



Hogeschool van Amsterdam

Instituut voor Media en Informatie Management

Datatopia

De toekomst in de visualisatie van statistische gegevens

Afstudeeropdracht

Nique Sanders

Tutor: Paul Reinierse

Docent-begeleidster: Annette Klein

Juni 2008

Voorwoord

Beste lezer,

Voor u ligt het rapport 'Datatopia', het resultaat van mijn afstudeeropdracht als afsluiting van mijn studie Media en Informatie Management aan de Hogeschool van Amsterdam. Ik heb voor dit onderwerp gekozen omdat het een verbindend element is tussen mijn eerdere opleiding als grafisch vormgever aan de Hogeschool van de Kunsten te Arnhem en het MIM. Het ligt tevens in het verlengde van mijn dagelijkse werkzaamheden als interactie ontwerper. En, als laatste, sluit het onderwerp prima aan bij een van de opdrachtgevers van Webtic, het bedrijf dat ik samen met Paul Jongsma heb opgericht, de Dienst Onderzoek en Statistiek. Het is hét onderzoeksbureau van de gemeente Amsterdam. O+S levert door middel van onderzoek informatie die behulpzaam is bij het ontwikkelen en evalueren van beleid.

Samenvatting

Terwijl de geschiedenis van de data visualisatie al eeuwen oud is staat de online interactieve applicatie die in staat is om statistische gegevens om te zetten naar een visuele weergave nog in de kinderschoenen. Al decennia lang zijn mensen bezig zijn met de theorie achter het maken van grafieken, tabellen of data kaarten maar toch zijn er nog maar weinigen die zich bezighouden met de integriteit en de esthetiek van een data visualisatie. Daarnaast maakt de markt een sterke ontwikkeling door op het gebied van de participerende gebruiker.

Aan de hand van bestaande theorieën wordt er in dit rapport een begin gemaakt met een theorievorming omtrent het ontsluiten van statistische data via een web gebaseerde applicatie. Hierbij is gebruik gemaakt van literatuurstudie, online onderzoek van bestaande applicaties en een interview. Vanuit de kennis verkregen via de literatuurstudie is er een puntenlijst gemaakt waaraan de bestaande online applicaties werden getoetst.

Er wordt betoogd dat interactiviteit, mits goed aangeboden, van meerwaarde kan zijn. Daarnaast wordt aan de hand van de theorie aangetoond dat esthetiek en integriteit in statistische data visualisatie onmisbaar is voor een goede weergave van gegevens.

Het wordt duidelijk dat tussen de onderzochte applicaties een enorm verschil zit. Terwijl de een op een eenvoudige manier patronen zichtbaar maakt in gegevens maakt de ander het combineren van gegevens tot kinderspel. Hierbij is geconstateerd dat teveel combineren van data leidt tot onleesbare visualisaties die eenvoudig verkeerd geïnterpreteerd kunnen worden.

Tevens is geconstateerd dat de mogelijkheid om data weer te kunnen geven in verschillende visualisatie opties zoals een data kaart of bijvoorbeeld een tijdreeks zorgt voor een andere interpretatie van de gegevens.

Tenslotte is er, aan de hand van twee online applicaties, de vraag geponen of een eenvoudige data set een interactieve data visualisatie rechtvaardigt. Dit rapport geeft hier geen eenduidig antwoord op maar biedt voldoende houvast voor verder onderzoek.

Inhoud

Inleiding-----	2
Achtergrond-----	2
Doelstelling-----	2
Onderzoeksvraag-----	2
Uiteindelijke uitwerking van de deelvragen-----	3
Onderzoeksmethode-----	4
Leeswijzer-----	4
Deel I: De geschiedenis van data visualisatie in vogelvlucht.....	6
Voor de 17e eeuw-----	6
De 17 e eeuw-----	7
De 18e eeuw: het experimenteren met vorm-----	8
Begin 19e eeuw-----	10
Tweede helft 19e eeuw: het gouden tijdperk voor datavisualisatie-----	11
Eind 19e, begin 20e eeuw-----	11
1950 -1975-----	11
Van 1975 tot nu toe-----	12
Conclusie-----	12
Deel II: De theoretische achtergrond van het visualiseren.....	14
De kracht van een data visualisatie-----	14
Visualisatie opties-----	14
De drie ingrediënten van een diagram-----	15
Subjectieve bewijsvoering-----	16
De vier grondbeginselen van een grafiek-----	17
Visualisatie is meer dan alleen een jasje-----	18
Encourage the eye-----	18
De vier basis vormen van een data visualisatie-----	19
De zes principes van grafische integriteit-----	23
De theorie van grafische data visualisatie-----	23
Conclusie-----	28
Inleiding voor deel III en IV.....	30
Deel III Online applicaties.....	32
Inleiding-----	32
Many Eyes-----	33
Quickstep (internet versie van Swing)-----	36
Conclusie-----	40
Deel IV Dienst Onderzoek en Statistiek.....	42
Inleiding-----	42

Amsterdamse Veiligheidsindex-----	42
StatStat-----	43
Bevindingen Amsterdamse veiligheidsindex en StatStat-----	44
Conclusie-----	46
Aanbevelingen.....	48
Bronnenlijst.....	52
Lijst van afbeeldingen.....	54
Verwijzingen.....	55
Bijlage I-----	56
Bijlage II-----	57
Bijlage III-----	58
Bijlage IV-----	59
Bijlage V-----	60
Bijlage VI-----	61
Bijlage VII-----	65

1 Inleiding

1.1 Achtergrond

Voor een opdrachtgever heeft Webtic een applicatie gebouwd die ervoor zorgt dat de statistische data direct gekoppeld kan worden naar een web gebaseerde omgeving. Toevoegen van meta informatie is mogelijk, indien gewenst kan ook de vormgeving van de data aangepast worden. Beperking van deze methode is dat de data niet online gefilterd kan worden door de gebruiker. De meerderheid van deze data wordt online gepresenteerd als een tabel, een enkele keer als afbeelding. Hierbij wordt alle data zonder interactiviteit gepresenteerd. Ontwikkelingen door de tijd, vergelijkingen van data, relaties tussen data, delen van een geheel of overeenkomsten kunnen binnen de web gebaseerde omgeving niet zichtbaar worden gemaakt. Juist het kunnen filteren van deze data zou een enorme rijkheid aan informatie kunnen opleveren. Data kan gefilterd worden zodat de uiteindelijke informatie beter toegesneden is op de gebruiker. Een voorbeeld: er is data beschikbaar over de openbare veiligheid in Amsterdam. Deze gegevens bevatten de datasets stadsdelen, aangiften, misdrijven, leeftijd en aanhoudingen. Terwijl de ene gebruiker vooral de misdrijven per stadsdeel zou willen bekijken, is het voor een ander juist interessanter om de misdrijven door de jaren heen per leeftijdsgroep te bekijken. Dit is nu niet mogelijk maar zou kunnen als data interactief gepresenteerd zou worden.

1.2 Doelstelling

Bovenstaande achtergrondinformatie resulteert in volgende doelstelling:

Het doel is om door praktijkgericht onderzoek inzicht te krijgen in het ontsluiten van statistische data via een web gebaseerde applicatie op een zodanige manier dat er een betere analyse door de gebruiker mogelijk is. Hierbij willen we veel aandacht besteden aan de visuele presentatie. Welke verschillende visualisatie mogelijkheden zijn er en welke is het meest geschikt om datastructuren binnen statistische data te presenteren.

1.3 Onderzoeksvraag

Voor het onderzoek is de onderzoeksvraag zo ruim mogelijk gedefinieerd.

Welke vormen van visualisatie zijn het meest bruikbaar voor het weergeven van statistische data die betrekking hebben op de regio Amsterdam via een web gebaseerde applicatie? Specifiek is de vraag aan de orde hoe interactiviteit met de huidige technologische mogelijkheden hierop van invloed kan zijn.

En om bovenstaande onderzoeksvraag te kunnen beantwoorden zijn onderstaande deelvragen beantwoord:

Met betrekking tot de theorie.

- Welke visualisatie opties zijn er?
- Wat is de historie van data visualisatie?
- Welke manieren zijn er om statistische data te visualiseren?
- Welke vormen van visualisatie zijn het meest bruikbaar voor de weergave van statistische data?
- Wat is de invloed van een interactieve legenda/applicatie?

Met betrekking tot de praktijk.

- Welke online applicaties bieden al een interactie met data?
- Welke functionaliteit bieden deze applicaties?
- Welke visualisatie mogelijkheden hebben ze?
- Welke variabelen spelen een rol bij het presenteren van statistische data?
- Wat is het resultaat van een interactieve legenda? Waarin verschilt dit van de traditionele benadering?
- Wat zijn de consequenties van een interactieve data presentatie voor de visualisatie (ontwerp benadering)?

Met betrekking tot de markt (Amsterdam in cijfers).

- Welke statistische data heeft een Dienst voor Onderzoek en Statistiek in Amsterdam zoal?
- Welke gebruikers van statistische data zijn er?
- Wie zijn de doelgroepen van een dienst voor onderzoek en statistiek in Amsterdam?

1.4 Uiteindelijke uitwerking van de deelvragen

Tijdens het onderzoek bleek als snel dat een grote uitdaging lag in het onder de knie krijgen van de terminologie. Woorden zoals sparklines, scatterplot, choropleet, kartogram of beeldstatistiek en andere standaard termen binnen de wereld van datavisualisatie waren totaal nieuw voor me. Het ontbreken van een professioneel woordenboek om de technische termen te vertalen naar het Nederlands maakte het er niet eenvoudiger op. Ook de, binnen de vakliteratuur, overlappingen van de verschillende vakgebieden zoals informatie visualisatie, datavisualisatie, geovisualisatie maakte het filteren van materiaal soms lastig. Dit zorgde voor een nog duidelijkere afbakening van mijn onderzoeksgebied. Zie bijlage I.

1.5 Onderzoeksmethode

De deelvragen met betrekking tot de theorie zijn beantwoord door het uitvoeren van een literatuurstudie. Alle deelvragen over de praktijk zijn door middel van observatie (online onderzoek) en inhoudsanalyse (applicaties) beantwoord. Om inzicht te krijgen in de markt is er een open interview met verschillende personen gevoerd binnen de Dienst Onderzoek en Statistiek. Daarnaast heeft is het bestaande materiaal (data dienst onderzoek en statistiek) geïnterviewd en geanalyseerd.

1.6 Leeswijzer

Het rapport is opgedeeld in vier delen.

- Deel I (*De geschiedenis van de data visualisatie*) geeft een kort overzicht van de ontwikkelingen tot nu toe. Dit is een samenvatting van de artikelen die ik tijdens het verzamelen van het materiaal op mijn blog <http://www.latebytes.nl> heb geplaatst. Per hoofdstuk is er een verwijzing opgenomen naar de gerelateerde artikelen op Latebytes. Ik wil hier laten zien wat de succesfactoren waren van de meest bekende visualisaties.
- Deel II gaat vooral over de theoretische achtergrond van het visualiseren van statistische data. Hierin maak ik een overzicht van alle mogelijkheden die er zijn om data te visualiseren.
- Deel III is een inventarisatie van de huidige stand van zaken op het gebied van online statistische data visualisatie. Aan de hand van een paar online applicaties kijk ik wat de mogelijkheden en wat de voor- en de nadelen zijn.
- Deel IV is een inventarisatie van de online applicaties van de Dienst Onderzoek en Statistiek.

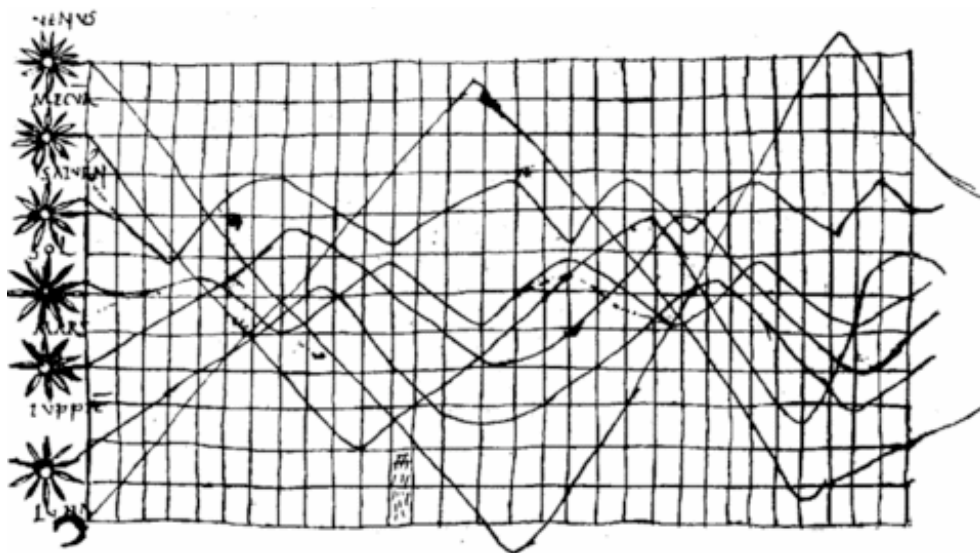
Deel I: De geschiedenis van data visualisatie in vogelvlucht

Dit hoofdstuk is bedoeld als fundament voor de overige hoofdstukken en is niet meer dan een kort overzicht van een paar hoogtepunten uit de geschiedenis. Het is een samenvatting van alle artikelen die al eerder door mij gepubliceerd zijn op de blog Latebytes.nl. In sommige gevallen heb ik dan ook een verwijzing hiernaar opgenomen. Ook zonder deze informatie is het een summier maar compleet overzicht.

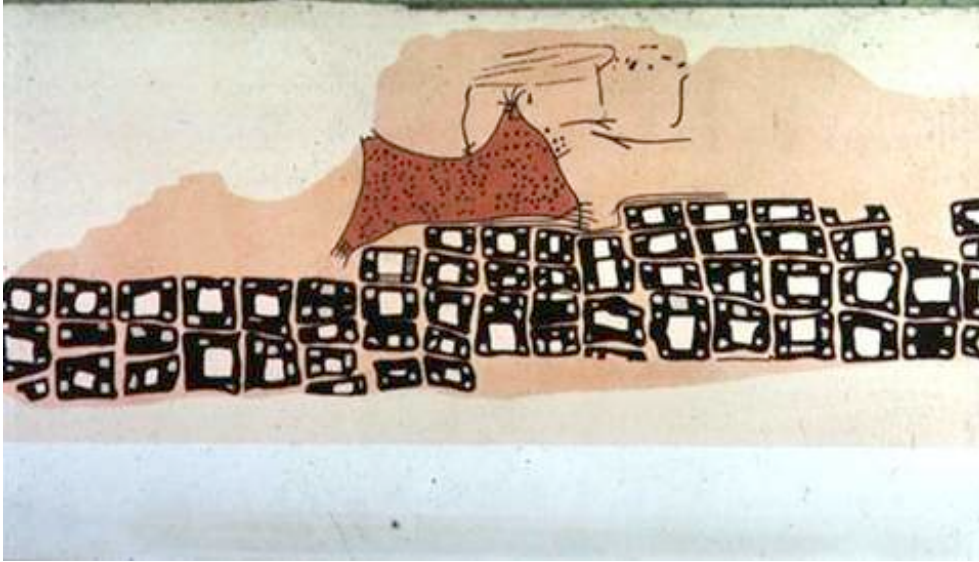
1 Voor de 17e eeuw

Michael Friendly (2007) heeft een chronologisch overzicht gemaakt van mijlpalen in thematische cartografie, statistische grafieken en andere datavisualisaties tot nu toe. Visualisaties vinden volgens hem vooral de oorsprong in geometrische diagrammen, kaarten en in tabellen met sterren standen. Hij noemt als reden de opkomst van nieuwe technieken en navigatie instrumenten in de 16e eeuw.

Voorbeelden zijn een diagram uit de 14e eeuw te bekijken waarin een onbekende astronoom de beweging van de planeten over een bepaalde periode heeft geïnventariseerd.



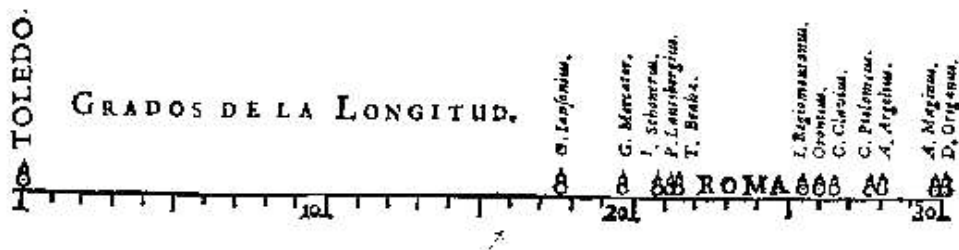
En de Catal Hyk stadsplattegrond als oudste (6.200 BC) welke te zien is in het Konya museum in Turkije (*afbeelding 2*).

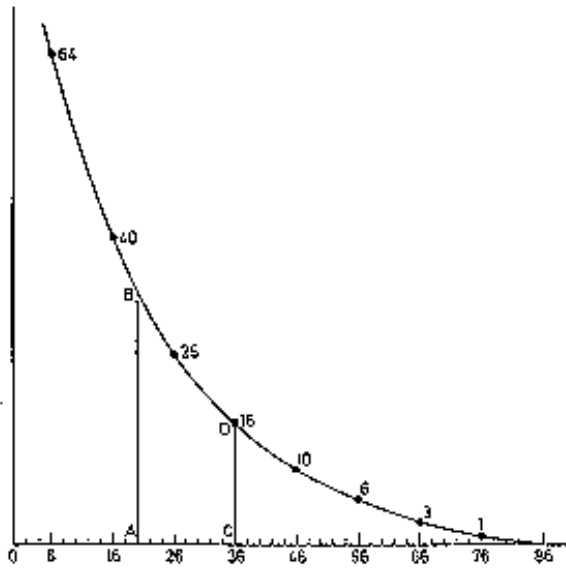


2 De 17 e eeuw

Deze eeuw kenmerkte zich vooral door het in kaart brengen van afstanden, territoriale afbakeningen, astronomie. Het gebied van schatting, waarschijnlijkheid (zie figuur 3), demografie en andere statistische begrippen ontwikkelde zich.

Voorbeeld is Michael van Langren's (1644) verbeelding van 12 verschillende bepalingen van de geografische lengte tussen Toledo en Rome. Het is waarschijnlijk de eerste weergave van statistische data (*afbeelding 3*).





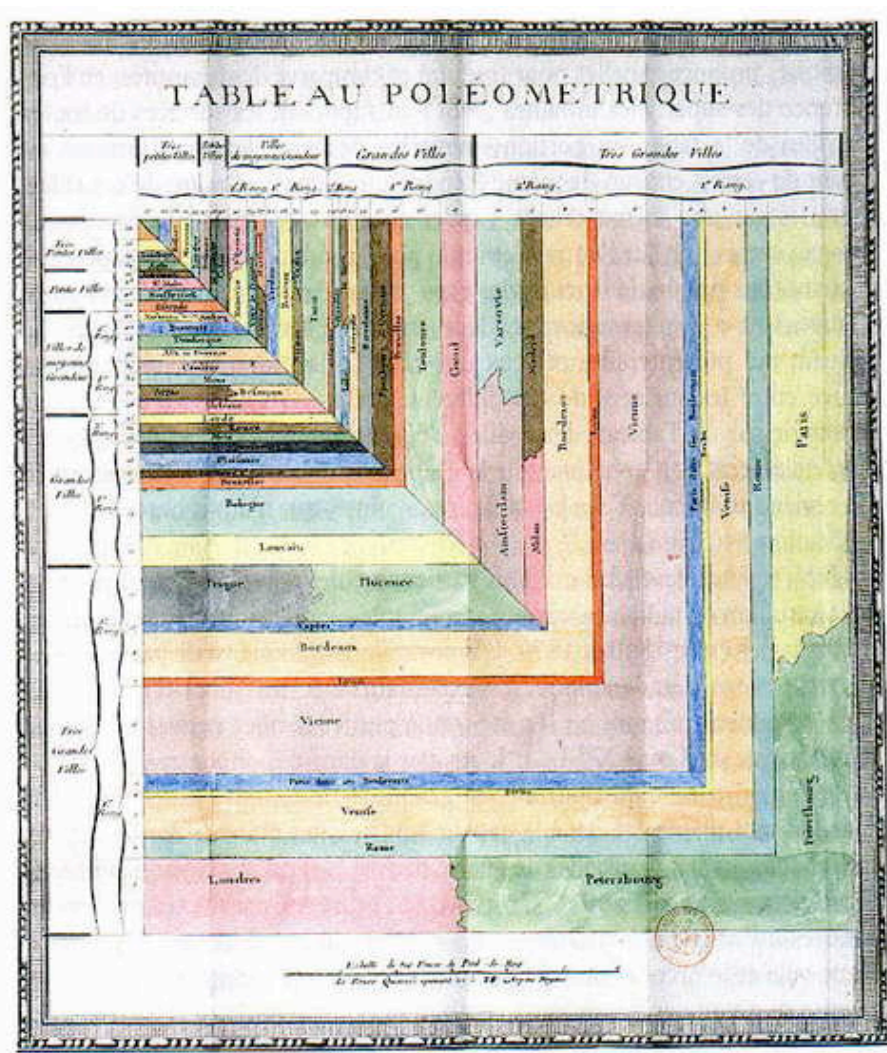
En Christiaan Huygens grafiek die de levensverwachting als grafiek weergeeft (*afbeelding 4*).

