

DATA ANALYTICS: EEN PLAATJE ZEGT MEER...

Kunnen we misschien nog even terug naar de selectie van de data? Ze vroeg het bijna verontschuldigend. Mijn Finse collega's en ik hadden de resultaten van ons visualisatie- en discoveryproject aan de klant gepresenteerd. Mooi werk, mooie bevindingen. De klant was verrast door de resultaten. Toch kwam het minder sterk over dan we hadden gehoopt. Pas later realiseerde ik me wat er aan mankeerde: onze presentatie droeg nog teveel sporen van de ontdekkingsreis die we hadden afgelegd.

Een beeld zegt meer dan duizend woorden. Door grootheden zorgvuldig te combineren en in een figuur te presenteren, maak je iets in één oogopslag duidelijk. Moeten lezers bij een verhaal in woorden of cijfers zelf dingen onthouden en in hun hoofd combineren, bij een grafische weergave zien ze meteen het verschil tussen groepen, een ontwikkeling in de tijd of een afwijking. Ze hoeven geen moeite te doen, ze zien het gewoon. Daarom kun je met datavisualisatie een heel krachtig verhaal vertellen.

Bij big data analytics ben je vaak juist op zoek naar dat verhaal. Je hebt meestal geen hypothese maar gaat op ontdekkingsreis door de data. Zoals ontdekkingsreizigers uitgerust zijn om de meest uiteenlopende omstandigheden te doorstaan, zo hebben data scientists een grote verzameling tools om de meest uiteenlopende soorten data te analyseren en te visualiseren. Maar ondanks hun goede uitrusting moeten ze onderweg nog veel improviseren.

De omstandigheden zijn natuurlijk altijd anders dan vooraf gedacht. Een weergave die vooraf zinnig leek, moet je eerst ontdoen van die nare verstorende factor. Of de ene weergave nodigt je uit om net nog een zijstapje te verkennen in een andere weergave. Voor je er erg in hebt, heb je een stapel grafieken opgetuigd waarin allerlei factoren zijn gecodeerd: assen, grootte, kleur, tint, symbolen en panelen; allemaal vertellen ze een deel van het verhaal dat zich in jouw hoofd ontspint. De vergezichten ontvouwen zich en opeens kan er dan iets in het oog springen, waarvan je voelt: dit is de moeite waard.

Omdat je zelf de reis hebt ondernomen, vergeet je



JAN-WILLEM LANKHAAR
Data scientist bij CGI

makkelijk dat anderen niet weten welke data je hebt geselecteerd, hoe je die hebt bewerkt en weergegeven. Wanneer je ze deelgenoot maakt van jouw ontdekking, ben je geneigd dezelfde route te nemen. Maar zijn al die gecodeerde factoren nog nodig om het verhaal te vertellen? Is er niet een snellere route naar hetzelfde vergezicht?

Hierbij helpen tools meestal niet, omdat tools (en opdrachtgevers soms ook) gericht zijn op meer toeters en bellen. Hier is minder het devies. Hier komt het aan op vakmanschap. Hoe vat je de essentie? Wat is ballast en wat moet er echt in? Vaak kan er veel meer weg dan je in eerste instantie denkt, maar dat is moeilijk: less is more... work! Als een opdrachtgever aan het eind van het verhaal een beetje vertwijfeld vraagt naar de selectie van de data, is dat kennelijk nog niet helemaal gelukt.

“ DATA SCIENTISTS HEBBEN
EEN GROTE VERZAMELING
TOOLS OM DATA TE ANALYSEREN
EN TE VISUALISEREN ”