

Stijlgids Datavisualisatie

Effectief visualiseren van gegevens binnen ProRail

Van	Michel Dekker
Eigenaar	[Eigenaar]
Kenmerk	VT20160142-1177460457-391
Versie	2.1
Datum	24 december 2021
Onderwerp	stijlgids prorail dashboard
Status	Definitief

Inhoudsopgave

1	Aanpak	10
2	Ontwerp standaarden	11
2.1	Layout	11
2.2	Kleur	13
2.3	Lettertype	18
2.4	Opmaak van nummers en datums	18
2.5	Visualisatie principes	20
2.5.1	Gestalt	20
2.5.2	Pre-attentieve waarneming	22
3	Visualiseren van data	24
3.1	De tabel	24
3.1.1	Onderscheiden van rijen en kolommen	24
3.1.2	Accentueren	26
3.1.3	Uitlijnen	27
3.1.4	(Verschillende) eenheden binnen een tabel	27
3.1.5	Grafische ondersteuning in de tabel	28
3.2	Het diagram	28
3.2.1	Keuzehulp	29
3.2.2	Basisvormen	31
3.2.2.1	Lijndiagram	31
3.2.2.2	Staafdiagram	33
3.2.2.3	Puntenwolk	38
3.2.2.4	Taartdiagram	38
3.2.2.5	Geografische kaarten	40
3.2.2.6	Boxplot	43
3.2.2.7	KPI Meter	44
3.2.2.8	Sparkline	46
3.2.2.9	Trellis chart (ook small multiples genoemd)	46
3.2.3	Ondersteunende elementen	48
4	Visualisatie voorbeelden	49
4.1	Reizigerspunctualiteit	49
4.2	Netto Apparaatskosten	52
4.3	Exploitatierekening	54
5	Bijlage 1: Design Reference Card	56
6	Bijlage 2: Diagram Selectie Matrix	57
7	Bijlage 3: Stijlgids samenvatting	58

Figurenoverzicht

Figuur 1 Definitie datavisualisatie	7
Figuur 2 Stadia van inzicht	7
Figuur 3 Toepassen van datavisualisatie	8
Figuur 4 Aandachtzones van een datavisualisatie	11
Figuur 5 ProRail dashboard kop	11
Figuur 6 1300px grid systeem	12
Figuur 7 layout met 2 componenten naast elkaar (diagram en tekst naast elkaar)	13
Figuur 8 layout met 3 componenten naast elkaar	13
Figuur 9 Standaard ProRail kleuren donkere variant	14
Figuur 10 Standaard ProRail kleuren lichte variant	14
Figuur 11 Voorbeeld met alle 12 kleuren	16
Figuur 12 Kleuren bij kleurenblindheid	16
Figuur 12 ProRail grijstinten	16
Figuur 13 ProRail rode kleuren	17
Figuur 14 Kwantitatief RoodGrijs pallet	17
Figuur 15 Kwantitatief Rood pallet	17
Figuur 16 Schreef en schreefloos lettertype	18
Figuur 17 Lettertype Arial	18
Figuur 18 Lettertype omvang	18
Figuur 19 Gestalt - Gelijkheid	20
Figuur 20 Gestalt - Nabijheid	21
Figuur 21 Gestalt - Continuïteit	21
Figuur 22 Gestalt - Eenvoud	21
Figuur 23 2D positie	23
Figuur 24 Lengte	23
Figuur 25 Dikte	23
Figuur 26 Oppervlakte	23
Figuur 27 Kleurintensiteit	23
Figuur 28 tabel met een lage data-inkt-ratio	25
Figuur 29 tabel met een hoge data-inkt-ratio	25
Figuur 30 Achtergrondkleur om de rijen te verduidelijken	26
Figuur 31 Subtiële lijnen om de rijen te verduidelijken.	26
Figuur 32 voorbeelden van accentueren in een tabel	26
Figuur 33 uitlijnen van numerieke data	27
Figuur 34 uitlijnen van tekst	27
Figuur 35 uitlijnen van datums	27
Figuur 36 Eenheden staan in iedere cel	28
Figuur 37 Eenheden staan in de koptekst	28
Figuur 38 Zonder grafische ondersteuning moet ieder getal gelezen worden	28
Figuur 39 Met grafische ondersteuning worden mogelijke patronen direct zichtbaar	28
Figuur 40 Lijndiagram	31
Figuur 41 Lijndiagram niet gebruiken voor categorische data	31
Figuur 42 Het afwijkende patroon van de stippellijn zorgt voor afleiding	32
Figuur 43 Het aangeven van een andere rol (doelstelling) kun je beter subtiel doen	32
Figuur 44 Data series zijn direct gelabeld	32
Figuur 45 Data series zijn indirect gelabeld in een legenda	33
Figuur 46 Inconsistente afstand in tijd	33
Figuur 47 "Smoothed lines"	33
Figuur 48 Staafdiagram	34

Figuur 49	Schaalverdeling begint niet met 0. Hierdoor kloppen de verhoudingen niet meer	34
Figuur 50	Als er niet begonnen kan worden met 0, kies dan een andere vorm	34
Figuur 51	Horizontale staafdiagram	35
Figuur 52	Gestapelde staafdiagram	35
Figuur 53	De witruimte tussen de staven is breder dan de staaf zelf	36
Figuur 54	Dit diagram ondersteund vooral het vergelijken binnen één maand	36
Figuur 55	Hier ligt de nadruk op het vergelijken van de verschillende maanden	36
Figuur 56	Zorg dat alle teksten horizontaal worden toegepast	37
Figuur 57	De positie in het diagram zorgt al voor onderscheid, kleur voegt dus niets toe	37
Figuur 58	Waterval diagram	37
Figuur 59	Voorbeeld van een puntenwolk	38
Figuur 60	Voorbeeld van een ballondigram	38
Figuur 61	Voorbeeld van een taartdiagram	39
Figuur 62	Alternatief voor de taartdiagram	39
Figuur 63	Label direct en beperk het aantal kleuren	40
Figuur 64	Donutdiagram	40
Figuur 65	voorbeeld van een geografische kaart	41
Figuur 66	voorbeeld van geografische kaart met sporen	42
Figuur 67	voorbeeld van geografische kaart met alleen sporen	43
Figuur 68	Onderdelen van een boxplot	44
Figuur 69	Vergelijken van verzamelingen met een boxplot	44
Figuur 70	Voorbeeld van een snelheidsmeter	45
Figuur 71	Bullet graph of KPI meter	45
Figuur 72	Hier wordt een pijl gebruikt voor de trend	46
Figuur 73	Door gebruik te maken van een sparkline geef je veel meer context	46
Figuur 74	Onduidelijkheid door het grote aantal data-series	47
Figuur 75	Trellis chart voorbeeld	47
Figuur 76	Voorbeelden van ondersteunende elementen bij een schermvullend diagram	48
Figuur 77	Reizigerspunctualiteit (huidig)	49
Figuur 78	Reizigerspunctualiteit (alternatief)	50
Figuur 79	Netto Apparaatskosten (huidig)	52
Figuur 80	Netto Apparaatskosten (alternatief)	53
Figuur 81	Exploitatierkening (huidig)	54
Figuur 82	Exploitatierkening (alternatief)	55

Versiebeheer

Versie	Datum	Omschrijving
0.1	20 september 2017	Eerste grove opzet van de stijlgids
0.2	29 september 2017	Verwerken feedback eerste review sessie
0.3	20 oktober 2017	Verwerken feedback tweede review sessie en toevoegen samenvatting
0.4	17 november 2017	Aangepaste inleiding
1.0	8 juni 2018	Aanpassingen kleuren nav feedback
1.1	14 december 2018	Aanpassingen nav feedback van communicatie
2.0	2021	Herzien opzet nav Qliksense & Power BI
2.1	24 december 2021	Aanpassen kleuren obv feedback

Voorwoord

Voor je ligt de stijlids van het ProRail Dashboard met de daaraan gelieerde producten zoals: ProRail maand en kwartaalrapportages, Keek op de Week, ExCo maandrapportages, rapportages voor stakeholders, etc.

Het doel van de stijlids is om richting te geven aan datavisualisaties. Dit is een ander doel dan bijvoorbeeld de corporate huisstijl van communicatie waar het meer om de esthetische kant en de beleving gaat (hierbij is wel zo veel mogelijk bij aan gesloten). Bij datavisualisaties gaat het om het goed waarnemen, begrijpen en interpreteren van data. Deze stijlids geeft richting hoe dat op een eenduidige en uniforme manier gewenst is.

Het uitgangspunt van deze stijlids is om de data zijn verhaal te laten vertellen en de ontwerper van de visualisatie te ondersteunen in het kiezen van een visualisatie-vorm die mensen het beste (eenvoudigst, snelst en nauwkeurigst) kunnen begrijpen.

Belangrijk daarbij is dat kleuren goed te onderscheiden zijn. En signaalkleuren, zoals rood en groen, zoveel mogelijk vermeden worden. Deze kleuren worden alleen ingezet als actie benodigd is. Daarnaast is voor de onderscheidende kleuren een palet aan kleuren gekozen die ook onderscheidend is voor mensen die kleurenblind zijn.

Deze gids is niet afhankelijk van specifieke software, want technologie is bij datavisualisatie meestal niet het probleem. De meeste winst valt te behalen in het ondersteunen van de ontwerper in zijn/haar ontwerpkeuzes. Voor het ProRail Dashboard wordt gebruikt gemaakt van Qlikview. Binnen Qlikview is een template beschikbaar met de visualisaties zoals beschreven in deze stijlids.

Geen datavisualisatiegids zonder voorbeelden. De meeste tips en richtlijnen worden geïllustreerd met een voorbeeld. Om verwarring te voorkomen staat naast ieder voorbeeld een icoon om aan te geven of het een



Goed, of



slecht voorbeeld betreft.

Op verschillende plekken in de stijlids staan tips, aanbevelingen en richtlijnen. De belangrijkste zijn herkenbaar door een groene achtergrond zoals bij deze alinea.

Mocht je vragen of opmerkingen hebben, aarzel niet om contact op te nemen met:

Corporate Control
Marco Uijl

Inleiding

Om deze stijlguides goed te kunnen gebruiken is het op de eerste plaats van belang dat we aangeven wat we met datavisualisatie bedoelen. Vandaar de onderstaande definitie van datavisualisatie.



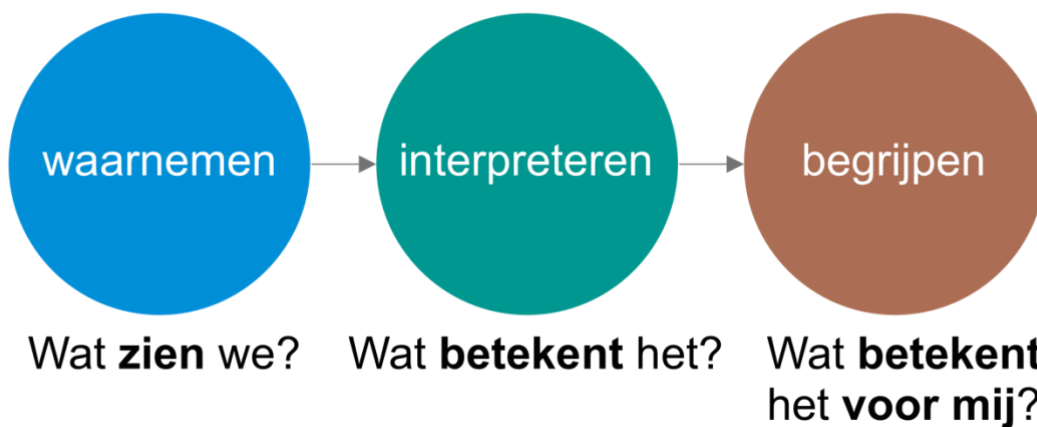
Het **visualiseren** van **gegevens** om **inzicht** te vergroten

Figuur 1 Definitie datavisualisatie

Ons doel van datavisualisatie is dus het vergroten van *inzicht*. Dit is dan ook direct het criterium dat we gebruiken om te bepalen of een datavisualisatie goed of niet goed is. Datavisualisatie is dus meer dan het toevoegen van een mooi plaatje om onze rapporten of dashboards te verfraaien.

Onze belangrijkste grondstof bij datavisualisatie zijn onze *gegevens*: de data. De gegevens spelen een hoofdrol en geen bijrol, dus moeten prominent en duidelijk afgebeeld worden.

Bij het *visualiseren* van de gegevens spelen allerlei aspecten een rol. Dit begint al met de keuze tussen een tabel of diagram. En als we kiezen voor een diagram, welke past dan het beste bij onze gegevens? Ook aspecten als kleur, indeling, lettertypes, etc. zijn hierbij van belang.



Figuur 2 Stadia van inzicht

Om inzicht te vergroten is het goed om eerst te kijken naar hoe wij mensen inzicht verkrijgen middels visuele prikkels: de stadia van inzicht.

