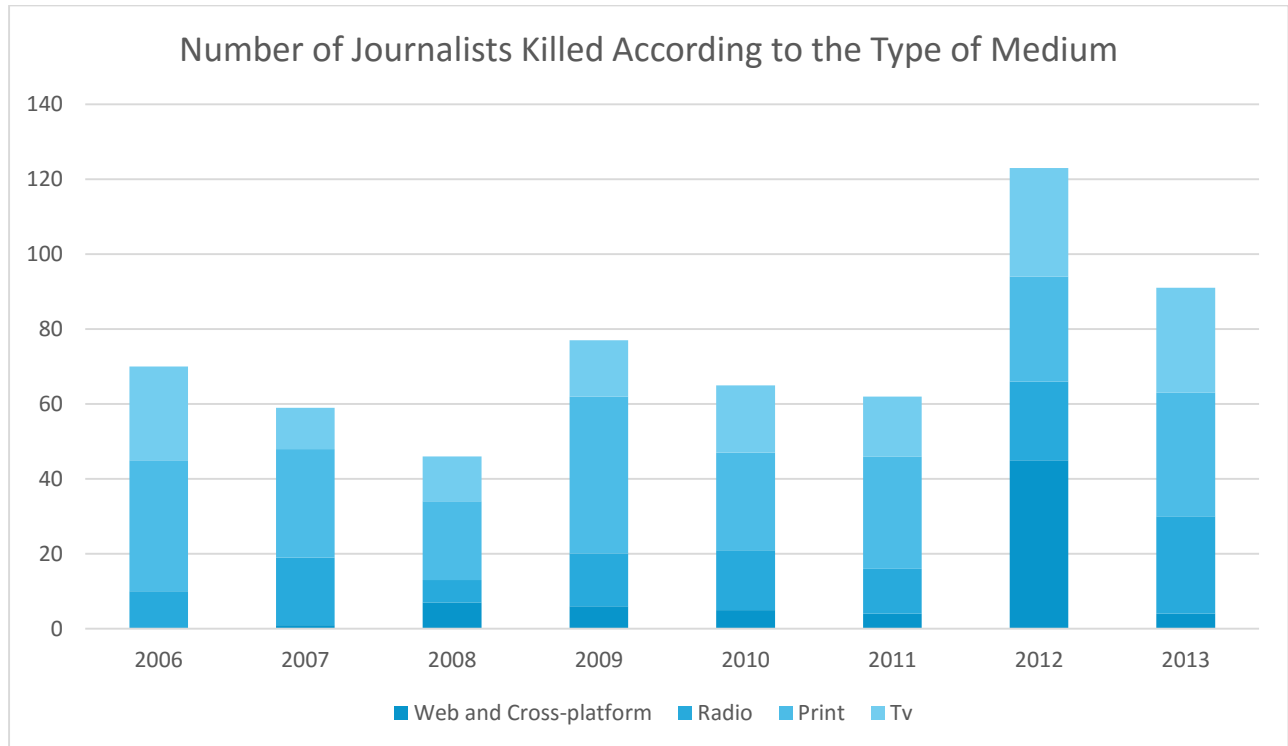


Regular column chart is usually used to represent total or relative amount and its change or trend in time. Regular column chart is useful in representing time series, but basically any other continuous variables on can be put on x-axis.

The column chart is usually alternative to line chart. They fit in same kind of situations, where bar chart does not. As the column chart is used in same kind of situations as the

line chart, there are same requirements to use them. So when using a column chart, there should be continuous value placed on x-axis such as time.

3.2.4. Stacked Column Bar Chart

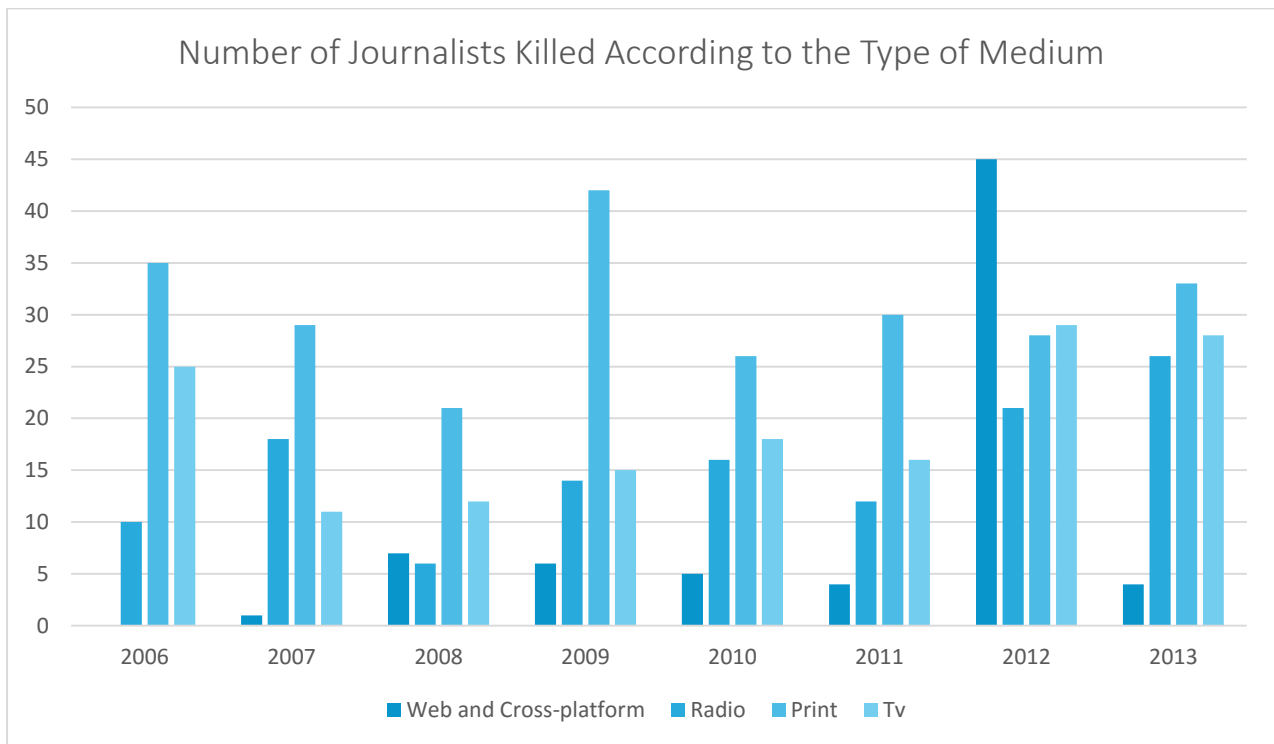


A stacked column chart represents the change of total progress in one chart. Stacked bar chart highlights the sum of its parts.

Bars in the stacked column chart can be made to have the same length even though they would show different order of quantity. Then bars portray relative quantities and the chart itself is good alternative to pie chart.

When colouring bars in stacked column chart, it is the best way to colour them so that the darkest or the most intense hues are at the bottom and the lightest ones are on the top.

3.2.5. Grouped Column Chart

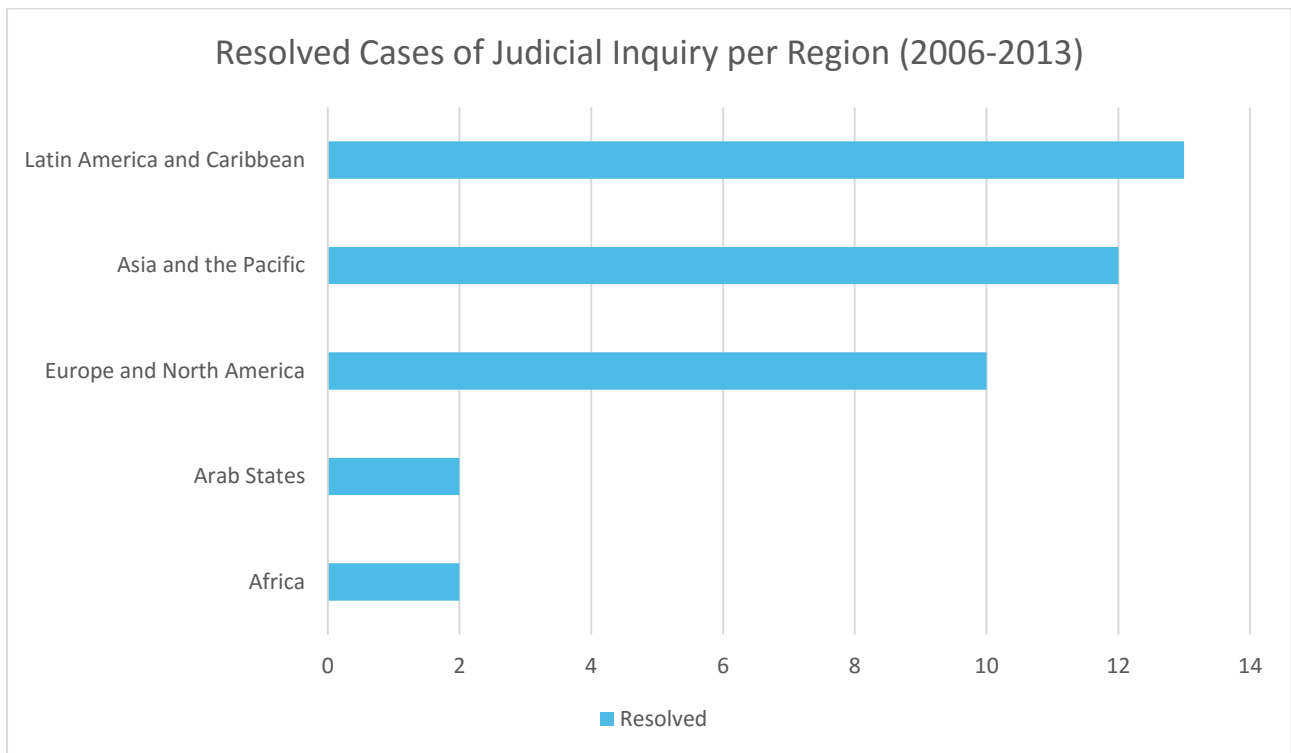


The basic idea behind the grouped column chart is basically same that what was in the regular column chart, except instead of one bar there is groups of two or three bars. With the grouped column chart it is possible to represent more time series in one chart.

Bars has to be in same order in every group. It is also preferable that bars in one group are ordered from highest to lowest if possible.

In the groups two and three bars in a group is still distinct. However, four bars in one group can be too much. If there are too many bars in group, it is hard to see a single group or bar.

3.2.6.Regular Bar Chart

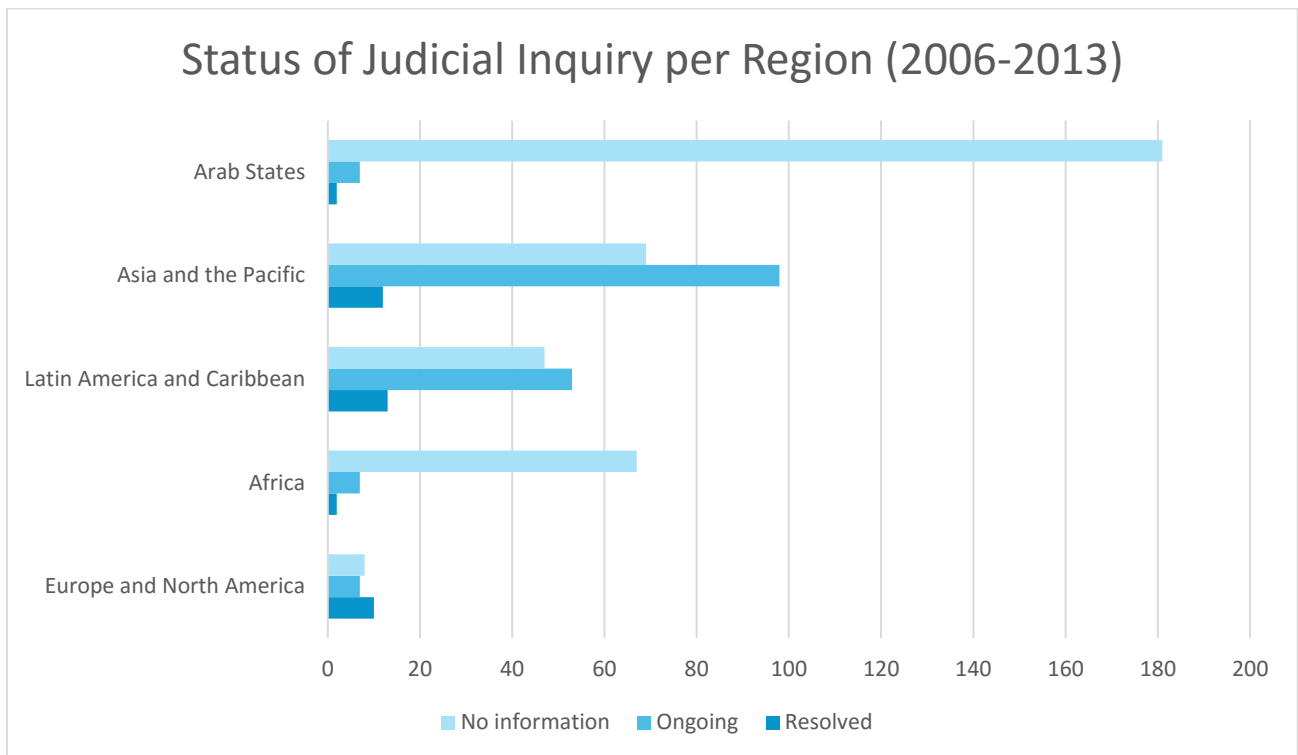


A regular bar chart is useful for portraying comparison of quantity or magnitude of changes of one thing or different classes.

A bar chart has not as tight requirements as the column bar chart – the continuation does not associate to y-axis as greatly it does on x-axis – as column bar chart, so it can be used more extensively.