3 De 18e eeuw: het experimenteren met vorm

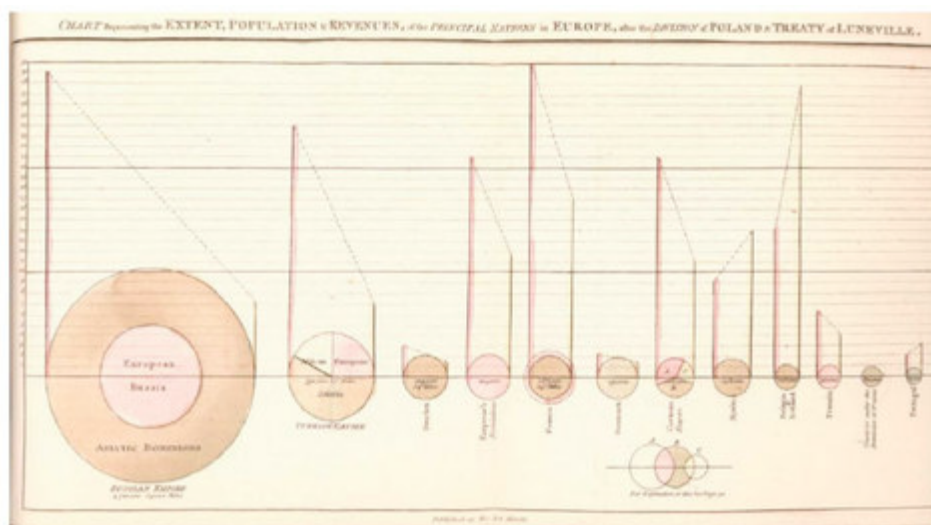
In deze eeuw probeert men vooral de gezondheid, geologie, economie en demografie in kaart te brengen. En doordat er meer data beschikbaar komt zie je nieuwe vormen van visualiseren ontstaan.

Voorbeelden, op de volgende bladzijde, zijn de 'Essai d'une table polomtrique' van Charles de Fourcroy (1782) waarbij proportionele geometrische figuren (rechthoeken) gebruikt worden om demografische hoeveelheden te vergelijken (*afbeelding 5*).

En een figuur van William Playfair welke in de linkeras en -lijn per cirkel de bevolking en in de rechteras en -lijn per cirkel de belasting laat zien. Zie artikel 'Mijlpaal in de data visualisatie' op *Latebytes*.



Essai d'une table polométrique. Charles de Fourcroy

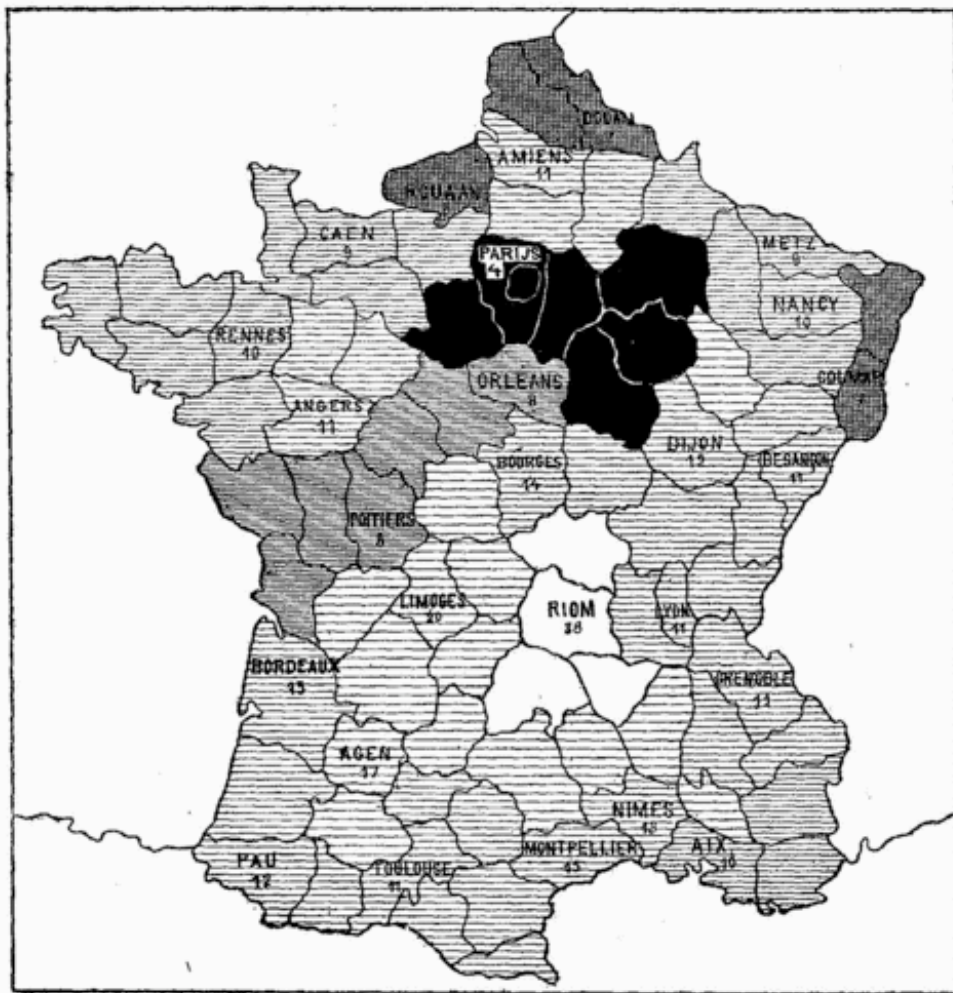


William Playfair.

4 Begin 19e eeuw

In deze periode zien we dankzij de ontwikkelingen in de 18e eeuw een explosieve groei van statistische grafieken en thematische kaarten.

Voorbeeld is de eerste choropleet kaart, dat is een kaart waarin kwantitatieve ruimtelijke gegevens met behulp van de arcering of het kleuren van bepaalde gebieden of percelen wordt weergegeven, is waarschijnlijk de 'Carte de la France obscure et la France éclairée' van Charles Dupin (1826) (afbeelding 7). Zie artikel 'Opkomst statistische visualisaties en thematische plattegronden (1800 - 1850)' op Latebytes.



5 Tweede helft 19e eeuw: het gouden tijdperk voor datavisualisatie

In deze periode heeft Minard de verliezen van Napoleons leger tijdens de tocht naar Moskou gevisualiseerd in de, volgens Edward Tufte, 'beste statistische grafiek ooit gemaakt'.

Het is een kaart, die berucht is om de hoeveelheid informatie (geografische locatie inclusief namen van steden en rivieren, tijd (alleen op de terugtocht), temperatuur (het onderste deel van de grafiek), route, richting van de troepen en het aantal (nog resterende) manschappen) die aangeboden wordt, uit 1869 laat de verliezen onder de manschappen (breedte van de banen), hun bewegingen (heen bovenste baan, terugtrekking onderste zwarte baan) en de buiten temperatuur tijdens de franse invasie naar Moskou (1812) zien. Zie artikel '*Minard, een terugkerend element*' en artikel '*Een interactieve Minard*' op Latebytes.

6 Eind 19e, begin 20e eeuw

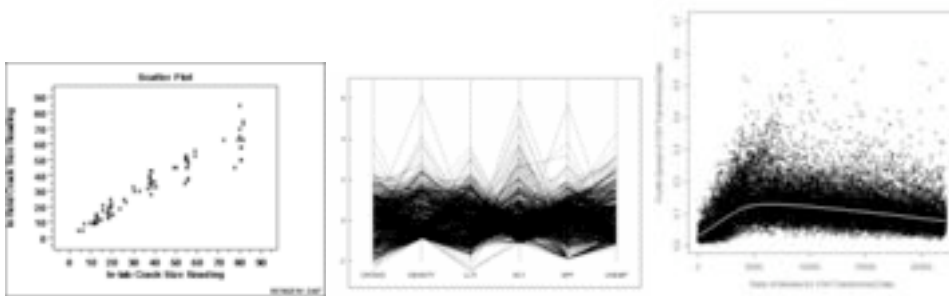
In deze periode ontstaan er overal in Europa officiële bureaus voor onderzoek en statistiek. Het belang van numerieke informatie voor handel, industrie, vervoer en sociale planning wordt onderkend. De organisaties die zich bezighouden met het verzamelen, organiseren en analyseren van officiële statistische informatie over bevolking, handel, sociale, politieke en morele zaken verspreiden zich snel over Europa. In Nederland werd in 1894 het Amsterdamse bureau voor Onderzoek en Statistiek opgericht. Het is het oudste statistische bureau van Nederland. Vijf jaar later, in 1899, werd bij Koninklijk Besluit het Centraal Bureau voor de Statistiek opgericht. In Nederland werd er nog hoofdzakelijk gewerkt met tabellen. Frankrijk, Hongarije, Duitsland en Finland werkte al veel meer met grafieken. Zie artikel '*Ontwikkelingen eind 19e, begin 20ste eeuw*'. Voor meer informatie over deze periode zie artikel '*Voor het grote publiek*' en artikel '*Sociale geografie*' op Latebytes.

7 1950 -1975

Na een stille periode begin 20e eeuw zien we zo rond 1950 een opleving in de ontwikkelingen binnen de data visualisatie. Reden hiervoor was onder andere de ontwikkeling van FORTRAN (The IBM Mathematical FORMula TRANslating System, 1957) een hogere programmeertaal speciaal ontwikkeld voor wetenschappelijke doeleinden. Tegen het einde van deze periode zien we een paar grote technische veranderingen die van grote invloed zijn op de data visualisaties zoals bijvoorbeeld nieuwe software pakketten, nieuwe computertalen, de computermuis, kleurenmonitoren, scanners, plotters. Zie artikel '*Opleving van data visualisatie (1950 -1975)*' op Latebytes.

8 Van 1975 tot nu toe

Deze technologische ontwikkelingen hebben zich sindsdien in hoog tempo doorgezet. Er komen steeds meer interactieve statistische data systemen dankzij. Nieuwe modellen en theorieën voor de directe manipulatie van visuele data analyses (linking, brushing, selection, focusing) worden ontwikkeld. Er ontstaan nieuwe manieren voor het visualiseren van high-dimensional data zoals scatterplot matrix (Tukey, 1981) parallel coördinates plot (Inselberg 1985, Wegman 1990), spreadplots (Young, 1994). Zie artikel ‘Van 1975 naar interactieve en 3D visualisaties’ van Latebytes.



Van links naar rechts: scatterplot matrix, coördinates plot en spreadplot.

9 Conclusie

Wanneer je de ontwikkelingen in datavisualisatie vergelijkt met die van de maatschappij zie je dat er vooral grote ontwikkelingen op het gebied van data visualisatie plaats vonden vlak na grote overgangen in de maatschappij. Vlak na het industriële tijdperk bijvoorbeeld, eind 18e - begin 19e eeuw, was er een enorme opbloei van data visualisatie. Zie artikel hoofdstuk 5 ‘Tweede helft 19e eeuw: het gouden tijdperk voor datavisualisatie’. En sinds het post-industriële tijdperk de opkomst van de personal computer (In 1983 introduceerde IBM zijn eerste Personal Computer in Nederland.) zie je een toename de verschillende visualisatie technieken (wordclouds, treemaps, spreadplots). Zie artikel hoofdstuk 8 ‘Van 1975 tot nu toe’.

Vandaag de dag volgen zowel de technologische als maatschappelijk ontwikkelingen elkaar in hoog tempo op. Computers worden steeds sneller, hebben steeds meer rekenkracht om ingewikkelde berekeningen uit te voeren. Het delen van informatie, het participeren hierin, rijke internet applicaties is steeds meer een gemeengoed. Het valt jammer genoeg buiten de scope van mijn rapport om in te gaan op de huidige enorme technologische ontwikkeling maar het zal zeker van grote invloed zijn op de datavisualisatie en het presenteren van gegevens. Het is een goed moment om nu te gaan kijken naar de nieuwe mogelijkheden.

Deel II: De theoretische achtergrond van het visualiseren

Dit onderdeel is een samenvatting van een uitgebreide literatuurstudie en is bedoeld als opbouw naar deel III en IV. Ook hier kan de lezer, indien hij of zij er behoefte aan heeft, meer informatie vinden op Latebytes.nl

1 De kracht van een data visualisatie

Een bak cijfers vertalen naar een visuele weergave kan ervoor zorgen dat statistische data beter gecommuniceerd wordt. Grafieken zijn zoals *Tufte (1997)*, een bekende auteur op het gebied van analytisch ontwerp, zegt 'instruments for reasoning'. Voordat het zover is zullen echter eerst een paar stappen genomen moeten worden. Ervan uitgaande dat de bak met data al verzameld is moet er gekeken worden welke cijfers met elkaar in verband kunnen worden gebracht. Is er een ontwikkeling in tijd, zijn er verschillen, is er een stijgende of dalende lijn in te ontdekken etcetra. Door deze data te rangschikken en te ordenen kunnen deze patronen zichtbaar gemaakt worden. Ontwikkelingen, schommelingen hierin, de frequentie hiervan, onderlinge relaties en een totaaloverzicht kunnen in beeld gebracht worden.

Data moet, volgens *Tufte (1997)*, zo in beeld gebracht worden dat

- vergelijkingen (het vergelijken van 2 of meer variabelen),
 - verspreidingen (hoe vaak verschillende waarden van één of meer variabelen voorkomen),
 - samenstellingen,
 - ontwikkelingen (variabelen door de tijd heen) en
 - onderliggende relaties
- zichtbaar gemaakt kunnen worden.

Dr. Andrew V. Abela (2006), een professor in marketing en marktonderzoek aan de Catholic University of America in Washington, gaat uit van 4 mogelijkheden: vergelijkingen, onderliggende relaties, verspreidingen en samenstellingen. Visueel is dit te zien in zijn 'Chart Suggestions—A Thought-Starter' (zie bijlage II).

2 Visualisatie opties

Om zinvolle informatie in geordende data zichtbaar te maken moeten deze dus met elkaar in verband worden gebracht. Je kunt dit doen door bijvoorbeeld het gemiddelde uit te rekenen, door het in groepen in te delen, door vergelijkingen of door verbanden te leggen. De statistiek kan laten zien of deze combinaties ook tot een conclusie kunnen leiden. Hiervoor kan geput worden uit een enorm aantal visualisatie mogelijkheden.

den. Het Visual Communication Lab van IBM bijvoorbeeld biedt in hun ‘levende laboratorium’ onderstaande mogelijkheden aan.

Locaties in kaart brengen (geografisch)	• Plattegrond • Kaart
Ontwikkelingen door de tijd heen zichtbaar maken	• Tijdslijn • Lijngrafiek of -diagram (Line graph) • Stapeldiagram (Stack graph)
Delen van een geheel zichtbaar maken	• Treemap • Taartdiagram of cirkeldiagram of schijfdiagram
Relaties zien tussen verschillende punten	• Netwerk diagram • Scatterplot
Vergelijken van een set variabelen	• Matrixkaart (Matrix chart) • Belgrafiek (Bubble chart) • Histogram of kolommendiagram • Staafdiagram

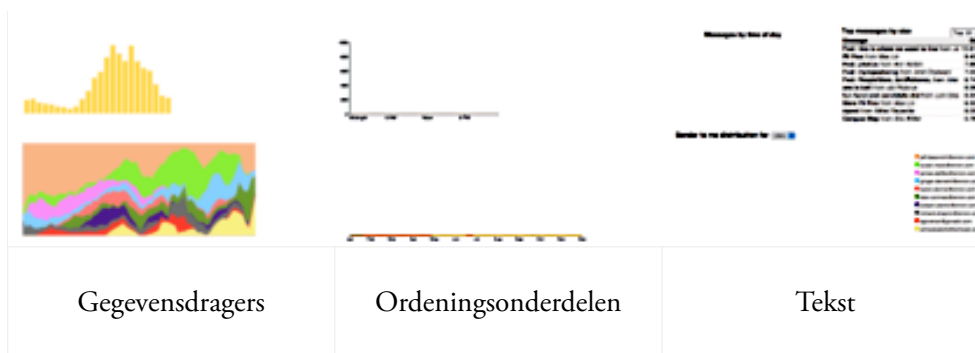
Dit is nog maar een kleine selectie van de verschillende mogelijkheden. *Robert Harris (1996)* heeft een overzicht gemaakt waarin ongeveer 4000 grafieken, tabellen en kaarten staan weergegeven (zie bijlage III).

3 De drie ingrediënten van een diagram

Een diagram of grafiek is opgebouwd uit 3 verschillende onderdelen.

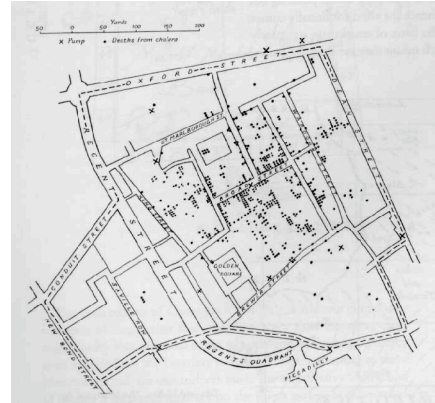
1. *Gegevensdragers* bijvoorbeeld de kolommen in een staafdiagram of de cirkels in een belgrafiek.
2. *Orderingsonderdelen* bijvoorbeeld de assen, rasterlijnen of stramien.
3. *Tekst* zoals titel, bijschriften, bronvermelding en legenda.

Soms komen daar de ondersteunende onderdelen zoals bijvoorbeeld de achtergrond of het kader nog bij. Visueel is dit het sommetje:



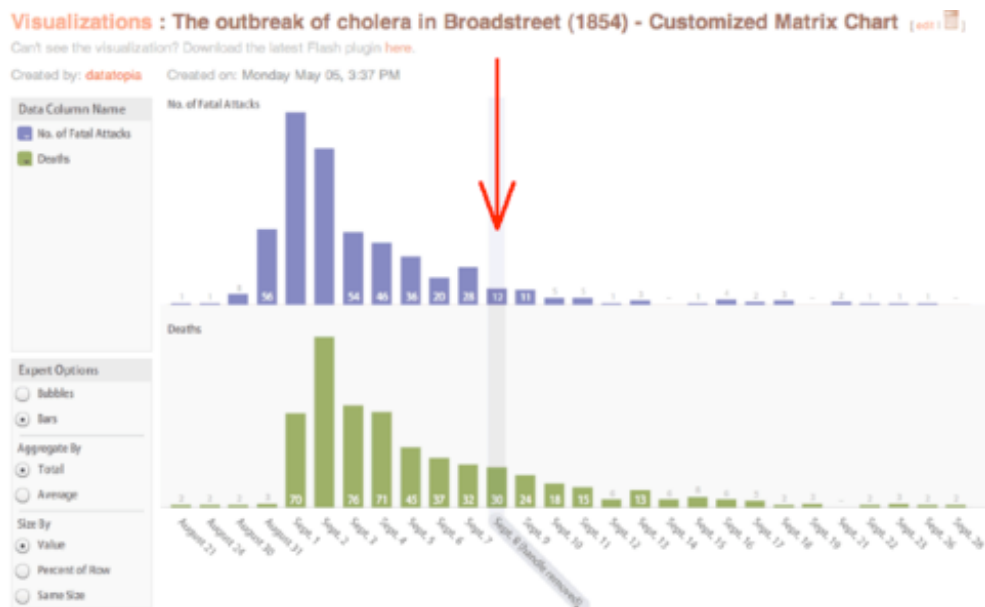
4 Subjectieve bewijsvoering

De keuze voor een visualisatie optie bepaald de uiteindelijke interpretatie van de data. Neem bijvoorbeeld de kaart visualisatie van John Snow. Door het weergeven van de sterfgevallen door cholera op een kaart van London kon hij inzichtelijk maken dat een waterbron in Broad Street de mogelijke bron van besmetting was. Op zich een knap stukje werk want men had toentertijd geen inzicht in de manier waarop de ziekte werd overgebracht. Zie bijlage IV voor een grotere afbeelding.



Uit het document 'On the mode of communication of cholera' van Snow heb ik cijfers uit 'Table I. exhibits the chronological features of this terrible outbreak of cholera.'

ingevoerd in Many Eyes. Daaruit blijkt dat zodra het aantal besmettingen met fatale afloop en het aantal sterfgevallen op een tijdslijn bekeken worden de besmettingen/sterfgevallen al aan het afnemen waren voordat de handel verwijderd werd.



Tufte zegt hierover (*Visual Explanations*) dat doordat Snow door eerdere onderzoeken al een idee had dat een waterput een mogelijke besmettingsbron zou kunnen zijn hij gericht op zoek is gegaan naar data om dit ook inzichtelijk te maken. Door zeer grondig en nauwkeurig gegevens te verzamelen over de locatie van de sterfgevallen, deze te analyseren, te ordenen en op een kaart te visualiseren kon hij aantonen dat de mensen ook daadwerkelijk water uit de bron in Broad Street gedronken hadden.

Daarnaast heeft hij ook onderzoek gedaan naar het tegenovergestelde gegeven namelijk naar de mensen die als potentiële gebruiker van de waterput besmet hadden moeten worden maar die uiteindelijk niet besmet zijn. Zoals een naastgelegen brouwerij waar de arbeiders omdat ze een bepaalde hoeveelheid van hun eigen brouwsel mochten drinken bijna niet besmet werden. Of een naastgelegen 'Work house' die zijn eigen bron had. Ook keek hij verder dan de radius rondom de bron. Hij kwam erachter dat meisjes die verder weg woonden en dus stierven op een locatie ver van de bron verwijderd daar wel in de buurt naar school gingen.

De methode waarop Snow zijn gegevens verzamelden waren, naast een goed idee op het juiste moment, volgens Tufte de bron van het succes. Het onderzoek van Snow kenmerkte zich door:

- 1 het plaatsen van dat in een geschikte omgeving (plattegrond) waardoor de impact van de gegevens versterkt werden,
- 2 hij maakte kwantitatieve vergelijkingen en stelde zich de fundamentele statistische vraag 'Vergeleken met wat?' (work house en brouwerij die wel in de gevarenzone liggen maar praktisch geen besmettingen kennen),
- 3 hij dacht na over alternatieve verklaringen en tegengestelde zaken (een dodelijk slachtoffer dat veel verder weg woonde, al in maanden niet in de buurt was geweest maar iemand dagelijks één fles water liet brengen gevuld bij de bron in Broad Street),
- 4 assesment van mogelijke fouten in de getoonde nummers in een grafiek (zijn begeleidende tekst handelde o.a. over mogelijke onnauwkeurigheden in de onderzoeksresultaten).

Kortom de manier van visualiseren maar dus ook de verzamelde gegevens zijn van invloed op de uiteindelijk conclusies die aan visualisaties ontleend kunnen worden.

5 De vier grondbeginselen van een grafiek

Tufte (1996) heeft de vier grondbeginselen waaraan een grafiek volgens hem moet voldoen op een rij gezet.

1. De legenda moet ten alle tijden zichtbaar en helder zijn.
2. De data moet de aandacht trekken, niet de vormgeving.
3. Oorzaak en gevolg moeten helder gecommuniceerd worden.
4. De data moet een goede ordening hebben.