De eerste stap is het *waarnemen*, met andere woorden “wat zien we?”. Tijdens deze stap zien we patronen en mogelijke afwijkingen. We krijgen in deze stap het antwoord op vragen als:

- Welke waarden zijn het grootste? Welke waarden zijn het kleinste?
- Hoe groot is een specifieke waarde ten opzichte van het totaal (deel van een geheel)?
- Op welke plaats staat dit resultaat ten opzichte van de andere waarden (rangorde)?
- Tussen welke waarden bestaat er een verband?

Vervolgens komen we in de tweede stap: *interpreteren* van het beeld. We proberen dan betekenis te geven aan wat we zien. Hiervoor stellen we vragen als:

- Is het hebben van een hoge waarde goed? Of is het beter om een lage waarde te hebben?
- Wat betekent een stijging of daling van een waarde over de tijd?
- Is de relatie tussen verschillende waarden betekenisvol en significant?
- Is de af- of toename van een waarde in een specifieke categorie bijzonder/verrassend?

Als laatste stap komen we bij het *begrijpen* van het beeld. Dit is heel persoonlijk, waarbij we trachten te bepalen wat het betekent voor ons. In deze stap proberen we dit te achterhalen met vragen als:

- Waarom is dit relevant?
- In welke behoefte wordt hierbij voorzien?
- Bevestigt dit resultaat mijn bestaande kennis en/of verwachtingen? Of krijg ik hiermee nieuwe inzichten?
- Word ik door dit resultaat emotioneel geraakt? Laat het mij onverschillig?
- Word ik door dit resultaat aangezet tot een actie? (actie=beslissing nemen of gedrag veranderen)
- Is dit resultaat extra informatie die mij wellicht in een later stadium kan helpen?

Als ontwerpers van datavisualisaties hebben we vrij veel invloed op de eerste stap: het waarnemen. Deze invloed komt neer op het volgende principe: laat de gegevens het verhaal vertellen, dus een resultaat dat 2 keer zo groot is wordt ook 2 keer zo groot afgebeeld. Hierbij maken we gebruik van basale kennis over de menselijke (visuele) waarneming (zie hoofdstuk 2.5).

Onze invloed op stap twee (interpreteren) is al veel minder groot, doordat de ontvanger en zijn/haar achtergrond en stemming een grotere invloed krijgen. Hierbij is de rol van de ontwerper meer ondersteunend. Deze ondersteuning bestaat vooral uit een heldere opbouw of structuur. Ook kan de ontwerper in deze fase helpen door te accentueren wat belangrijk is.

Bij de laatste stap (begrijpen) is de invloed van de ontwerper minimaal. Het is een stap waar de ontwerper rekening moet houden met de achtergrond van zijn publiek: wat vindt het publiek "normaal" en wat is het publiek gewend?



Figuur 3 Toepassen van datavisualisatie

Datavisualisaties kunnen op een drietal manieren gebruikt worden om het inzicht van de lezer te vergroten. Op de eerste plaats is dit het *verklaren* van een bepaalde status of

situatie. Voorbeelden hiervan zijn statusrapporten die beschrijven wat er gebeurd is. Het is vaak het antwoord op een vooraf gespecificeerde vraag.

Een datavisualisatie kan ook gebruikt worden om gegevens te *onderzoeken*. De lezer wil dan vaak de gegevens vanuit verschillende gezichtspunten bekijken en vergelijken. Een interactief managementdashboard waarbij de lezer zelf kan filteren en inzoomen is hiervan een bekend voorbeeld.

Als laatste worden datavisualisaties ook gebruikt om lezers te “verleiden” gegevens te bekijken door deze te *exposeren*. Hierbij wordt veel meer een beroep gedaan op de emotie van de lezer door het gebruik van uitnodigende illustraties en opmaak. Voorbeelden hiervan zijn de zogenaamde infographics uit de reclame of op sociale media.

In de dagelijkse praktijk binnen ProRail komen vooral de eerste twee toepassingen veel voor. Meestal hoeven lezers van datavisualisaties niet verleid te worden, want ze hebben de betreffende data hard nodig om hun werk uit te kunnen voeren. Vandaar dat de rest van deze stijlgids zich voornamelijk richt op de eerste twee manieren (verklaren en onderzoeken) door leesbaarheid voorop te zetten.

1 Aanpak

Het ontwerpen van een goede datavisualisatie kan lastig zijn. Naast het verzamelen van uiteenlopende gegevens is het ook een continu afwegen van (soms tegenstrijdige) belangen. De volgende tips bieden wat houvast in het ontwerpen van datavisualisaties¹:

1. Bepaal de centrale boodschap
De datavisualisatie moet een verhaal vertellen. Om dit goed te kunnen doen moet je op de eerste plaats je doel vaststellen: waarom maak je deze datavisualisatie? Ook moet je duidelijk krijgen voor wie je de datavisualisatie maakt, je publiek. Maar ook wat zij er mee gaan doen: willen zij alleen geïnformeerd worden over een status of ook door de data “navigeren”?
2. Verzamel de data
Verzamel de data die je in staat stellen het “waarom” te beantwoorden;
3. Schets de informatie
 - I. Schets met een potlood de onderdelen van je visualisatie op papier: bepaal de beste plaats voor tekst, tabellen, diagrammen, etc.
 - II. Bepaal de beste manier om de data te visualiseren (tabel of diagram, welk type diagram, etc.).
 - III. Ondersteun je publiek door het toevoegen van teksten.
 - IV. Verwijder alle elementen die niet noodzakelijk zijn. Beperk je tot het minimum dat nodig is om de boodschap over te brengen: de visualisatie is pas af als er niets meer af kan.
 - V. Creëer een informatiehiërarchie zodat de belangrijkste elementen de meeste aandacht vragen.
4. Ontwerp
Creëer een visuele hiërarchie door de belangrijkste informatie te accentueren. Gebruik hiervoor bijvoorbeeld afwijkende kleuren en lettertypes. Maak de minst-belangrijke informatie licht, bijna onzichtbaar, voor een zo groot mogelijk contrast met de belangrijke informatie.

Het vervolg van dit document gaat dieper in op de verschillende ontwerpkeuzes. Het zijn daarmee eerder richtlijnen dan wetten. Je mag de richtlijnen dus wel buigen, maar zeker niet breken. De praktijk is te divers en dynamisch om te vatten in starre wetten. Het oordeel en het talent van de ontwerper zullen van doorslaggevend belang blijven. Maar afwijken van de richtlijnen moet wel een bewuste keuze zijn.

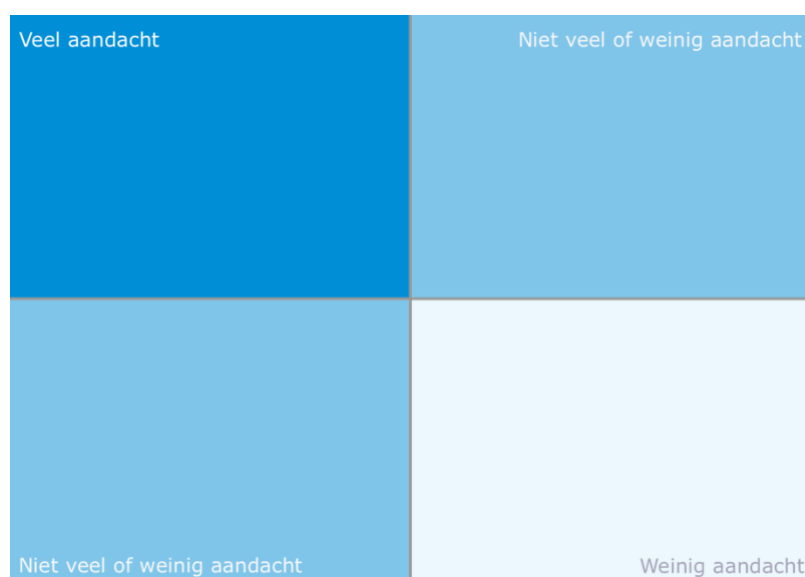
¹ Zie bijlage 1: Design Reference Card

2 Ontwerp standaarden

Een deel van de onderwerpen in dit hoofdstuk zijn reeds beschreven in algemene huisstijl documenten van ProRail. Voor de leesbaarheid zijn een aantal onderdelen hier herhaald, maar de meeste onderdelen zijn een toevoeging aan de bestaande documenten. De toevoegingen zijn dan specifiek gericht op datavisualisaties.

2.1 Layout

Bij het ontwerpen van de datavisualisatie kun je de aandacht trekken naar bepaalde informatie door het gebruik van bijvoorbeeld kleur of lettertype-opmaak. De layout is over het algemeen statisch, deze verandert niet als de data verandert. Uit onderzoek² blijkt echter dat niet ieder gebied op een scherm of pagina van nature evenveel aandacht krijgt. Vooral de linkerbovenhoek krijgt veel aandacht.



Figuur 4 Aandachtzones van een datavisualisatie

Gebruik de linkerbovenhoek dus bij voorkeur voor de belangrijkste informatie. Logo en contextgegevens (als bijvoorbeeld een paginanummer) horen hier meestal niet bij, dus plaats die bij voorkeur ergens anders.



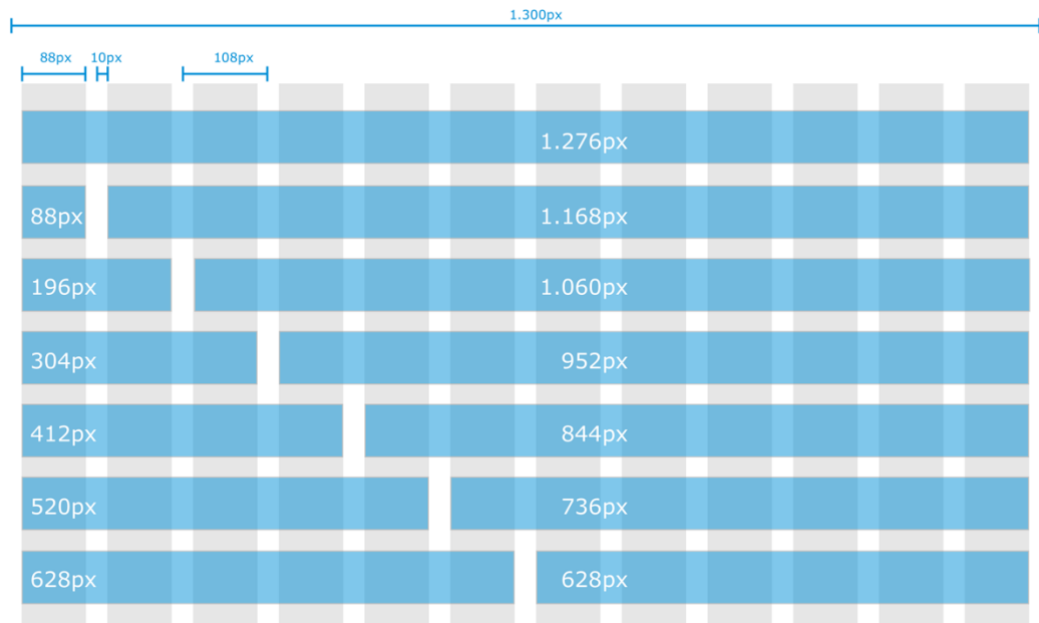
Figuur 5 ProRail dashboard kop

Naast de specifieke inhoud heeft ieder ProRail dashboards een standaard kop. Waarbij linksboven de belangrijkste functies zijn geplaatst (Home, Menu en een viertal functie iconen). Daarnaast staat de titel van het dashboard met daaronder één of meerdere data-filters (hier Jaar-Kwartaal). Tenslotte staat rechts het ProRail logo.

² Bron: Show me the numbers – S. Few

Verder zijn de dashboards binnen ProRail geoptimaliseerd voor een scherm van 1300 pixels breed en 780 pixels lang (1300x780). Om optimaal gebruik te maken van deze ruimte is het goed om gebruik te maken van een standaard grid. Zo'n grid zorgt voor rust en structuur door alle componenten (titels, teksten, tabellen, grafieken, illustraties, etc.) te plaatsen op dit grid.

Het gebruikte grid bestaat uit 12 kolommen. Iedere kolom heeft aan beide zijden witruimte, ook wel de *gutter* genaamd. Deze gutter zorgt dat de onderlinge afstand van de te plaatsen componenten altijd gelijk is. Op basis van het eerdergenoemde schermformaat (1300x780) ziet het grid er dus als volgt uit.