Bar chart should be used over a column bar chart when the y-axis is not perceived as continuum as strongly as the x-axis, bar chart is at least as illustrative as column chart, You can use informative means that work poorly on column chart, and there is no particular reason to use column bar chart.

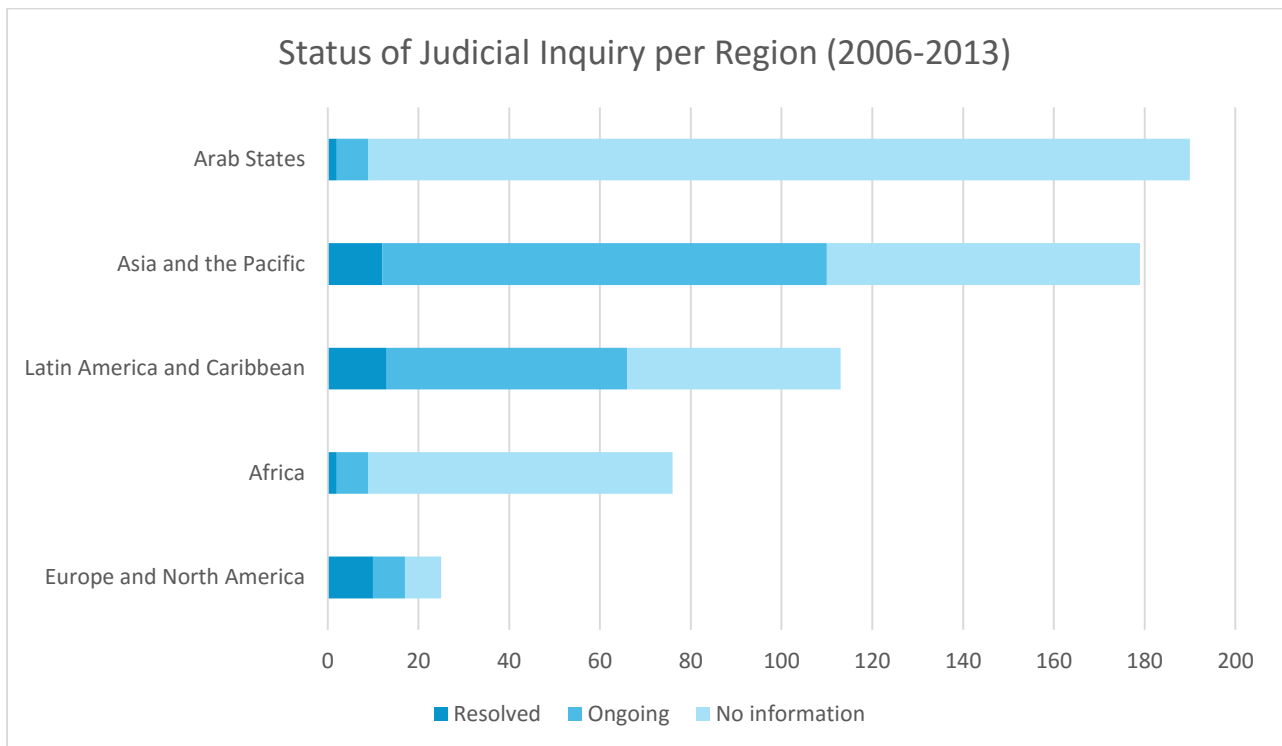
3.2.7. Grouped Bar Chart



In a grouped bar chart, there are two or three subdivisions in each class. Three or more bars should be used only when it is absolutely necessary. If too many bars are used, it will be difficult to compare different classes and the bars will be too narrow.

It is preferred that the group which has the longest bar is placed uppermost or lowest. After that the rest of the groups are placed in length order.

3.2.8. Stacked Bar Chart

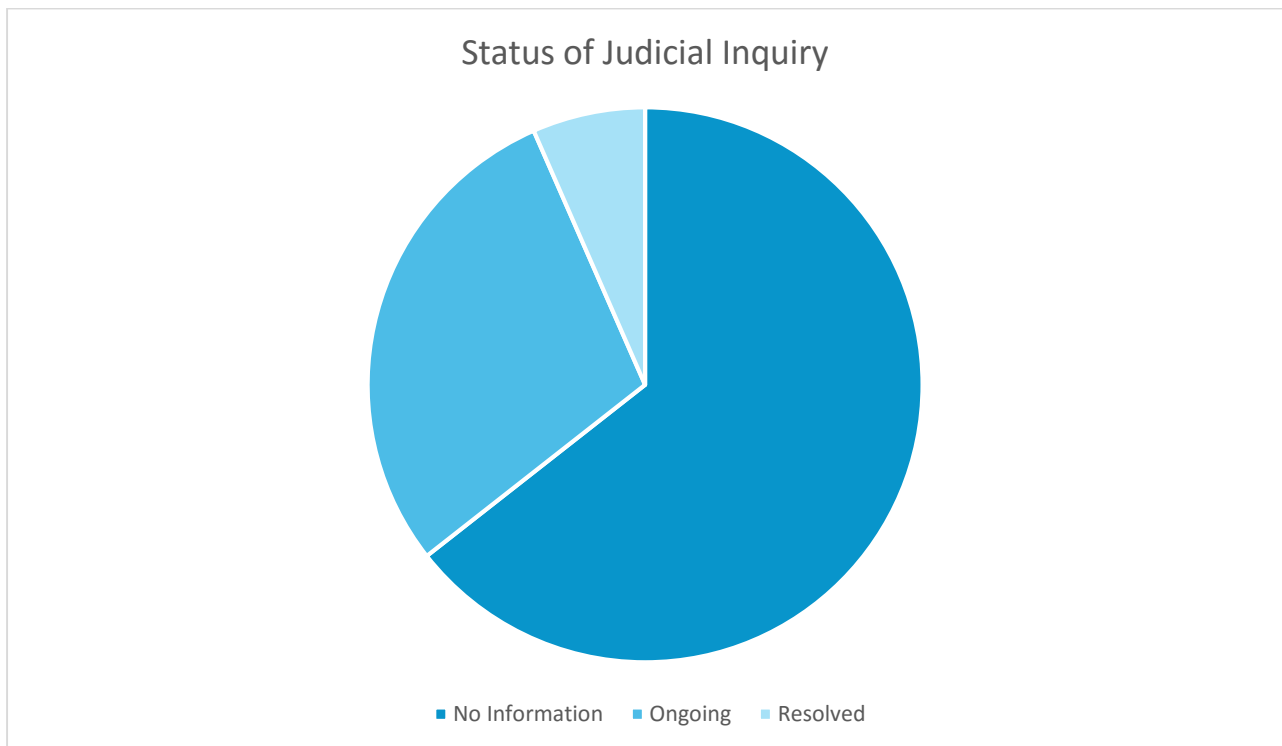


Stacked bar chart can fit in to same kind of situations like a grouped bar chart, though stacked bar should be used over a grouped bar chart when representation of total quantity is more important than the portion of single parts.

A few more subdivisions can be plotted in a stacked bar chart than in grouped bar chart, without ruining the whole chart. Six subdivisions is usually the maximum.

Stacked bar chart is also a good alternative to pie chart when having many components to compare. When stacked bar chart is used over a pie chart, every bar is made to be as lengthy. Then chart portrays relative distribution of subdivisions and the unit is percent just like in pie chart.

3.2.9. Pie Chart



Pie chart represents on how large pieces something has divided. Alone pie chart can't detail anything else than division of whole as the composition of the chart simply doesn't support other interpretations.

From the aesthetical point of view pie chart is better than bar or column chart. Although the examination of stacked bar and column chart is quicker and more precise as the viewer can see the length of the bars easier than examine the size of sectors in pie chart. In addition, it is faster to compare the bars than sectors.

There are different principles how to place each sector. Usually the largest sector is placed to 12 o'clock and then continued clockwise with placing the second largest sector next to the first sector.

The largest segment can also be placed on 12 o'clock and then continue counter clockwise. The most important slices comes on top and the less significant ones fall to the bottom.

When the sectors are ordered by their size from the largest to smallest, the result is more illustrative.

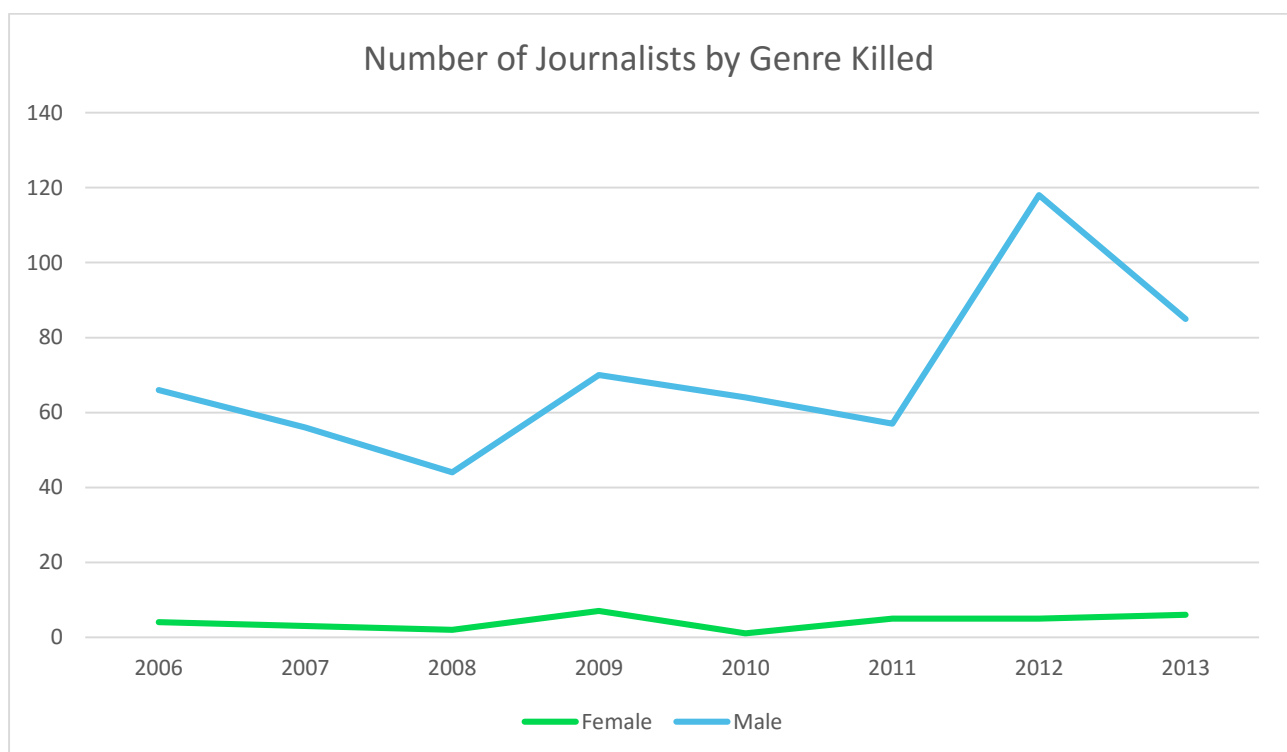
You can also mark the numbers to near sectors. Marking the percentage values is not necessary although it is common, since the pie chart itself tells them. If the reason to add percentage values is precision, it should be thought, is the pie chart the right way of representation.

The common mistake with pie chart is to divide it to too many sectors in which case the visual comparison is very hard. The maximum amount of sectors is five and sometimes even that is too much. If there are more than five segments, the less significant segments should be combined to one slice and label it as "other".

Pie chart is very common due it is seen more appealing than the other charts. Still it has been critiqued for being too imprecise. If you put too pie charts next to each other, you can't really tell the difference.

There are some variations of pie chart, one being the doughnut chart. It acts the same way like pie chart and its composition is also the same. Doughnut chart is used for the sake of variation, but it is even more imprecise than the pie chart because the compared sectors are even further from each other.

3.2.10. Line Chart



A line chart is basically used to represent the change and progress over a period of time. Line chart emphasizes a trend line and the relationship of change to it. The line

chart can show massive amount of data in a small space. It is best used with continuous data series. The line chart suites for showing of trend, acceleration, deceleration, and volume

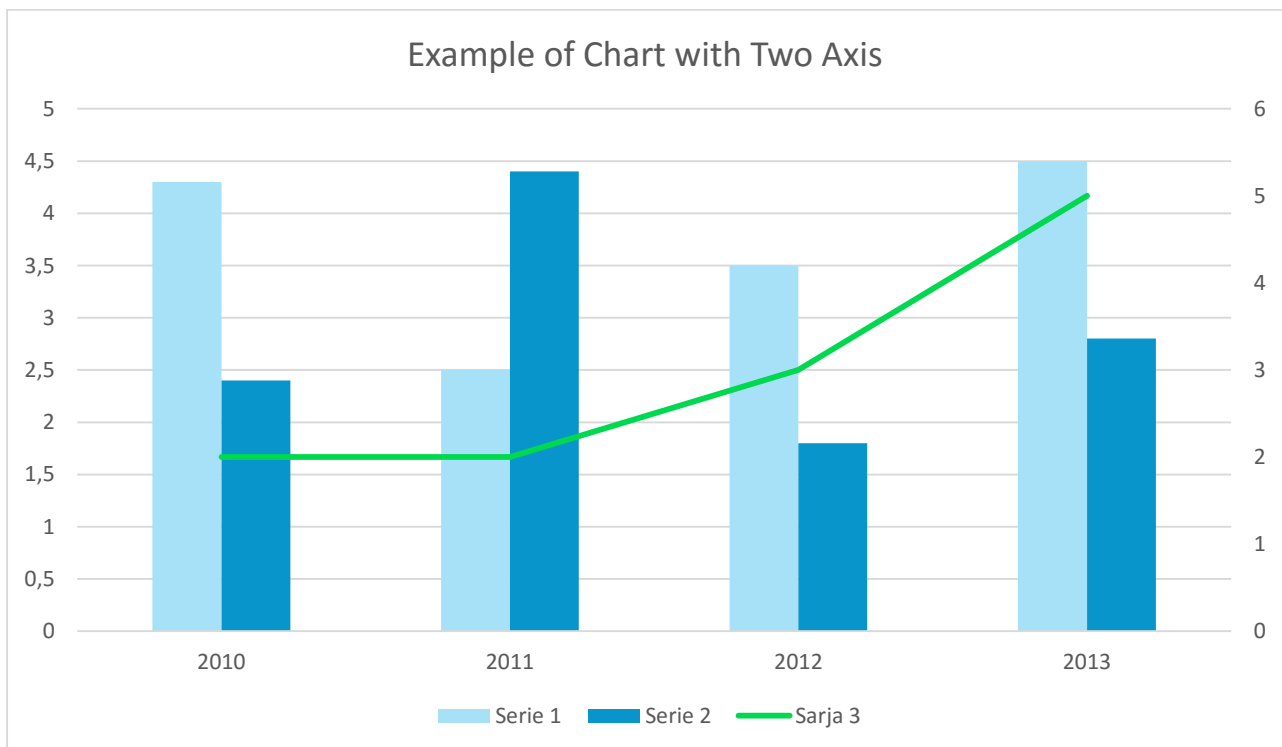
Like in bar chart the Y-axis represents quantity and X-axis time. Axes must be set with carefulness. As the purpose of a line chart is to show trend so choosing a y-axis that yields a flat line (obscures the message) or an exaggerated line (overstates the trend) must be avoided.