Aan de hand van visualisaties gemaakt bij de nabeschouwingen van de ramp met de Challenger (1986) laat hij zien hoe hij tot bovenstaande opsomming is gekomen. Zie artikel '*De 4 grondbeginselen van een grafiek*' op Latebytes.

6 Visualisatie is meer dan alleen een jasje

Uit voorgaande twee hoofdstukken 'Subjectieve bewijsvoering' en 'De 4 grondbeginselen van een grafiek' blijkt dat wanneer de weergave van gegevens waar en onthullend moet zijn de logica van het ontwerp de logica van de analyse moet reflecteren. De juiste en rationele opeenvolging van oorzaak en gevolg moet zowel in het ontwerp als in de analyse van de data gelijk zijn. Tufte (1997) zegt hierover

“Visual representation of evidence should be governed by principles of reasoning about quantitative evidence. For information displays, design reasoning must correspond to scientific reasoning. Clear and precise seeing becomes as one with clear and precise thinking. p. 53)”

De kwaliteit van de vormgeving komt voort uit de intellectuele kwaliteit. De principes, nodig voor zowel het beredeneren van statistische bewijsvoering als voor het ontwerpen van statistische grafieken, omvatten

1. het documenteren van bronnen en van de karakteristieken van de data,
2. het doorlopend afdwingen van de meest geschikte vergelijkingen,
3. het demonstreren van mechanismen van oorzaak en gevolg,
4. deze mechanismen kwantitatief in beeld brengen,
5. het herkennen van de inherent zijnde multivariate (de samenhang tussen een groot aantal waarnemingen) natuur van analytische problemen en
6. het inspecteren en evalueren van alternatieve verklaringen.

Visualisaties van informatie moeten, mits een correcte weerspiegeling van inhoud en kern van de zaak,

- documentair,
- de oorzaak insluitend en verklarend,
- vergelijkend,
- in hoeveelheden uitgedrukt,
- multivariate (de samenhang tussen een groot aantal waarnemingen) en
- verklarend zijn.

7 Encourage the eye

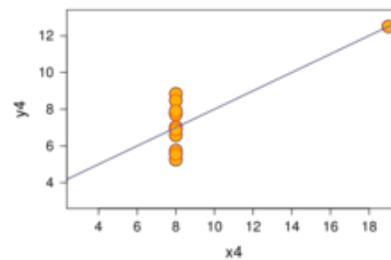
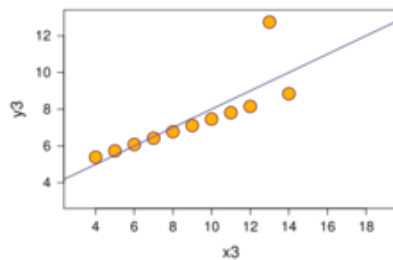
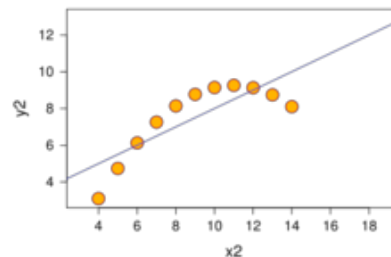
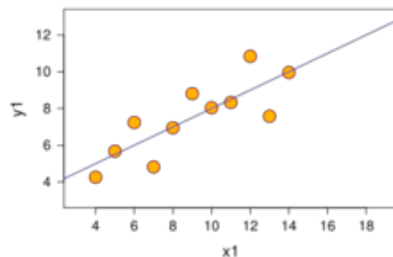
Tabellen zijn taaie kost, een rauwe opsomming van getallen. De visualisatie van de gegevens moeten 'het oog aanmoedigen' om deze data te interpreteren'. Tufte (1991) zegt dat grafische weergaven meerder functies hebben. Ze moeten:

- de data tonen;
- voorkomen dat data vervormd wordt weergegeven of geïnterpreteerd;
- een grote hoeveelheid cijfers op een beperkt oppervlak weergeven;
- een coherent geheel maken van de omvangrijke data sets;

- het oog uitnodigen om de verschillende brokken data met elkaar te vergelijken;
 - de data op verschillende niveaus onthullen, van een breed overzicht tot in het kleinste detail;
 - een duidelijke doel dienen: beschrijving, exploratie, tabuleren of decoratie;
 - nauw verweven zijn met de statistische en verbale beschrijvingen van de data set.
- Grafieken onthullen data. Tufte haalt om dit te ondersteunen Anscombe's quartet aan waarbij 4 data sets er bijna gelijk uitzien. Pas nadat deze grafisch worden weergegeven zijn de trends direct zichtbaar.

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

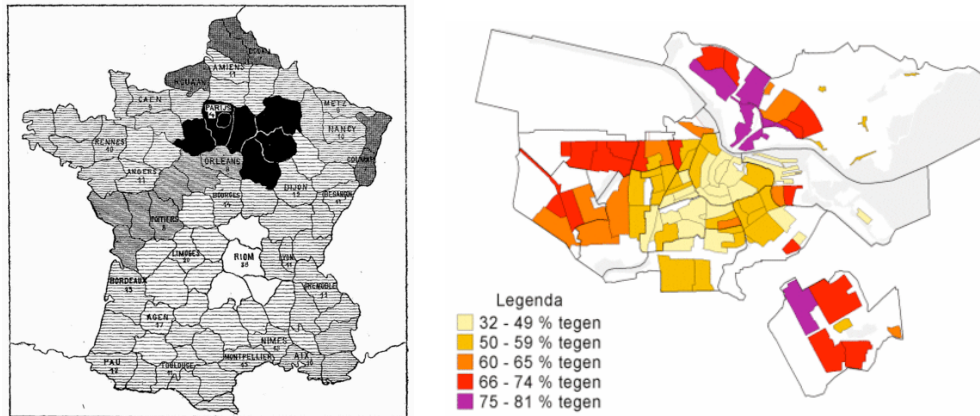


8 De vier basis vormen van een data visualisatie

Bij het communiceren van complexe kwantitatieve ideeën zijn er vier basisvormen te onderscheiden data kaarten, tijdsreeksen, plaats-tijd verhalende ontwerpen en relationele grafieken.

1. Data kaarten

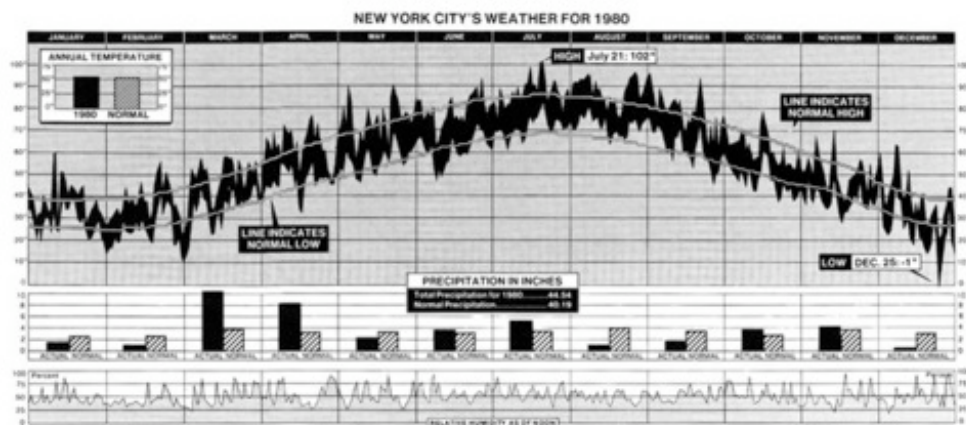
Een afbeelding kan zeer veel data op een klein oppervlak overzichtelijk presenteren. Het is ook een zeer herkenbare vorm voor het menselijke oog. De aandacht wordt direct op de inhoud gericht. Nadeel van deze methode is dat niet het aantal mensen maar een geografische grens de inhoud bepaald. De verhouding oppervlakte - aantal mensen wisselt per regio. Zoals bijvoorbeeld (afbeelding 6, 7)



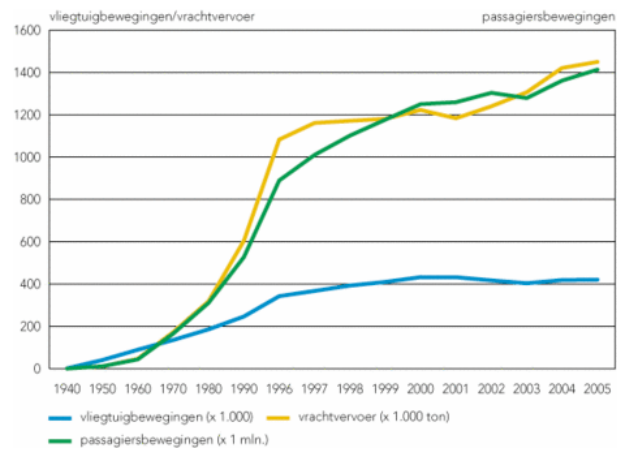
2. Tijdreeksen

Bij deze grafieken is één as altijd gereserveerd voor een tijdseenheid zoals seconden, minuten, uren, weken, maanden, eeuwen of millennia. De natuurlijke ordening naar tijd zorgt voor een krachtig, efficiënt ontwerp. Deze vorm is uitermate geschikt voor de meer complexere data. Het nadeel met tijdsreeksen is dat het verstrijken van tijd geen goede verklarende variabele is. Beschrijvende chronologie is veelal geen oorzakelijke verklaring. Uitzonderingen hierop bevestigen deze regel.

Voorbeelden zijn de New York weather in 1980. Deze grafiek uit de New York Times en de vliegtuigbewegingen, passagiersbewegingen en vrachtvervoer op Amsterdam Airport Schiphol, 1940-2005 (afbeelding 8, 9).



New York weather in 1980. Deze grafiek uit de New York Times (Jan. 11, 1981, p. 32)



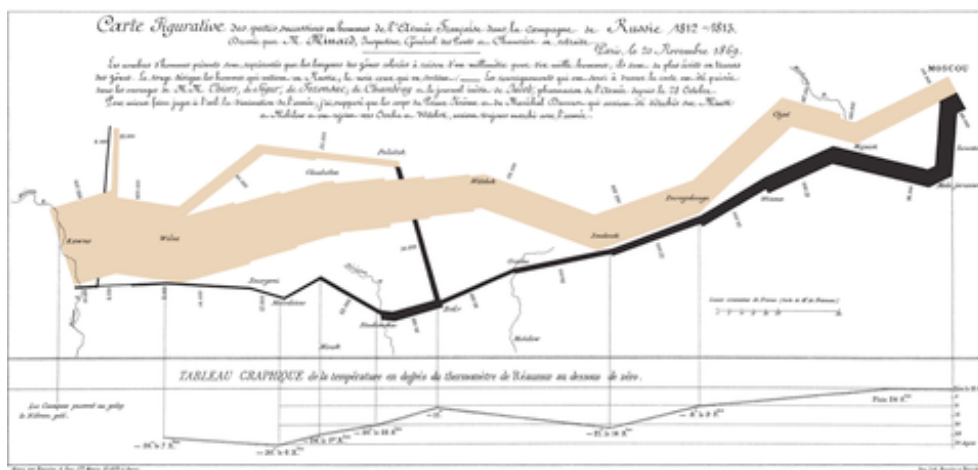
Vliegtuigbewegingen, passagiersbewegingen en vrachtvervoer op Amsterdam Airport Schiphol, 1940-2005.

3. Plaats-tijd verhalende ontwerpen

Het verschil met de tijdreeks is de ruimtelijke dimensie dat toegevoegd is aan het ontwerp. De data wordt zowel in tijd als in ruimte (2 of 3 dimensionaal) weergegeven.

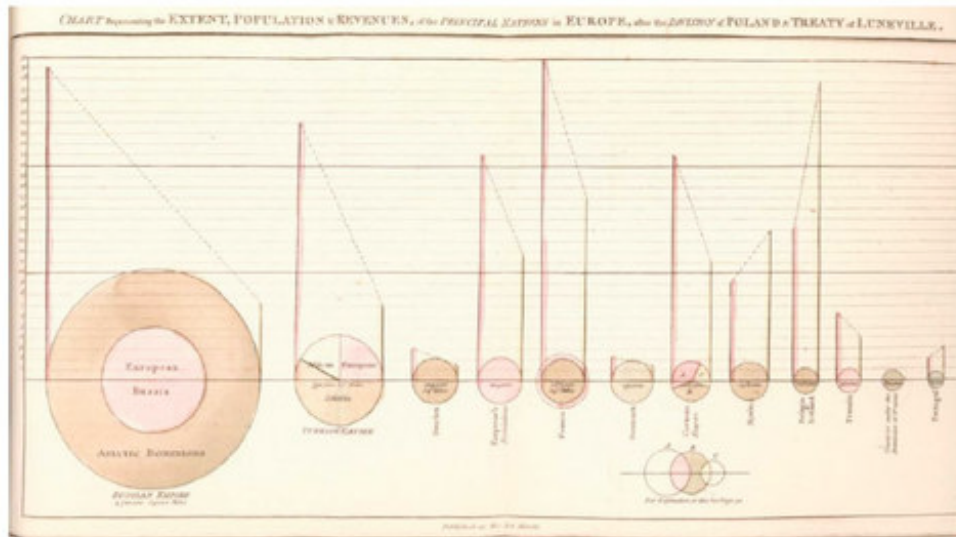
Een voorbeeld hiervan is de kaart van Minard. Tufte noemt dit zelfs 'probably the best statistical graphic ever drawn...' (Tufte, 1991).

Deze kaart, die berucht is om de hoeveelheid informatie (geografische locatie inclusief namen van steden en rivieren, tijd (alleen op de terugtocht), temperatuur (het onderste deel van de grafiek), route, richting van de troepen en het aantal (nog resterende) manschappen) die aangeboden wordt, uit 1869 laat de verliezen onder de manschappen (breedte van de banen), hun bewegingen (heen bovenste baan, terugtrekking onderste zwarte baan) en de buiten temperatuur tijdens de franse invasie naar Moskou (1812) zien. Minard gebruikt hier een voor die tijd innovatief idee. De breedte van elementen is proportioneel en is daardoor direct te interpreteren.

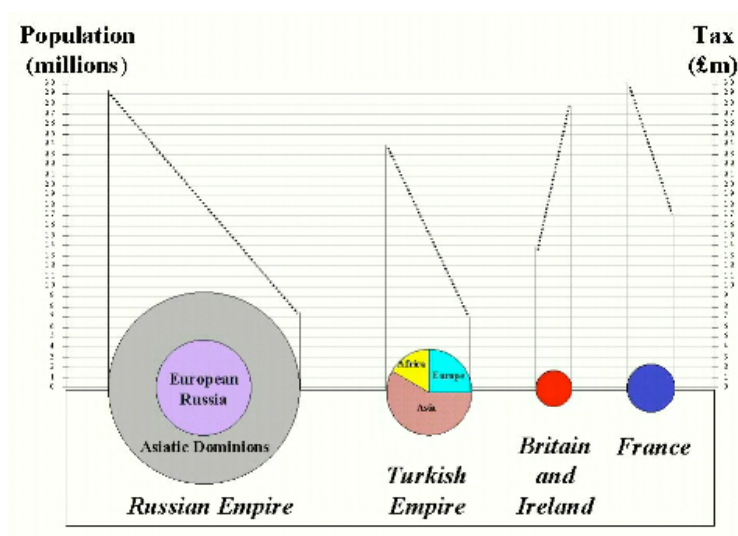


4. Relationale grafieken

Zodra er sprake is van meer abstracte data en er geen link meer is met de fysieke wereld van tijd of locatie is de relationele grafiek een uitermate geschikte vorm. Een voorbeeld is de grafiek van Playfair. Het (afbeelding 10) laat in de linker- en -lijn per cirkel de bevolking en in de rechter- en -lijn per cirkel de belasting zien.



De uitvergroting op de volgende bladzijde laat dit beter zien. De bedoeling is dat de helling die deze linker- en rechterlijn met elkaar verbindt de hoogte van belasting direct zichtbaar maakt. Alleen wordt de helling in deze afbeelding ook bepaald door de diameter van de cirkel. Het is in ieder geval wel zichtbaar dat de helling behorende bij 'Britain and Ireland' de andere kant op gaat. In de huidige statistische visualisaties is het een regel om nooit twee verticale assen voor verschillende waarden (bevolking en belasting) te gebruiken.



9 De zes principes van grafische integriteit

In het vertalen van data naar vorm kan veel, bewust of onbewust verkeerd gaan. *Tufte (1991)* heeft het daarom over 'Graphical integrity'. Volgens hem kenmerkt dit zich door een zestal principes:

1. Zorg dat de weergave van nummers, zoals daadwerkelijk gemeten binnen een grafiek, moeten gelijk zijn aan de gevisualiseerde numerieke hoeveelheden.
2. Gebruik heldere, grondige en gedetailleerde etikettering, om grafische vervorming en dubbelzinnigheid te voorkomen.
3. Toon variatie in data, niet in het ontwerp.
4. Geef geld in tijdslijnen gecorrigeerd voor inflatie en in standaardeenheden weer. Dat is veelal beter dan een nominale weergave van het geld.
5. Gebruik geen twee of driedimensionale weergave om eendimensionale data te visualiseren. Het aantal getoonde informatie-dragende dimensies mag niet het aantal dimensies in gegevens overschrijden.
6. Laat geen data zien die buiten de context vallen.

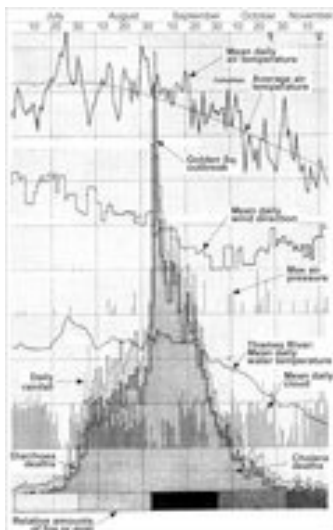
Zie artikel 'De 6 principes van grafische integriteit' op Latebytes voor voorbeelden.

10 De theorie van grafische data visualisatie

Het is me opgevallen dat bijna alle boeken en artikelen die ik heb gelezen verwijzen naar de ideeën van Tufte. Met het risico van blikvernauwing verwijs ik voor een theoretische achtergrond dus nogmaals naar Tufte. *Tufte (1991)* verdeelt het onderdeel 'theorie van grafische data visualisatie' in vijf onderdelen.

1. Data-inkt en grafische restyling

De inkt gebruikt voor de gehele grafiek mag nooit meer zijn dan de inkt gebruikt voor de gegevens.



Cholera deaths per day in London, August-September 1854. The peak reflects a doubling of deaths during the ten-day outbreak in Golden Square. Bron: The John Snow archive and research company. (Zie bijlage V voor een uitvergroting)

De data moet de aandacht van de gebruiker trekken, niet de overige onderdelen. Hij laat voorbeelden zien waarin het grid een verspilling is van inkt, waarin een staafgrafiek terug gebracht wordt tot een paar lijnen (een staaf heeft een vulling, een linker-, een rechter- en een bovenlijn, hij brengt dit terug tot 1 simpele lijn). In een ander voorbeeld laat hij zien dat indien sprake van symmetrie je eigenlijk de helft van een afbeelding

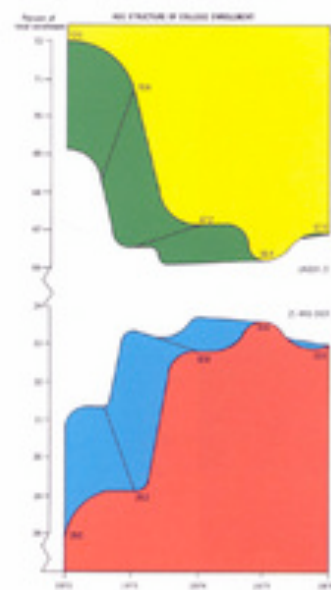
kunt weglaten (kijkers zien vaak maar de helft). De 5 principes hierin zijn:

1. Toon boven alles de data.
2. Maximaliseer de data-inkt ratio (= data-inkt - totale hoeveelheid inkt gebruikt bij het ontwikkelen van de grafiek).
3. Wis alle data-inkt welke niet gebruikt wordt voor de gegevens.
4. Wis alle overbodige data-inkt.
5. Verbeter en publiceer.

2. Chart-junk: vibraties, grids en overdesign

Waarschijnlijk "the worst graphic ever to find its way into print" (*Tufte, p. 118*). Chart junk bestaat uit decoratieve elementen die geen informatie bevatten en vooral voor verwarring zorgen.

1. Arceringen zorgen voor moiré effecten en visuele rommel.
2. Een opvallend grid bijvoorbeeld door een donkere lijnvoering zorgt voor visuele onrust en is dus chart junk.
3. Decoratieve vormen, elementen van Grafische Stijl die de kwantitatieve informatie verbergen.



Allen dragen ze bij aan de zogenoemde 'chart junk'. Technieken om chart junk te voorkomen zijn bijvoorbeeld het vervangen van arceringen en rasters door volvlakken of grijstonen, het gebruik van een directe etikettering in plaats van een legenda en het vermijden van overheersende gegevensdragers.