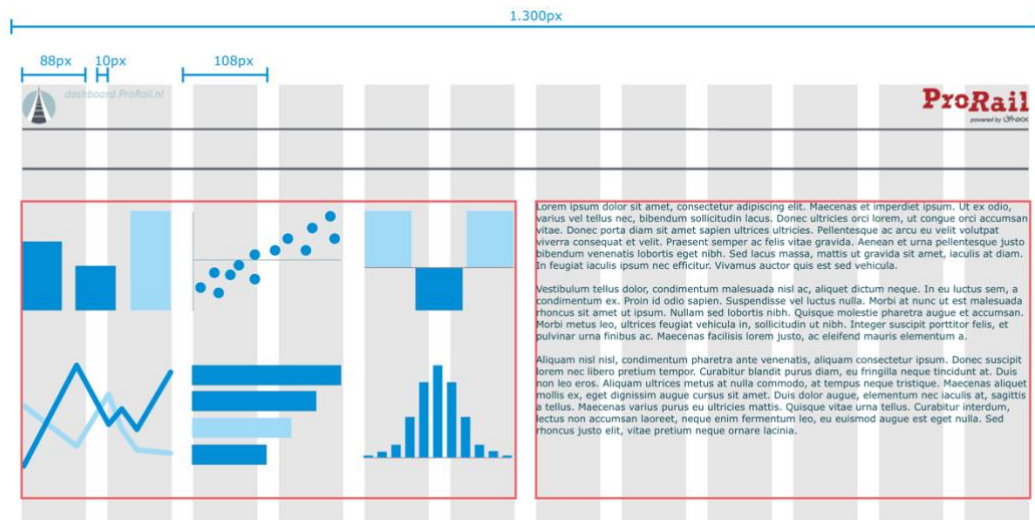


Figuur 6 1300px grid systeem

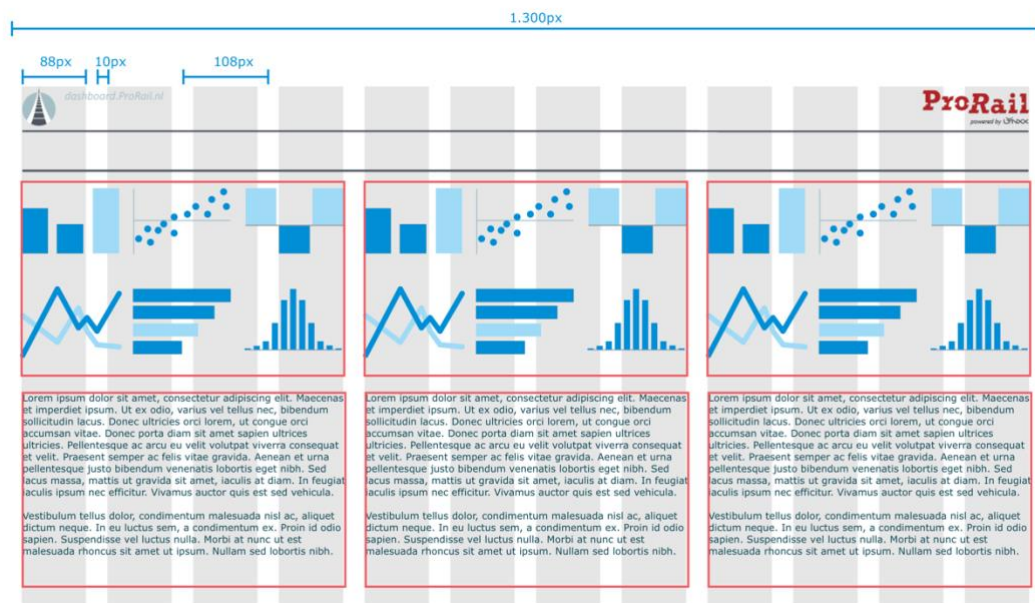
Mocht je slecht één component plaatsen op een regel, dan kan deze maximaal 1276 pixels breed gemaakt worden. Het smalste component is 88 pixels.

Ieder component dat op een dashboard of rapport wordt geplaatst wordt uitgelijnd op één van de kolommen. Een component kan meerdere kolommen breed zijn. Mocht op een regel meerdere componenten naast elkaar geplaatst worden dan wordt er altijd een *gutter* aan witruimte tussen de componenten gehanteerd.

Bij ProRail worden de dashboard standaard uitgerust met een 1, 2 of 3 kolommen indeling. De volgende plaatjes illustreren hoe deze geplaatst kunnen worden op het grid.



Figuur 7 layout met 2 componenten naast elkaar (diagram en tekst naast elkaar)



Figuur 8 layout met 3 componenten naast elkaar

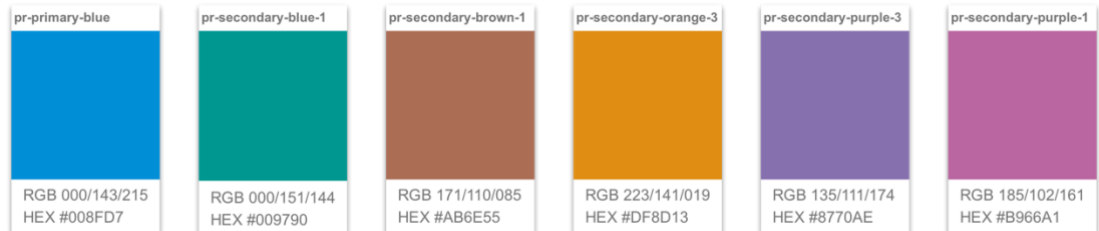
De grid kolommen worden dus alleen gebruikt om de elementen te plaatsen, het is niet de bedoeling dat de grid kolommen ook worden afgebeeld in de uiteindelijke datavisualisatie.

2.2 Kleur

In datavisualisaties zijn er vele verschillende onderdelen met ieder een eigen kleur of kleuren. Kleuren hebben een grote invloed op de perceptie van orde, rust en uniforme uitstraling. Vandaar dat het van groot belang is om zo veel mogelijk de volgende kleuren toe te passen en zo min mogelijk hiervan af te wijken.

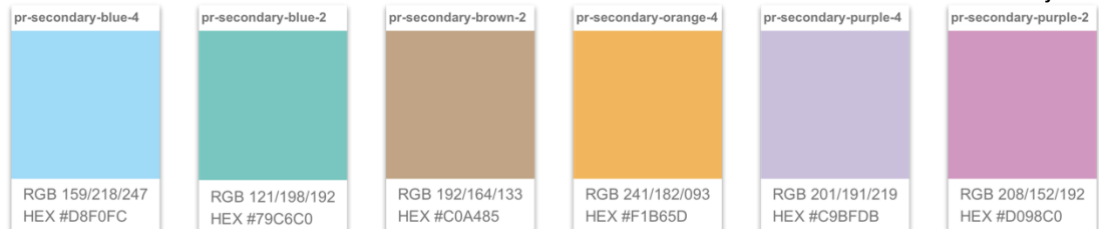
Verder proberen we aan te sluiten bij de algemene ontwerprichtlijnen binnen ProRail. Vandaar dat de belangrijkste datavisualisatie kleuren gebaseerd zijn op de kleuren uit de ProRail-Mendix stijlgids (<https://prorailstyleguide.mendixcloud.com>).

Voor het identificeren van data-elementen (data-series) maken we gebruik van de volgende zes kleuren:



Figuur 9 Standaard ProRail kleuren donkere variant

Verder hebben we voor ieder van deze kleuren ook een lichte variant beschikbaar. Dit zijn:



Figuur 10 Standaard ProRail kleuren lichte variant

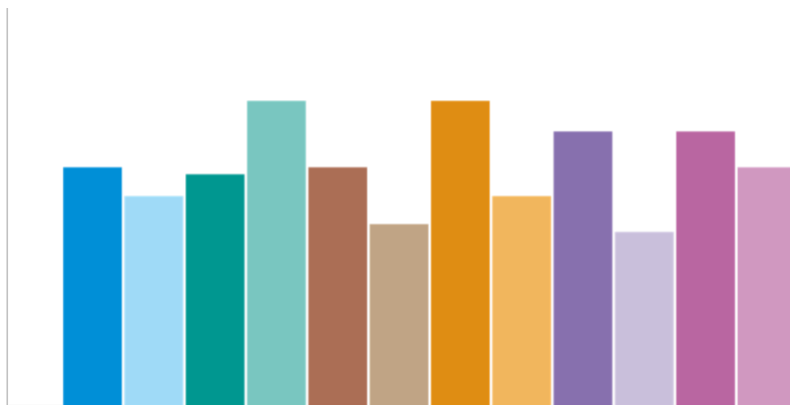
Afhankelijk van de situatie worden deze kleuren als volgt toegepast:

1. Bij maximaal 6 data-series kunnen de donkere varianten gebruikt worden voor het identificeren van de huidige periode (jaar/maand/etc.) en de lichte varianten voor de voorgaande periode. De kleuren worden gebruikt in de volgorde zoals ze hier staan afgebeeld (van links naar rechts);
2. Voor alle overige gevallen gebruik je telkens eerst een donkere variant gevolgd door de lichtere variant. De kleuren worden gebruikt in de volgorde zoals ze hier staan afgebeeld (van links naar rechts). Ter illustratie een Kolomdiagram met alle 12 kleuren in de juiste volgorde.
3. Voor het identificeren van regio's wordt zoveel mogelijk gebruik gemaakt van dezelfde kleur per regio. De standaardkleuren per regio zijn te zien in de onderstaande kaart



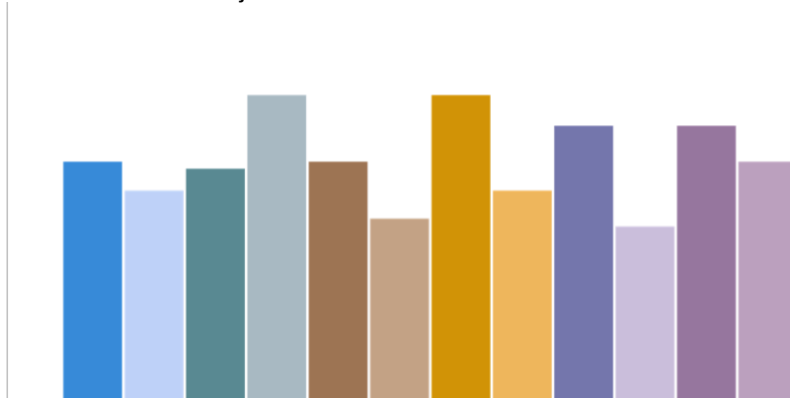
Voor het identificeren van deze regio's dus bij voorkeur de volgende kleuren gebruiken:

- Midden = pr-secondary-blue-2
- Noord = pr-secondary-brown-2
- Noord-West = pr-primary-blue
- Oost = pr-secondary-orange-3
- Zuid-Holland-Noord = pr-secondary-blue-4
- Zuid-Holland-Zuid = pr-secondary-blue-1
- Zuid-Oost = pr-secondary-orange-4
- Zuid-West = pr-secondary-brown-1
- Zuid-Zevenaar = pr-secondary-purple-4



Figuur 11 Voorbeeld met alle 12 kleuren

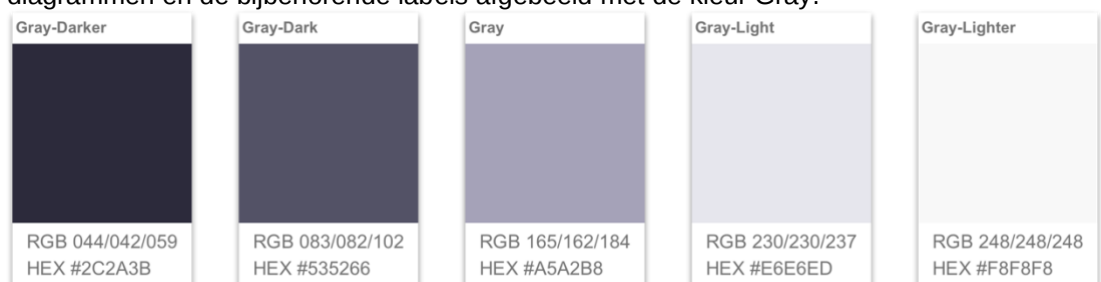
Ook hebben we deze kleuren getest op de meest voorkomende vormen van kleurenblindheid. Bij kleurenblindheid zullen deze kleuren er als volgt uitzien:



Figuur 12 Kleuren bij kleurenblindheid

Bovenstaand plaatje laat duidelijk zien dat het onderscheid tussen de 12 kleuren duidelijk aanwezig blijft.

Naast de data-elementen zelf zijn er een groot aantal ondersteunende elementen binnen een datavisualisatie zoals schaalverdelingen, labels, titels, etc. Deze geven context aan de datavisualisatie, dus hiervoor gebruiken we een neutrale kleur. Zo worden alle assen in de diagrammen en de bijbehorende labels afgebeeld met de kleur Gray.



Figuur 13 ProRail grijstinten

Andere contextgegevens op een datavisualisatie (datum opgemaakt, gebruikersnaam, etc.) krijgen ook de kleur Gray. De kleur Gray-Dark wordt toegepast bij verklarende teksten.

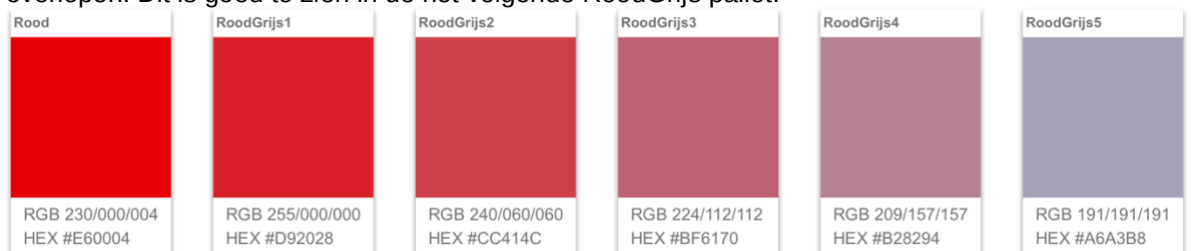
Mocht het nodig zijn om binnen een verklarende tekst de nadruk op een woord of groep woorden te leggen, dan kan hiervoor de kleur Gray-Darker gebruikt worden.