The right height for the line chart is two-thirds of the chart area. The scale of the Y-axis should adjusted so that the height of the line occupies roughly a two-thirds of the chart area. Unlike a bar chart, line chart doesn't require a zero baseline.

More than three or four lines should not be plotted on a single chart. The purpose of multiple-line chart is to compare and contrast different data series.

The most import line should be one colour and the other line should be shades of a second colour. With more colours, chart will be confusing may be illegible for colour-blind people.

3.2.11. Multiple Axes



Usage of equidistant scales is one of the primal requirements of visual comparison. This applies to all charts which use the coordinate system. This applies also when there are two different scales on a y-axis in one chart.

Generally usage of two different scales in one chart should be avoided, because it easily leads to a chart, where reader has to view long before understands it. In the worst case, reader understands the chart wrongly.

Despite of the dangers, usage of two axes is sometimes necessary. With most phenomena – especially correlation or lack of correlation of different time series is easier to see if they are plotted one on the other. It is a fact that when visualizing two interrelated matters in one chart, it is possible to reveal more from their reciprocal relation than with two charts. If two different phenomena which are measured with different units, are plotted on same chart, two different scales are naturally needed.

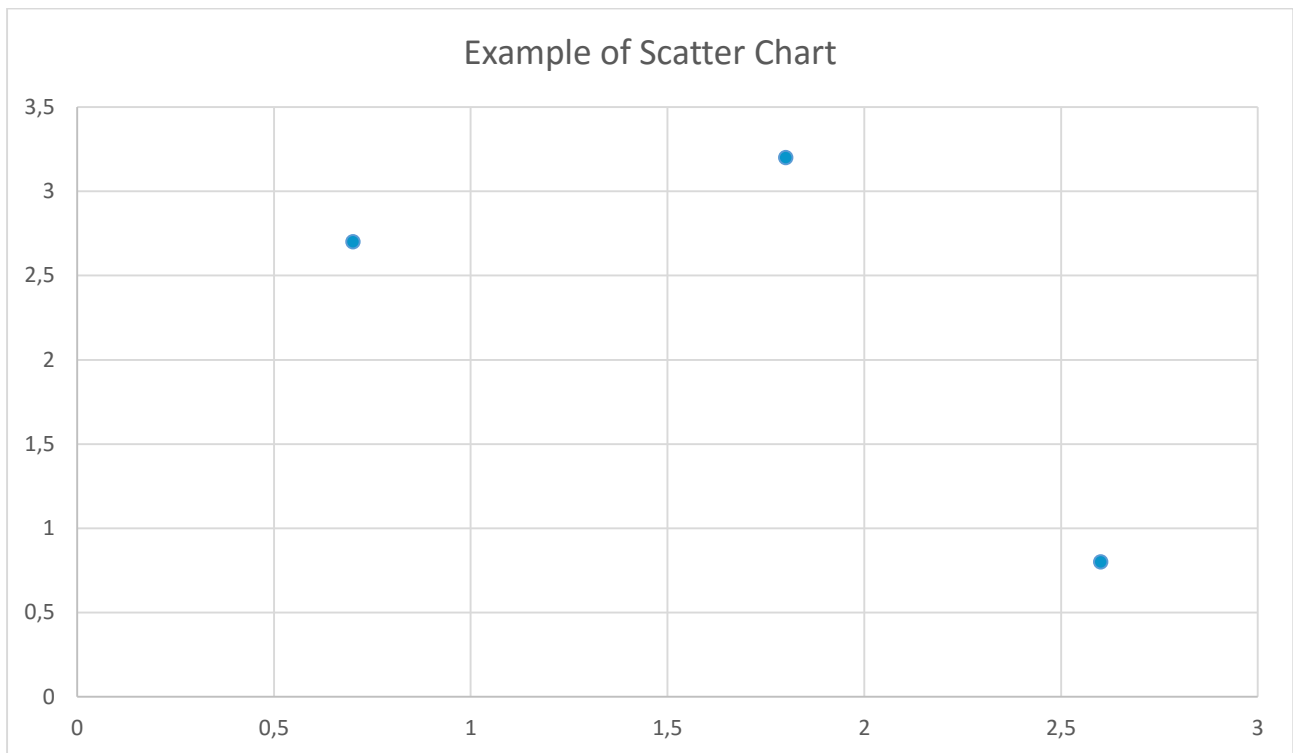
The line should be used for continuous data and bars for discrete quantities. Deviation should not be used for stylistic reasons. The only exception is when both data series call for a chart with vertical bars. In such occasion line can be used.

When two scales are used, two different graph types are also used. Usually line and column bar chart are used. To avoid any misinterpretations – which scale is for which graph – should be mentioned in chart.

Usage of two different scales can be made with two different principles. Either so that the line and column graphs are clearly apart or so that the graphs are one on the other. In the latter situation both graphs have to be seen.

When two scales are used, the final result should be always review critically.

3.2.12. Scatter Chart



Scatter chart is one the simplest charts. In scatter chart every variable is placed with a simple mark on coordinate system. Chart type is efficient because the human eye is precise to perceive the shapes, regulars, and irregulars that develop on the chart after all variables have been placed.

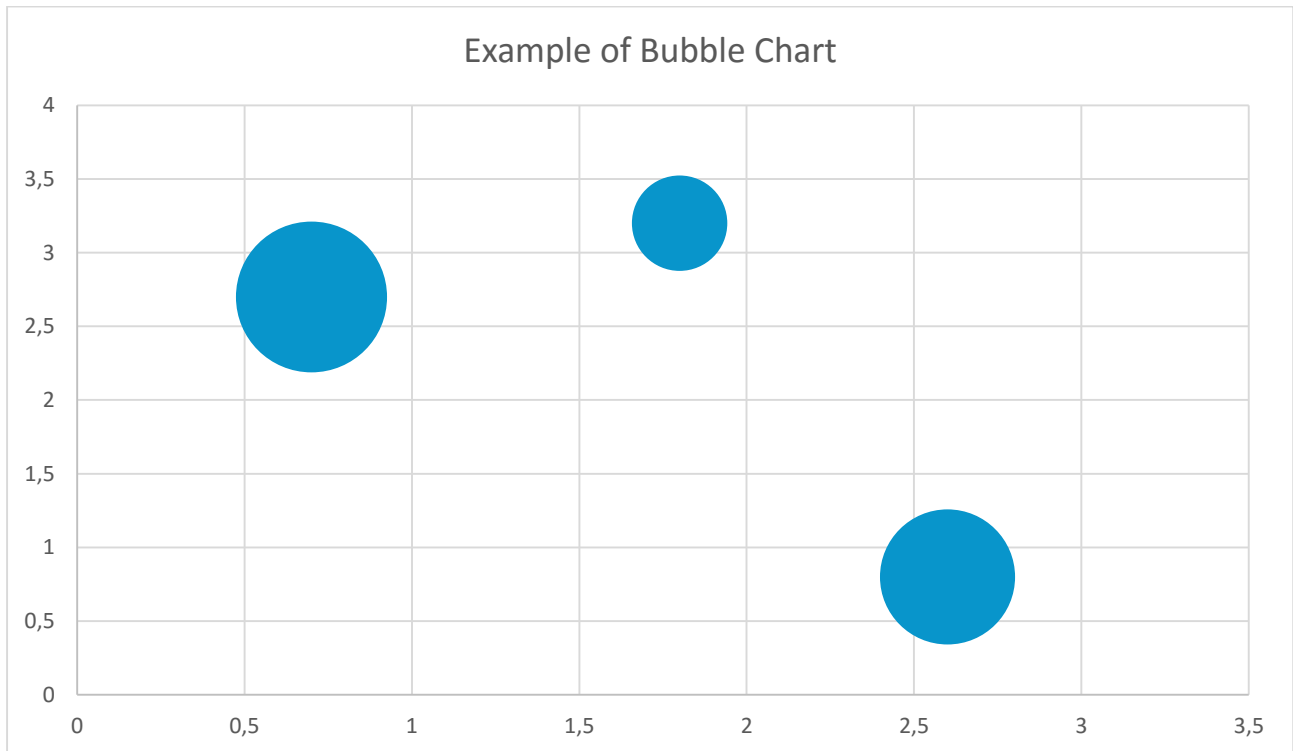
If there are lot of variables to be plotted on chart, scatter chart is not the best choice. Since the dots can be set on same spot, they do not stand out.

With scatter chart it is easy to state correlations and density and it easy to understand even with little knowledge of statistics. In addition scatter chart can be used to bring out the groups in data and describe their differences.

When scatter chart is used to visualize causal connection, x-axis should be used for phenomenon that is believed to be the reason and the y-axis for what is believed to be consequence. Otherwise the connection between reason and consequence may fade out.

Scatter is not quite common due it is not as visual as other charts. Still it is a supreme chart to visualize correlation.

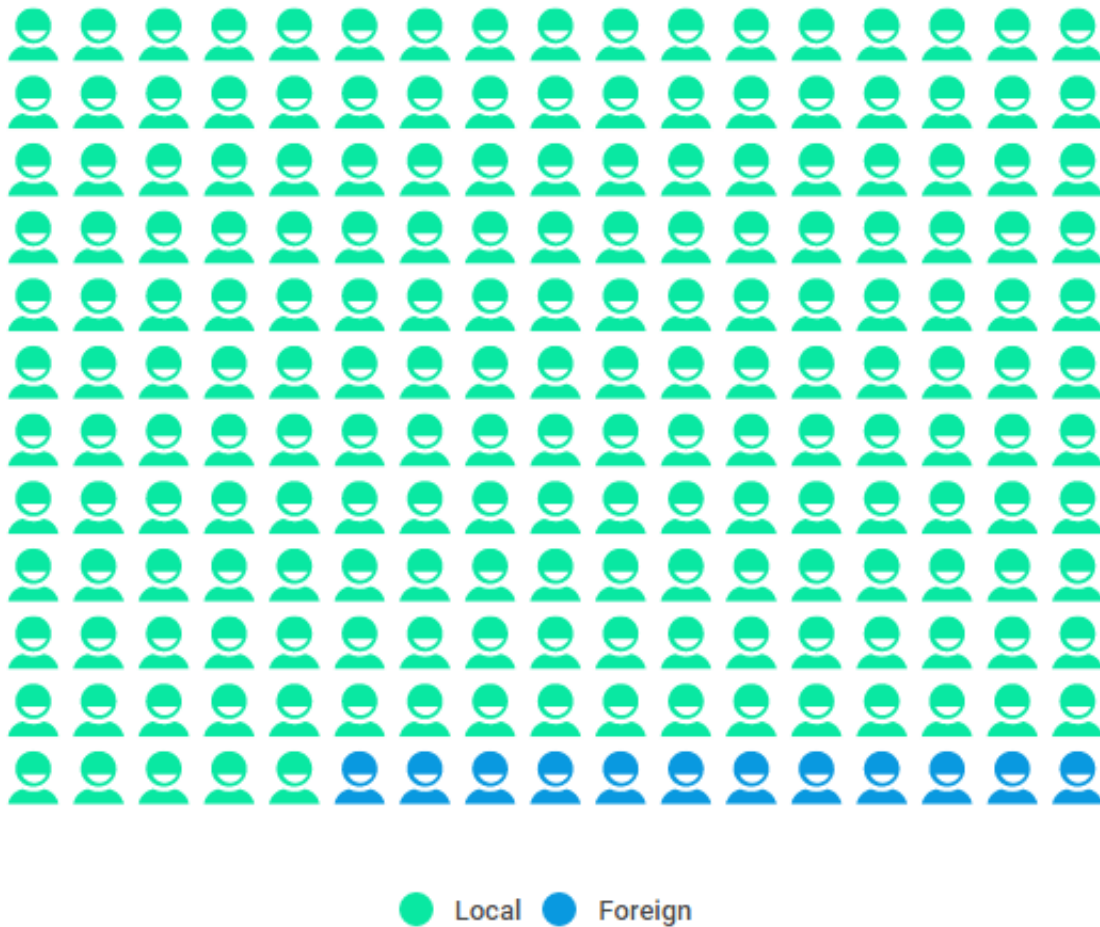
3.2.13. Bubble chart



Sometimes it is necessary to visualize great amount observations with a scatter chart. Because in basic scatter chart there is a danger that many observation sets in exact same place, it is worthwhile to visualize large data set with a bubble chart.

Bubble chart works with same logic as scatter chart, but in addition it has a third dimension: area. Numbers are detailed as bubbles or circles and their area grows according to count. With bubble chart three or – if colours are used as well – even four different things can be visualized at the same time.