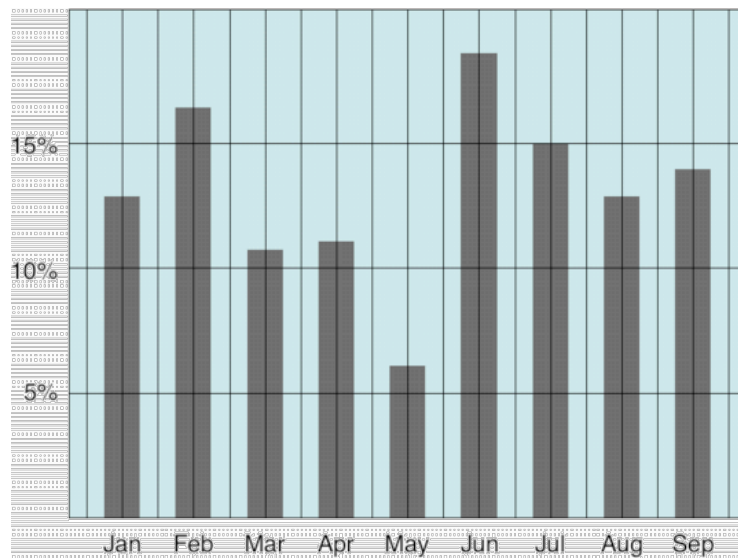
3. Data-inkt maximalisatie en grafische vormgeving

Aan de hand van voorbeelden toont Tufte (1991) hoe je de bestaande basis grafieken zou kunnen verbeteren. Het komt er vooral neer op een vermindering van het aantal gebruikte lijnen. Zijn conclusies zijn

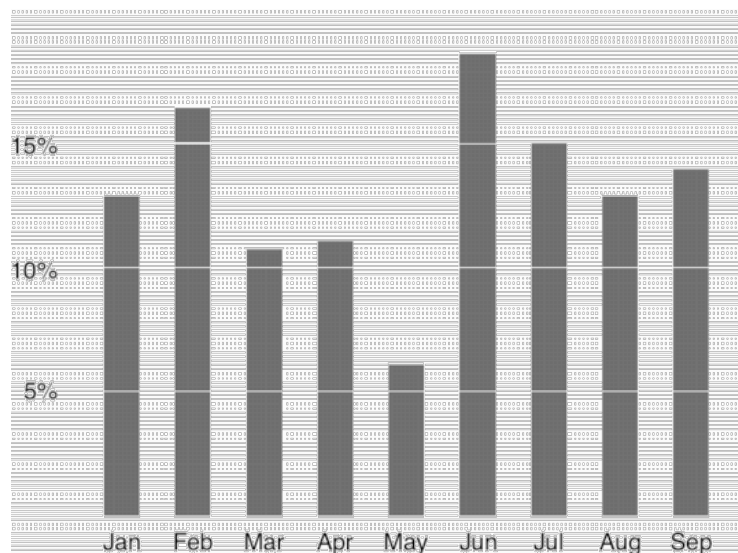
1. Sommige van deze aangepaste grafieken ogen op het eerste gezicht vreemd maar dat is dan vooral doordat we er nog niet aan gewend zijn.
2. Het optimaliseren van de data-inkt is maar een beperkt onderdeel van een complexe en multivariate ontwerp taak.
3. Grafieken zullen bijna altijd verbeteren dankzij het bewerken, reviseren en testen van verschillende ontwerpen.
4. Hou rekening met de kijker. Zal deze de nieuwe ontwerpen begrijpen of deze juist verwarrend vinden.

Een voorbeeld:

Een grafiek voor data-inkt maximalisatie:



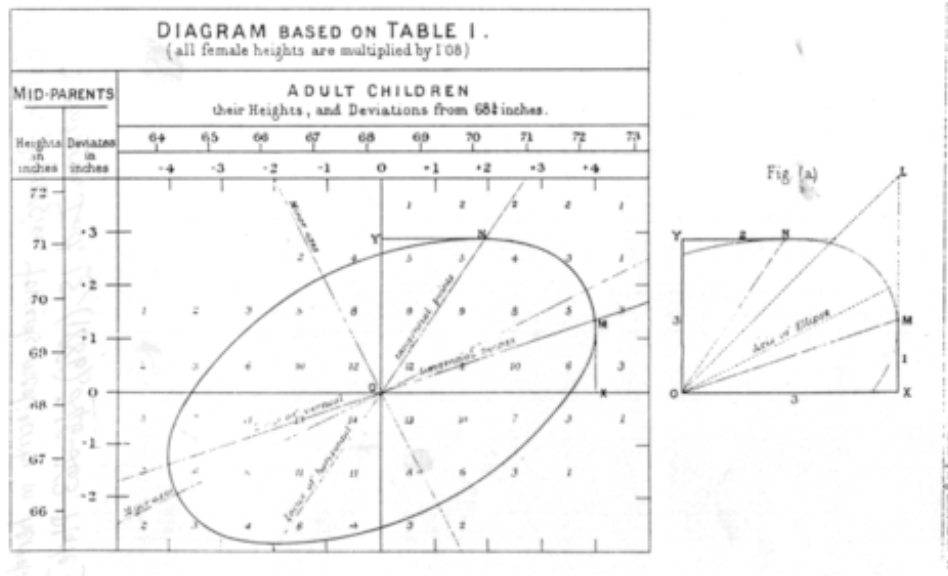
Dezelfde grafiek na data-inkt maximalisatie:



4. Multifunctionele grafische elementen

Hierin laat hij onder andere zien:

- dat de gegevensdragers (Zie '*De drie ingrediënten van een diagram*') opgebouwd kunnen worden met de data zelf,
- dat het grid zelf de gegevens zichtbaar kan maken,
- of dat de etikettering een dubbele functie kan hebben.



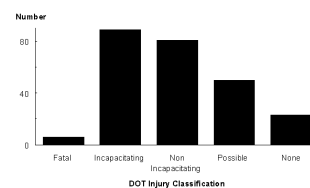
Afbeelding 11

5. Data dichtheid en kleine veelvouden

Onze ogen kunnen op een klein oppervlak heel veel verschillen in structuur onderscheiden. *Tufte (1991)* vraagt zich in het hoofdstuk '*Data density and small multiples*' dan ook af hoe je hiervan voordeel kunt behalen in een statistische grafiek en hoeveel informatie een grafiek minimaal moet bevatten om nog duidelijk te zijn. Hij heeft hiervoor een formule ontwikkeld:

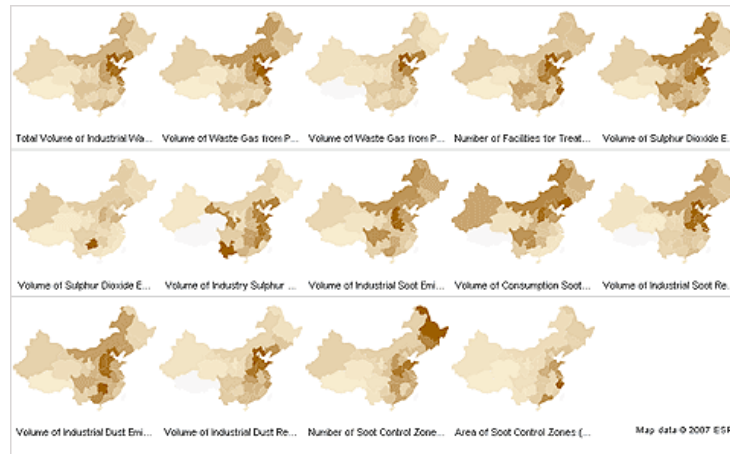
$$\text{Data density of a graphic} = \frac{\text{number of entries in data matrix}}{\text{area of data graphic}}$$

Een voorbeeld: Stel dat een staafdiagram een oppervlakte heeft van 50 vierkante centimeter en 8 ingangen heeft dan zou de data dichtheid $8 : 50 = 0,16$ (waardes per vierkante centimeter) zijn. Dat is een zeer lage dichtheid. Dit zou zeker hoger kunnen. Een wetenschappelijk tijdschrift heeft grafieken die variëren van 50 - 200 nummers per vierkante inch (1 inch = 16,387 centimeter).



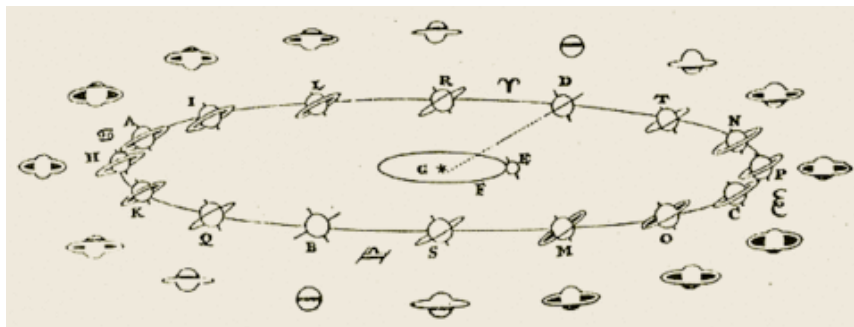
Data grafieken kunnen in het algemeen beter gebaseerd zijn op een grote data matrix waardoor ze veel informatie kunnen bevatten. Data rijke ontwerpen zorgen voor context en geloofwaardigheid van statistische bewijzen. De manier om de data density bij eenvoudige grafieken (weinig informatie) te verhogen is de matrix te verkleinen.

Small multiples of kleine veelvouden zijn een serie van grafieken die bij een gelijkblijvende variable een andere, wisselende variable in kaart brengen.



Small-multiples view: map of China showing several dimensions of air pollution.

Bron Many Eyes Blog, IBM.



Bron: Un demi siècle pour découvrir l'anneau de Saturne

Het grote voordeel van goed ontworpen kleine veelvouden is dat je makkelijk kunt vergelijken. Het zijn kleine grafieken met een hoge data dichtheid die daarnaast ook nog eens heel verhalend kunnen zijn.

6. Esthetiek en techniek in grafische data visualisatie

“Graphical elegance is often found in simplicity of design and complexity of data”

Tufte, The visual display of quantitative information, blz. 177

Een paar richtlijnen voor een goede esthetiek die hij geeft zijn

- kies voor een geschikt formaat en ontwerp
- gebruik woorden, nummers en lijnen als één geheel
- zorg voor evenwichtige en relevante schaalverdeling
- toon een toegankelijk detailleringsniveau
- laat de data het verhaal vertellen

- teken met aandacht en professioneel
- vermijd chart junk en informateloze decoratie
- gebruik verschillende lijndiktes als een aantrekkelijke en compacte manier om data te tonen
- een liggende grafiek verdient de voorkeur boven een staande grafiek (mits de content dit toelaat)

11 Conclusie

De vormgeving van een grafiek is meer dan alleen een mooi jasje, het gaat ook om de integriteit en om de doelstelling van de visualisatie. Je moet zelf voortdurend de vraag stellen wat je met de visualisatie wilt bereiken. Is het een bewijs voor een bepaald idee zoals Snow het bijvoorbeeld deed met zijn cholera map. Of wil je trends ontdekken.

We weten dat er verschillende visualisatie mogelijkheden zoals data kaarten, tijdsreeksen, plaats-tijd narrative ontwerpen en relationele grafieken zijn. We hebben geleerd welke het meest geschikt is om datastructuren binnen statistische data te presenteren. Er zijn een paar standaard richtlijnen zoals legenda kloppend en zichtbaar, data die niet onderdoet voor de vormgeving, etcetera behandeld. Maar niet alleen de manier van visualiseren is essentieel. Ook de verzamelde gegevens zijn van invloed op de uiteindelijk conclusies die aan visualisaties ontleend kunnen worden. Het meest waardevol is die visualisatie waaruit conclusies getrokken kunnen worden.

Inleiding voor deel III en IV

Tot zover dus de theorie. Maar hoe zit het nu met de markt, welke applicaties bieden online tools om data te visualiseren? En wat gebruikt een Dienst Onderzoek en Statistiek als online applicatie voor het visualiseren van data? Om een uniforme conclusie te kunnen trekken heb ik aan de hand van de theorie zoals beschreven in deel II een lijst met aandachtspunten opgesteld. Ik wil toetsen in hoeverre bestaande online applicaties uitgaan van bestaande theoretische bevindingen. Per applicaties gebruik ik volgende vragen als uitgangspunt:

- ***Orderingsonderdelen***
 - *Worden oorzaak en gevolg helder gecommuniceerd?*
 - *Wordt het oog uitgenodigd om de verschillende brokken data met elkaar te vergelijken?*
 - *Is de weergave van nummers, zoals daadwerkelijk gemeten binnen een grafiek, gelijk aan de gevisualiseerde numerieke hoeveelheden?*
 - *Heeft de data een goede ordening?*
 - *Wordt de data op verschillende niveaus onthult, van een breed overzicht tot in het kleinste detail?*
 - *Wordt de data vervormd weergegeven of kan deze verkeerd geïnterpreteerd worden?*
 - *Vormen de omvangrijke data sets een coherent geheel?*
 - *Wordt er overtollige data getoond die buiten de context valt?*
 - *Is er een evenwichtige en relevante schaalverdeling?*
- ***Tekst***
 - *Hebben de visualisaties een duidelijke titel?*
 - *Is er een duidelijke bronverwijzing?*
 - *Is de legenda ten alle tijden zichtbaar en helder?*
- ***Gegevensdragers***
 - *Laat de vormgeving toe dat de data tot op verschillende niveaus onthult wordt, van een breed overzicht tot in het kleinste detail?*
 - *Is de etikettering helder, grondig en gedetailleerd, zodat grafische vervorming en dubbelzinnigheid voorkomen worden?*
 - *Trekt de data de aandacht en niet de vormgeving?*
 - *Esthetiek en techniek*
 - *Zijn woorden, nummers en lijnen als één geheel gebruikt?*
 - *Is de data dichtheid optimaal?*
 - *Is er gekozen voor een geschikt formaat en ontwerp?*
 - *Is de vormgeving nauw verweven met de statistische en verbale beschrijvingen van de data set?*
 - *Dient de vormgeving een duidelijke doel: beschrijving, exploratie, tabuleren of decoratie?*

- Worden er verschillende lijndiktes als een aantrekkelijke en compacte manier om data te tonen gebruikt?
- Een liggende grafiek verdient de voorkeur boven een staande grafiek (mits de content dit toelaat)! Is dit principe toegepast?
- Is de data-inkt en grafische styling in verhouding?
- Ontbreekt Chart-junk (vibraties, grids) en informatieloze decoratie?
- Is de data-inkt maximalisatie en grafische vormgeving optimaal?
- **Interactiviteit**
 - Is er sprake van interactiviteit?
 - Kan gekozen worden voor verschillende vormen?
 - Kan men verschillende data sets met elkaar combineren?
 - Is er een groot gebruikersgemak?

(Zie bijlage VI)

In deel III bekijk ik twee standaard applicaties die er op de markt zijn. Dit zijn

- *Many Eyes*, een laboratorium van IBM dat zich bezighoudt met nieuwe, sociale manier, om data te analyseren te ontwikkelen en
- *Quickstep*, de online visualisatie applicatie behorende bij Swing. Swing combineert de eigenschappen van een datawarehouse, een geografisch informatiesysteem en een statistisch pakket. Swing kan op een eenvoudige wijze data ontsluiten en presenteren.

Specifiek voor O+S wordt bekeken:

- de Veiligheidsindex van Amsterdam en
- de recentelijk verschenen StadStat.

Beide applicaties visualiseren met behulp van een data kaart en een tabel thematische (veiligheid in Amsterdam en statistische gegevens over de stad aan de hand van zeven thema's) statistische data verzameld door O+S.

In deel III en IV ga ik dieper in op deze applicaties en waarom ik voor ze gekozen heb.

Voordat ik aan deze delen begin laat ik alvast de resultaten per applicatie op bovenstaande vragen zien. Hieruit blijkt dat er een groot verschil zit tussen de markt en O+S. Zowel *Many Eyes* als *Quickstep* laten een gebruiker veel zelf bepalen. Verschillende visualisatie opties kunnen zelf gekozen worden, data kan naar eigen inzicht gecombineerd of gefilterd worden. Dit is maar in heel beperkte mate mogelijk binnen de applicaties van O+S.

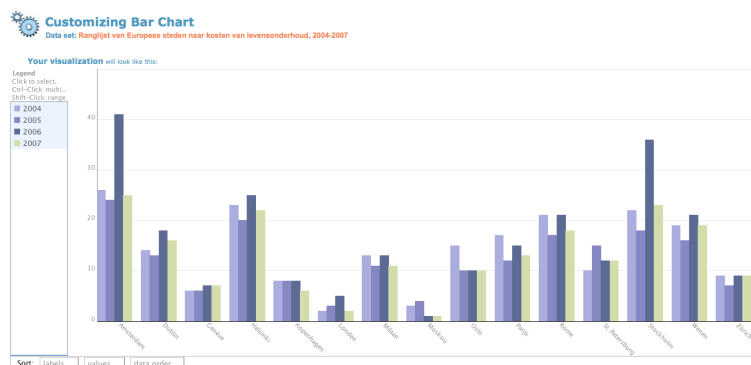
Deel III Online applicaties

1 Inleiding

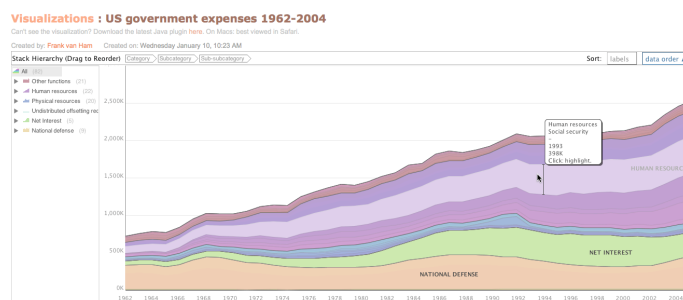
Zoals al gezegd in de conclusie van deel I is het, wanneer je de lijn der ontwikkelingen doortrekt, een moment voor data visualisatie om te groeien en te veranderen. Je ziet al steeds meer online applicaties die het mogelijk maken statistieken in te voeren. Je kunt als gebruiker je data sets uploaden en vervolgens kiezen of je je data dan als een lijndiagram, een taart of een scatterplot wil weergeven. Many Eyes van IBM is hier bijvoorbeeld mee bezig.

20.2.2 Ranglijst van Europese steden naar kosten van levensonderhoud, 2004-2007				
	2004	2005	2006	2007
Moskou	3	4	1	1
Londen	2	3	5	2
Kopenhagen	8	8	8	6
Genève	6	6	7	7
Zürich	9	7	9	9
Oslo	15	10	10	10
Milaan	13	11	13	11
St. Petersburg	10	15	12	12
Parijs	17	12	15	13
Dublin	14	13	18	16
Rome	21	17	21	18
Wenen	19	16	21	19
Helsinki	23	20	25	22
Stockholm	22	18	36	23
Amsterdam	26	24	41	25

wordt dan



waarbij de gebruiker kan sorteren op label of waarde en kan filteren op tijd. De meeste mooie visualisaties zijn mogelijk.

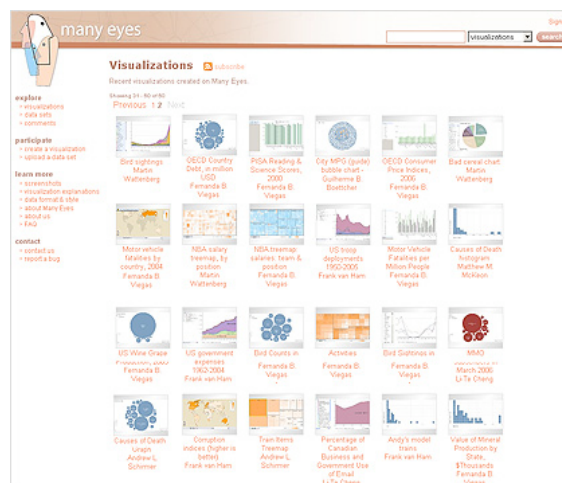


Omdat Many Eyes de grenzen van interactieve online applicaties opzoekt is dit een mooi referentiekader. Ik heb daarnaast gekozen voor QuickStep van Swing omdat deze nu al gebruikt wordt door andere gemeenten om hun data weer te geven en omdat er binnen O+S al kennis aanwezig is over dit systeem.

2 Many Eyes

Many Eyes wil de kracht van de menselijke visuele intelligentie om patronen te ontdekken zichtbaar maken. Het is hun doel om visualisaties te ‘democratiseren’ en zo een nieuwe, sociale manier, om data te analyseren te ontwikkelen.

Visualization is traditionally viewed as an efficient way of transferring a large amount of information from a database into an individual's head. We believe that visualizations become even more powerful when multiple people access them for collaborative sensemaking. By building graphical displays that spark the exchange of ideas and insights, we investigate the social and communicative value of information visualization. Our living laboratory for exploring these concepts is Many Eyes. We develop visualizations that help people see and exchange information in novel ways. Our designs aim to transform visualization from a solitary activity into a collaborative one. Some application areas are citizen advocacy, social network analysis, software development, and executive decision-making. By allowing people to observe and orient themselves in complex information landscapes, our inventions enable faster, more insightful decisions. (IBM, 2008)

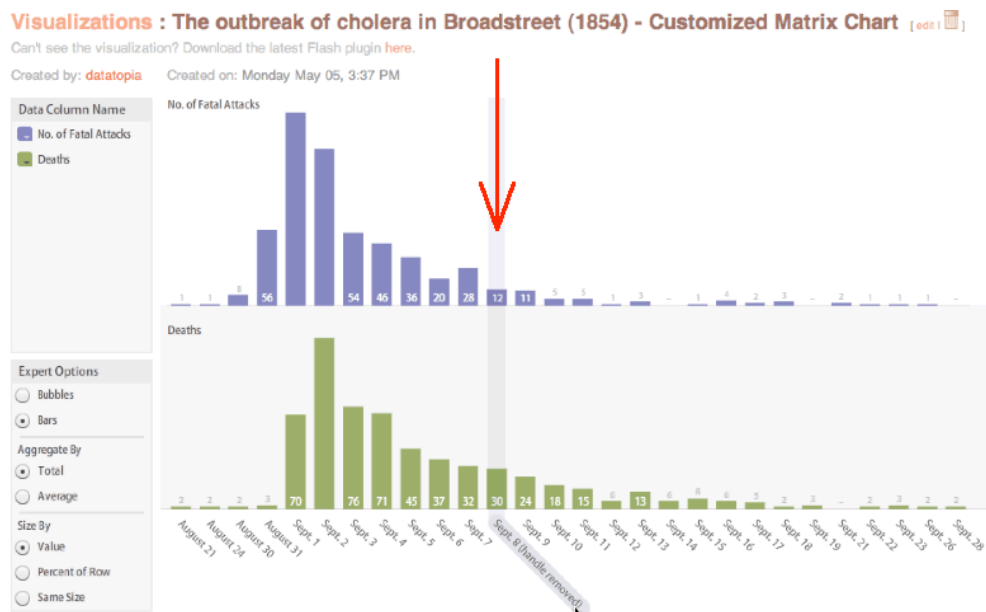


Many Eyes biedt een zeer gevarieerde visualisatie opties (zie bijlage VII). Data sets kunnen hier eenvoudig geupload en gevisualiseerd worden. Het is een prima start om de verschillende visualisatie opties te beoordelen en te testen. Mogelijkheden om verschillende data sets met elkaar te combineren zijn er niet. Als test heb ik bijvoorbeeld met bestaande data een eenvoudige visualisatie gemaakt met opzienbarende resultaten als gevolg.