Om de lezer van een datavisualisatie te helpen kan er een signaalkleur gebruikt worden. De signaalkleur bij uitstek is de kleur Rood. Om zeker te zijn dat het signaal wel blijft opvallen moet deze kleur beperkt worden toegepast. Om ook duidelijk een onderscheid te kunnen maken met het ProRail Logo Rood is een felle kleur rood gekozen. De kleur ProRail Logo Rood wordt niet gebruikt in visualisaties, maar is exclusief voor het logo.



Figuur 14 ProRail rode kleuren

In datavisualisaties worden kleuren niet alleen gebruikt om onderscheid te maken in categorieën (afdelingen, regio's, onderdelen, etc.), maar ook om een bepaalde waarde te tonen. Doordat er voor de meeste van ons geen vaste volgorde in kleuren bestaat is het lastig om verschillende kleuren te gebruiken om een kwantitatief verschil te tonen. Vandaar dat er gekozen is om één of maximaal twee kleuren te gebruiken en deze in elkaar te laten overlopen. Dit is goed te zien in de het volgende RoodGrijs pallet.



Figuur 15 Kwantitatief RoodGrijs pallet

Dit pallet begint met standaard Rood wat de hoogste waarde aangeeft. Vervolgens gaat de kleur in gelijke stappen over naar de standaardkleur LichtGrijs.

Dit kan ook gedaan worden binnen één enkele kleur. In het volgende voorbeeld is dit gedaan met alleen Rood.



Figuur 16 Kwantitatief Rood pallet

2.3 Lettertype

Grofweg worden lettertypen onderverdeeld in lettertypen met- (serif) en zonder schreef (sans-serif). In de illustratie³ rechts staat boven een lettertype zonder schreef en onder met schreef (de schreef is gemarkeerd in rood).



Figuur 17 Schreef en schreefloos lettertype

Een schreefloos lettertype wordt voornamelijk gebruikt voor getallen, losse woorden en korte zinnen. Voor langere teksten wordt vaak gebruik gemaakt van een lettertype met schreef, omdat dit de leesbaarheid vergroot.

Doordat in datavisualisaties lange teksten weinig voorkomen wordt binnen ProRail bij voorkeur gebruik gemaakt van het schreefloze lettertype Arial. Dit is in lijn met de huistsijl.

AaBbCcDdEeFfGgHhIiJjKkLlMmNn
OoPpQqRrSsTtUuVvWwXxYyZz01
23456789€£\$%+=+*/;,:."{}[]!?'&

Figuur 18 Lettertype Arial

De standaard puntgrootte voor alle teksten is 10 pixels. De titels worden groter en vet afgedrukt (20, 12 en 10 pixels). Voor de rapportageteksten in PowerPoint wordt voor de leesbaarheid een iets groter standaard lettertype gebruikt. Voor deze teksten wordt een puntgrootte van 10,5 gebruikt.

Primaire titel (Arial 20px Bold)

Caption-grafiek	(Arial 16 px)
Algemene tekst	(Arial 14px)
Secundaire titel	(Arial 12px Bold)
Tertiaire titel	(Arial 10px Bold)
As-waarden	(Arial 12px)
Kleine tekst	(Arial 9px)

Figuur 19 Lettertype omvang

2.4 Opmaak van nummers en datums

Er zijn vele verschillende manieren om nummers en datums op te maken. Om de leesbaarheid te vergroten is het goed om één en dezelfde manier te hanteren binnen een visualisatie. Voor de meest voorkomende gevallen wordt in de volgende tabel aangegeven wat de voorkeur heeft. De "Nederlandse" notatie wordt gebruikt voor het onderscheiden van duizendtallen (met een punt) en decimalen (met een komma).

³ Bron: <https://nl.wikipedia.org/wiki/Schreefloos>

Data type	Opmerking	Opties
Getal in duizenden	Gebruik 'k' om de lengte van het getal te verkorten.	32.924 wordt 32,9k
Getal in miljoenen	Gebruik 'M' of 'mln' om de lengte van het getal te verkorten.	10.465.209 wordt 10,5M of 10,5mln
Getal in miljarden	Gebruik 'G' of 'mld' Om de lengte van het getal te verkorten.	34.665.829.774 wordt 34,6G of 34,6mld
Negatieve getallen	Voorafgaand aan het getal de tekens -/-	Voorkeur: -/- 234,56 Ruimtegebrek: - 234,56
Gebruiken van het jaar als een label	Schrijf het hele jaar uit (4 posities). Alleen bij gebrek aan ruimte afkorten naar 2 posities.	Voorkeur: 2017 Ruimtegebrek: '17
Gebruiken van het kwartaal als een label	Gebruik 'Q' gevolgd door het kwartaalnummer.	Q1, Q2, Q3 of Q4
Gebruiken van de maand als een label	Gebruik de eerste drie letters van de maand. Alleen bij gebrek aan ruimte afkorten met het maandnummer.	Voorkeur: okt Ruimtegebrek: 10
Gebruiken van de maand en het jaar als een label	Vaak wordt dit gebruikt om een periode van meerdere aaneengesloten maanden aan te duiden.	Voorkeur: jan 2017-jun 2017 Ruimtegebrek: 01/17-06/17
Gebruiken van dag, maand en eventueel jaar als een label	Gebruik alleen het jaar als er meer dan 1 jaar aan data zichtbaar is. Zorg dan wel dat er op een andere plek duidelijk vermeld staat welk jaar het is.	Voorkeur: 19-mrt of 19-mrt-2017 Ruimtegebrek: 19-03 of 19-03-2017
Decimalen	Rond alle getallen af naar het minimum aantal decimalen die nodig zijn om het onderscheid te kunnen maken tussen de waarden.	32,9241 wordt 32,9
Geldbedragen	Geldbedragen worden voorafgegaan door het valutateken (€) of de ISO 4217 afkorting (EUR)	Voorkeur: € 826,97 Alternatief: EUR 826,97

Mocht in een visualisatie (tabel of diagram) een grootte (bijv. geldbedragen) veelvuldig voorkomen zorg dan dat alle waarden dezelfde eenheid (bijv. miljoenen Euro's zonder decimalen: € 234M) en notatiewijze hebben. Vermijd ook het herhalen van eenheden zoals beschreven in sectie 3.1.4.

2.5 Visualisatie principes

Volgens Jock Mackinlay van de Stanford University moet een datavisualisatie zowel *Expressief* als *Effectief* zijn. Met deze twee termen vat Mackinlay de belangrijkste visualisatieprincipes goed samen.

Een visualisatie is *Expressief* als deze alle feiten uit de data weergeeft, maar ook alleen de feiten. Met andere woorden: vertel de waarheid en niets anders dan de waarheid (lieg niet, ook niet door het weglaten van data).

Een visualisatie is *Effectief* als deze gebruik maakt van vormen die onze visuele waarneming het beste kan interpreteren. Het beste betekent hier sneller en accurater begrip van de data.

Om een visualisatie *Effectief* te maken is vooral kennis over “Gestalt” en “Pre-attentieve waarneming” nuttig. In de volgende paragrafen worden beide onderwerpen kort toegelicht.

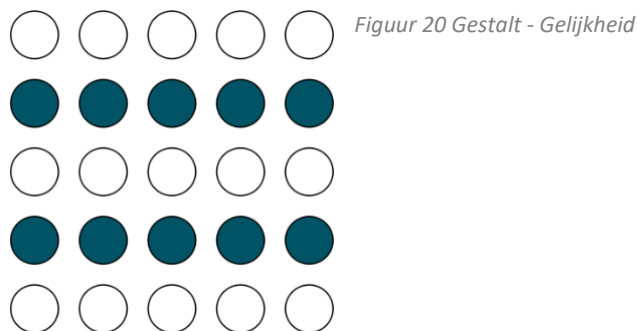
2.5.1 Gestalt

Gestalt is de Duitse term voor "georganiseerde" of "gehele" vorm. Het centrale idee achter Gestalt is dat de som der delen groter is dan de afzonderlijke delen. Met andere woorden: het geheel (de auto) heeft een andere en ook grotere betekenis dan de afzonderlijke onderdelen (banden, lak, ramen, metaal, etc.).

Ondanks dat deze wetten al bijna 100 jaar oud zijn worden ze nog steeds gezien als de verklaring voor het functioneren van onze visuele waarneming. Deze wetten zijn daardoor ook de basis van de meeste ontwerpprincipes van visualisaties⁴.

Voor het effectief ontwerpen van een visualisatie zijn de volgende Gestalt-wetten van belang:

- Wet van gelijkheid: Elementen die op elkaar lijken worden gegroepeerd. Elementen die veel van elkaar verschillen worden als aparte elementen waargenomen;

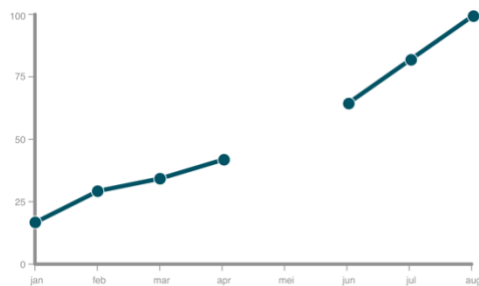


⁴ Information Visualization, Second Edition: Perception for Design – Colin Ware

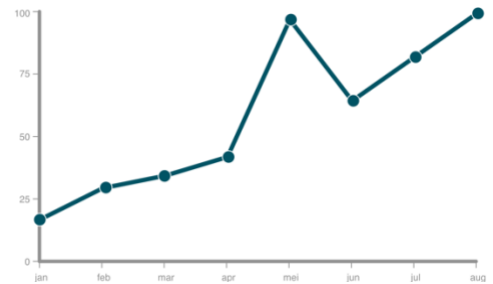
- Wet van nabijheid: Elementen die zich dicht bij elkaar bevinden, worden als groepen gezien. Elementen die zich ver van elkaar bevinden worden als onafhankelijk waargenomen;



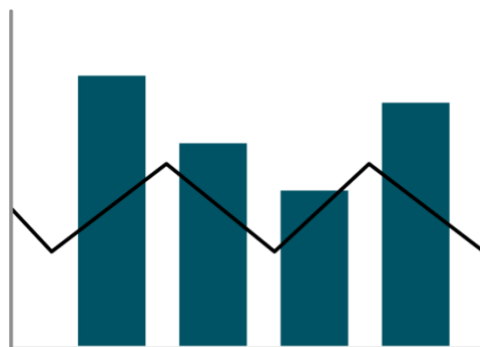
- Wet van geslotenheid: Elementen die omrand worden door een lijn worden als geheel gezien. Elementen die doorkruist worden door een lijn worden als aparte elementen waargenomen. Een voorbeeld is Figuur 20: er staan in deze figuur 2 assen welke de indruk geven dat er een "gesloten" rechthoek bestaat waarbinnen de grafiek afgebeeld is;
- Wet van continuïteit: De vorm van een element wordt voortgezet zoals deze is begonnen;



Figuur 22 Gestalt - Continuïteit



- Wet van eenvoud: Elementen worden waargenomen in de meest eenvoudige vorm: minste cognitieve inspanning heeft de voorkeur;



Figuur 23 Gestalt - Eenvoud



Voor het ontwerp van datavisualisaties moeten we dus rekening houden met:

- Mensen groeperen elementen op basis van kleur, vorm, richting, nabijheid en geslotenheid;
- Mensen hebben een voorkeur voor eenvoudige, gesloten en symmetrische vormen die weinig inspanning kosten om te verwerken;
- De mogelijk negatieve gevolgen van de wetten: voorkom dat mensen groepen gaan zien die er eigenlijk niet zijn (volgens de onderliggende data).

2.5.2 Pre-attentieve waarneming

Onderzoek gedurende de afgelopen decennia heeft aangetoond dat er een beperkte set van visuele eigenschappen bestaat die we kunnen waarnemen zonder dat we er (lang) over hoeven na te denken. Dit zijn de zogenaamde pre-attentieve eigenschappen⁵.

Belangrijke doelen van datavisualisatie zijn het informatief en betekenisvol presenteren van data, maar tegelijkertijd moet de presentatie ook intuïtief zijn en weinig moeite kosten. Hieraan kunnen deze eigenschappen een belangrijke bijdrage leveren.

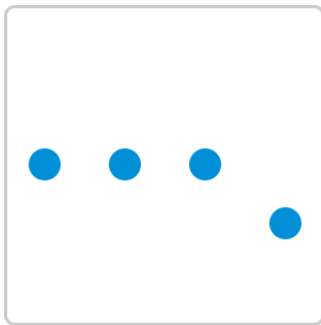
Van de pre-attentieve eigenschappen zijn er een aantal die we kunnen gebruiken in datavisualisaties. De lijst is gesorteerd op nauwkeurigheid. Hoe hoger op de lijst hoe nauwkeuriger de waarden waargenomen worden⁶ (zie voorbeeld 23 tm 27).

