

# DE ENE VISUALISATIE IS DE ANDERE NIET...

*Ooit al eens blindgestaard op een rapport zonder echt te begrijpen wat het echt wil overbrengen aan de lezer? Als de auteur gewild had dat je er iets van had gesnapt, had hij of zij er zeker veel meer moeite voor gedaan om het begrijpelijk op te bouwen. Lekker niet dus...*

✍ DRIES VAN NIEUWENHUYSE

We kunnen ruwweg vier fundamentele soorten actie hebben waarbij we info willen visualiseren: we willen zien hoe veranderlijk één maatstaf is, we willen de samenstellende delen van een totaal in kaart brengen, we willen de relatie tussen twee veranderlijken illustreren of we willen evoluties zien in de cijfers.

## DISTRIBUTIES

De verdeling van één variabele kan worden aangetoond door een histogram voor zoverre niet te veel verschillende waarden voorkomen. Een histogram kan best beschreven worden als de schoenen van een willekeurige groep mensen die per schoenmaat verzameld en vervolgens per maat gesorteerd worden. Per maat krijg je een staaf en de gesorteerde staven geven een helder zicht van de variabiliteit. De meest voorkomende schoenmaat of de modale klasse is de grootste hoop en ligt nogal eens in het midden. Bij natuurlijke processen, zoals schoenmaat per geslacht, bekomen we op die manier een mooie, symmetrisch glooiende heuvel die ook een gauss-curve wordt genoemd. Detectie van foute data en uitbijters is zeer gemakkelijk op deze manier. Ze springen letterlijk uit de grafiek.

## ONTRAFELN VAN EEN TOTAAL

Als we de samenstelling van een geheel willen ontrafelen, dan kunnen we best een watervaldiagram gebruiken eerder dan een taartdiagram. Als de taartstukken allemaal bijna even groot zijn,

is het relatieve verschil quasi niet op te merken. Staven van licht verschillende grootte daarentegen zijn door het oog veel beter te onderscheiden. Vandaar dat we voor het tonen van verandering doorheen de tijd best gestapelde blokken afbeelden. Per deel van het geheel kunnen we een andere kleur gebruiken. De hoogte van de stapel gekleurde blokjes toont de totale hoeveelheid. Per tijdsperiode wordt een stapel toegevoegd. In absolute waarden zien we de evolutie van de totalen, een procentuele breakdown levert per periode even hoge totalen op (100 procent) waardoor de evolutie van de relatieve samenstelling duidelijk wordt. De evolutie doorheen de tijd van de bijdrage in de omzet per productklasse kan duidelijk seizoensgebonden schommelingen blootleggen. De respectievelijke kleuren kunnen verbonden worden met lijntjes zodat de trend nog meer wordt geaccentueerd. Als we maanden vergelijken, zetten we staven naast elkaar, als we dagen willen vergelijken tekenen we oppervlaktes die gelaagd op elkaar liggen, omdat de staafjes te klein zouden zijn en het zicht vertroebelen.

## RELATIES

Als we zeer veel maatstaven hebben, passen we best een of andere datacompressie-methode toe. Factoranalyse bijvoorbeeld identificeert correlaties tussen verschillende variabelen en combineert correlerende maatstaven tot nieuwe,

verwijdert synoniemen en identificeert waar dubbele data voorkomt. Die inzichten laten toe bewust metrieken al dan niet mee te nemen in de visualisaties of analyses. Redundante info kan verwijderd worden. We kunnen ons focussen op wat er echt toe doet...

De verdeling van observaties met twee veranderlijken kan best via een x-y-diagram worden getoond. Enig lijnvormig verband tussen beide variabelen is gemakkelijk te zien door een sigaarvormige puntenwolk die van links naar rechts naar boven wijst bij een positief verband en naar beneden bij een negatieve relatie.

Scatterdiagrammen kunnen karakteristieken van observaties in een vlak voorstellen, waarbij elke observatie getoond wordt op de x- en de y-coördinaat. Een derde dimensie kan worden toegevoegd door elk punt als een bol voor te stellen. Dergelijke bubble chart illustreert zodoende drie dimensies op een vlak. Een alternatief zou een 3D-representatie zijn die door middel van een netwerkstructuur een oppervlakte weergeeft. Dit kan met geografisch data zinvol zijn als terreinmodel doch zelden voegt dit waarde toe aan de interpretatie. Complexere relaties met meer dan drie variabelen zijn zelden nog visueel eenvoudig te illustreren.

## EVOLUTIES

Willen we info vergelijken dan zien we twee opties, ofwel vergelijken we

doorheen de tijd ofwel vergelijken we alternatieven. Heel wat visualisaties zijn relevant voor items die we ten opzichte van elkaar bestuderen. Gecombineerde histogrammen voor mannen en vrouwen bijvoorbeeld tonen zowel de heterogeniteit van beide klassen als de relatieve positie ervan. Evoluties doorheen de tijd nemen een of meerdere maatstaven en illustreren vooral de trends. Dit kan door kolomdiagrammen als er een beperkt aantal periodes vergeleken wordt zoals maanden. Bij dagelijkse metingen worden typisch lijnen gebruikt in plaats van blokken. Sparklines zijn hiervan excellente voorbeelden, ze zijn bijzonder compact, een fijn lijntje met aanduiding van de uitbijters als knipperlichtjes via een rood bolletje. Een grijze achtergrond toont de 95% confidentie-interval

“Een goede tip blijft alvast om per rapport in je eigen woorden uit te leggen welke beslissing je wil onderbouwen en hoe je daar naar kijkt. Uitstekende manier om empathie met de lezer te bekomen maar minder makkelijk dan het lijkt...”

waarvan we weten dat de kans dat een observatie erbuiten ligt, slechts 5% is, statistisch eerder uitzonderlijk dus.

#### LEZER CENTRAAL

Dank zij de mobiele technologie komen nog steeds nieuwe visualisaties ter beschikking. Ofwel leidt dit tot nog meer keuzestress voor de rapportontwikkelaar, ofwel scheidt dit mogelijkheden om

nog spraakmakender illustraties te ontwerpen. Dit dwingt ons nog meer tot nadenken over welk signaal de visualisaties nu echt willen geven aan de lezer. Een goede tip blijft alvast om per rapport in je eigen woorden uit te leggen welke beslissing je wil onderbouwen en hoe je daar naar kijkt. Uitstekende manier om empathie met de lezer te bekomen maar minder makkelijk dan het lijkt...

*Hans Rosling was een Zweedse arts en professor die vooral bekend stond vanwege de manier waarop hij alledaagse statistische informatie wist te visualiseren en concretiseren. Hij overleed op 7 februari 2017.*