De Engelse arts John Snow maakte in 1854 een data kaart waarop alle sterfgevallen in een bepaalde buurt in London zichtbaar waren. Daaruit bleek dat een waterpomp in Broad Street de bron van besmetting was. Door het verwijderen van de handel van de pomp kon er niemand meer besmet raken. Hij wilde vooral locaties in beeld brengen (zie bijlage IV).

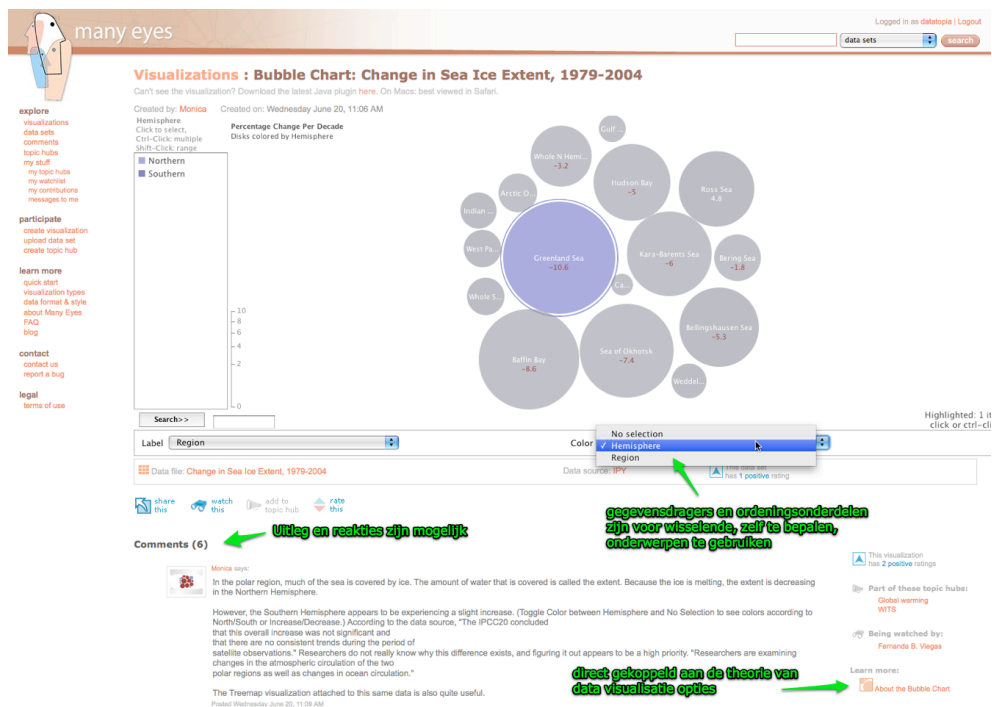
Een experiment

Ik heb dezelfde data (verkregen op de John Snow website) die Snow als basis voor zijn cholera kaart heeft gebruikt, ingevoerd in Many Eyes. Ik heb gekozen voor een kolommen diagram omdat ik een ontwikkeling door de tijd zichtbaar wilde maken. Daaruit blijkt dat zodra het aantal besmettingen met fatale afloop en het aantal sterfgevallen op een tijdslijn bekeken worden de besmettingen/sterfgevallen al aan het afnemen waren voordat de hendel verwijderd werd. Zie artikel 'Subjectieve bewijsvoering' op *Latebytes*.



Hieruit blijkt dus de kracht (het ontdekken van patronen) van het kunnen kiezen uit verschillende visualisatie opties.

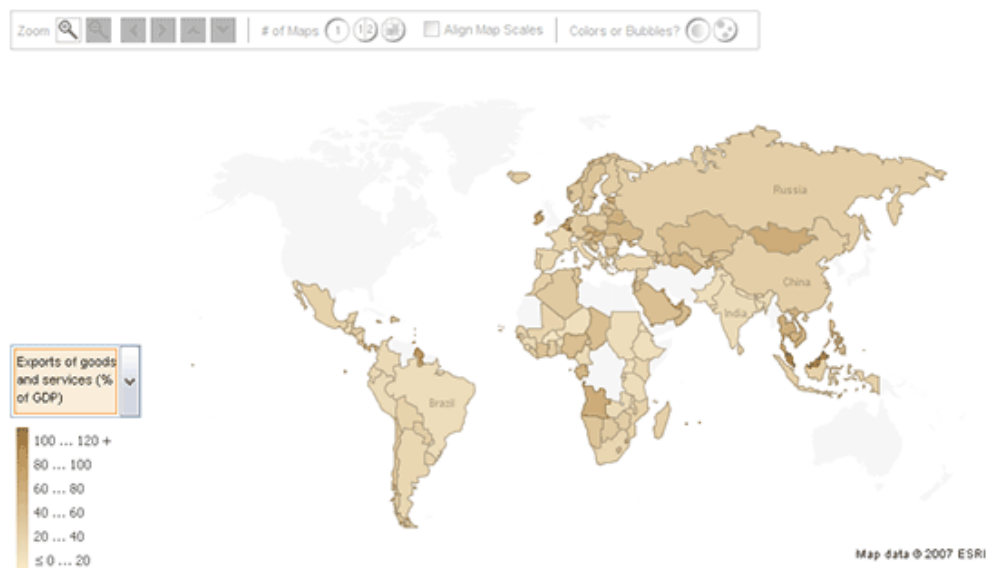
Naast het bieden van meerdere visualisatie opties kent Many Eyes veel interactiviteit. Een gebruiker kan verschillende soorten gegevens aan de ordeningsonderdelen toekennen (zie afbeelding op volgende bladzijde). Het biedt tegelijkertijd een directe link naar basisinformatie nodig voor het visualiseren van data.



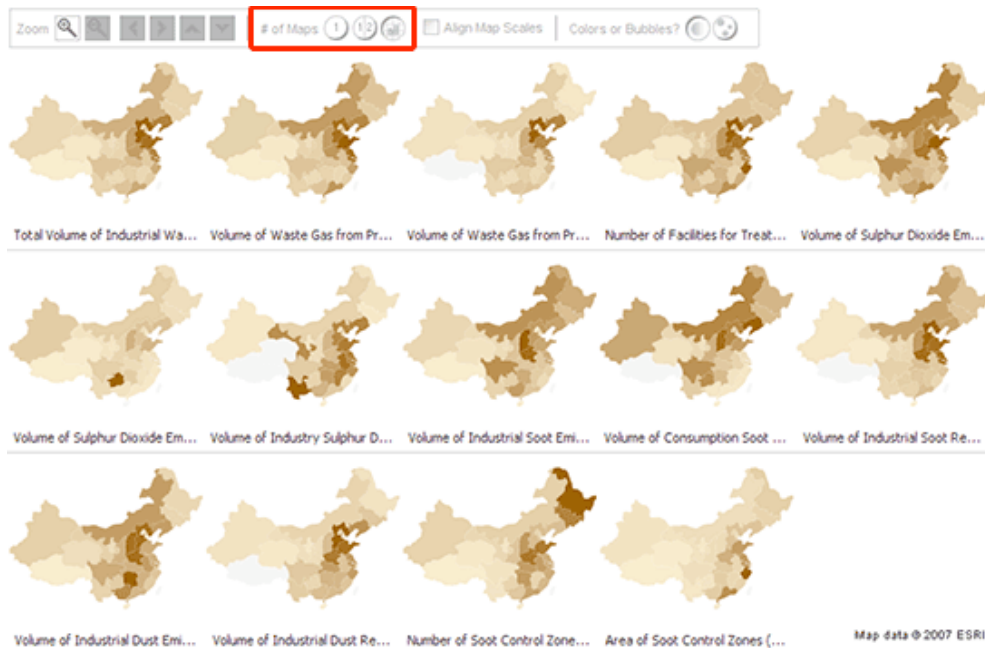
Many Eyes visualisatie opties voldoen tot op zeer hoog niveau (zie *testrapport pagina 32*) aan de grafische integriteit genoemd in deel II hoofdstuk 9 en 10. Alle voorbeelden worden gelinkt naar pagina's met zeer duidelijke uitleg en randvoorwaarden voor de data.

De mogelijkheden van bijvoorbeeld hun data kaarten zijn talrijk.

Je kunt kiezen voor de wereld of voor landen:



Je kunt zoomen, kiezen voor kleurvlakken of cirkels als gegevensdragers, je kunt gebieden selecteren en belangrijkste functie, je kunt kaarten vergelijken:



Door de grote mate van vrijheid die een gebruiker heeft zie je dat sommige visualisaties absoluut onzinnig zijn. Ze houden geen rekening met de doelstelling (*Hfdst. 5 'De vier grondbeginselen van een grafiek', deel II*) of kiezen een visualisatie optie die niet geschikt is voor de gegevens (*Hfdst. 8 'De vier basis vormen van een data visualisatie', deel II*). Daartegenover zitten er ook informatierijke toppers tussen.

Ze houden ook rekening met de ontwikkeling van het sociale web waar mensen steeds meer participeren in informatie in plaats van consumeren.

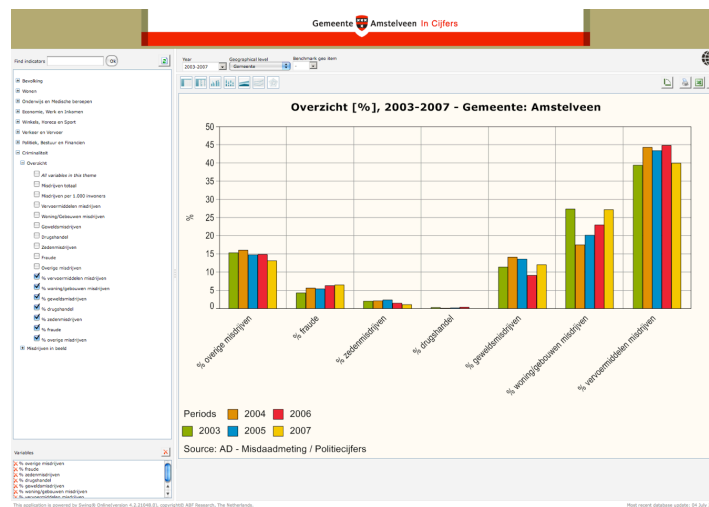


Kortom een pracht online applicatie van hoog niveau dat goed te gebruiken is als referentiekader.

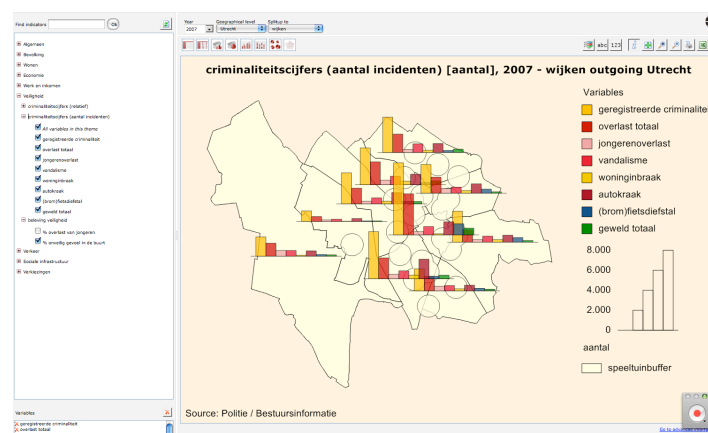
3 Quickstep (internet versie van Swing)

Met Quickstep kunnen eenvoudig statistische gegevens opgezocht en gepresenteerd worden. Het vinden van de juiste onderwerpen gebeurt aan de hand van de thema-indeling of door een zoekterm op te geven. Wanneer een onderwerp is aangevinkt wordt er direct een presentatie weergegeven. Vervolgens zijn er verschillende presentatie vormen beschikbaar waarmee de gegevens kunnen worden weergegeven en geanalyseerd. Quickstep is de internet versie van Swing. Het software programma Swing is ontwikkeld voor de opslag en presentatie van geografisch georiënteerde gegevens. Het pakket voorziet op zeer gebruiksvriendelijke wijze in de meest voorkomende functies ten aan-

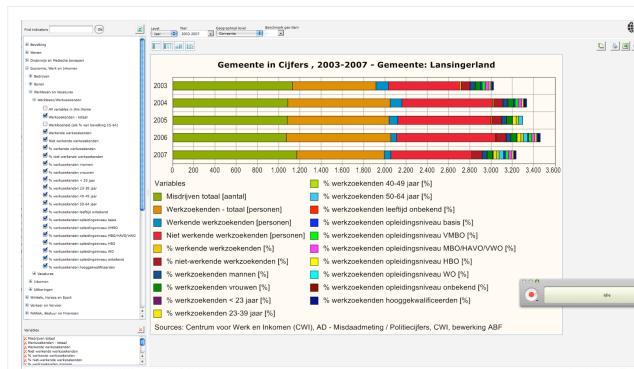
zien van beheer, bewerking, analyse en presentatie van gegevens. Het pakket houdt het midden tussen een statistisch pakket, een datawarehouse en een geografisch informatiesysteem. Tijdreeksen van verschillende lengte voor verschillende gebiedsniveau's kunnen worden samengebracht in Swing. Het kan duizenden onderwerpen aan. En is een snel en gebruikersvriendelijk informatiesysteem. Er zijn al meerdere gemeentes (Amstelveen, Rotterdam, Den Bosch, Groningen en nog veel meer) die Swing in combinatie met Quickstep gebruiken.



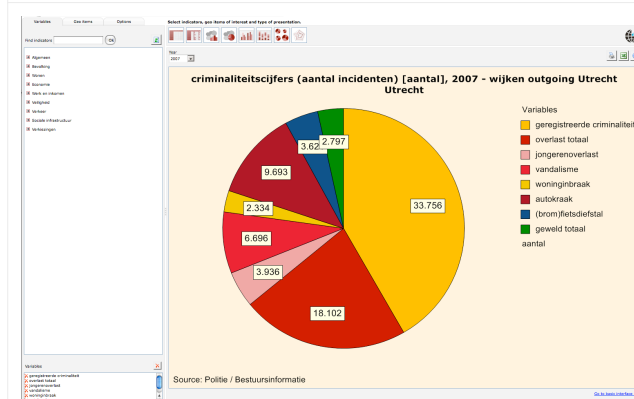
Quickstep van Swing is een goed voorbeeld van de mogelijkheid om statistische data te combineren en te visualiseren. Er zijn meerdere visualisatie opties. En meerdere data sets kunnen gecombineerd worden. Ook zou O+S nog enige filtering kunnen aanbrengen op de mogelijkheden tot visualiseren en combineren. Ook hier blijft de mogelijkheid voor een gebruiker om de getoonde data te downloaden.



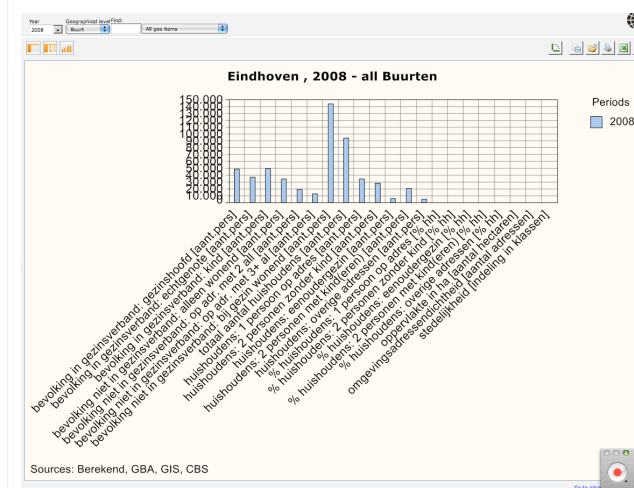
Ondanks dat de vormgeving in veel gevallen er goed uit ziet gaat het toch af en toe verkeerd. Moeilijk leesbare grafieken tot resultaat hebbend:



Gegevensdragers op een kluitje waardoor het geheel onleesbaar wordt.



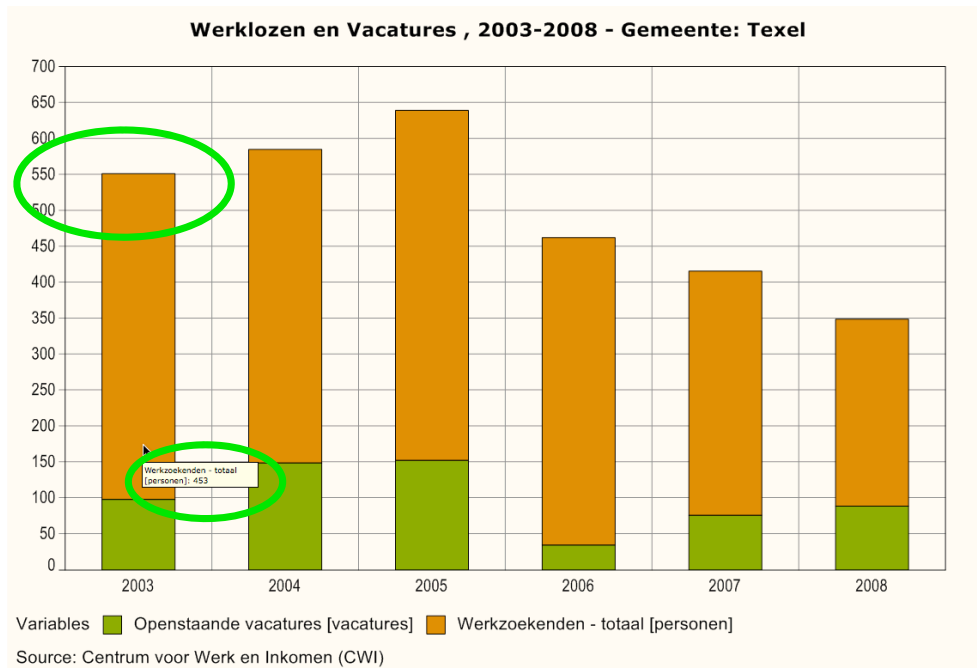
Nietszeggende vormgeving. Kolomdiagram zou al beter zijn.



Onleesbare ordeningsonderdelen waardoor gegevens verkeerd geïnterpreteerd kunnen worden.

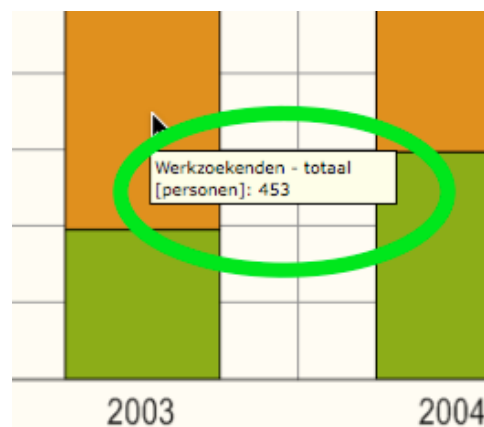
Deze onleesbaarheid wordt vaak veroorzaakt doordat de applicatie teveel visualisatie opties biedt en doordat er teveel gegevens die eigenlijk weinig zin hebben worden gecombineerd. Er is nader onderzoek mogelijk om de mogelijkheid tot optimaliseren in beeld te brengen.

Quickstep heeft een paar duidelijke verbeter punten. Ik heb bijvoorbeeld bij de gemeente Texel gekeken hoeveel mensen er werkloos zijn en hoeveel vacatures er zijn. Het resultaat is:



Legenda is duidelijk zichtbaar, ook de schaalverdeling klopt. Maar als je nu naar de gegevensdragers kijkt begint de oranje staaf (vacatures) niet op nul. Het aantal werkzoekenden in 2003 bedroeg 453 personen. De staaf eindigt echter op 550!

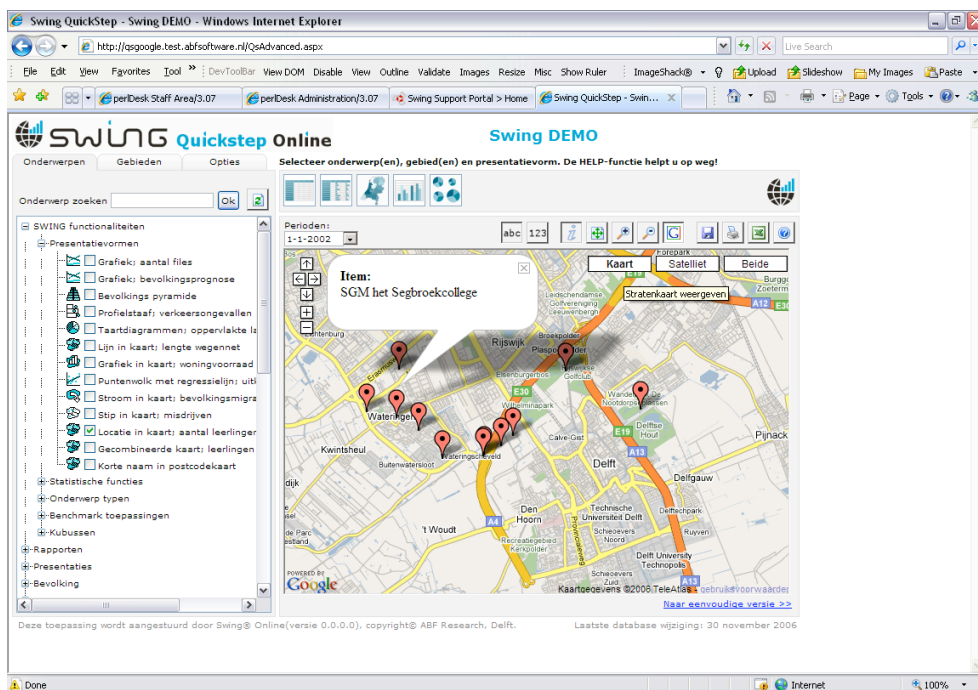
Quickstep voldoet duidelijk niet aan alle punten van grafische integriteit zoals beschreven in deel II hoofdstuk 9. Punt 1 (*Zorg dat de weergave van nummers, zoals daadwerkelijk gemeten binnen een grafiek gelijk zijn aan de gevisualiseerde numerieke hoeveelheden*), punt 2 (*Gebruik heldere, grondige en gedetailleerde etikettering, om grafische vervorming en dubbelzinnigheid te voorkomen*) en punt 3 (*Toon variatie in data, niet in het ontwerp*) worden veelvuldig overtreden. Gegevensdragers zijn regelmatig frivole onleesbare elementen geworden, teksten worden zonder problemen over elkaar heen geplaatst of onleesbaar schuin gezet.