1. 2D Positie
2. Lengte
3. Dikte
4. Oppervlakte
5. Kleurintensiteit

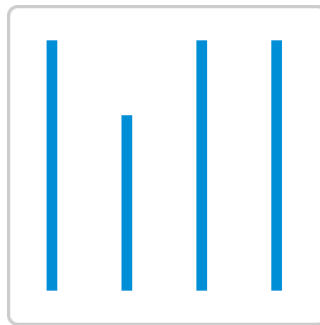
Hierbij lenen alleen de eerste twee (2D positie en lengte) zich goed als nauwkeurigheid een vereiste is. Vandaar dat in de meeste gevallen een lijndiagram, staafdiagram of puntenwolk de voorkeur hebben bij visualisaties.

⁵ High-Speed Visual Estimation Using Preattentive Processing – Healey, Booth, Enns – juni 1996

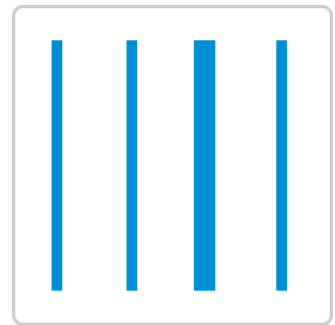
⁶ The Functional Art – Alberto Cairo - 2013



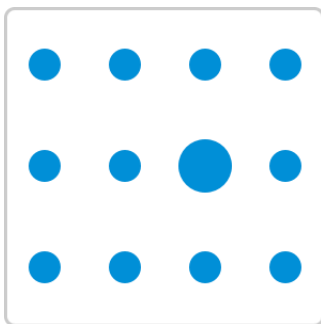
Figuur 24 2D positie



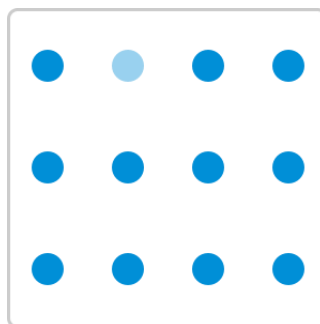
Figuur 25 Lengte



Figuur 26 Dikte



Figuur 27 Oppervlakte



Figuur 28 Kleurintensiteit

3 Visualiseren van data

Er zijn oneindig veel manieren om data te visualiseren, van eenvoudige tabellen tot uiteenlopende diagramvormen. Dit hoofdstuk zal de verschillende manieren beschrijven en aangeven wanneer welke vorm het beste werkt. De keuze tussen de verschillende vormen maken we op basis van de theorie en principes zoals beschreven in het vorige hoofdstuk.

Verder is het van groot belang om rekening te houden met de kennis en ervaring van de doelgroep. Een visualisatievorm kan in theorie wel effectief zijn, maar als deze voor de doelgroep niet duidelijk is schieten we ons doel voorbij.

Bij het ontwerpen van datavisualisaties is het goed om verschillende benaderingen uit te proberen (experimenteren) en te schetsen. Het beste resultaat ontstaat door veel oefening en met vallen en opstaan.

3.1 De tabel

Wellicht is het niet het eerste waar je aan denkt bij datavisualisatie, maar de tabel is waarschijnlijk de belangrijkste basisvorm. Het is vooral sterk in het vergelijken (met grote precisie) van individuele waarden. De tabel maakt het ook mogelijk om waarden van verschillende meeteenheden (bijv. percentage van het totaal, gemiddeld, geldbedragen, aantallen, etc.) te vergelijken.

3.1.1 Onderscheiden van rijen en kolommen

Het onderscheiden van rijen en kolommen wordt vaak gedaan met lijnen. Hierdoor heeft dit vaak een negatieve invloed op de zogenaamde data-inkt-ratio. De data-inkt-ratio is een concept geïntroduceerd door Edward Tufte. Hij beschrijft het concept als volgt: "Een groot gedeelte van de gebruikte inkt voor het afbeelden van een diagram (of tabel) is bedoeld voor de data-informatie. Dit is dus de inkt die meeverandert als de data verandert. Data-inkt is de kern van de visualisatie die niet weggelaten kan worden zonder betekenis te verliezen. De rest is non-data-inkt, welke geminimaliseerd dient te worden. De data-inkt-ratio is de data-inkt gedeeld door de totaal gebruikte inkt (data-inkt+non-data-inkt)".

Het onderscheiden van rijen en kolommen is dus een voorbeeld van non-data-inkt. Hiervoor moeten we dus een vorm kiezen die zo min mogelijk inkt kost. Voor het onderscheid zijn in het algemeen 3 manieren:

- Witruimte;
- Achtergrondkleur;
- Lijnen.

Dit is ook de voorkeursvolgorde: begin altijd met witruimte. Maar zoals de term (witruimte) al aangeeft kost het wel ruimte. Dus mocht deze ruimte niet beschikbaar zijn gebruik dan een subtiele achtergrondkleur. Deze kleur moet net zichtbaar zijn, maar mag zeker niet teveel afleiden. De laatste optie is een lijn in een subtiele kleur.

In de praktijk lijken lijnen vaak de voorkeur te krijgen om onderscheid te maken tussen rijen en kolommen. Een voorbeeld hiervan is de onderstaande tabel met een lage data-inkt-ratio.

	januari	februari	maart	april	mei	juni
produkt 1	267	357	587	320	268	398
produkt 2	365	387	401	406	421	404
produkt 3	554	582	561	551	583	600
totaal	1.186	1.326	1.549	1.277	1.272	1.402



Figuur 29 tabel met een lage data-inkt-ratio

Door de data naar voren te halen en de non-data te verwijderen of minder prominent te maken kan de data-inkt-ratio worden verhoogd.

	januari	februari	maart	april	mei	juni
produkt 1	267	357	587	320	268	398
produkt 2	365	387	401	406	421	404
produkt 3	554	582	561	551	583	600
totaal	1.186	1.326	1.549	1.277	1.272	1.402



Figuur 30 tabel met een hoge data-inkt-ratio

Echter, als de tabel veel breder (meer kolommen) of langer (meer rijen) wordt volstaat witruimte vaak niet meer. Het wordt dan vooral lastig om te bepalen bij welke rij een bepaald getal thuishoort. Dan is het subtiel toepassen van achtergrondkleur of een lijn de enige optie. Probeer dit tot een minimum te beperken, maar de leesbaarheid staat altijd voorop.

	januari	februari	maart	april	mei	juni	juli	augustus	september	oktober	november	december
produkt 1	2.267	265	2.267	265	2.267	265	2.267	265	2.267	265	2.267	265
produkt 2	3.365	231	3.365	231	3.365	231	3.365	231	3.365	231	3.365	231
produkt 3	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354
produkt 4	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942
produkt 5	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153
produkt 6	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756
produkt 7	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354
produkt 8	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942
produkt 9	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153
produkt 10	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756

Figuur 31 Achtergrondkleur om de rijen te verduidelijken

In het bovenstaande voorbeeld (Figuur 31) wordt een subtiele achtergrondkleur gebruikt om het onderscheid te maken tussen de verschillende regels. Deze achtergrondkleur kan ook worden gebruikt om regels met subtotalen te onderscheiden van de normale regels, zoals te zien is in het voorbeeld in Figuur 83.

	januari	februari	maart	april	mei	juni	juli	augustus	september	oktober	november	december
produkt 1	2.267	265	2.267	265	2.267	265	2.267	265	2.267	265	2.267	265
produkt 2	3.365	231	3.365	231	3.365	231	3.365	231	3.365	231	3.365	231
produkt 3	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354
produkt 4	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942
produkt 5	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153
produkt 6	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756
produkt 7	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354	1.554	354
produkt 8	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942	1.357	942
produkt 9	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153	2.987	153
produkt 10	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756	2.582	756

Figuur 32 Subtiele lijnen om de rijen te verduidelijken.

3.1.2 Accentueren

Door onderdelen van een tabel (bijvoorbeeld een cel of groep van cellen) te accentueren kunnen we hiervoor extra aandacht vragen. Er zijn drie effectieve manieren om te accentueren:

- Afwijkende achtergrondkleur;
- Tekst vet afdrucken;
- Tekst kleuren.

In de onderstaande figuur zijn deze drie manieren gebruikt als voorbeeld. Het is echter niet de bedoeling om alle manieren tegelijkertijd toe te passen. Kies er één en probeer je daartoe te beperken.

	januari	februari	maart	april	mei	juni
produkt 1	267	357	587	320	268	398
produkt 2	365	387	401	406	421	404
produkt 3	554	582	561	551	583	600
totaal	1.186	1.326	1.549	1.277	1.272	1.402

Figuur 33 voorbeelden van accentueren in een tabel

3.1.3 Uitlijnen

Numerieke data moet zoveel mogelijk (rechts) uitgelijnd worden op de decimale komma. In de meeste software is het niet mogelijk om op de decimale komma uit te lijnen. Zorg in dat geval dat alle getallen dezelfde hoeveelheid decimalen hebben en lijn deze getallen rechts uit. Als er geen komma staat dan wordt er uitgelijnd op het meest rechtse getal. Hiermee staan de getallen die met elkaar vergeleken mogen worden ook onder elkaar.

correct uitgelijnd	incorrect uitgelijnd	incorrect uitgelijnd
9.453,54	9.453,54	9.453,54
100,5437	100,5437	100,5437
5,6	5,6	5,6
1.004.236	1.004.236	1.004.236
0,27	0,27	0,27

Figuur 34 uitlijnen van numerieke data

Doordat we lezen van links-naar-rechts moet tekst vrijwel altijd links uitgelijnd worden. Onder tekst vallen ook nummers die geen kwantitatieve waarde vertegenwoordigen, zoals een baanvaknummer of afdelingscode. Uitzondering op deze richtlijn zijn de teksten die gebruikt worden als label van een kolom (zie bijvoorbeeld de maanden in Figuur 33). Deze teksten worden op dezelfde manier uitgelijnd als de inhoud van de kolom (in het geval van dit voorbeeld dus rechts).

correct uitgelijnd	incorrect uitgelijnd	incorrect uitgelijnd
Dienst A	Dienst A	Dienst A
Categorie B	Categorie B	Categorie B
235-65-5432	235-65-5432	235-65-5432
ProRail BV	ProRail BV	ProRail BV
Moreelsepark 3	Moreelsepark 3	Moreelsepark 3

Figuur 35 uitlijnen van tekst

Door datum- en tijdsaanduidingen is er geen expliciete voorkeur voor rechts of links uitlijnen. Zolang binnen een visualisatie iedere aanduiding een gelijk aantal posities gebruikt is het goed.

correct uitgelijnd	correct uitgelijnd	correct uitgelijnd
01-01-2015	01-01-2015	01/01/15
05-12-2016	05-12-2016	05/12/16
23-06-1995	23-06-1995	23/06/95
31-03-2014	31-03-2014	31/03/14
16-11-2011	16-11-2011	16/11/11

Figuur 36 uitlijnen van datums

3.1.4 (Verschillende) eenheden binnen een tabel

Meestal heeft een rij of een kolom één eenheid. Bijvoorbeeld allemaal Euro's of aantal treinen of een bepaald percentage. De eenheid staat dan vaak ook in de betreffende rij of kolom.

	omzet	aandeel	⊗
produkt 1	€ 267	3,3%	
produkt 2	€ 3.365	41,0%	
produkt 3	€ 554	6,7%	
produkt 4	€ 1.357	16,5%	
produkt 5	€ 87	1,1%	
produkt 6	€ 2.582	31,4%	
totaal	€ 8.212	100,0%	

Figuur 37 Eenheden staan in iedere cel

	omzet(€)	aandeel(%)	⊙
produkt 1	267	3,3	
produkt 2	3.365	41,0	
produkt 3	554	6,7	
produkt 4	1.357	16,5	
produkt 5	87	1,1	
produkt 6	2.582	31,4	
totaal	8.212	100,0	

Figuur 38 Eenheden staan in de koptekst

Door de eenheden in iedere cel te plaatsen ontstaat er veel dubbele data die vaak niets toevoegt. Sterker nog: bij bijvoorbeeld het euroteken (de omzet kolom) lijdt de dynamische positie alleen maar af. Vandaar dat het beter is om eenheden in de koptekst te plaatsen. Enige uitzondering hierop is als er binnen één rij of kolom verschillende eenheden voorkomen.

3.1.5 Grafische ondersteuning in de tabel

Zoals al aangegeven zit de kracht van een tabel in het exact vergeleken van individuele waarden. Het kan zijn dat de hoeveelheid waarden zo groot is dat het vergelijken bemoeilijkt wordt. In deze gevallen kunnen kleine grafische toevoegingen het vergelijken vereenvoudigen. Een vaak toegepast voorbeeld is de liggende staafdiagram. Door de waarden uit een kolom ook als staaf op te nemen worden afwijkingen sneller zichtbaar.