3.2.14. Pictorial charts / Pictograms



Pictorial chart are visually more engaging than traditional charts, but also less effective for comparing a large amount of data. The purpose of pictograms is to create a snapshot of data or they can use pictorial elements like male or female figures same way as column and bar charts use bars. A pictogram should be used only when comparing a few simple data series.

Pictorial charts are used because of their appeal so they are a method to attract those attention who otherwise would not notice charts. An effective pictogram is visually engaging and gives a quick comparison of the variables.

Pictogram can give a quick snapshot of quantity and volume, but do not suit for visualizing massive amounts of data. A bar chart is far more effective than a pictogram when comparing discrete quantities of several complex data series.

Pictorial chart is good when one figure represents certain quantity. Thus figures are kind of units. When figures are placed one after another or side by side, a pictorial chart is made. For example one figure can represent 10,000 people and when figures are used 100,000 is equal to 10 figures.

Choice of symbol should be considered with care. Certain symbols suit in different situations, but above all, symbol is just a method of calculation so it should be efficient in that task. Symbol should be some common shape or construct like male or female figure.

Icons that are visually interesting don't necessarily make good symbols for pictograms. A symbol with too many details hinders the readers from comparing data. When these symbols are used in multiples, they create a cluttered and busy picture.

Icons in pictograms should be simple because when these symbols are used in multiples, they still maintain a clear picture and present the data in an attractive and efficient way.

Usage of partial icons – like a truncated person – in pictograms should be avoided as they add confusion.

One symbol should be used with different shades to represent the variables. The readers can focus on comparing and contrasting the data and not on the different styles of the icon. A good symbol is simple, symmetrical and clear.

The most data points should be multiples of a complete icon. Each icon should represent a natural counting increment, such as 1, 2, 5 or 10. Data values should always be labelled. Readers should not make the count.

Pictorial charts can be informative, stylish and easy to interpret, but it also can be very hard to compare two complex figures. Reason is that comparison of two patterns is based on their areas. If the patterns are complex it is inaccurate to review them by their areas.

The thing is more problematic because figures are enlarged or reduced to correlate with their numerical values so that the aspect ratio is retained. When enlargement or reduction is made like this, figure will not only be taller but also wider therefore area of the figure will eventually be four times bigger. If figures are stretched only horizontally or vertically they will look unnatural.

The biggest problem in pictorial charts is, that if they are made falsely, the visual comparison disappears and in most cases distorts information. Pictogram should not be used if the data points are close together. It is difficult to contrast and discern subtle differences in a busy picture.

3.2.15. Thematic Map



Maps can be used to visualize practically anything location related data. Thematic maps are relatively efficient tool to represent territorial data. In many situations map can adduce geographic correlations which would be in other ways hard or unclear to visualize.

A choropleth map is a one type of thematic map. It bases on categorised data and colours. Different areas of the map are coloured so that each hue or colour represent one category of data. Areas are usually administrative areas because data is usually available as sorted by administrative areas.

Choropleth maps can represent only one thing at the time and only relative variation is possible to disclose. Uniting many different things is not possible as then hue and colour categories won't visually work anymore.

If it is wanted that map forwards information effectively, it is capable to use only one colour and it's different hues. Using different colours in representation is not justifiable because usually there is no common hierarchy between the different colours.

Nonetheless, situation where the data can be divided to two opposite class is exception. For example when visualizing positive and negative stance with map. In that case areas below average or midpoint are tinted with one colour and the areas above with other colour. Also, in this case darkness of the colour means quantity.

One of the main points when designing choropleth map is how to class the data. Changing class and thus altering colouring causes quite different visualizations from same subject.

There are different rules to calculate classes. Usually five or six different class is maximum, but even ten or twelve class can create good visualization. In most cases the amount of classes is the cube root of amount of observations.

$$\sqrt[3]{\text{amount of observations}}$$

For example if we would have data set containing observations, ergo numbers, from different countries which we want to visualize, we would calculate the cube root from the amount of countries we have data.

With choropleth map we can make classes with different intervals. The can be even or be based on some natural class border. It also to possible to define classes so that each class have same amount of units. Although the intervals between the classes can vary so that eventually the map does not represent true distribution. The puzzle is that there is no commonly applying criteria to define classes, so it must be done specifically each time.

A common principle is that choropleth map transmits information about quantities with intensity and brightness of colour. The darker the hue is, the more there are things quantitatively.

Solid shades of a colour should be used to highlight an area in map. Thematic map should not be used, unless geography is relevant to the message.

3.3. Colours

The colour palette for the Youth Newsroom (from the lightest to darkest hue):

- Blue: (#A6E1F7, **#73CDEF**, #28AADC, #0895CB)
- Green (#A4F9C4, #70F3A1, #22E56B, **#00D850**)
- Darker Blue with a glimpse of purple (#AEBAF8, #8092F2, #3A54E1, #1836D3)
- With charts mainly blue and green and the bolded hues are used.

Colours should be used gracefully in charts. All the colours in palette should not be applied at the same time. Usage of too many colours is confusing. Instead, harmonic combinations should be chosen, such as different shades of the same colour or colours on the same side of the colour wheel. Even if colour is available, it is okay not to use it at all. Every time a colour is changed, it signifies a change in information or added layer of data. In the end, the presented information should determine every colour that is chosen for chart.

Warm colours are those in the red are of the colour spectrum such as red, orange and yellow and the cool colours are the blue side of the spectrum including blue, green and neutral grey. Warm colours appear larger than cool colours, therefore red can visually overpower blue though they would be used in equal amounts. Warm colours appear closer and cool colours visually recede.

Hex value ergo hex triplet is a six-digit hexadecimal number or triple which is used to define colours in web design. Colours can be specified in the format of #RRGGBB where the letters are hexadecimal values for the red, green and blue values of the colour. The range of colour components is from #00 to #FF which is the highest value.

A chart is only successful if the reader can understand the content. A colour change in any chart element signifies a change in information or and added layer of data. If a colour is carrier of information and is not seen, the translation of information is severely impeded.

Multiple colours must not be used to represent the same kind of data. However, same colour should be used to represent the same variable so the readers can focus on comparing data.

If a different colour is used to distinguish different chart elements or signify a change in data, a lighter or darker shade of that second colour should be used. It is easier for the eyes to differentiate lightness and darkness. Sufficient contrast in values makes the chart more accessible.

Different colours, or colours on the opposite side of the colour wheel – such as red and green – must not be used in a multiple-bar chart because the contrast distracts the reader from data. Instead graduating shades of one colour, or colours on the same side of the colour wheel should be used to keep a multiple-bar chart clean.

The colour scale should graduate from the lightest to darkest or vice versa, regardless of the colour. A simple test is to convert the colour scale to black and white and check for smooth progression from light to dark. The eyes can't draw meaningful comparison jumping between light and dark shades. Therefore, a scale with alternating light and dark colours in the middle of the scale should not be set. The lack of contrast in lightness makes chart virtually unreadable for colour-blind users.

Colours can reflect the tone for instance, deep blue for conservative and bright colours for something cheerful. Generally, thematic representation of colours should be avoided.

It is wise to label directly on chart elements if it possible because it is helpful for readers. If a legend is used, a colours with a high contrast value should be used. A legend that relies on colour alone to convey information can be hard for general users and impossible to interpret for colour-blinds.

3.4. Fonts

Typography of charts should be chosen from the perspective that type in charts is meant to describe the information, not to adorn. Typography should not be on the centre stage in charts. The data is the focus. Type in charts is there to describe the chart clearly and to evoke emotions.

Typography should be kept simple. The headline can be either bold or a couple of sizes larger.

Fonts should never be too small or too condensed. Style should be kept simple. Bold or italic can be used to emphasize a point, but they should never be used at the same time.

All the text should never be totally uppercase because it is hard to read. Hyphenation and highly stylized fonts should be avoided.

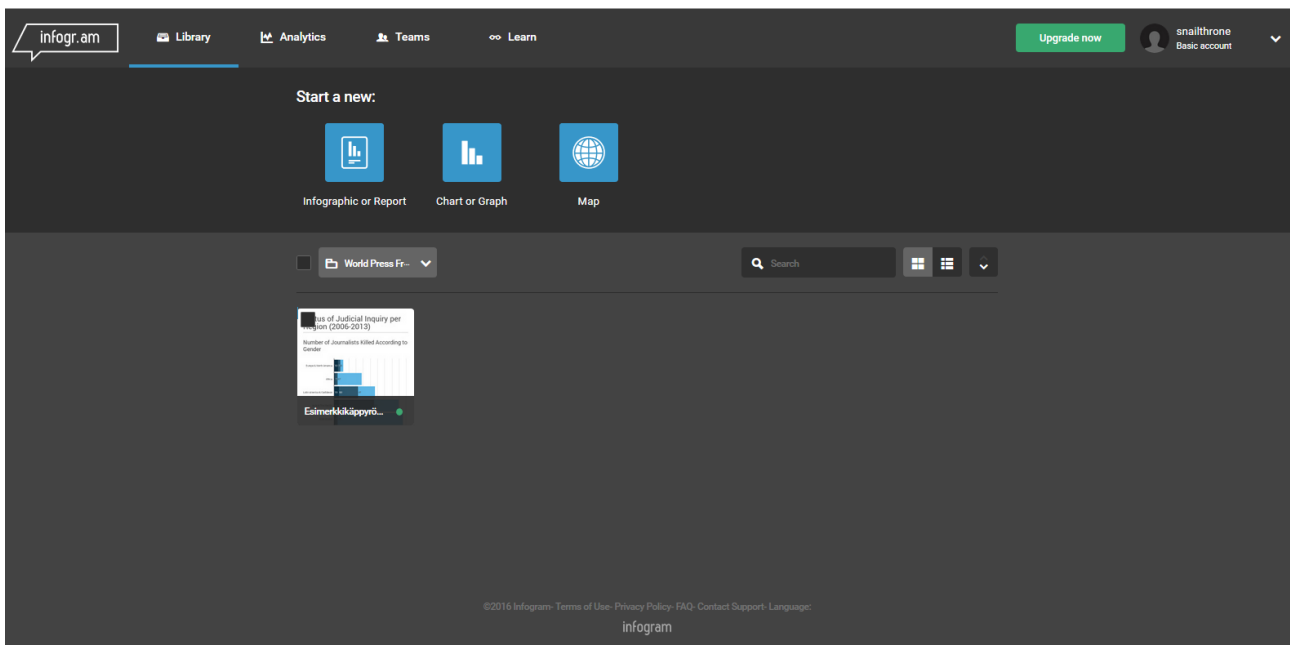
Poor typography draws undue attention away from the underlying data, which carries the main message.

4. How to Visualize Data with Infogr.am

Infogr.am is a Latvia and USA based platform for making of interactive data visualizations. Usage of it is easy and quick to learn.

Infogr.am is easy for beginners as it does a lot for user, for example it makes scale points and grid lines automatically.

4.2. The First View



The view from library that open when logged in Infogr.am. Library, Analytics, Team, Learn, and User Profile are found on header bar.

The first thing you see after you have logged in Infogr.am is the library. Here are all your visualizations and folders. At the top of the page are library, analytics, team, and learn buttons.

Selection of library icon takes you back to your library. Selecting analytics icon opens a tab where you can follow how much audience your charts have got, from which part of world your audience comes from or who is referring to your charts.

On teams tab you can set up a team and share your library for example with your work pair, but this requires an update for your account. Behind the learn tab are a few fast and easy tutorials for Infogr.am made by its creators.

On the upper right corner you can find link to upgrade page and your own profile tab. On your profile tab you can find settings for account and infographics.

4.3. Upgrades

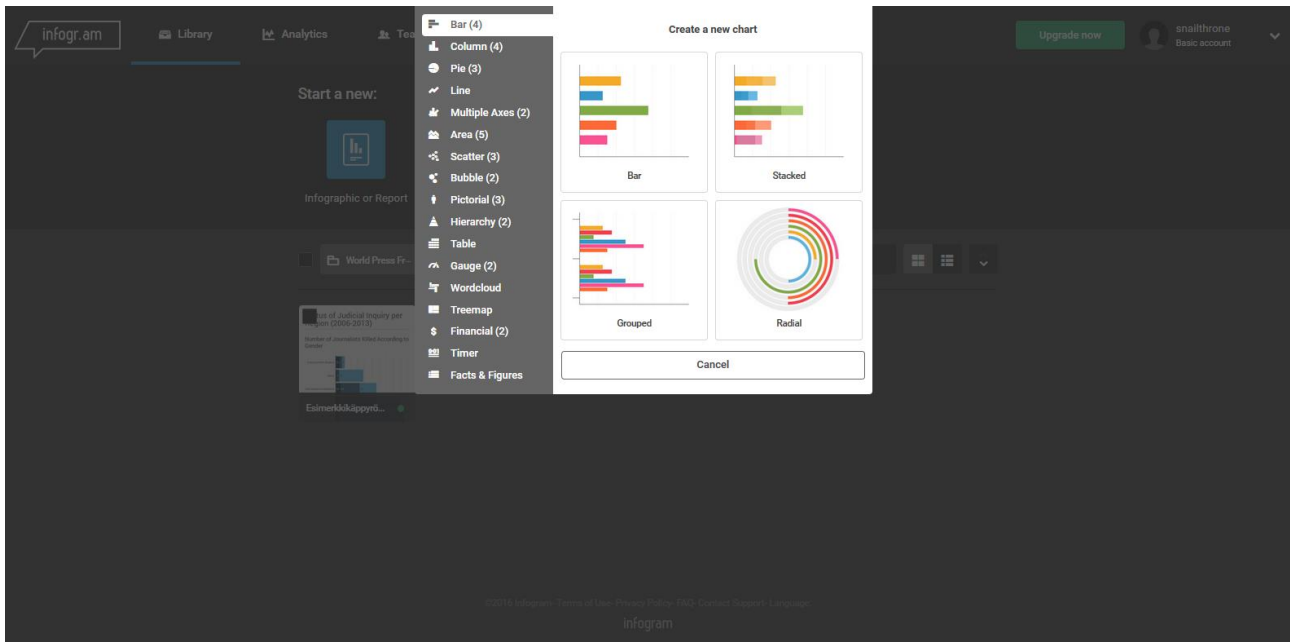
There are three different Upgrade stages on Infogr.am: Pro (\$18 / month), Business (\$50 / month) and Enterprise (\$166 / month). Enterprise upgrade also includes custom designed themes starting from \$800.