4 Conclusie

Many Eyes laat zien dat online interactieve data visualisaties van hoge kwaliteit kunnen zijn. Door het koppelen naar de theorie van data visualisatie is het zelfs voor een leek eenvoudig te gebruiken. Echter informatieve, kwalitatieve data visualisaties vereisen inzicht in en kennis van de materie. Er moet nagedacht worden wat de doelstelling is van een data visualisatie. Maar de kern van het visualiseren, het kunnen ontdekken van onverwachte patronen, wordt in deze applicatie (met name dankzij de mogelijkheid tot interactie met de data) heel dicht benaderd (zie deel II hoofdstuk 1).

Quickstep (Swing) kan kwalitatief zeer slechte visualisaties genereren en het is binnen dit onderzoek niet in te schatten of dit, door iemand met meer kennis van data visualisatie, beter te stroomlijnen is. Het grootste probleem zit erin dat een gebruiker teveel gegevens kan combineren, er is teveel interactie met de data mogelijk. Het resultaat is ten alle tijden een visuele brei die in geen enkel opzicht voldoet aan grafische integriteit. Ook is er een vertekening van de gegevens. Of dit nu komt door degene die de data in Quickstep invoert of door gebreken in de applicatie is mij nog niet duidelijk. Noemenswaardig is nog dat op de website van Swing gezegd wordt dat 'Swing (en dus ook Quickstep) in 2007 gekoppeld is aan Google Maps en Google Earth'. De voordelen hiervan zijn talrijk maar dit valt buiten de scope van dit rapport.



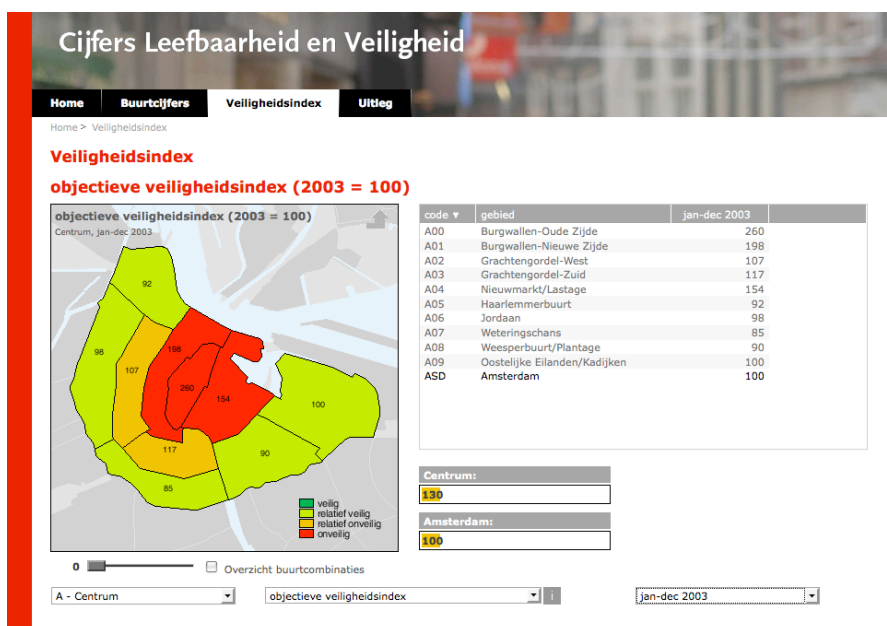
Deel IV Dienst Onderzoek en Statistiek

1 Inleiding

In het vorige hoofdstuk heb ik gekeken naar de huidige online mogelijkheden. Many Eyes heeft qua vormgeving en visualisatie opties het meest te bieden. Quickstep sluit het dichtst aan bij de data die de Dienst Onderzoek en Statistiek heeft. Beiden bieden gebruikers de mogelijkheid zelf datasets interactief kunnen combineren en gegevens kunnen filteren. Voor O+S heb ik gekozen uit de twee meest recente online voor het publiek toegankelijke applicaties. Deze bieden niet echt de mogelijkheid om datasets te combineren maar zijn wel, zij het in beperkte mate, interactief. Natuurlijk is het een beetje appels met peren vergelijken omdat de doelstellingen van de vier gekozen applicaties verschillen. Maar omdat ik niet inga op de techniek en de data strategieën die erachter liggen is het verantwoord.

2 Amsterdamse Veiligheidsindex

De Amsterdamse veiligheidsindex is gebaseerd op twee belangrijke informatiebronnen. Enerzijds is dat het bevolkingsonderzoek waarin zo'n 10.000 Amsterdammers hun mening geven over de veiligheid in Amsterdam en in hun buurt. Anderzijds zijn dat de gegevens van de politie Amsterdam-Amstelland over inbraak, diefstal, geweld, overlast, vandalisme en drugs. De index laat zien hoe de veiligheid zich op deze zeven onderdelen ontwikkelt (objectief) en hoe de veiligheidsgevoelens van de Amsterdamse bewoners zich ontwikkelen (subjectief). Je kiest interactief voor buurt, jaartal en onderdeel. Ook de achterliggende politie- en bevolkingsonderzoekcijfers zijn te bekijken.



Zo op het eerste gezicht is het prettig om de cijfers in verhouding tot een locatie gepresenteerd te krijgen. Het ziet er goed uit. De gegevens worden op een klein oppervlak overzichtelijk presenteren. Het is ook een zeer herkenbare vorm voor het menselijke oog. De aandacht wordt direct op de inhoud gericht. Nadeel van deze methode is dat niet het aantal mensen maar een geografische grens de inhoud bepaald. De verhouding oppervlakte - aantal mensen wisselt per regio.

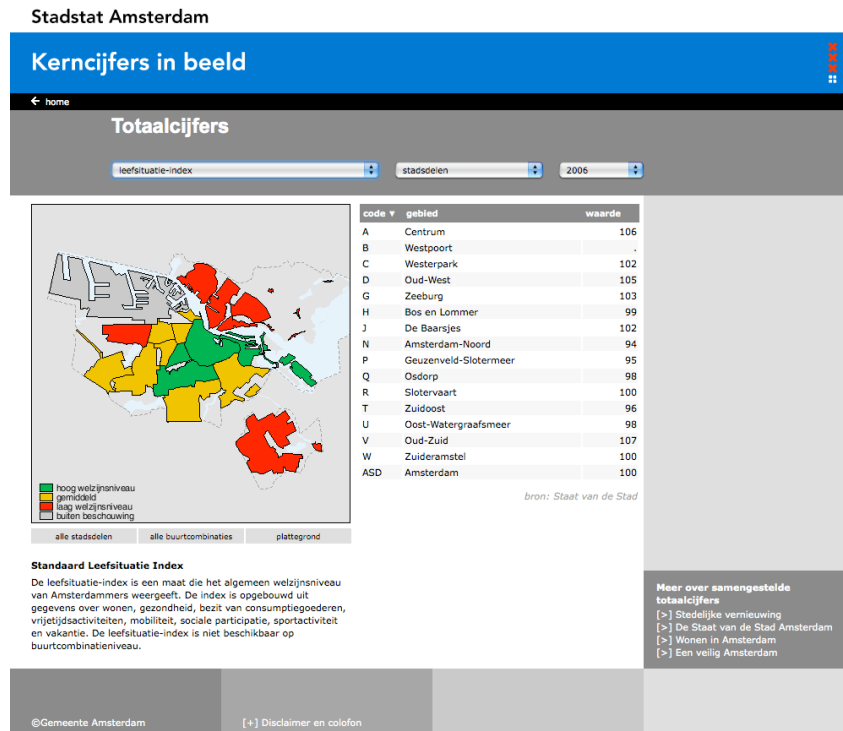
De applicatie bevat daarnaast geen 'rijke' informatie. Het is niet mogelijk om oorzaak en gevolg te achterhalen. Er zijn geen patronen te ontdekken. Je kunt wel in de tabel twee jaargangen met elkaar vergelijken. Jammer genoeg niet in een, voor deze data geschikte visualisatie, tijdreeks. En het is ook niet mogelijk om een langere periode (2003 - 2007) in één keer te vergelijken.

Ik mis hier de mogelijkheid van kleine veelvouden waardoor ik bijvoorbeeld de kaart als basis voor vergelijking zou kunnen gebruiken.

Ik heb niet kunnen controleren of de weergave van de nummers, zoals daadwerkelijk getoond binnen de kaart en de tabel, gelijk is aan de gevisualiseerde numerieke hoeveelheden (*zie deel II hoofdstuk 9*). De cijfers zijn niet als data in te kijken (of zoals bijvoorbeeld het CBS of alle gemeentes die Quickstep gebruiken als Excel te downloaden) en de bronvermelding, essentiële informatie, vind ik pas nadat ik op een vraagtekentje heb geklikt.

3 StadStat

Stadstat volgt de voortgang van de beleidsdoelstellingen van het gemeentebestuur. Het Programma akkoord 2006-2010 is ingedeeld in zeven thema's: Amsterdam Topstad, armoedebestrijding, basisvoorwaarden, gezonde lucht, Kinderen Eerst, overlast/veiligheid en sociale cohesie. Je kunt zien hoe de Amsterdamse buurten scoren op diverse kwaliteitsindexen. Deze indexen beschrijven meerdere thema's tegelijk.



De data visualisatie binnen deze omgeving komt in grote mate overeen met de Amsterdamse Veiligheidsindex. Het sterke van deze omgeving is dat het bieden van thema's als ingang veel houvast biedt. Jammer genoeg kunnen ook hier ontwikkelingen, vergelijkingen en onderliggende relaties niet zichtbaar gemaakt worden.

De overeenkomsten in vormgeving tussen Amsterdamse veiligheidsindex en StadStat zijn zo groot dat ik om het overzichtelijk te houden mijn bevindingen over beide applicaties als één geheel presenter.

4 Bevindingen Amsterdamse veiligheidsindex en StadStat

Ik heb de applicaties onderzocht heb aan de hand van de vragenlijst (zie deel II hfdst. 11). Hierbij viel het me op dat ze de beschikbare **visualisatie opties** (zie deel II hfdst. 8) eigenlijk niet benutten. Veel informatie zoals trends of ontwikkelingen door de tijd heen vragen om een relationele grafiek. Geen van beide omgevingen biedt echt een mogelijkheid tot filteren of combineren van informatie.

De **kracht van een data visualisatie** (deel II hfdst. 1) is om te ontdekken of er een ontwikkeling is in tijd, of er verschillen zijn, of er een stijgende of dalende lijn in te ontdekken is etcetra. Door data te rangschikken en te ordenen kunnen patronen zichtbaar gemaakt worden. Ontwikkelingen, schommelingen hierin, de frequentie

hiervan, onderlinge relaties en een totaaloverzicht kunnen in beeld gebracht worden. Dat principe komt niet overtuigend naar voren in de beide applicaties van O+S.

In *hoofdstuk 2 van deel II* heb ik beschreven dat er voor de verschillende doelstellingen bepaalde **visualisatie opties** het meest geschikt zijn. Om locaties in kaart te brengen is de data kaart zeer geschikt, om ontwikkelingen door de tijd heen te laten zien is er de tijdslijn, de lijngrafiek of de stapeldiagram. Voor het vergelijken van een set variabelen kun je bijvoorbeeld een kolommen diagram of staafdiagram gebruiken. O+S maakt in haar andere uitingen zoals website en drukwerk wel gebruik van bredere visualisatie opties. Waarom dan in beide applicaties alleen één data kaart en een simpele tabel. Het principe van kleine veelvouden en tijdslijnen zouden uitermate geschikt zou zijn om data van O+S optimaal te communiceren.

In *hoofdstuk 4 ‘Subjectieve bewijsvoering’ van deel II* konden we lezen dat het bieden van een alternatieve visualisatie een enorme meerwaarde kan hebben. Hier bleek namelijk dat de data gevisualiseerd op een tijdslijn in plaats van een op een data kaart het aantal besmettingen met fatale afloop en het aantal sterfgevallen en de besmettingen/sterfgevallen al aan het afnemen waren voordat de handel verwijderd werd!

In *hoofdstuk 6 van deel II ‘Visualisatie is meer dan alleen een jasje’* hebben we gezien dat het demonstreren van mechanismen van oorzaak en gevolg en deze mechanismen kwantitatief in beeld brengen onderdeel zijn van de basis waarop het ontwerpen van statistische grafieken gebaseerd is. O+S levert door middel van onderzoek informatie die behulpzaam is bij het ontwikkelen en evalueren van beleid. Of zoals ze zelf zegt : *O+S beschikt over uitgebreide en actuele - al dan niet statistische - gegevens over Amsterdam op het gebied van bijvoorbeeld wonen, leefbaarheid, bevolking en werkgelegenheid, veiligheid en infrastructuur.*

en

O+S bezit en verzamelt op allerlei manieren een schat aan informatie, die in de meest klantvriendelijke vorm bij de klant terecht moet komen. Alle marketingactiviteiten zijn gebaseerd op het principe dat de opdrachtgever centraal staat.

Maar het oorzaak-gevolg principe of de mogelijkheid om alternatieve verklaringen te inspecteren en te evalueren wordt nergens geboden.

Gezien de huidige ontwikkelingen op het gebied van data visualisatie en de veranderende maatschappij waarin de markt actief met informatie omgaat en hier ook in participeert (zie *deel I hoofdstuk 8 en 9*) is het een meerwaarde om hier met een online applicatie op in te springen. O+S biedt zowel in haar publicaties als op haar site deze - al dan niet statistische - gegevens over Amsterdam op het gebied van bijvoorbeeld wonen, leefbaarheid, bevolking en werkgelegenheid, veiligheid en infrastructuur. Op de site zijn deze gegevens zelfs als Excel te downloaden. Een interactieve applicatie met

meerdere visualisatie opties zou deze gegevens juist enorm verrijken. Wat er nu is is eigenlijk niets meer dan een alternatief jasje (data kaart) van hetzelfde.

5 Conclusie

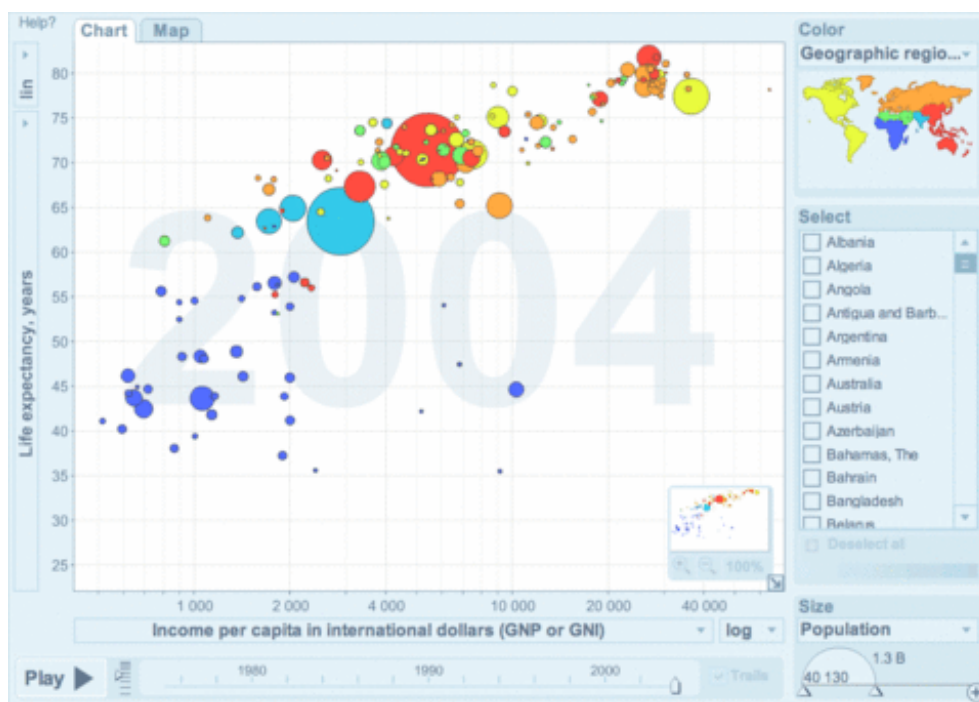
Terwijl aan de grafische integriteit in grote mate wordt voldaan wordt aan de essentie van interactieve data visualisatie voorbijgegaan. Natuurlijk wordt er een beperkt patroon zichtbaar gemaakt. Maar dat is het topje van de ijsberg. Alles onder water wordt niet getoond. Is niet bereikbaar voor een gebruiker. Is waarschijnlijk ook niet gekoppeld. Ik stel me, na het bekijken van beide omgevingen, oprecht de vraag of het alle inspanningen waard is geweest. Was een verhaal met visualisaties zoals bijvoorbeeld gebruikt in alle print uitingen van O+S niet veel krachtiger geweest?

Zoals blijkt uit Many Eyes en Quickstep biedt interactiviteit met als doelstelling ontwikkelingen, schommelingen hierin, de frequentie hiervan, onderlinge relaties en een totaaloverzicht kunnen in beeld kunnen brengen een enorme meerwaarde. De data zo kunnen rangschikken en ordenen dat deze patronen zichtbaar worden. Dàt is de toekomst van interactieve online applicaties.

In mijn afsluitende aanbeveling introduceer ik Gapminder. Niet omdat dit precies is wat O+S nodig heeft. Daar is nader onderzoek voor nodig. Het geeft alleen naast dit rapport een kader waarbinnen nieuwe stappen genomen zouden kunnen worden.

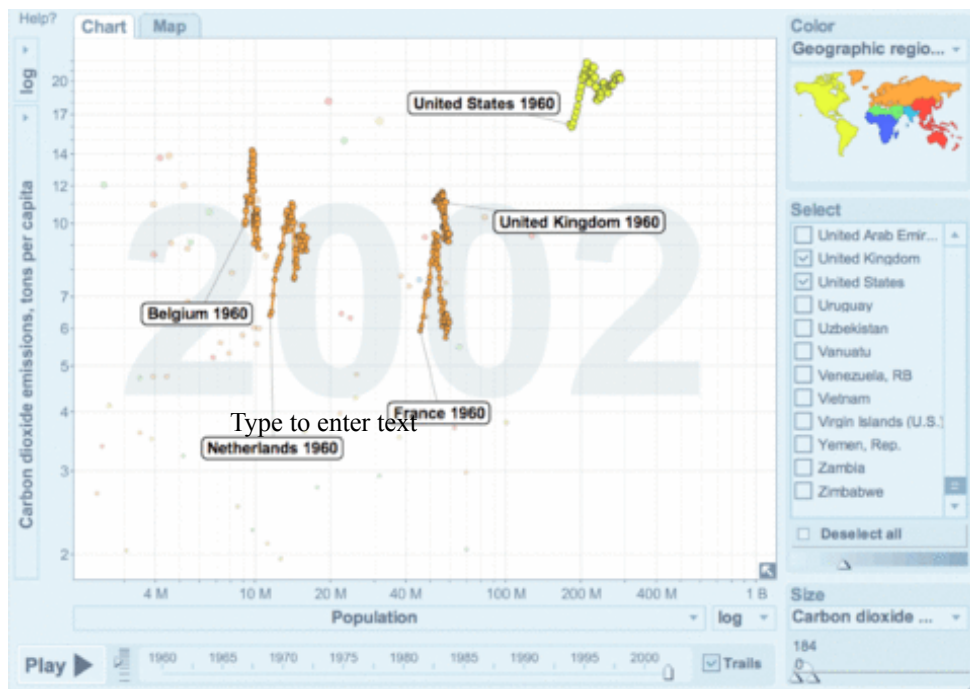
Aanbevelingen

Verder onderzoek is hard nodig. De huidige online omgeving van O+S zou ter discussie moeten worden gesteld. Quickstep van Swing zou vooralsnog een goed alternatief kunnen zijn. Dit rapport biedt duidelijk niet het antwoord dat nodig is om het beleid van online data visualisatie op te starten. Het geeft alleen houvast voor verder onderzoek. Zoals in de conclusie van deel III al geïntroduceerd ga ik kort in op Gapminder. Het ontdekken van patronen, mits in alle redelijkheid gebruikt, is bij Gapminder zeer eenvoudig.



Februari 2006 heeft Hans Rosling tijdens TED talks laten zien dat data uit verschillende bronnen gekoppeld zou kunnen worden aan visualisaties. Een belangrijk onderdeel de mogelijkheid tot zoeken niet achterwege latend. Dit is gebaseerd op Trendalyzer software die internationale statistieken omzet in bewegende, interactieve en plezierige grafieken. (*Latebytes, Gapminder, april 2008*) Het doel was om begrip te kweken bij het grote publiek. En het inzicht in de wereldproblematiek te vergroten door het aanbieden van vrij toegankelijke openbare statistieken. In zijn lezingen gebruikt hij deze grafiek 'Gapminder' om wereldontwikkeling te visualiseren.

Gapminder biedt de mogelijkheid om de gegevens welke getoond worden op x- en y-as zelf aan de hand van een paar voor gedefinieerde data sets zelf te combineren. Ook de omvang van de gegevensdrager kan gekozen worden. Er kan een spoor aangezet worden om ontwikkelingen door de tijd zichtbaar te maken.