	Regio A	Regio B	⊗
produkt 1	2.267	265	
produkt 2	3.365	231	
produkt 3	1.554	354	
produkt 4	1.357	942	
produkt 5	2.987	153	
produkt 6	2.582	756	
totaal	14.112	2.701	

Figuur 39 Zonder grafische ondersteuning moet ieder getal gelezen worden

	Regio A	Regio B	⊙
produkt 1	2.267	265	
produkt 2	3.365	231	
produkt 3	1.554	354	
produkt 4	1.357	942	
produkt 5	2.987	153	
produkt 6	2.582	756	
totaal	14.112	2.701	

Figuur 40 Met grafische ondersteuning worden mogelijke patronen direct zichtbaar

3.2 Het diagram

Zoals eerder aangegeven is de tabel een krachtige basis voor de meeste data visualisaties. Echter, het heeft ook zijn beperkingen doordat de getallen in een tabel gelezen moeten worden. Dit dient de nauwkeurigheid, maar maakt het lastig om het grotere geheel te zien.

Als de boodschap meer in de vorm van de data zit wordt het beoordelen van alle data in één keer van groter belang. Hiervoor kunnen we gebruik maken van diagrammen. Aan diagramtypes zeker geen gebrek, maar in de meeste gevallen zijn de basisvormen in het vervolg van dit hoofdstuk voldoende.

Om een goede keuze te maken in de beschikbare diagrammen kijken we eerst naar de mogelijke relaties in data. Deze bepalen welke diagrammen nuttig zijn, waardoor de keuze eenvoudiger wordt.

3.2.1 Keuzehulp

Bij het herkennen van relaties in data moeten we beginnen met de twee basis types: categorieën en hoeveelheden. Voor een datavisualisatie bestaat de relatie minimaal uit één hoeveelheid en één categorie. Bijvoorbeeld het aantal treinen (hoeveelheid) per baanvak (categorie).

Zowel de hoeveelheden als de categorieën worden in een datavisualisatie getoond op een schaalverdeling. Bij categorieën onderscheiden we drie type schalen:

- **Nominale schaal**
Bij een nominale schaal bestaat er geen natuurlijke volgorde van de elementen. Een voorbeeld daarvan zijn afdelingen (financiën, inkoop, HR, ICT, etc). De term (nominaal) geeft aan dat de elementen alleen verschillende namen hebben zonder een bijzondere relatie onderling;
- **Ordinale schaal**
In tegenstelling tot een nominale schaal heeft een ordinale schaal wel een natuurlijke volgorde. Bijvoorbeeld het alfabet (de B komt na de A). De volgorde op een ordinale schaal mag je dus niet zomaar veranderen. De bekendste ordinale schaal die we tegenkomen in datavisualisaties is tijd.
- **Interval schaal**
Deze schaal ontstaat vaak bij het samenvatten van een hoeveelheid-schaalverdeling. Het kan nuttig zijn om hoeveelheden in te delen in groepen (bijvoorbeeld aantal storingen: 0, 1-4, 5-9, 10-14, >14). Deze groepen zijn een interval schaal.

Met het onderscheid tussen categorieën en hoeveelheden enerzijds en drie categorie-schalen anderzijds kunnen we de relaties beschrijven die de keuze voor een visualisatievorm bepalen. De volgende tabel laat per relatie de voorkeur voor een basisvorm zien (gebaseerd op de diagram-selectie-matrix uit Bijlage 2).

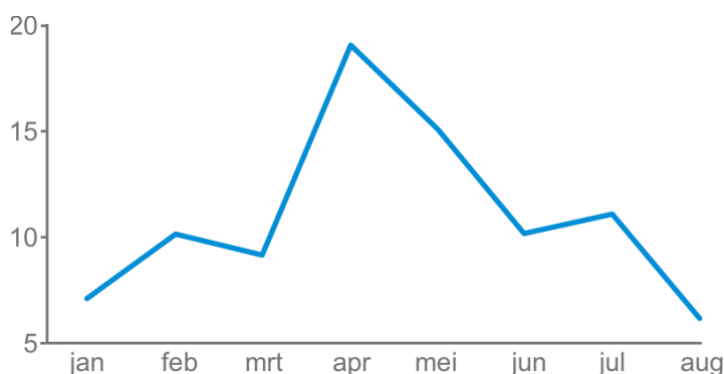
Relatie / beschrijving	Diagramvorm
 Nominaal vergelijk Een eenvoudig vergelijk van hoeveelheden per categorie zonder een bepaalde volgorde	De voorkeur gaat uit naar een staafdiagram. Mocht de numerieke schaal niet op nul (kunnen) beginnen gebruik dan een puntendiagram (zie Figuur 51).
 Tijdreeksen Hoeveelheden op verschillende tijdstipmomenten	De eerste voorkeur gaat uit naar het lijndiagram, vooral voor het zichtbaar maken van trends. Als de nadruk ligt op het vergelijken van verschillende categorieën op één tijdstip krijgt de verticale staafdiagram de voorkeur.
 Rangorde Een categorische onderverdeling van een hoeveelheid georderd op waarde (van groot naar klein of andersom)	Een horizontale staafdiagram is meestal de beste keuze. Mocht de numerieke schaal niet op nul (kunnen) beginnen gebruik dan een puntendiagram.
 Deel van groter geheel De verdeling van een categorie naar rato (als percentage) van een hoeveelheid	Op de eerste plaats kies je hier voor een horizontale staafdiagram. Zeker als er verschillende verdelingen met elkaar vergeleken moeten worden. Mocht dit niet het geval zijn en zijn er maximaal 4 delen, dan kan een taartdiagram gebruikt worden
 Afwijking Een categorische onderverdeling waarbij de ene hoeveelheid wordt vergeleken met een andere hoeveelheid	Afwijking leent zich goed voor een staafdiagram. In combinatie met een tijdreeks is ook een lijndiagram mogelijk.
 Correlatie Vergelijk van twee hoeveelheden, waarbij gekeken wordt hoe de ene hoeveelheid zich gedraagt tov de ander	De puntenwolk is specifiek hiervoor ontstaan en is ook de eerste keus.
 Verdeling Een telling van het aantal hoeveelheden per categorie (vaak een interval)	Het lijndiagram (frequentie polygoon) en staafdiagram (histogram) zijn de beste keuze als het gaat over een enkele verdeling. Als verschillende verdelingen met elkaar vergeleken moeten worden is de boxplot de beste keuze
 Geospatial Hoeveelheden afgezet tegen een categorie met een geografische eigenschap	Met kleurintensiteit op een kaart kan de relatie tussen een waarde(bereik) en een locatie goed gevisualiseerd worden. Om meer precisie te verkrijgen kunnen ook cirkels gebruikt worden, waarbij de omvang van de cirkels varieert met de betrokken waarde.

3.2.2 Basisvormen

Met de volgende basisvormen kunnen de meeste data-relaties goed gevisualiseerd worden. Verder zijn de meeste zo bekend bij een breed publiek dat uitleg tot een minimum beperkt kan worden. Belangrijke stelregel is wel dat ze alleen in de getoonde 2D variant gebruikt worden. Geen enkel diagram wordt beter (leesbaar) door het gebruik van 3D (effecten).

3.2.2.1 Lijndiagram

Het lijndiagram leent zich bij uitstek goed voor het visualiseren van continue data (vaak tijd). Hierbij worden waarden op verschillende tijdstippen verbonden met een lijn. Hierdoor ontstaat er een beeld van de relaties tussen de verschillende waarden: gaat het omhoog of omlaag, is er een trend, etc.



Figuur 41 Lijndiagram

Voor categorische data is het lijndiagram een minder goede keuze. Vaak omdat er tussen de categorieën geen natuurlijke volgorde bestaat, waardoor de lijnen tussen de waarden geen betekenis hebben.

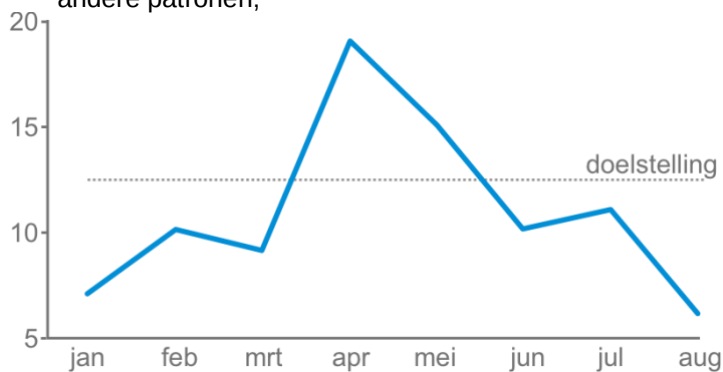


Figuur 42 Lijndiagram niet gebruiken voor categorische data

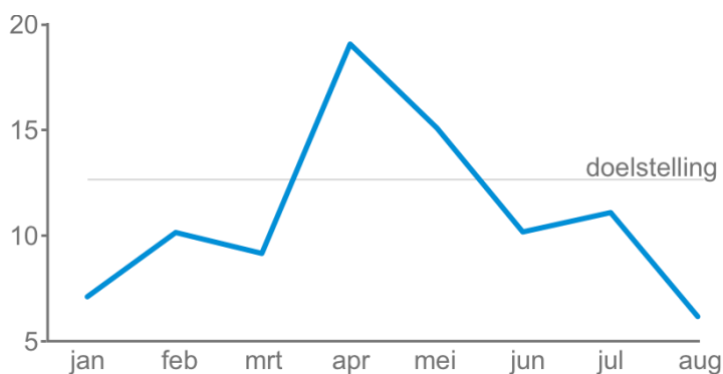
Verder is het van belang om de volgende richtlijnen te volgen bij het toepassen van een lijndiagram:

- Nooit meer dan 4 lijnen. Indien er meer dan 4 series nodig zijn probeer dan meerdere diagrammen te gebruiken of een trellis chart (zie Figuur 76);

- Gebruik één lijnstijl, bij voorkeur een niet-onderbroken lijn. Dus geen stippellijnen of andere patronen;

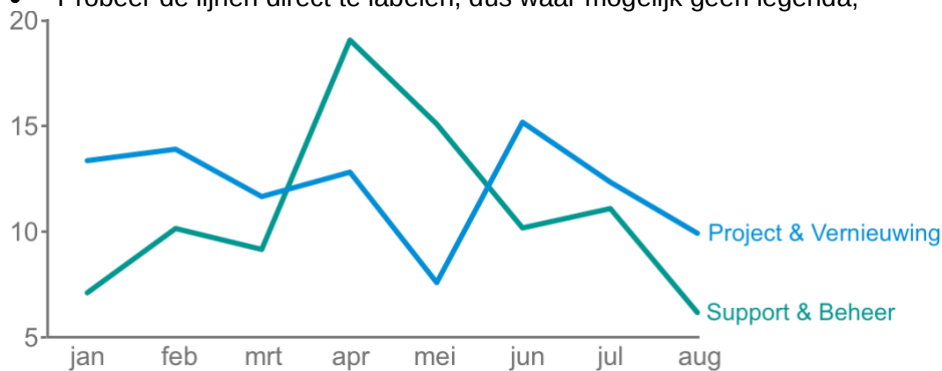


Figuur 43 Het afwijkende patroon van de stippellijn zorgt voor afleiding

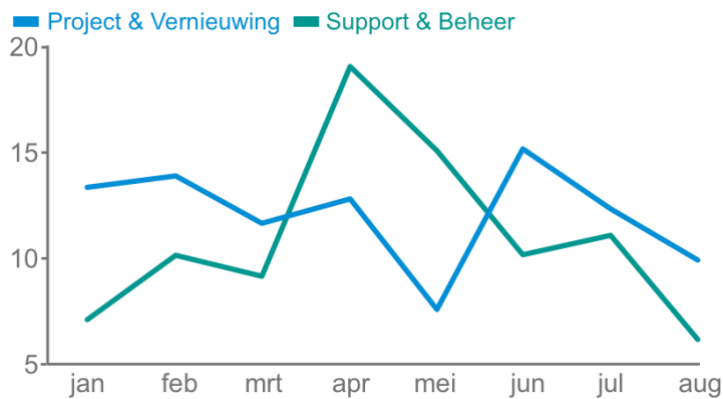


Figuur 44 Het aangeven van een andere rol (doelstelling) kun je beter subtiel doen

- Probeer de lijnen direct te labelen, dus waar mogelijk geen legenda;

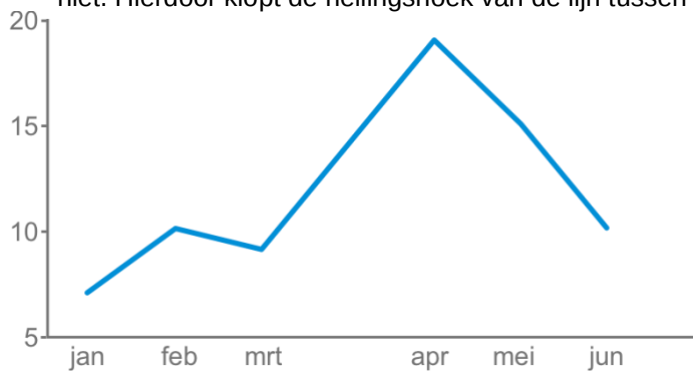


Figuur 45 Data series zijn direct gelabeld



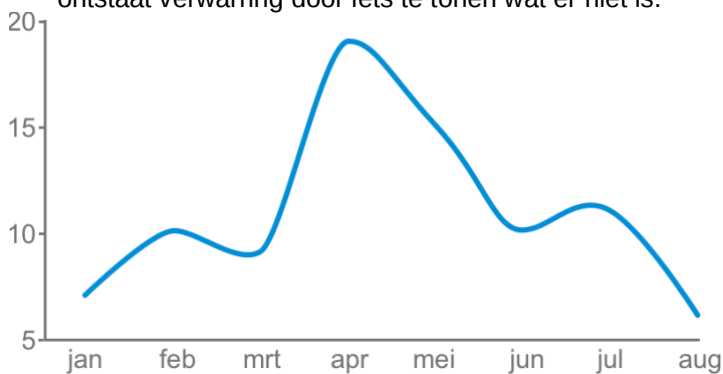
Figuur 46 Data series zijn indirect gelabeld in een legenda

- Zorg voor een consistente afstand met tijd. In het onderstaande voorbeeld is de afstand tussen de kwartalen vergroot, maar in werkelijkheid bestaat die extra afstand er in tijd niet. Hierdoor klopt de hellingshoek van de lijn tussen maart en april niet meer.