	Free	Pro	Business	Enterprise
Price	Free	\$18 / month and \$15 / annually	\$50 / month and \$42 / annually	\$166 / month
Infographics limit	10	100	1,000	10,000
Chart types	[laske]	35+	35+	35+
Maps	2	200+	200+	200+
Image upload limit	10	100	1,000	10,000
Rich Text Editing		✓	✓	✓
Import CSV, XLS		✓	✓	✓

Connect live data		✓	✓	✓
Private Sharing		✓	✓	✓
Download PNG, PDF		✓	✓	✓
API		✓	✓	✓
SSL Encryption		✓	✓	✓
Premium Themes		✓	✓	✓
Your colours and logo			✓	✓
Custom links and sharing			✓	✓
Analytics Tracking			✓	✓
Training				✓
Dedicated Account Manager				✓
Custom-designed themes				from \$800

Infogr.am itself recommends the business version, but is up to you what upgrade you want if any. This guideline is made with free version.

4.4. Let's Visualize



Full infographics and simple chart and elements can be made with Infogr.am. It is easiest to start with a single chart or graph element because if needed, different elements like pictures, videos, maps and text can be added later.

These chart type are most common and traditional ones. Infogr.am offers some additional chart types, but they are not needed as we can visualize data same way already or the chart elements are too inaccurate, for example the doughnut chart.

Start selecting icon on top of the "Chart or Graph" –text. A "create new chart" –window appears. If wrong chart element is picked, it can be changed easily.

Settings

Before adding any data to the editor, a few things should be configured first.

There are some buttons at the top of the screen. With clicking "untitled" the chart can be named. At settings, theme of the visualization can be chosen. By default Infogr.am sets up one with a white background and black text colour, which is embeds greatly on web site that uses white background colour.

The next one is width. There are some things to be considered when setting it up. First of course, the width setting defines how wide your chart elements will be. On Infogr.am two or more elements can be set vertically and two element horizontally. If two chart

elements are set, for example bar or line charts, horizontally, the height of chart must be fixed.

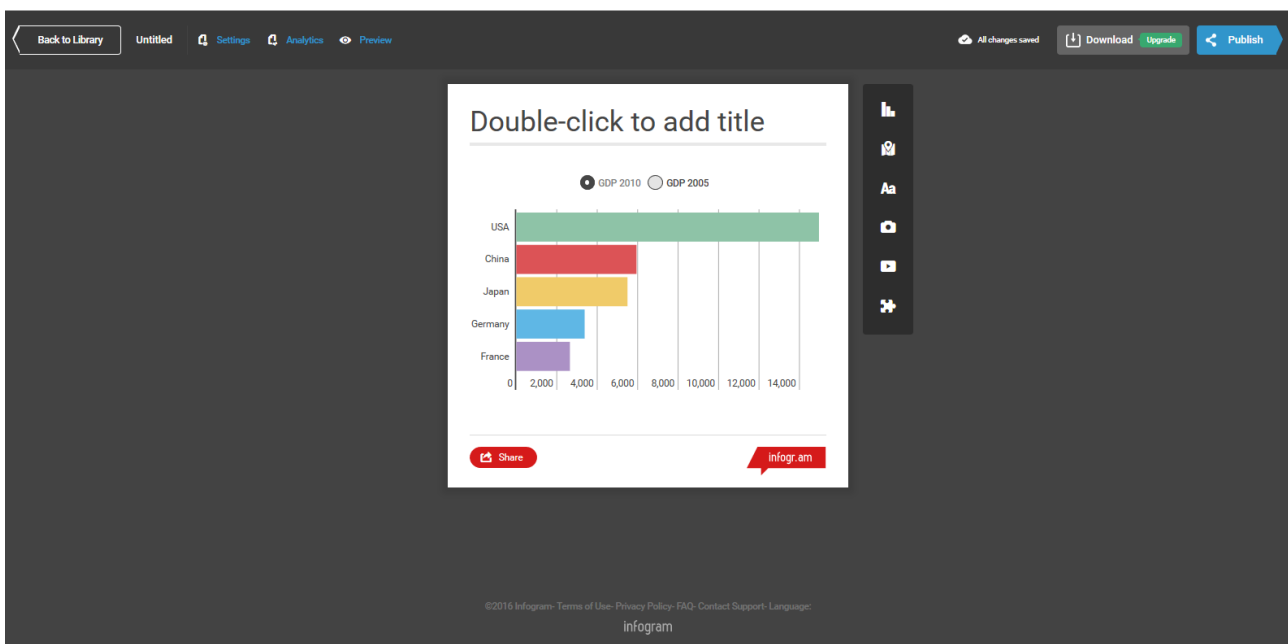
For example, if the width is 600 pixels and two elements which height is 400 pixels are placed horizontally, their size will eventually be 300 pixels wide and 400 pixels high. In this case a correct fix would be to set height of both charts to 200 pixels so that aspect ratio is equal to the original.

After width there are some premium settings. If needed upgrades have been acquired, Infogr.am's logo can be replaced. The social media share button can also be replaced. Below the share button option a custom hyperlink to visualization can be set.

At last there is option to replace the traditional left-to-right text to right-to-left text.

Next to the settings page there is analytics tab and preview button. Preview is useful to examine your chart before publishing it.

4.4.1. Making Things Visual



By default Infogr.am adds the picked chart element and also a title and a body text elements. Elements can be edited by clicking them. Different elements can be added from the toolbar which is located right next to the layout.

Toolbar:

1. Chart menu

2. Map menu
3. Text elements
4. Pictures (Upgrade needed JPG, PNG and GIF supported)
5. Video (has to be added from third source like YouTube or URL)

4.4.1.1. Text Elements

There are four different text elements on Infogr.am: Headline, subtitle, quote, and body text. Eventually, if the account is upgraded enough, text elements can be edited.

Infogr.am can be used to make a whole story, but without the rich text editor, it is better to make text elements on the web page and handle just the chart elements on Infogr.am. Still, body text elements can be used in making of description, headline, or subtitle element to add caption to the chart, if the default fonts fit in visually with other content.

4.4.1.2. Pictures and Videos

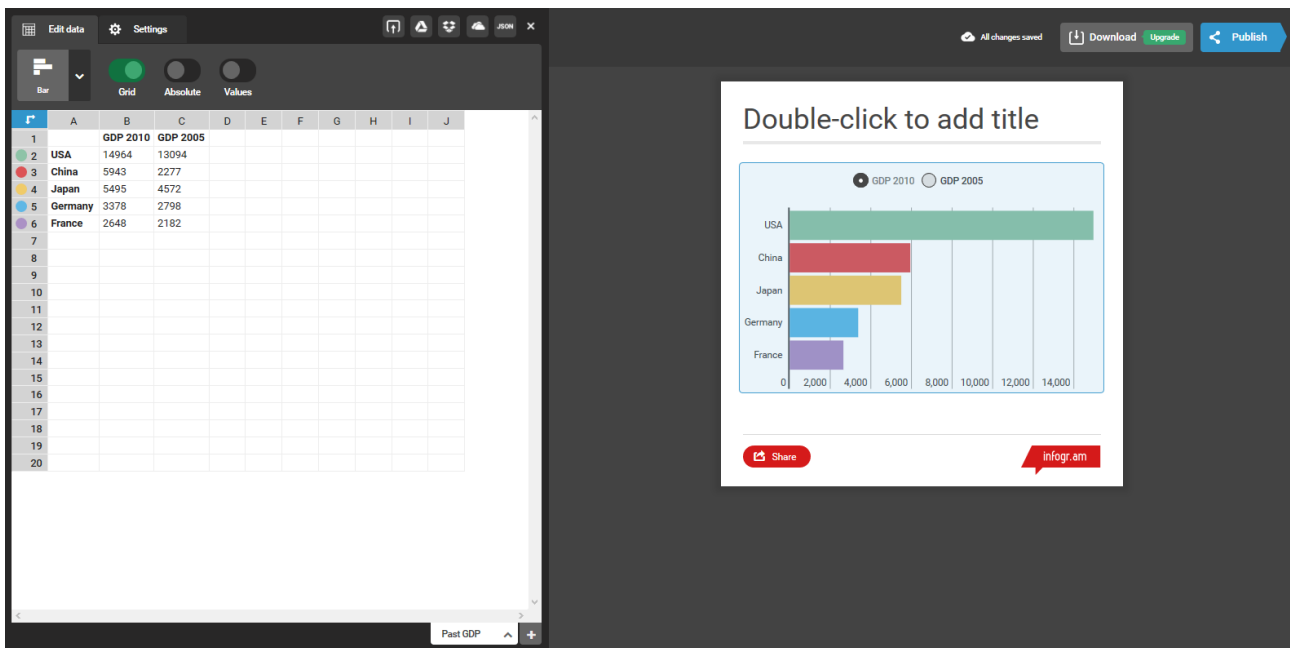
Additional pictures and videos can be considered as chart junk. They are distractive elements which take attention from visualizations. There is a place for pictures and videos, but not next to the chart.

4.4.1.3. Facts & Figures

Facts and figures element is a simple pictorial element which can be used like floating quote to raise one numerical fact from the text. Fact & figures -element is composed from a symbol, numerical fact and a short text which unwraps the fact.

Fact & Figures elements can be used are more visual method to tell a simple numerical based fact.

4.4.1.4. Editor



The core of the editor is a simple spreadsheet application where the data is inserted. A more than one spreadsheet layer can be added. With additional layers a slide show like charts that have can be made.

The editor of Infogr.am usually changes a little bit with different chart as does the way the data is plotted on the table. Still, there are some universal options which should be remembered.

Grid lines and legend should be always used, though pie chart is an exception. Instead of legend, callouts and their names should be shown in pie chart.

Checking the absolute option allows you to visualise as many icons as the value you have entered in the editing table. In addition if you have different layers on same chart, the absolute option keeps the scale same. Absolute option should always be used.

On the settings page chart colours, chart category height, axis range and decimal separator can be modified. Chart colours can be set manually or by pasting the html code of the colour.

Chart height can be set in two ways. Either the chart category height or the chart height can be modified. Chart category height effects on how high the different categories will be and the chart height to the highness of whole element. It can be faster and easier

to set the height with chart height option but the adjusting categories can be more detailed way.

Nevertheless, the end result should be approximately 3:2. For example if the chart is 600 pixels wide the chart will be 400 pixels high. Remember that if two chart elements are plotted horizontally, they split the width of the whole chart.

Naming of the labels should not be forgotten. Although with line and column chart it may be obvious that the x-axis represents time. It should be written on the labels what they represent.

The scales should be adjusted so that chart will not be differed anyhow. Infogr.am sets the scale so that the highest bar tops the highest grid line. The scale can be adjusted so that there is an extra grid line.

Decimal separator should be dot if the chart is published on page that uses English as language.

4.4.1.5. Exceptions in editors and how the values are plotted in table editor.

Column bar chart and bar chart: "A" column contains time series. The first row has the different layers. With grouped and stacked column chart, "A" column defines the bars or parts of them and the first row the labels on x-axis.

Pie chart: The first row defines the different layers and "A" column the segments of the pie.

Line and area chart: Time series go to the first row and the different lines to "A" column.

It may be tempting to put smooth curves beside the straight lines but this option should never be used as the smooth curved line is not as accurate as a straight line. Also, there is no reason to use reverse axis. You get the same result by adding minus sign before your numbers.

Multiple axes chart: "A" column contains the time series, "B" column the bar element, and "C" column the line element.

Scatter and bubble chart: "A" column defines different labels, "B" column x-axis units, "C" column y-axis unit, "D" column tooltip and area, and "E" column the groups.

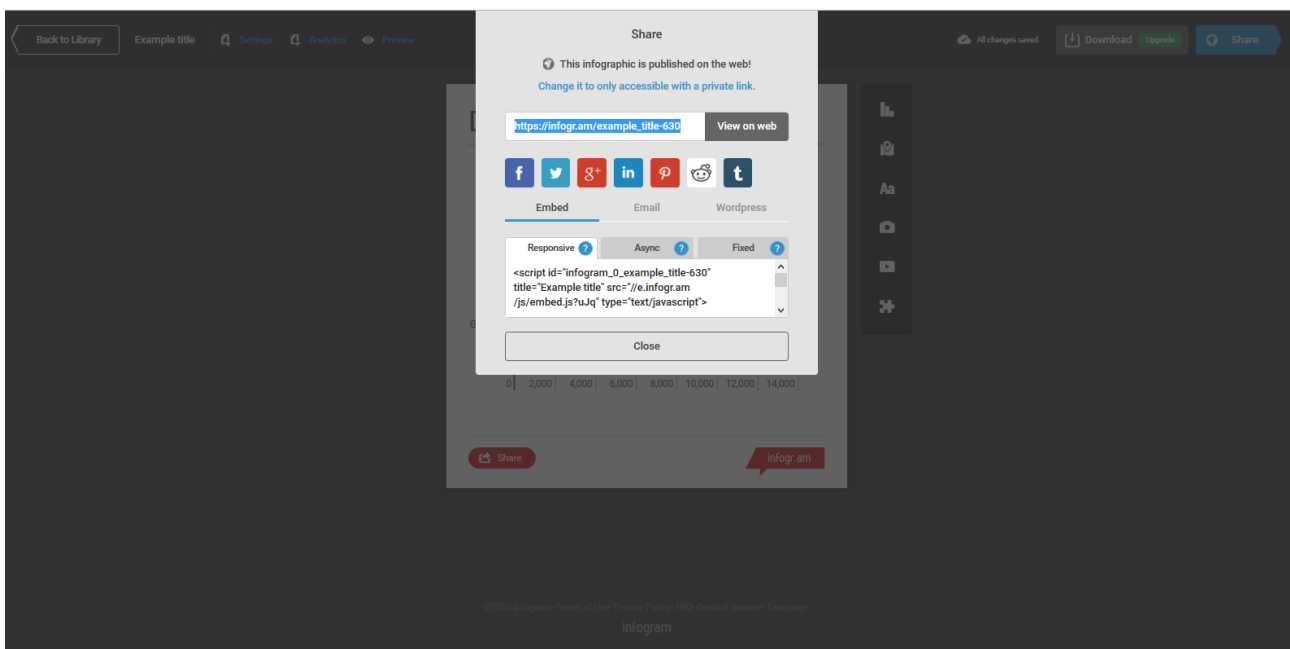
Pictorial chart: The first row defines the layers, and "A" column labels and their units.