Maart 2007 heeft Google Trendalyzer overgenomen van Gapminder zodat ontwikkelingen versneld kunnen worden en er een verbeterde versie voor het publieke beschikbaar komt. Recentelijk is daar de eerste versie van online gekomen. Hier kan op een eenvoudige manier uitgeprobeerd worden wat de meerwaarde zou moeten zijn van een online interactieve applicatie:

<http://www.gapminder.org/blog/gapminder-foundation-blog/make-your-own-graph-google-announces-motion-chart.html>

Bronnenlijst

- Michael Friendly (May 1, 2007). *Milestones in the history of thematic cartography, statistical graphics, and data visualization*.
- Michael Friendly. *Gallery of Data Visualization The Best and Worst of Statistical Graphics*. Geraadpleegd tussen februari en juni 2008, van: <http://www.math.yorku.ca/SCS/Gallery/>
- Edward Tufte. *Charles Joseph Minard, "Graphic Tables and Figurative Maps"*. Geraadpleegd op 14 april 2008, van: <http://www.edwardtufte.com/tufte/minard-maps>.
- Edward Tufte. *The Visual Display of Quantitative Information, Second Edition*, Graphics Press, USA, 1991.
- Edward Tufte. *Computing Lie Factor by Dividing Percentages*. Geraadpleegd op 15 mei 2008, van: http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00002o
- Mark Zachry and Charlotte Thralls. *Cross-Disciplinary Exchanges . An Interview with Edward R. Tufte*. Technical communication quarterly, 13(4), p. 447–462, 2004.
- Edward Tufte. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, USA, 1997.
- Chaomei Chen. *Information Visualization: Beyond the Horizon*. Springer; 2nd ed. 2004. 2nd printing edition. 2006.
- Edward Tufte. *Beautiful Evidence*. Graphics Press, USA, 2006.
- Robert L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, USA; 1 edition. 2000.
- Dr. Andrew V. Abela. *Chart Suggestions—A Thought-Starter*. 2006. Geraadpleegd op 21 april 2008, van: <http://extremepresentation.typepad.com/blog/2008/06/visualization-taxonomies.html>
- *Van Taarten En Balken*. Grafisch Nederland - Kerstnummer Grafisch Nederland 1996.
- IBM. *Many Eyes*. Geraadpleegd tussen februari en juni 2008, van: http://services.alphaworks.ibm.com/manyeyes/page/Visualization_Options.html
- Wade Robinson et al. *Representation and Misrepresentation: Tufte and the Morton Thiokol Engineers on the Challenger*. Science and Engineering Ethics (2002) 8, 59–81.
- Ron Graham, Wolfgang Hees, Lisa Henn. *Viewgraphs*. Geraadpleegd tussen februari en juni 2008, van: <http://www.tcnj.edu/~rgraham/rhetoric/viewgraphs.html>
- The John Snow archive and research companion. Geraadpleegd op 5 mei 2008, van: <http://matrix.msu.edu/~johnsnow/>

- IBM. Visual Communication Lab. Geraadpleegd tussen februari en juni 2008, van <http://www.research.ibm.com/visual/>.
- Moritz Stefaner. Visual tools for the socio-semantic web. Master's Thesis Interface Design Programme. University of Applied Sciences Potsdam. June 2007.
- Latebytes. *Nieuwe manieren om globale statistische data te presenteren*. 2008. <http://www.latebytes.nl/archives/2008/04/post-1.html>.
- Latebytes. *Intelligente legenda's*. 2008. <http://www.latebytes.nl/archives/2008/01/intelligente-le.html>.
- Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, Matt McKeon. *Many Eyes: A Site for Visualization at Internet Scale*. Infovis, 2007.
- Frank van Ham. *Interactive Visualization of Large Graphs* PhD thesis, Technische Universiteit Eindhoven, 2005.
- Frank van Ham. *Collaborative Visualization : IBM's Many Eyes project*. Video gezien op http://www.seeconference.org/#/en/Livestream/?stream=seeconference3_frankvanham. See #3 conference, Wiesbaden, Germany. April 19 2008.
- ABF Research. *Swing Quickstep 4.2 demo*. Geraadpleegd in mei en juni 2008 op <http://qslight.demo.abfsoftware.nl/>.
- ABF Research, Swing. *Buurtmonitor*. Geraadpleegd in mei en juni 2008 op <http://www.buurtmonitor.nl/>
- ABF Research. *Gemeente in cijfers. Statistische informatie over uw gemeente op internet*. Geraadpleegd in mei en juni 2008 op <http://www.gemeenteincijfers.nl/>
- Dienst Onderzoek en Statistiek. *StadStat Amsterdam. Kerncijfers in beeld*. Geraadpleegd juli 2008 op <http://stadstat.osamsterdam.nl/>
- Gemeente Amsterdam: Dienst O+S, Directie Openbare Orde en Veiligheid. *Veiligheidsindex Amsterdam*. Geraadpleegd juni 2008 op <http://buurtcijfers.eenveiligamsterdam.nl/veiligheidsindex/>
- Bestuursakkoord stad en stadsdelen 2006 - 2010. Gemeente Amsterdam. *Bijlage 2: Op weg naar Stadstat*.
- Jan Kraak. *Waarom werkt informatie- visualisatie zo goed?* Pictogram 1. Februari, maart 2005
- Chris Houwing. *Visualisatie Statistische gegevens*. Afstudeerscriptie Master Software Engineering. Universiteit van Amsterdam. 21 juni 2004.
- The Economist (print edition). *Charts. Worth a thousand words*. Dec 19th 2007.

Lijst van afbeeldingen

- Afbeelding 1: Funkhouser, H.G. (1936). 'A note on a tenth century graph'. *Osi-
ris* 1 (1936) pp. 260-262.
- Afbeelding 2: *Catal Hyk stadsplattegrond*. Geraadpleegd op 10 april, van
<http://www.atamanhotel.com/catalhoyuk/oldest-map1.html>.
- Afbeelding 3, 4, 6, 7: Michael Friendly. Gallery of Data Visualization The Best
and Worst of Statistical Graphics. Geraadpleegd tussen februari en juni 2008,
van: <http://www.math.yorku.ca/SCS/Gallery/>
- Afbeelding 5: Palsky (1996): *Des Chiffres et des Cartes*, p.51-52.
- Afbeelding 5: Choropleth map with shadings from black to white (distribution
and intensity of illiteracy in France), the first (unclassified) choropleth map, and
perhaps the first modern statistical map. Bron: "Milestones in the History of
Thematic Cartography, Statistical Graphics, and Data Visualization" van Michael
Friendly en Daniel J. Denis, York University, Canada.
- Afbeelding 7: Aandeel nee-stemmers Referendum Europese Grondwet per buurt-
combinatie. Bron: Dienst Onderzoek en Statistiek. Gemeente Amsterdam.
- Afbeelding 8: New York weather in 1980. Deze grafiek uit de New York Times
(Jan. 11, 1981, p. 32; Tufte (1983), p. 30) [1] toont, op een succesvolle uitda-
gende manier, 2200 cijfers die de trend en patronen van het weer in New York in
1980 samenvat. De drie gecombineerde grafieken tonen temperatuur, neerslag en
relatieve vochtigheid. In de temperatuur grafiek is het gebied tussen het dagelijk-
se minimum en maximum gevuld.
- Afbeelding 9: Vliegtuigbewegingen, passagiersbewegingen en vrachtvervoer op
Amsterdam Airport Schiphol, 1940-2005. Bron: Dienst Onderzoek en Statistiek.
Gemeente Amsterdam.
- Afbeelding 10: Florence Nightingale's Statistical Diagrams. Hugh Small.
[http://web.utk.edu/~leon/stat571/Nightingale/Florence%20Nightingale%20Mus-
eum%20Web%20Site_files/small.htm](http://web.utk.edu/~leon/stat571/Nightingale/Florence%20Nightingale%20Mus-
eum%20Web%20Site_files/small.htm)
- Afbeelding 11: Bron: Pearson, K. (1920). Notes on the history of correlation.
Biometrika, 13, Figure 2: Galton's "Table of Correlation".

Verwijzingen

- Latebytes: <http://www.latebytes.nl/>.
Mijn blog over informatie in een netwerktydperk.

Latebytes...

Informatie in een netwerktydperk

Categories

- Data en visualisatie (50)
 - Applicaties (4)
 - Historie (15)
 - Theorie (12)
- De metadata van informatie (10)
- Informatie en architectuur (4)
- Informatie en ordening (7)
- Informatie en privacy (7)
- Informatie en visualisatie (35)
- Informatie management (8)
- Informatie- en mediarecht (8)
- Kennismanagement (4)
- Trends en maatschappelijke ontwikkelingen (33)
- Usability & user-centered design (6)
- Werkveld Informatiemanager (8)
- Zoeken naar informatie (18)
- Boeken (9)
- Events (7)
- Glossary of terms (6)
- HVA MIM (9)
- Knipselkrant (1)
- Links (35)

Archives

Select a Month...

Recent Assets

Kijken achter de cijfers

June 24, 2008 1:21 PM | Permalink | Comments (0) | TrackBacks (0)

Afgelopen zaterdag (21 juni 2008) is in het Parool een uitgebreid artikel verschenen over de publicaties 'Kijken achter de cijfers' en 'Criminogeniteitsbeeld Amsterdam 2008'.

Dragen van slag- en steekwapens

Indexscores

- > 130
- 110-130
- 90-110
- < 90

Kaarten op basis van recentste cijfers

100 = 0.544 incidenten per 1000 inwoners

Bron: O+S/VU © Het Parool

Onderzoekers hebben voor het eerst tot op wijkniveau in kaart gebracht waar in Amsterdam het risico op criminaliteit het grootst is - en door welke factoren. Nieuw is dat niet zozeer is gekeken naar de misdaadcijfers, maar juist naar omstandigheden die criminaliteit in de hand werken: het aantal laagopgeleide ouders in een buurt, allochtonen, krappe huizen, uitvallers op school. [Lees verder op de site van het Parool.](#)

Gerelateerde informatie

Dienst Onderzoek en Statistiek. Publicaties: [Kijken achter de Cijfers](#).
Vrije Universiteit Amsterdam. Onderzoekspublicaties: [Kijken achter de cijfers en criminogeniteitsbeeld](#).

Categories: [Data en visualisatie](#), [Informatie en visualisatie](#)

Een experiment in 'organische'* software visualisatie

June 21, 2008 12:19 PM | Permalink | Comments (0) | TrackBacks (0)

De visualisatie, Code_swarm genoemd, toont een overzicht van de bijdragen van verschillende personen aan de ontwikkeling van software door de tijd heen. Een bijdrage is wanneer een persoon een verandering aanbrengt in de code of in een document en deze toevoegt aan de centrale bewaarplaats. Zowel ontwikkelaars als documenten worden als bewegende objecten weergegeven. Zodra een ontwikkelaar een document bewerkt ticht het op en verschuift naar de ontwikkelaar. De document zijn kleur gecodeerd, afhankelijk van of het code is of gewoon tekst. Wanneer gedurende een langere periode documenten niet wijzigen of ontwikkelaars geen bijdrage leveren vervagen ze naar de achtergrond. Een histogram onderaan laat zien wat er eerder is

Search

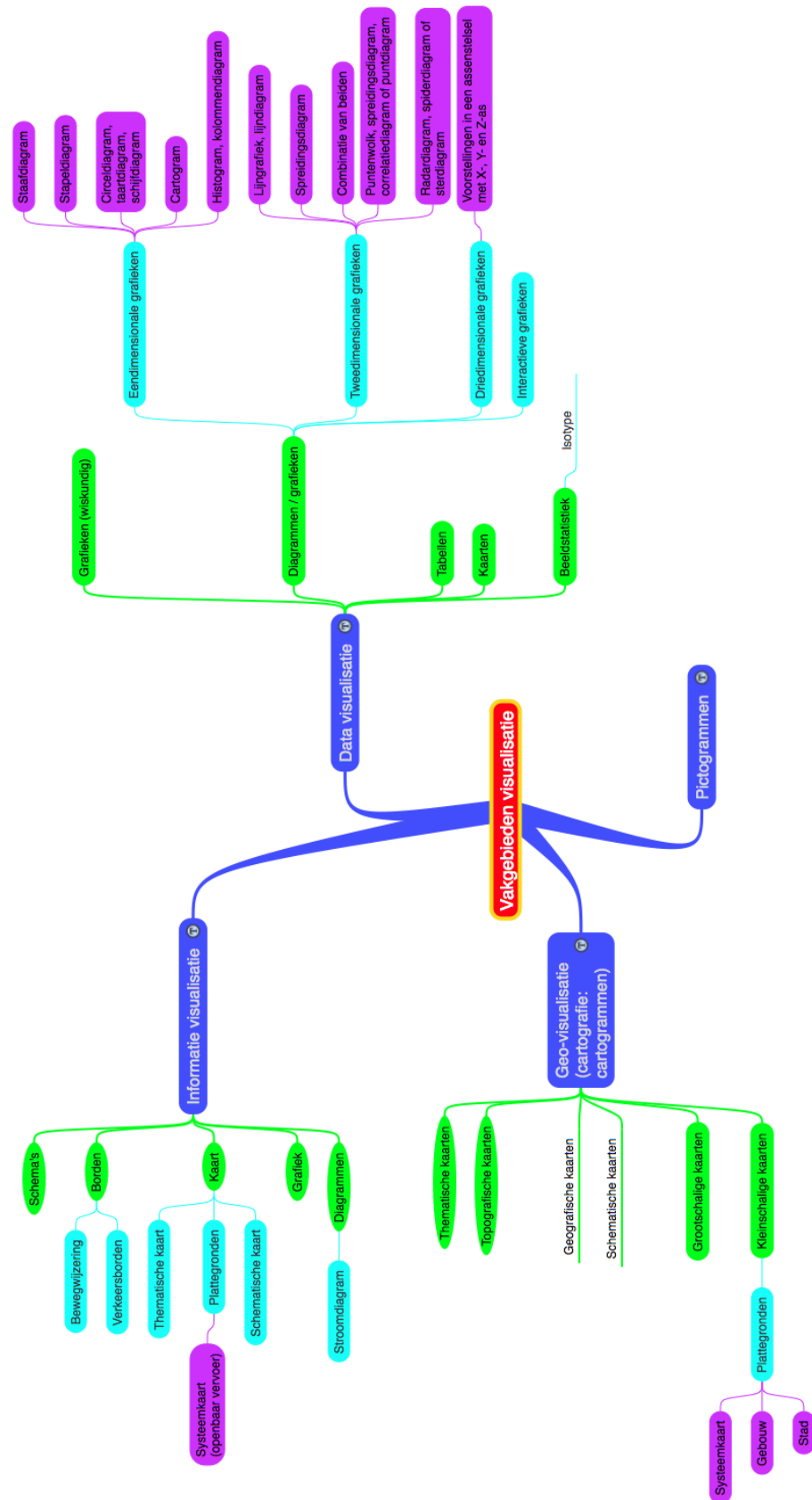
Pages

[Over Nique](#)

Tags

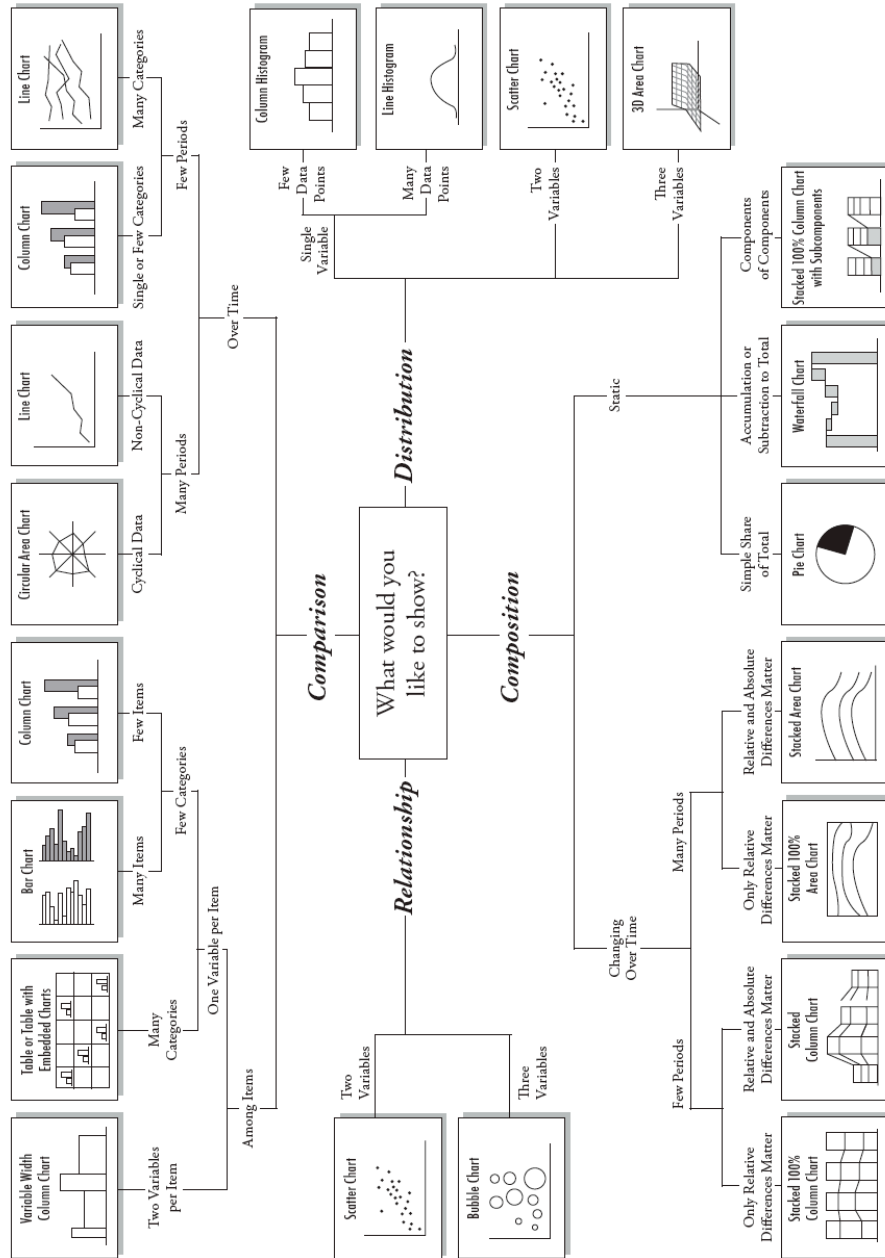
- Alma (1)
- Arntz (1)
- Ben Fry (1)
- Data360 (1)
- Dupin (1)
- Fourcroy (1)
- Friendly (4)
- Funkhouser (1)
- Gapminder (2)
- Grimwade (1)
- Guerry (1)
- Holmes (1)
- ISOTYPE (2)
- Kraak (1)
- LATCH (1)
- ManyEyes (2)
- Minard (3)
- Neurath (1)
- Nightingale (1)
- Playfair (1)
- Priestley (1)
- Processing (1)
- Quételet (1)
- Reidemeister (1)
- Rosling (2)
- Sensemaking (1)

1 Bijlage I



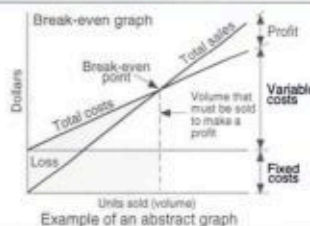
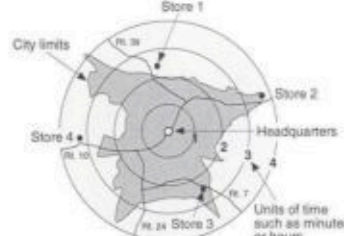
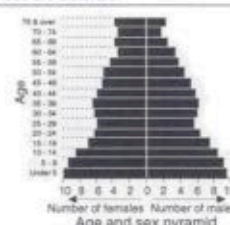
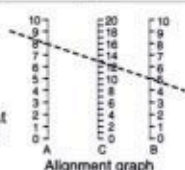
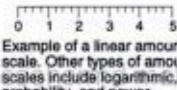
2 Bijlage II

Chart Suggestions—A Thought-Starter



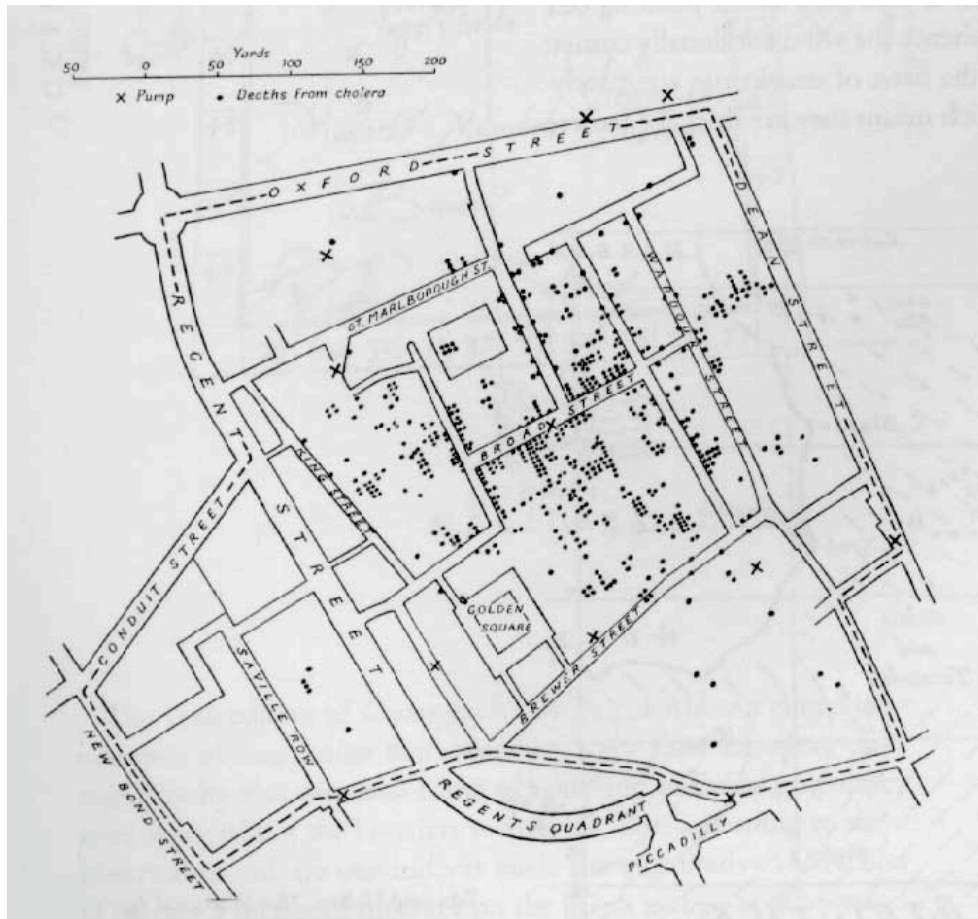
3 Bijlage III

Een voorbeeldpagina:

Abscissa	The distance a data point is located from zero along the horizontal axis of a graph.								
Abscissa Axis	A technical name for the horizontal or X-axis on a two-dimensional graph.								
Abstract Graph	Sometimes referred to as a diagram. An abstract graph typically has no scales, tick marks, or grid lines and is used primarily to investigate, illustrate, and convey ideas and concepts rather than specific values. A break-even graph, shown at right, is one of the better-known examples of an abstract graph. This graph illustrates some of the basic concepts associated with sales, costs, and profits without using numeric values.  Example of an abstract graph								
Abstract Map	A map that conveys a focused but limited amount of specific information such as location, direction, relationships, time, or information by area with little concern for the presence or accuracy of information outside its focus. For instance, in the example at the right, the distances on the map show equal driving times between locations. Most of the actual ground distances are distorted. Abstract maps generally convey one or at most a few sets of facts, ideas, concepts, etc. Many distorted maps, strip maps, cartograms, and diagrammatic maps qualify as abstract maps.  An abstract map indicating points of equal driving time from a central location, taking into consideration type of roads, terrain, congestion, etc. Most of the actual ground distances are distorted.								
Age and Sex Pyramid	Also referred to as a population pyramid. A specific application of a pyramid graph or two-way histogram for studying a population by age and sex. Age intervals are plotted on the vertical axis and the number of males and females in each age interval on the horizontal axis. See Population Pyramid.  Age and sex pyramid								
Alignment Graph	Sometimes referred to as a nomograph, nomogram, or calculation graph. An alignment graph is designed to solve an equation involving three or more variables. Such graphs consist of three or more scales arranged so that a straight line crossing all the scales intersects the scales at values that satisfy the equation. The graph at right satisfies the equation $A + B = C$ (e.g., 8 on scale A plus 5 on scale B equals 13 on scale C). See Nomograph.  Alignment graph								
Alignment of Text	A term used to describe how text is formatted. There are five major types of alignment: the four shown below plus alignment on the decimal point. See Text. <table><tr><th>Aligned flush left</th><th>Aligned flush right</th><th>Centered</th><th>Justified</th></tr><tr><td>Sometimes referred to as justified left and/or ragged right. When text is aligned left, the left edges of all lines are even. The right edge of each line is determined by the full words, partial words, and spacings in the line.</td><td>Sometimes referred to as justified right and/or ragged left. When text is aligned right, the right edges of all lines are even. The left edge of each line is determined by the full words, partial words, and spacings in the line.</td><td>Sometimes referred to as ragged right and left. Uniform spacings are used between words and letters. Excess space is distributed equally at the two ends of each line. Generally used only with a few lines of text.</td><td>Sometimes referred to as flush right and left. Text that is justified normally has the right and left edges aligned. This is accomplished by varying the space sizes between words and letters and the frequent use of hyphens.</td></tr></table>	Aligned flush left	Aligned flush right	Centered	Justified	Sometimes referred to as justified left and/or ragged right. When text is aligned left, the left edges of all lines are even. The right edge of each line is determined by the full words, partial words, and spacings in the line.	Sometimes referred to as justified right and/or ragged left. When text is aligned right, the right edges of all lines are even. The left edge of each line is determined by the full words, partial words, and spacings in the line.	Sometimes referred to as ragged right and left. Uniform spacings are used between words and letters. Excess space is distributed equally at the two ends of each line. Generally used only with a few lines of text.	Sometimes referred to as flush right and left. Text that is justified normally has the right and left edges aligned. This is accomplished by varying the space sizes between words and letters and the frequent use of hyphens.
Aligned flush left	Aligned flush right	Centered	Justified						
Sometimes referred to as justified left and/or ragged right. When text is aligned left, the left edges of all lines are even. The right edge of each line is determined by the full words, partial words, and spacings in the line.	Sometimes referred to as justified right and/or ragged left. When text is aligned right, the right edges of all lines are even. The left edge of each line is determined by the full words, partial words, and spacings in the line.	Sometimes referred to as ragged right and left. Uniform spacings are used between words and letters. Excess space is distributed equally at the two ends of each line. Generally used only with a few lines of text.	Sometimes referred to as flush right and left. Text that is justified normally has the right and left edges aligned. This is accomplished by varying the space sizes between words and letters and the frequent use of hyphens.						
Amount Scale	Sometimes referred to as a quantitative, numeric, value, or interval scale. An amount scale consists of numbers organized in an ordered sequence with meaningful and uniform spacing between them. Quantitative variables are typically measured along amount scales. See Scale.  Example of a linear amount scale. Other types of amount scales include logarithmic, probability, and power.								

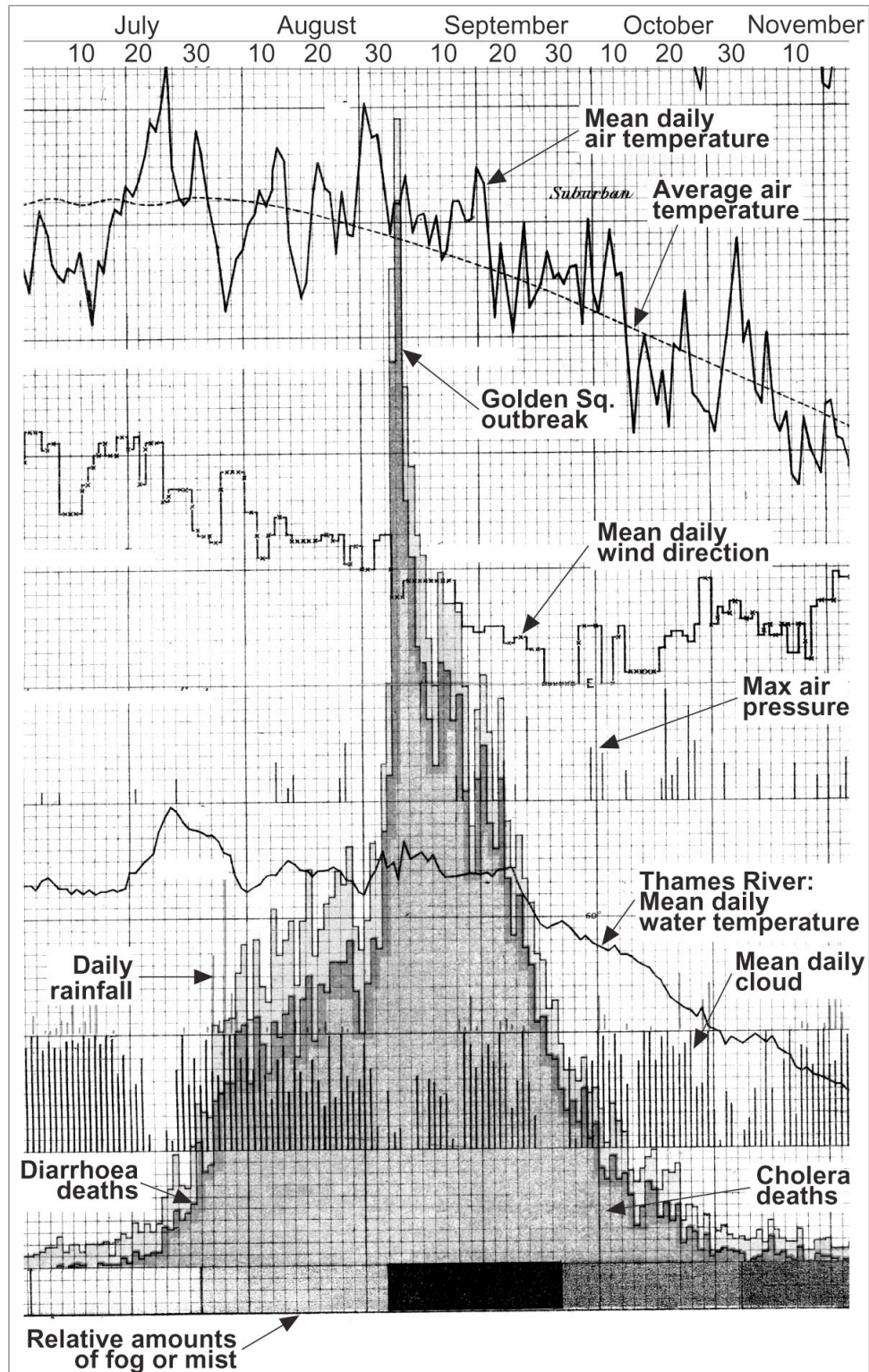
Het boek gaat in op alle gegevensdragers behorende bij tabellen, grafieken etcetera

4 Bijlage IV



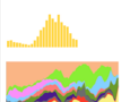
Bron: Geography Department, Rutgers University. <http://mapmaker.rutgers.edu/>

5 Bijlage V



Cholera deaths per day in London, August-September 1854. The peak reflects a doubling of deaths during the ten-day outbreak in Golden Square. Bron: The John Snow archive and research company.

6 Bijlage VI

		++ = zeer goed + = goed o = redelijk - = slecht -- = zeer slecht te bepalen door gebruiker = kan variëren van -- tot ++			
	Orderingsonderdelen	Many Eyes	Swing [0]	Veiligheidsin dex (O+S)	StadStat (O+S)
	Worden oorzaak en gevolg helder gecommuniceerd?	--	--	n.v.t. [1]	n.v.t. [1]
	Wordt het oog uitgenodigd om de verschillende brokken data met elkaar te vergelijken?	+	o	-- [2]	-- [2]
	Is de weergave van nummers, zoals daadwerkelijk gemeten binnen een grafiek, gelijk aan de gevisualiseerde numerieke hoeveelheden?	++	++	n.v.t. [3]	n.v.t.
	Heeft de data een goede ordening?	te bepalen door gebruiker	te bepalen door gebruiker	+	+
	Wordt de data op verschillende niveaus onthult, van een breed overzicht tot in het kleinste detail?	o	o	o [4]	o [4]
	Wordt de data vervormd weergegeven of kan deze verkeerd geïnterpreteerd worden?	--	--	++	++
	Vormen de omvangrijke data sets een coherent geheel?	te bepalen door gebruiker	te bepalen door gebruiker	-- [5]	-- [5]
	Wordt er overtollige data getoond die buiten de context valt?	te bepalen door gebruiker	--	++	++
	Is er een evenwichtige en relevante schaalverdeling?	-	--	--	--
	Tekst				
	Hebben de visualisaties een duidelijke titel?	te bepalen door gebruiker	++	++	++
	Is er een duidelijke bronverwijzing?	te bepalen door gebruiker	++	o	o
	Is de legenda te allen tijde zichtbaar en helder?	++	++	++	++
	Gegevensdragers				
	Laat de vormgeving toe dat de data tot op verschillende niveaus onthult wordt, van een breed overzicht tot in het kleinste detail?	-	--	o	o

		++ = zeer goed + = goed o = redelijk - = slecht -- = zeer slecht te bepalen door gebruiker = kan variëren van -- tot ++			
	Is de etikettering helder, grondig en gedetailleerd, zodat grafische vorming en dubbelzinnigheid voorkomen worden?	++	--	-- [6]	-
	Trekt de data de aandacht en niet de vormgeving?	++	+	--	--
	Esthetiek en techniek				
	Zijn woorden, nummers en lijnen als één geheel gebruikt?	++	+	o	o
	Is de data dichtheid optimaal?	-	--	+	+
	Is er gekozen voor een geschikt formaat en ontwerp?	te bepalen door gebruiker	te bepalen door gebruiker	o [7]	o
	Is de vormgeving nauw verweven met de statistische en verbale beschrijvingen van de data set?	te bepalen door gebruiker	o	o [8]	o [8]
	Dient de vormgeving een duidelijke doel: beschrijving, exploratie, tabuleren of decoratie?	te bepalen door gebruiker	te bepalen door gebruiker	++ [9]	++ [9]
	Worden er verschillende lijndiktes als een aantrekkelijke en compacte manier om data te tonen gebruikt?	o	o	o	o
	Een liggende grafiek verdient de voorkeur boven een staande grafiek (mits de content dit toelaat)! Is dit principe toegepast?	++	++	n.v.t.	n.v.t.
	Is de data-inkt en grafische styling in verhouding?	o	--	+	+
	Ontbreekt Chart-junk (vibraties, grids en eenden) en informateloze decoratie?	++	+	++	++
	Is de data-inkt maximalisatie en grafische vormgeving optimaal?	o	--	o [10]	o [10]
	Interactiviteit				
	Is er sprake van interactiviteit?	++	++	o [11]	o [11]
	Kan gekozen worden voor verschillende vormen?	++	++	--	--
	Kan men verschillende data sets met elkaar combineren?	--	++	--	--
	Is er een groot gebruikersgemak?	+	+	-	-

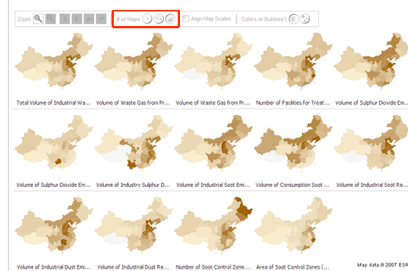
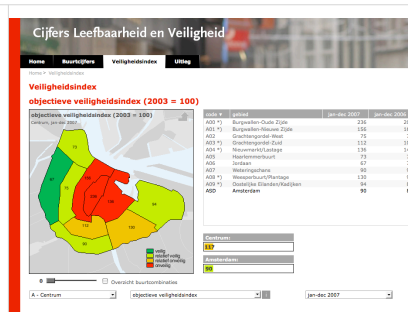
[0] Bronnen Swing http://www.buurtmonitor.nl/ http://www.gemeenteincijfers.nl/																																																																	
[1] Aangezien bij zowel de Amsterdamse Veiligheidsindex als StadStat geen sprake is van een weergave van oorzaak en gevolg, van statistische bewijsvoering. Zie artikel 'Conclusie' op Latebytes. http://www.latebytes.nl/archives/2008/05/conclusie.html.																																																																	
[2] combineren van verschillende gegevens niet mogelijk (jaartallen vergelijken, combinaties maken met andere data-sets zodat oorzaak en gevolg zichtbaar kan worden). (Zie deel II hfdst. 4)																																																																	
[3] Doordat weergave altijd geografisch is (zie deel II hfdst. 2) is de zinvolle informatie die verkregen kan worden beperkt tot één thema waarbinnen oorzaak en gevolg niet gecommuni-ceerd worden.																																																																	
[4] Doordat er een nogal beperkte data set wordt getoond kan er geen echte data rijkheid ontstaan. Hiervoor moet data multivariate gepresenteerd worden. Mits de doelstelling van O+S niet is het presenteren van de kale cijfers is dit een gemiste kans. Is de doelstelling juist wel het presenteren van puur de cijfers dan is de interactieve vormgeving grote over-kill. Juist dan verwacht je als gebruiker wel multivariate data. Zie deel II hfdst. 6																																																																	
[5] De doelstelling van StadStat is in kaart brengen van on- twikkelingen door de tijd heen.* Deze vorm van visualisatie is dan uitermate ongeschikt. Beter zijn tijdslijnen, lijn- grafieken of -diagrammen of stapeldiagrammen. Zie deel II hfdst. 2 • "Via Stadstat kan elke Amsterdammer eenvoudig zien hoe de stad en de eigen buurt zich ontwikkelen: een belangrijke voorwaarde voor maatschappelijk debat en effectief beleid."	< Een goed voorbeeld is de stelling 'In de Vondelbuurt is het aantal fysiotherapeuten per bewoner zes keer hoger dan gemid-deld'. Er is geen link gelegd tussen bijvoor-beeld het percentage ouderen, of de reik-wijde van een praktijk. Een praktijk kan net zo goed zijn praktijk op de grens van de Vondelbuurt hebben liggen, zijn werkge-bied vooral in een naaast gelegen buurt hebben. Maar wordt toch nog tot de Von-delbuurt gerekend.																																																																
[6] Doordat de gegevens versnipperd over de pagina (Bovenin de titel, onderaan kaart de opties, in kaart de leg-enda) worden weergegeven is een gebruiker snel het over-zicht kwijt.																																																																	
[7] Er wordt alleen gebruikt gemaakt van een geografische visualisatie, rechts van de kaart staat een tabel waarin twee jaar met elkaar vergeleken kan worden. Deze worden in een tabel gepresenteerd. Beter zou bijvoorbeeld een staafdia-gram zijn waarin een ontwikkeling over meerdere jaren ge-toond kan worden.	<table><tr><th>code ▼</th><th>gebied</th><th>jan-feb 2008</th><th>jan-feb 2007</th></tr><tr><td>A *)</td><td>Centrum</td><td>315</td><td></td></tr><tr><td>C</td><td>Westerpark</td><td>95</td><td>84</td></tr><tr><td>D</td><td>Oud-West</td><td>68</td><td>72</td></tr><tr><td>G</td><td>Zeeburg</td><td>97</td><td>71</td></tr><tr><td>H</td><td>Bos en Lommer</td><td>99</td><td>67</td></tr><tr><td>J</td><td>De Baarsjes</td><td>95</td><td>79</td></tr><tr><td>N</td><td>Amsterdam-Noord</td><td>183</td><td>181</td></tr><tr><td>P</td><td>Geuzenveld-Slotermeer</td><td>105</td><td>113</td></tr><tr><td>Q</td><td>Osdorp</td><td>157</td><td>144</td></tr><tr><td>R</td><td>Slotervaart</td><td>86</td><td>172</td></tr><tr><td>T</td><td>Zuidoost</td><td>233</td><td>262</td></tr><tr><td>U</td><td>Oost-Watergraafsmeer</td><td>185</td><td>251</td></tr><tr><td>V</td><td>Oud-Zuid</td><td>169</td><td>172</td></tr><tr><td>W</td><td>Zuideramstel</td><td>81</td><td>78</td></tr><tr><td>ASD</td><td>Amsterdam</td><td>1975</td><td>2064</td></tr></table>	code ▼	gebied	jan-feb 2008	jan-feb 2007	A *)	Centrum	315		C	Westerpark	95	84	D	Oud-West	68	72	G	Zeeburg	97	71	H	Bos en Lommer	99	67	J	De Baarsjes	95	79	N	Amsterdam-Noord	183	181	P	Geuzenveld-Slotermeer	105	113	Q	Osdorp	157	144	R	Slotervaart	86	172	T	Zuidoost	233	262	U	Oost-Watergraafsmeer	185	251	V	Oud-Zuid	169	172	W	Zuideramstel	81	78	ASD	Amsterdam	1975	2064
code ▼	gebied	jan-feb 2008	jan-feb 2007																																																														
A *)	Centrum	315																																																															
C	Westerpark	95	84																																																														
D	Oud-West	68	72																																																														
G	Zeeburg	97	71																																																														
H	Bos en Lommer	99	67																																																														
J	De Baarsjes	95	79																																																														
N	Amsterdam-Noord	183	181																																																														
P	Geuzenveld-Slotermeer	105	113																																																														
Q	Osdorp	157	144																																																														
R	Slotervaart	86	172																																																														
T	Zuidoost	233	262																																																														
U	Oost-Watergraafsmeer	185	251																																																														
V	Oud-Zuid	169	172																																																														
W	Zuideramstel	81	78																																																														
ASD	Amsterdam	1975	2064																																																														

[8] De keuze voor een data kaart is zeer geschikt omdat deze afbeelding zeer veel data op een klein oppervlak overzichtelijk kan presenteren. Het is ook een zeer herkenbare vorm voor het menselijke oog. De aandacht wordt direct op de inhoud gericht. En daarnaast handelt de data over Amsterdam. Nadeel van deze methode is dat niet het aantal mensen maar een geografische grens de inhoud bepaald. En door alleen te kiezen voor een data kaart en andere visualisatie opties niet te benutten blijft het geheel erg eenzijdig. *Zie deel II hfst. 8 'De vier basis vormen van een data visualisatie'.*

[9] De data kaart is duidelijk beschrijvend bedoeld.

[10] De tekst (pulldown menu, titel, cijfers) zou veel meer als een geheel gepresenteerd kunnen worden. De ordeningsonderdelen (cijfers in de kaart) zijn ondergeschikt aan de vorm. Dit zou andersom beter zijn.

[11] Eigenlijk is er betrekkelijk weinig interactiviteit. Een gebruiker kan geen data combineren, kan geen filtering aanbrengen. Omdat het natuurlijk bij zowel StadStat als de Veiligheidsindex om beperkte onderwerpen gaat is dit natuurlijk minder relevant. Vraag die gesteld mag worden is of de doelstelling een dergelijke 'interactieve' applicatie wel rechtvaardigt. Een overzicht van kwalitatief beter vormgeven data kaarten zou in beide gevallen waarschijnlijk veel meer gerechtvaardigd zijn. Daarnaast zou het principe van 'small multiples' een enorme bijdrage kunnen zijn >



7 Bijlage VII

See the world

-  World map
-  Country maps

Track rises and falls over time

-  Line graph
-  Stack graph
-  Stack Graph for Categories

Compare a set of values

-  Bar chart
-  Block histogram
-  Bubble chart
-  Matrix chart

See relationships among data points

-  Scatterplot
-  Network Diagram

See the parts of a whole

-  Pie chart
-  Treemap
-  Change Treemap

Look for common words in a text

-  Tag Cloud
-  Word Tree

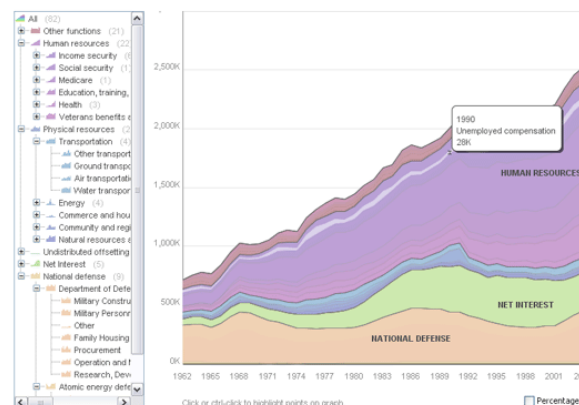
Stack Graph for Categories

When to use a stack graph for categories

A stacked graph is meant for visualizing the total change over time of a group of quantities. Because items are "stacked," this type of graph is especially useful when it makes sense to add up the underlying data points. For example, a stacked graph is good for sales data, since you may be interested not only in sales of individual items but also want to know how total sales varies over time. For simple data sets you may be able to use a simple [stack graph](#). But if your data items are arranged into categories and subcategories, it makes sense to use a stack graph for categories.

How a stacked graph for categories works

The Many Eyes stacked graph for categories is shown below.



Each ribbon of color in the graph represents a data item changing over time. In this graph the ribbons represent United States government spending on different budget items. The height, or thickness, of each ribbon represents the dollar figure for spending. The overall height of the graph shows the total spending.

In many data sets, such as this one, items are arranged into categories and subcategories. The tree outline at left provides a color key to these categories. It allows the user to drill down into the data by expanding and collapsing