Figuur 47 Inconsistente afstand in tijd

- Gebruik geen bochten ("smoothed lines") in het verbinden van de datapunten. Hierdoor ontstaat verwarring door iets te tonen wat er niet is.

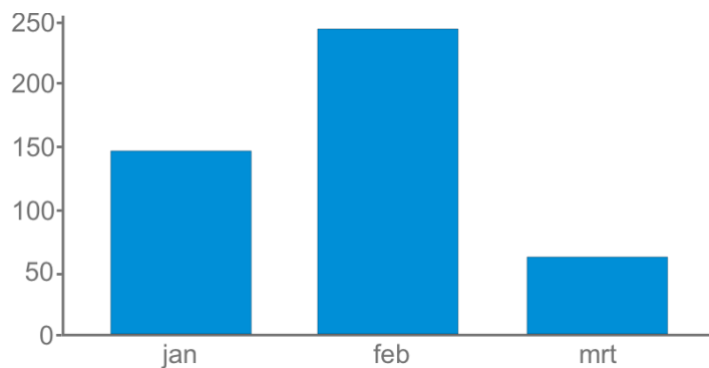


Figuur 48 "Smoothed lines"

3.2.2.2 Staafdiagram

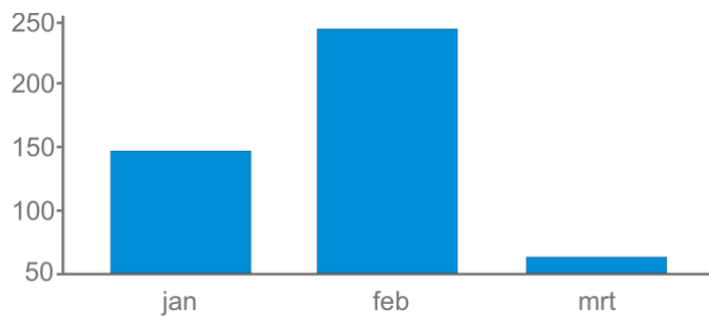
Vaak worden staafdiagrammen vermeden omdat ze zo gewoontjes zijn, maar dit is juist de reden om ze wel te gebruiken. Het gaat immers om de data zo effectief mogelijk te

communiceren. Door gebruik te maken van gewenning voorkom je de focus op de vorm in plaats van de inhoud.

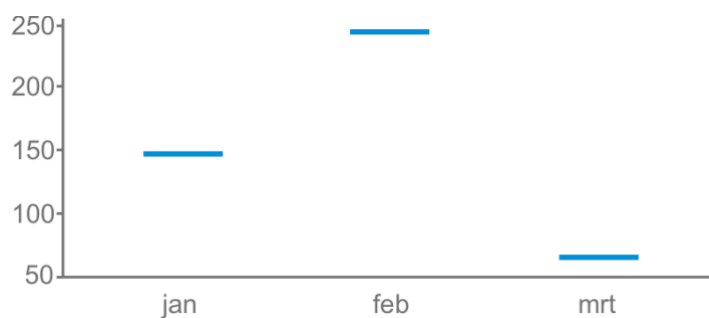


Figuur 49 Staafdiagram

Bij deze vorm van diagrammen wordt de lezer uitgenodigd om de relatieve lengte van de staven te vergelijken. Om dit mogelijk te maken is het van belang om de schaal van de staafdiagram altijd met 0 te laten beginnen. Als hierdoor het verschil tussen de staven niet meer duidelijk wordt (lees: te klein) is de enige oplossing het kiezen van een ander type diagram. In het volgende voorbeeld is de staaf vervangen door een teken (hier een lijn).



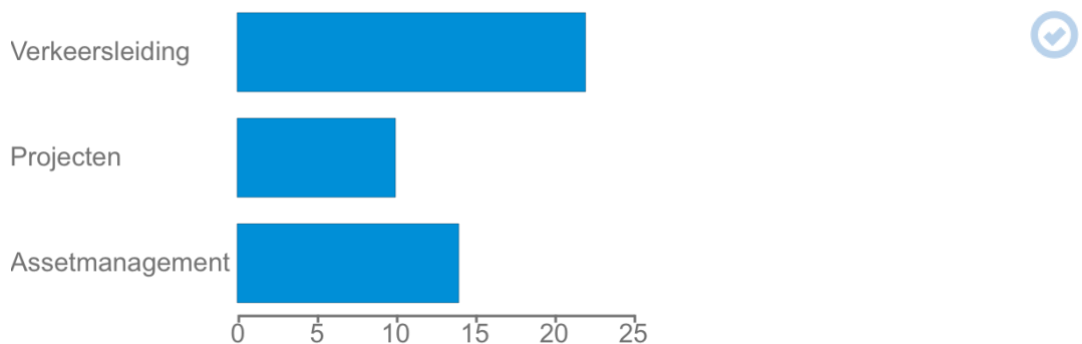
Figuur 50 Schaalverdeling begint niet met 0. Hierdoor kloppen de verhoudingen niet meer



Figuur 51 Als er niet begonnen kan worden met 0, kies dan een andere vorm

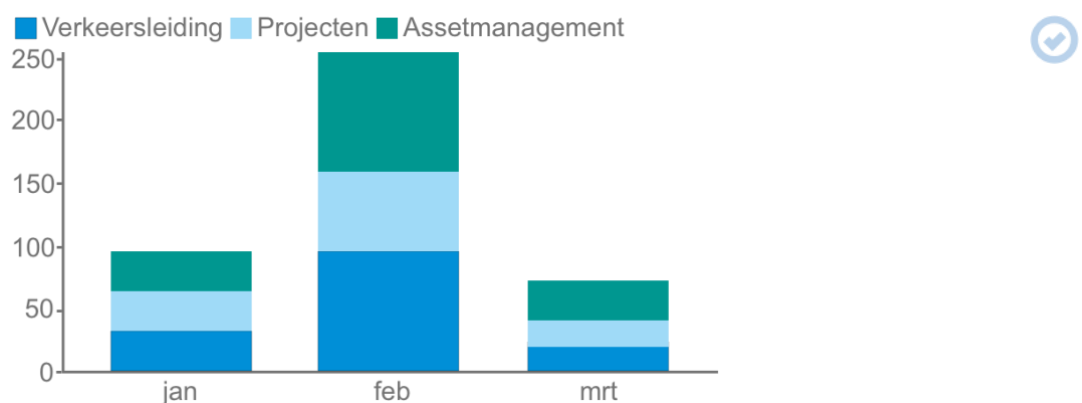
Verder kan het staafdiagram zowel horizontaal als vertikaal worden afgebeeld. Vertikaal (ook wel kolomdiagram genoemd) heeft de voorkeur als er een chronologische volgorde bestaat tussen de staven. Ook als er negatieve waarden worden afgebeeld heeft vertikaal de

voorkeur. Horizontaal heeft de voorkeur als de labels van de staven lang zijn of bij een groot aantal staven.



Figuur 52 Horizontale staafdiagram

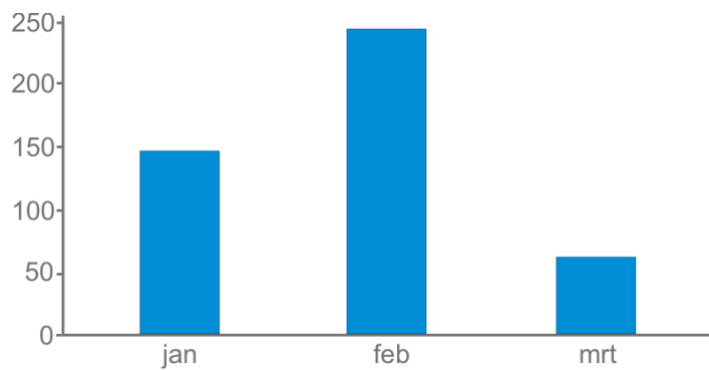
Een andere variant van dit type diagram is het gestapelde staafdiagram. Deze wordt vooral toegepast om per staaf een bepaalde onderverdeling zichtbaar te maken. Voor het vergelijken van de verschillende onderdelen binnen één staaf werkt dit goed, maar voor het vergelijken van één onderdeel over verschillende staven niet. Dit komt doordat een dataelement over de verschillende kolommen heen op verschillende hoogtes begint (met uitzondering van de eerste die op de x-as begint).



Figuur 53 Gestapelde staafdiagram

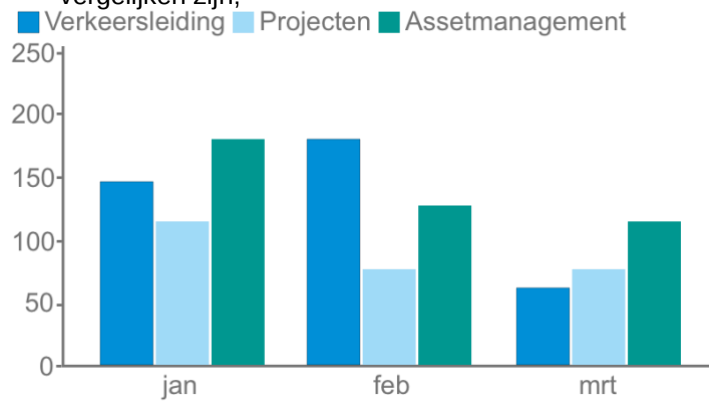
Verder is het van belang om de volgende richtlijnen te volgen bij het toepassen van een staafdiagram:

- Zorg dat de staaf breder is dan de witruimte tussen de staven;

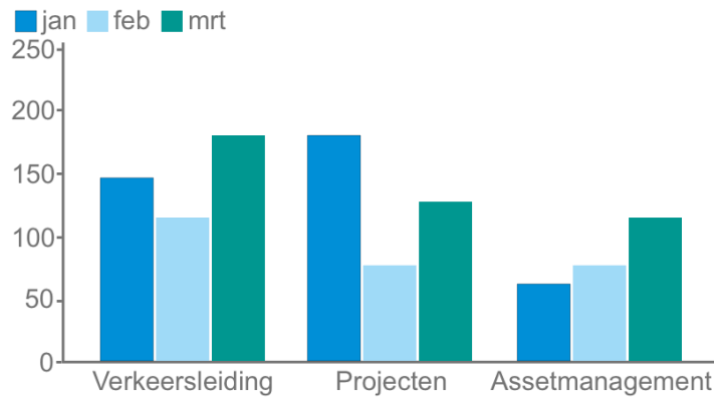


Figuur 54 De witruimte tussen de staven is breder dan de staaf zelf

- Als er meerdere series worden getoond bepaald de groepering welke series goed te vergelijken zijn;

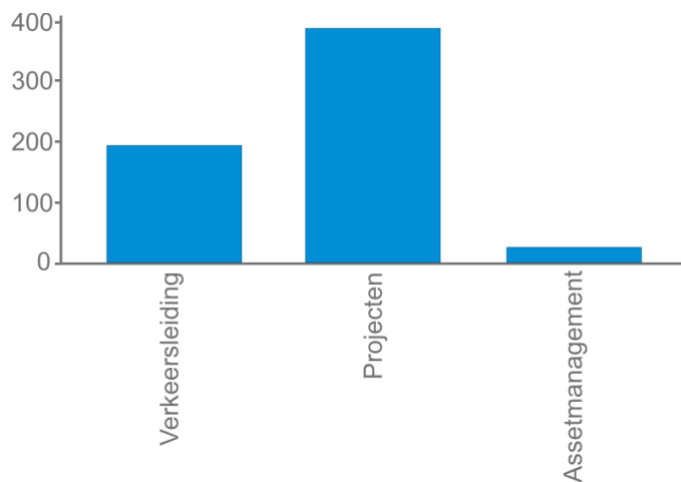


Figuur 55 Dit diagram ondersteund vooral het vergelijk binnen één maand



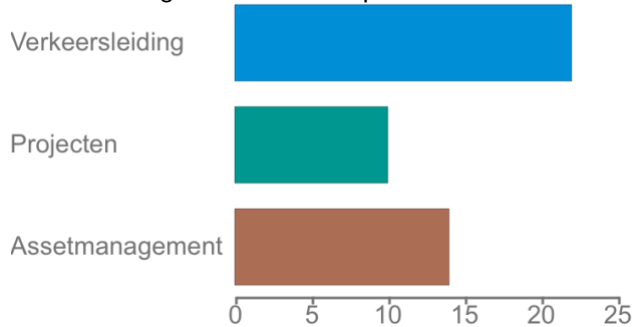
Figuur 56 Hier ligt de nadruk op het vergelijken van de verschillende maanden