Table: Works like the editor.

Maps: The Map has six different columns: "A" column contains the English name of the countries. "B" column contains the values. "C" column contains the group of the countries by continent. "D" column contains the coordinates and the "E" column the name of the label which is by default the same as the English title. In addition there are five different text boxes where can be written desired text which appears when the mouse is hovered on the country.

Heatmap option turns the maps to heatmap. Grouped data option colours the countries with colour that is assigned to the group. Point option turns the map to a dot density map. Hide blanks option can be selected only with heat map option and it hides the countries with no value. Lat/lon option sets the map to equal correct latitude and longitude values.

4.5. Time to Publish



When the visualization is ready it can be published with pressing "publish" button which can be found from upper right corner. Selecting it opens the share window.

It is preferable to place some tags so the charts can be found. After tagging, the format of the visualization – interactive or static – can be chosen.

Static image should be chosen if the visualization is published on paper. Static visualization can be used on internet too, but if the interactivity is wanted, interactive option should be chosen, as the static format does not support interactive elements.

Infogr.am offers direct hyperlink and different ways to embed visualizations to the web page. Visualization element can be embed with three different scripts and can be send with email. There also direct share buttons for different social media applications like Facebook and Twitter.

Using the responsive script, the element will change its size with user viewport. For example if the element is embed with to the web page and it is browsed with smartphone – and with smaller screen – the visualization scales to the small screen.

“Async” should be chosen if there are multiple charts placed on the same page. Async option loads each chart separately at the precise moment when the reader reaches the place where the chart is embedded. In addition, it loads the page faster.

Fixed script embeds the visualization with an iframe tag. This is not recommend by Infogr.am as the visualization’s size may vary across different platforms, but the element not.



Guideline for Data Visualizations

Synopsis for Guideline for Data Visualizations

Eemeli Martti

Introduction

This synopsis is made in case of quick checks for data visualisation team of Youth Newsroom. This synopsis covers the formats, charts, colours, and example charts.

Formats

In Youth Newsroom, visualizations are part of a story. They pass on information and can demonstrate the phenomenon better than text, but also act like pictures – as elements that complete and variate the whole story.

- **Static visualizations** (Used in print)
- **Interactive visualizations** (Used in web page)

Settings

- Width is 1000 pixels wide at maximum.
- Height is adjusted so that the aspect ratio is about 3:2.

Regular Column Chart

- Usually used to represent total or relative amount and its change or trend in time.
- Useful in representation of time series, but basically any other continuous variable can be put on x-axis.
- The column chart is usually an alternative for a line chart. They fit in same kind of situations, where bar chart does not. As the column chart is used in same kind of situations, there are same requirements to use them. So when using a column chart, there should be continuous value placed on x-axis such as time.
- One of the most common mistakes made with bar and column bar chart is to cut of the scale. This is done because it is thought that the line on top of the bar is the part that transmits the information. Cutting the scales vanishes the visual comparison from the chart.

Stacked Column Chart

- Highlights the sum of its parts.
- Usage: the change of total progress in one chart.
- The best way to colour is that the most intense hues are at the bottom and the lightest ones are on the top.

Grouped Column Chart

- The basic idea is somewhat same than in regular column chart, but with grouped column chart it is possible to represent more time series in one chart.
- Bars have to be in the same order in every group. It is also preferable that bars in one group are ordered from the highest to lowest if possible.
- If there are too many bars in group, it is hard to see a single group or chart. Three or four bars are usually the maximum.

Regular Bar Chart

- Suitable for portray of comparison of quantity or magnitude of changes of one thing or different classes.
- In regular column and bar charts all the bars should be the same colour and shade since they measure the same variable. However, a darker hue can be used to highlight a particular bar. With grouped and stacked charts the shading of the bars should move from lightest to darkest for easy comparison.

Grouped Bar Chart

- 2-3 subdivisions in each class.
- Three or more bars should be used only when it is absolutely necessary.
- It is preferred that the group which has the longest bar is placed uppermost or lowest and the rest of the groups are placed in length order.

Stacked Bar Chart

- Should be used over grouped bar chart when the representation of total quantity is more important than the portion of single parts.

- Six subdivisions is usually the maximum amount.
- A good alternative to pie chart when having many components to compare.

Pie Chart

- Represents how many large pieces something has divided.
- Is better than bar or column bar chart from aesthetical point of view.
- Usually the largest sector is placed to 12 o'clock on top and then continued clockwise with placing the second largest sector next to the first sector.
- Numbers can be marked to near sectors. Marking of the percentage values is not necessary.
- A common mistake is to divide the pie to too many sectors. The maximum amount of the sectors is five and sometimes even this is too much. If there are too many sectors, the less significant segments should be combined and label it as "other".

Line Chart

- Used to represent change and progress over a period of time.
- Emphasizes trend line and the relationship of change to it.
- Can show massive amount of data in a small space.
- Best used with continuous data series.
- Suites for showing of trend, acceleration, deceleration, and volume.
- 3-4 lines is maximum on single chart. Otherwise the chart is too confusing.
- The most important line should be one colour and the other line(s) should be shades of a second colour. With more colours, the line chart will be confusing and possibly illegible for colour-blind people.

Multiple Axes

- Generally usage of two different scales in one chart should be avoided, but sometimes it is necessary.

- Use always bars and lines. Never two bars or two lines.
- Line should be used for continuous data and bars for discrete quantities.

Scatter Chart

- With scatter chart it is easy to state correlations and density and it is easy to understand even with little knowledge of statistics.
- Scatter chart can be used to bring out the groups in data and describe their differences.
- When scatter chart is used to visualize causal connection, x-axis should be used for phenomenon that is believed to be the reason and the y-axis for what is believed to be consequence.
- If there are lot of variables to be plotted on char, scatter chat is not the best choice since the dots can be set on the same spot, they don't stand out.

Bubble Chart

- Sometimes it is necessary to visualize great amount observations with scatter chart. Because in basic scatter chart there is danger that many observation sets in exact same place, it is worthwhile to visualize large data set with bubble chart.
- Works with the same logic as the scatter chart, but it has also a third dimension: area.
- With bubble chart you can visualize three or – if colours are used as well – even four different things at the same time.

Pictorial Chart

- The purpose of pictograms is to create a snapshot of data or they can use pictorial elements like male or female figures same way as column and bar charts use bars.
- Pictorial chart are visually more engaging than traditional charts, but also less effective for comparing a large amount of data.
- Symbol should be some common shape or construct like male or female

figure.

Thematic Map

- Maps can be used to visualize practically anything location related data.
- A choropleth map bases on categorised data and colours. Different areas of the map are coloured so that each hue or colour represent one category of data.
- If it is wanted that map forwards information effectively, it is capable to use only one colour and its different hues. Using different colours in representation is not justifiable because usually there is no common hierarchy between the different colours.
- Nonetheless, situation where the data can be divided to two opposite class is exception. (For example when visualizing positive and negative stance with map.)
- Usually five or six different class is maximum, but even ten or twelve class can create good visualization. In most cases the amount of classes is the cube root of amount of observations.

Colours

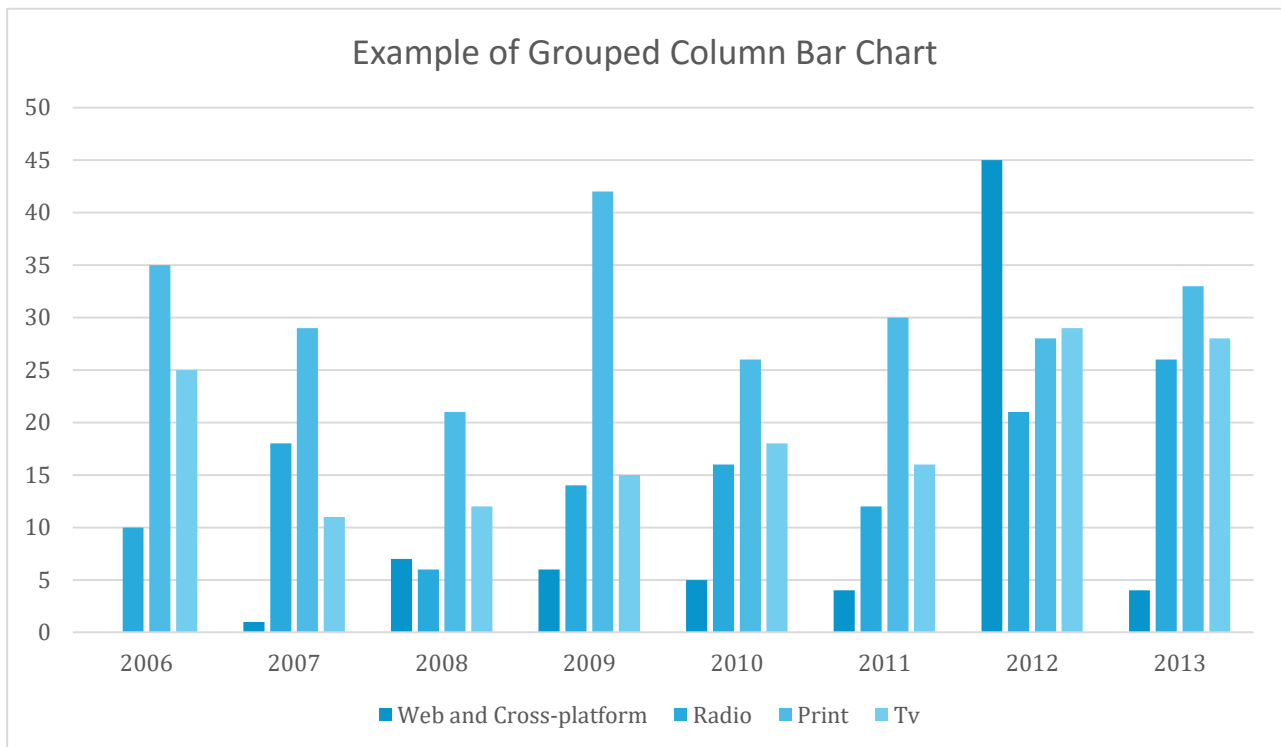
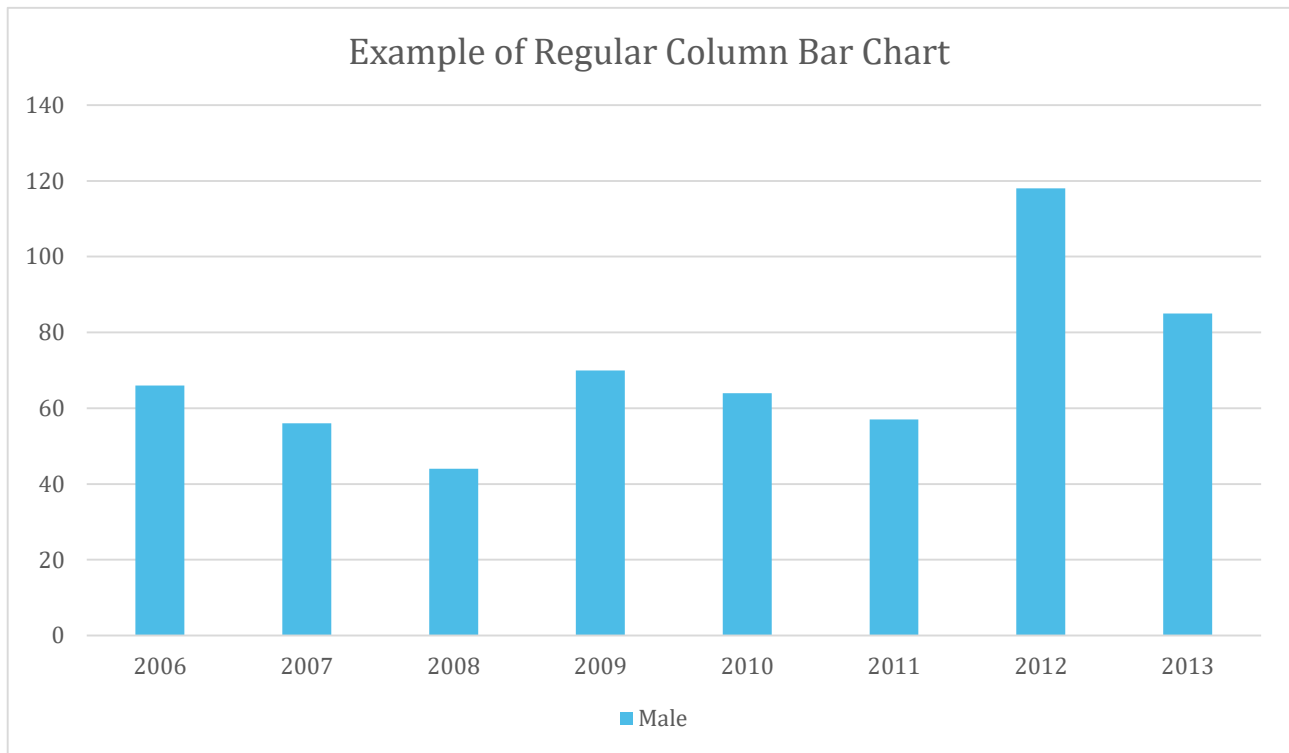
- The colour palette is (from the lightest to darkest hue):
 - Blue: (#A6E1F7, **#73CDEF**, #28AADC, #0895CB)
 - Green (#A4F9C4, #70F3A1, #22E56B, **#00D850**)
 - Darker Blue with a glimpse of purple (#AEBAF8, #8092F2, #3A54E1, #1836D3)
 - With charts we use mainly blue and green and the bolded hues.
- All the colours in palette should not be applied at the same time as the usage of too many colours is confusing. Instead, harmonic combinations, such as different shades of the same colour or colours on the same side of the colour wheel.
- Multiple colours must not be used to represent the same kind of data.

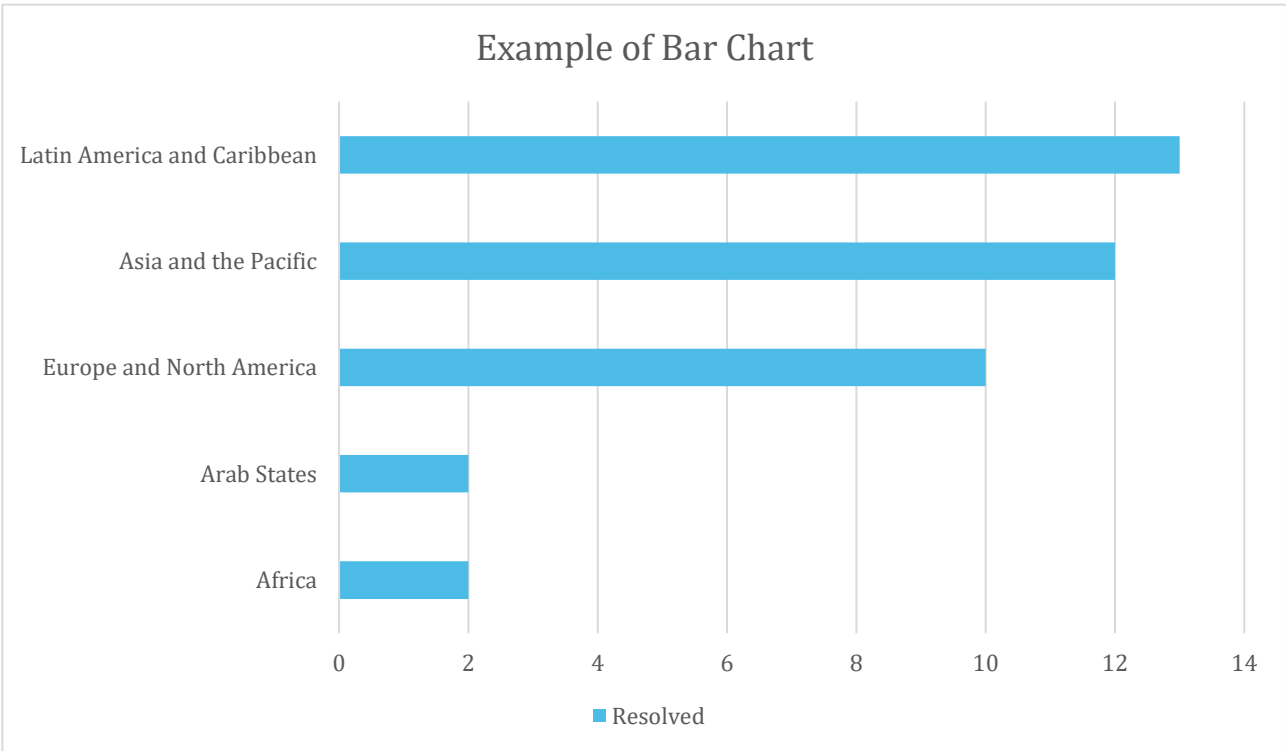
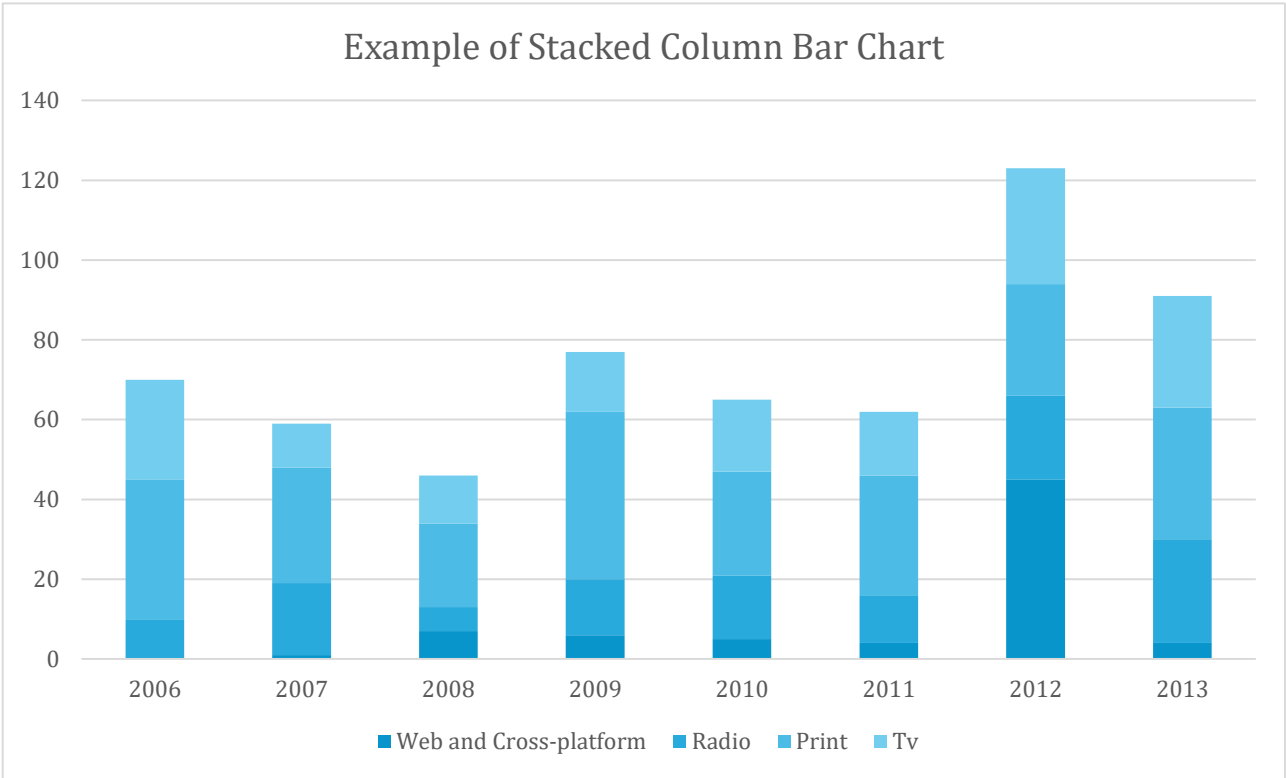
- A colour change in any chart element signifies a change in information or an added layer of data.
- If a different colour is used to distinguish different chart elements or signify a change in data, a lighter or darker shade of that second colour should be used.
- The lack of contrast in lightness makes chart virtually unreadable for colour-blind users.

Fonts

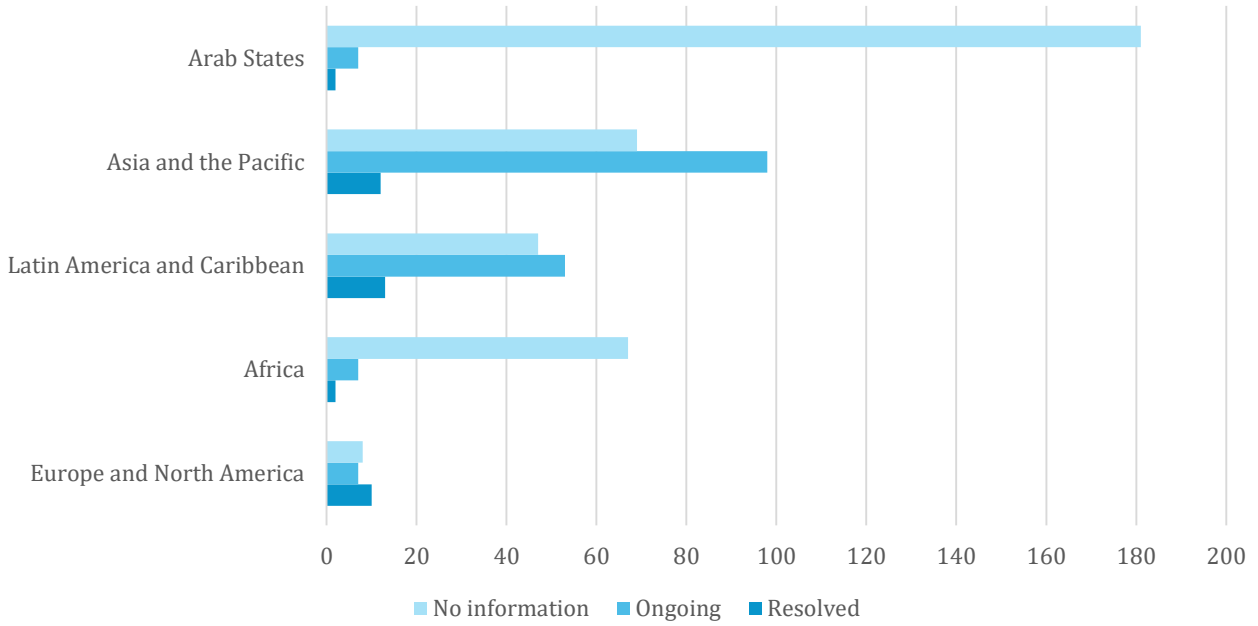
- Typography should be kept simple. The Headline can be either bold or a couple of sizes larger.
- Fonts should never be too small or too condensed. Style should be kept simple. Bold or italic can be used to emphasize a point, but they should never be used at the same time.
- All the text should never be totally uppercase because it is hard to read. Hyphenation and highly stylized fonts should be avoided.

Examples

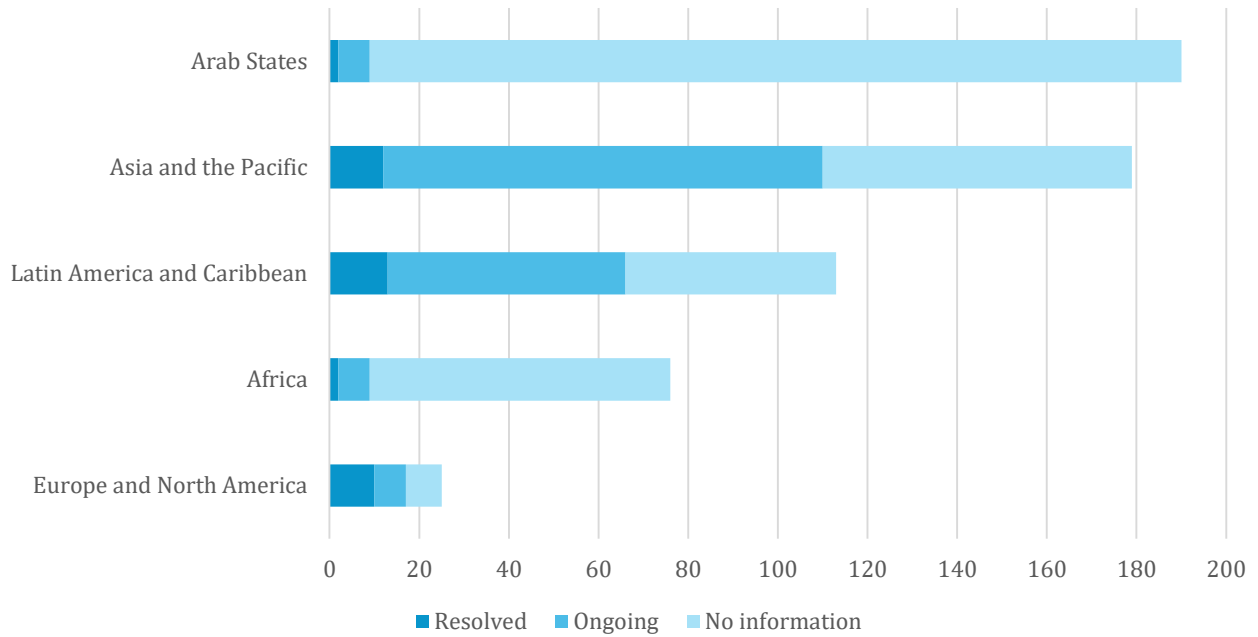




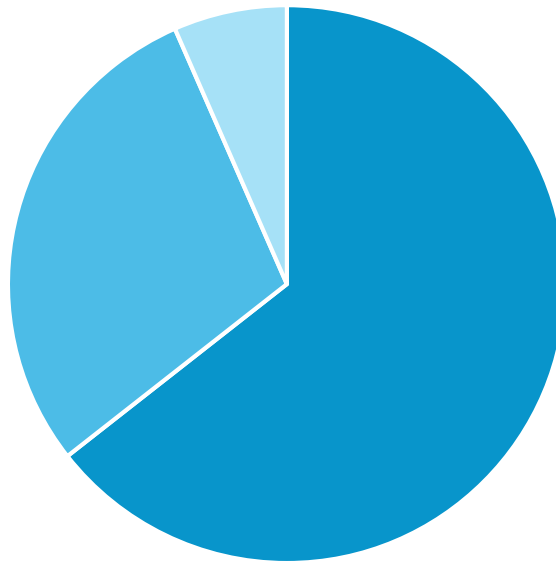
Example of Grouped Bar Chart



Example of Stacked Bar Chart

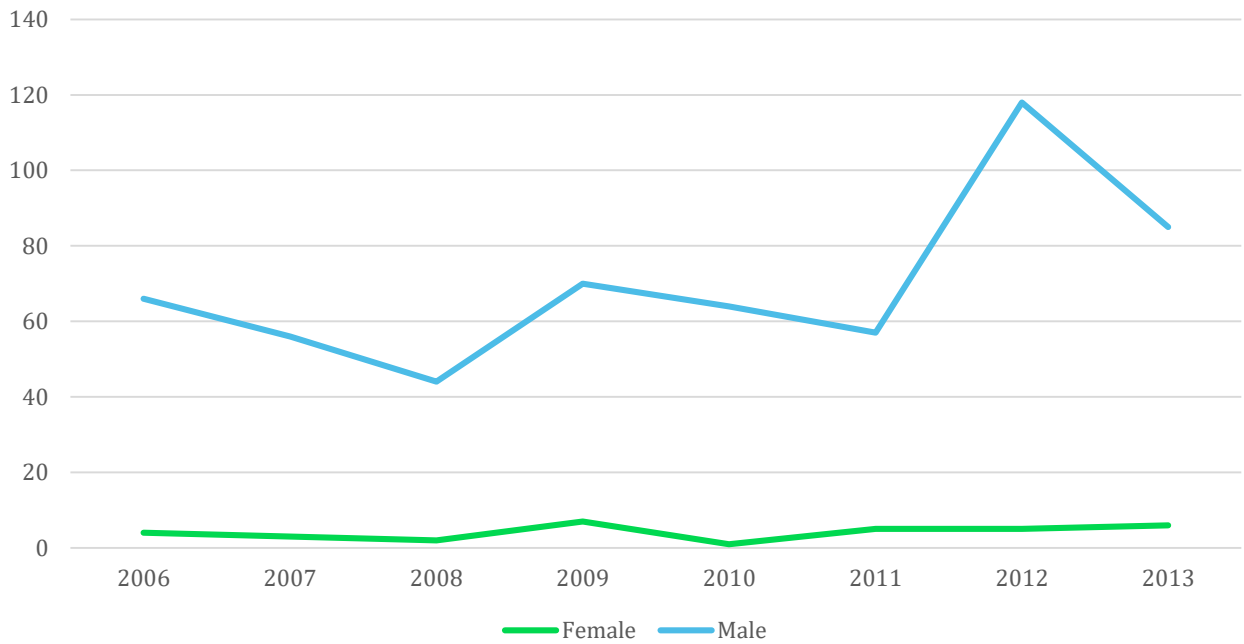


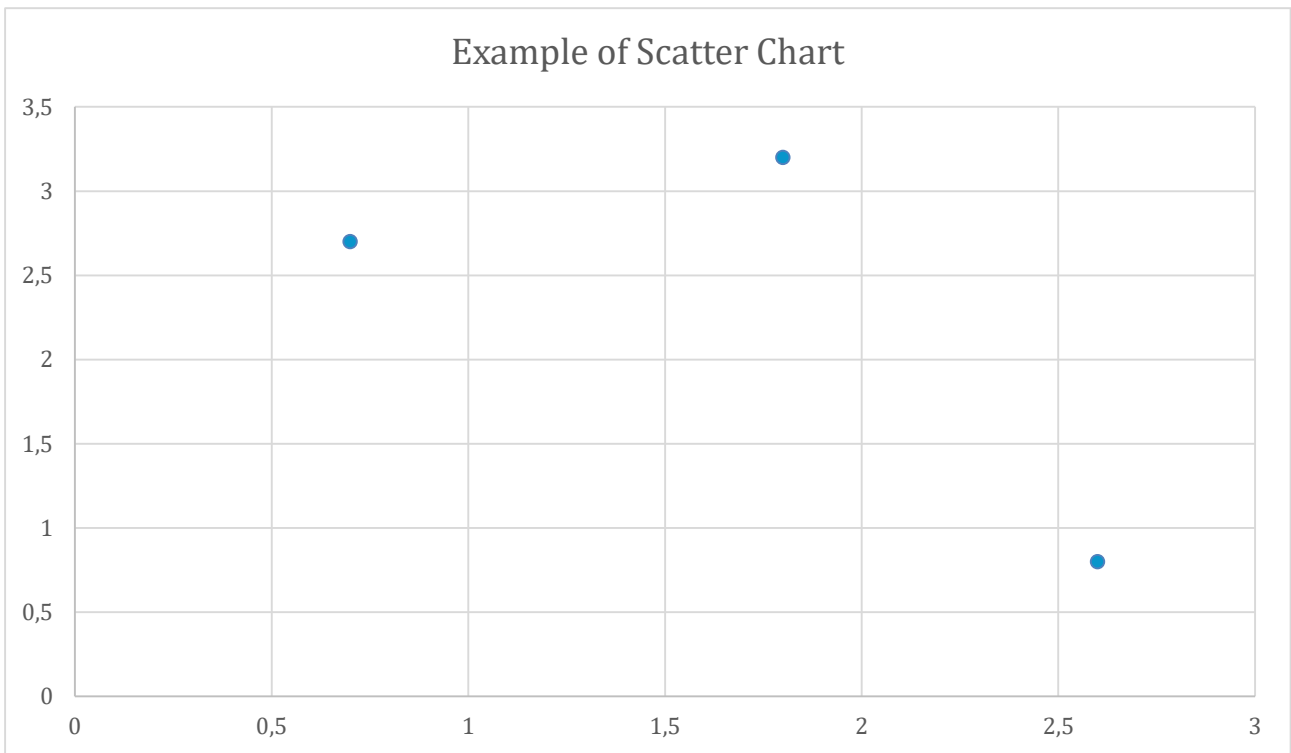
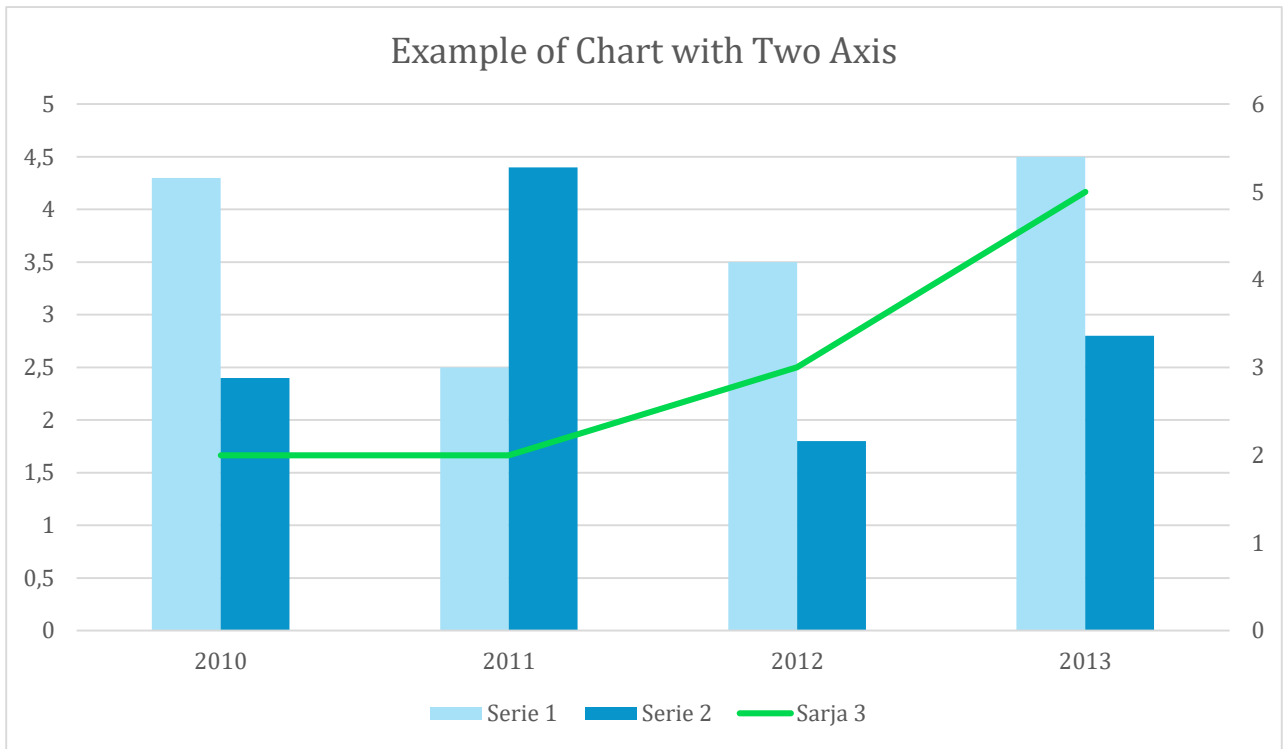
Example of Pie Chart



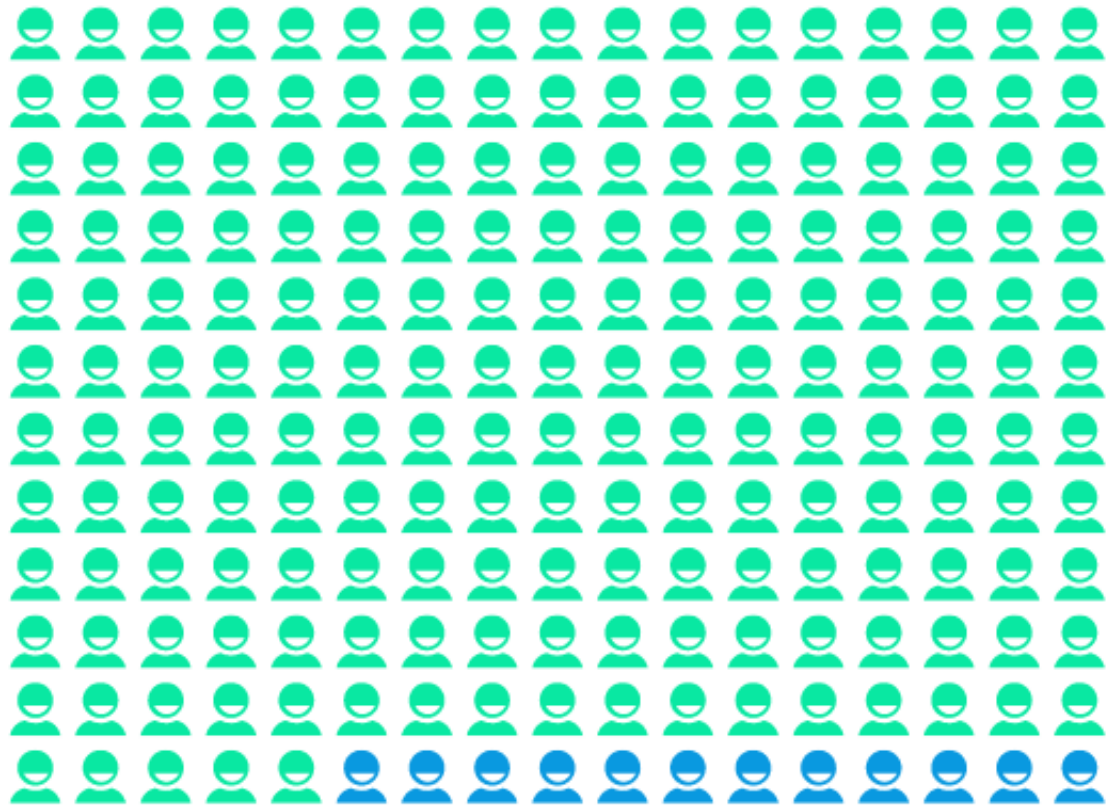
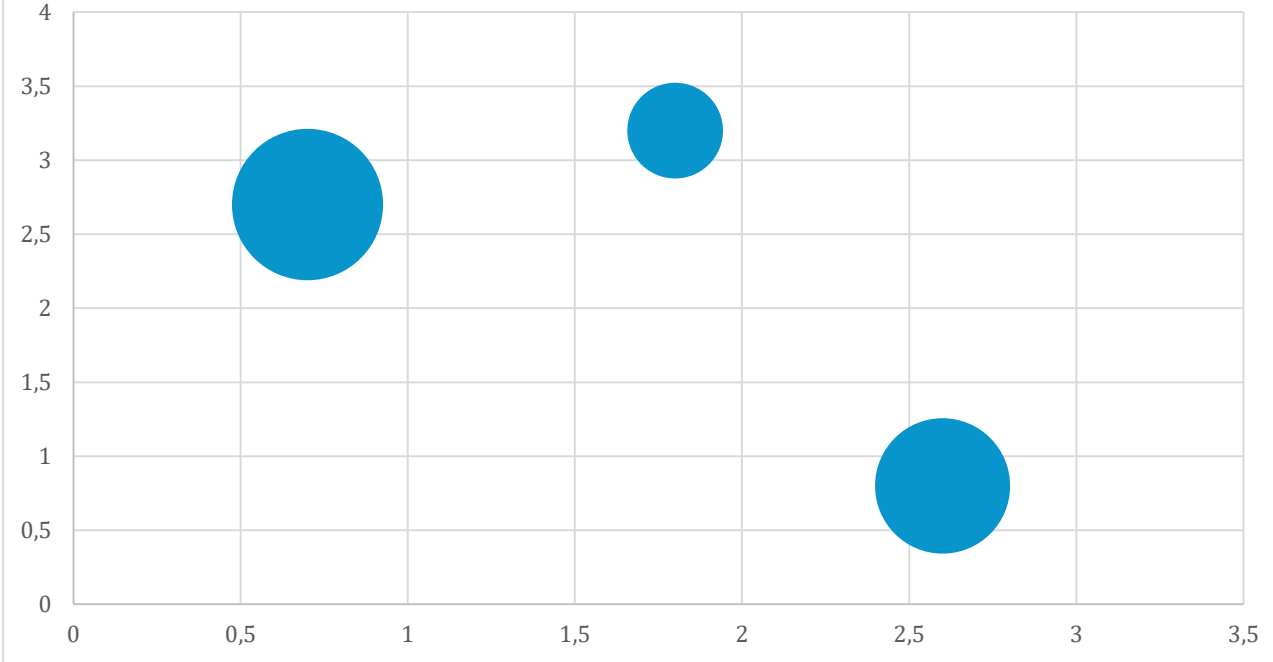
■ No Information ■ Ongoing ■ Resolved

Example of Line Chart





Example of Bubble Chart



● Local ● Foreign