- Zorg dat alle labels horizontaal staan;



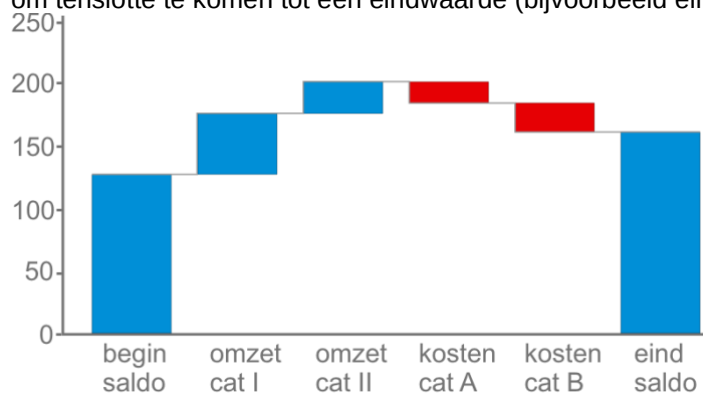
Figuur 57 Zorg dat alle teksten horizontaal worden toegepast

- Gebruik geen kleuren als positie al onderscheidend is;



Figuur 58 De positie in het diagram zorgt al voor onderscheid, kleur voegt dus niets toe

Als afsluiter van deze paragraaf over staafdiagrammen nog een bijzondere vorm: het watervaldiagram. Deze vorm komt vooral vaak voor in financiële omgevingen. Het toont een initiële waarde (bijvoorbeeld beginsaldo) en de opvolgende positieve en negatieve invloeden om tenslotte te komen tot een eindwaarde (bijvoorbeeld eindsaldo).



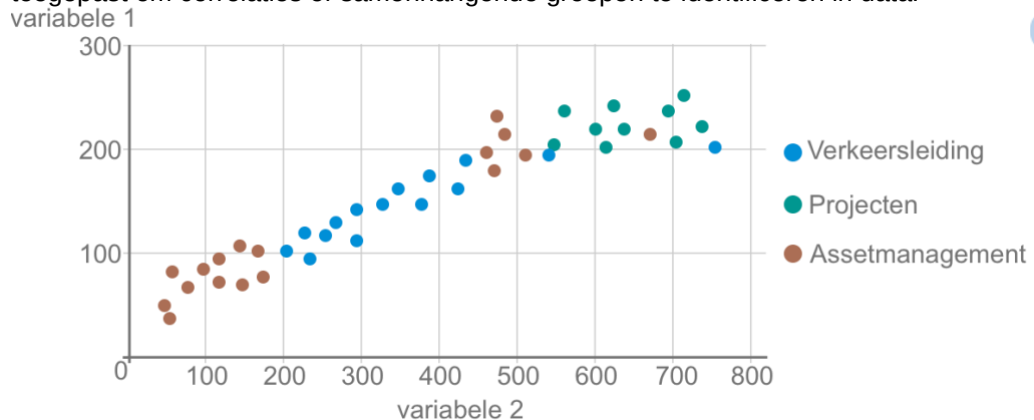
Figuur 59 Waterval diagram

De eerste staaf van een watervaldiagram toont de beginwaarde en de laatste staaf de eindwaarde. De tussenliggende staven de opvolgende wijzigingen. De staven die de

wijziging tonen beginnen telkens op het punt waar de voorafgaande staaf is geëindigd. Om de leesbaarheid te vergroten wordt voor een positieve wijziging (hier lichtblauw) een andere kleur gebruikt dan een negatieve wijziging (hier lichtrood). Ter ondersteuning van deze stapsgewijze veranderingen kan er eventueel een lijn getekend worden langs de uiteinden van de staven.

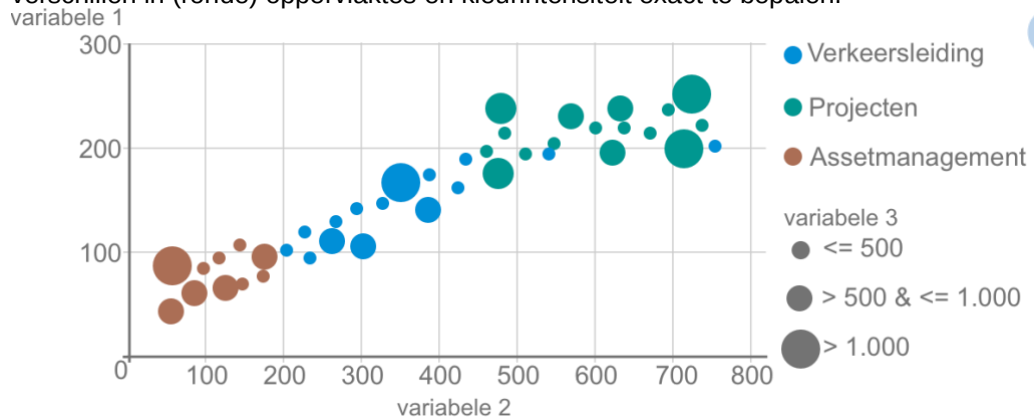
3.2.2.3 Puntenwolk

De puntenwolk is vooral goed in het vergelijken van 2 variabelen. Vaak worden deze toegepast om correlaties of samenhangende groepen te identificeren in data.



Figuur 60 Voorbeeld van een puntenwolk

Een variant op de puntenwolk is het ballondigram. Hierbij wordt nog een derde variable toegevoegd die de omvang van het punt bepaald. Naast de omvang van het punt kan ook de kleurintensiteit gebruikt worden om een derde variabele toe te voegen. Met deze varianten voeg je wel variabele toe, maar dit gaat dan ook ten koste van de precisie. Het is lastig om verschillen in (ronde) oppervlaktes en kleurintensiteit exact te bepalen.



Figuur 61 Voorbeeld van een ballondigram

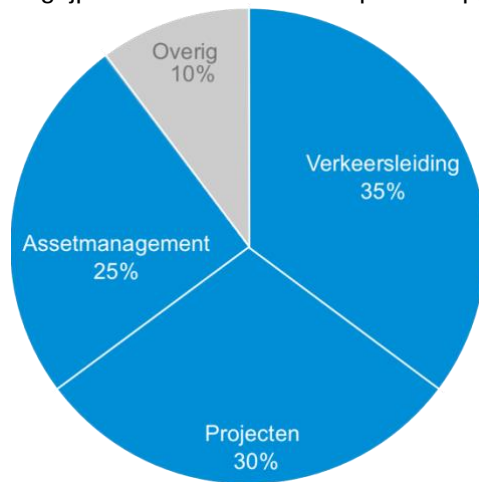
3.2.2.4 Taartdiagram

Dit is waarschijnlijk de meest beruchte basisvorm in datavisualisatie. Experts als S. Few⁷ en E. Tufte⁸ vermijden ieder gebruik, maar in de praktijk is het een zeer populaire vorm. Verder

⁷ Save the Pies for Dessert – Stephen Few – August 2007

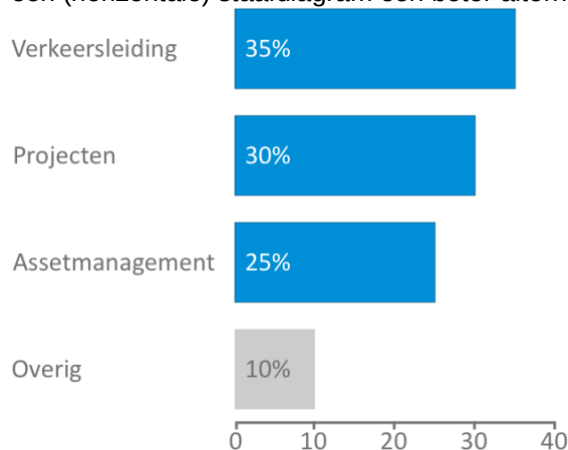
⁸ The worst chart in the world – Walter Hickey – Business Insider – June 2013

is er geen diagram welke zo sterk een “deel-van-een-geheel” relatie symboliseerd. Iedereen begrijpt instinctief dat de taartpunten opgeteld de gehele taart vormen.



Figuur 62 Voorbeeld van een taartdiagram

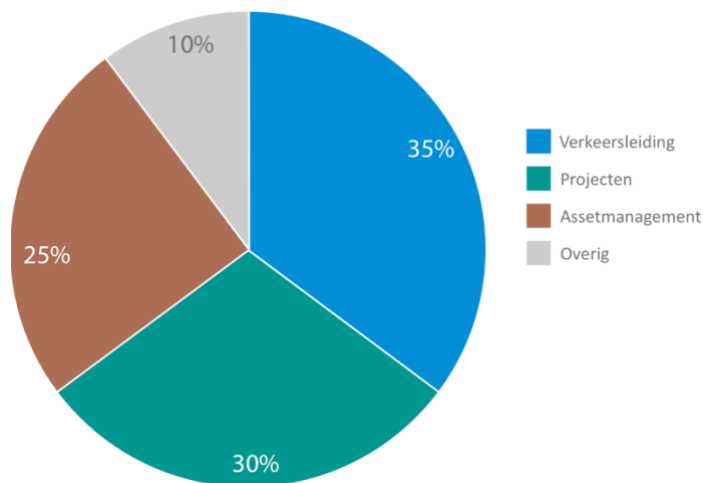
De kritiek op deze vorm bestaat vooral uit het probleem dat we niet goed zijn in het vergelijken van ronde oppervlaktes en dat de ronde vorm veel ruimte inneemt. Dus als precisie een vereiste is en de ruimte beperkt dan is deze vorm geen goede keuze. Meestal is een (horizontale) staafdiagram een beter alternatief.



Figuur 63 Alternatief voor de taartdiagram

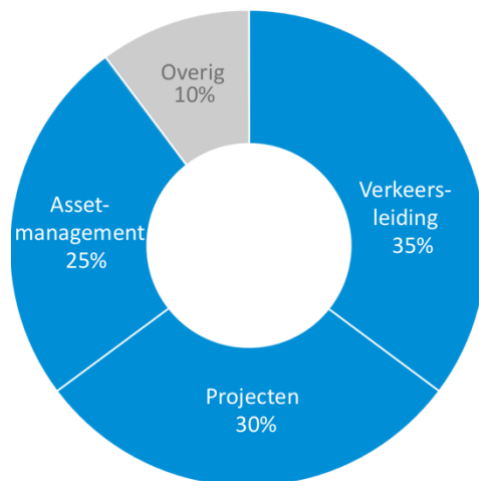
Verder is het van belang om de volgende richtlijnen te volgen bij het toepassen van een taartdiagram:

- Maximaal 4 categorieën (taartpunten) per diagram. Mocht je meer dan 4 categorieën hebben toon dan de top 3 en voeg de rest samen in een Overig categorie;
- Gebruik geen taartdiagram als je meerdere diagrammen met elkaar wilt vergelijken;
- De som van alle taartpunten is altijd gelijk aan 100%;
- Hanteer altijd dezelfde volgorde van taartpunten: begin op 12 uur met de grootste punt en plaats vervolgens met de klok mee de rest van groot naar klein;
- Label de taartpunten direct, waardoor het eenvoudig is om te lezen en je het aantal kleuren beperkt;



Figuur 64 Label direct en beperk het aantal kleuren

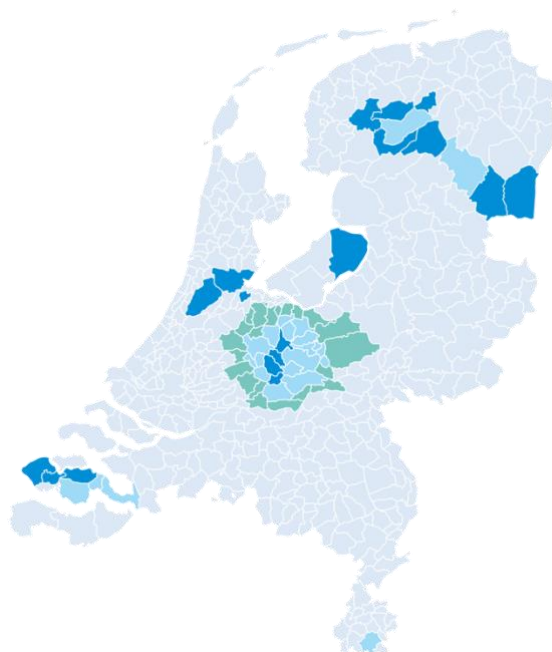
- Gebruik slechts één basiskleur, eventueel aangevuld met een kleur om accenten aan te brengen;
- Gebruik geen donutdiagram. Dit is een taartdiagram met een gat in het midden.



Figuur 65 Donutdiagram

3.2.2.5 Geografische kaarten

Als fysieke locatie een belangrijke factor is zijn (geografische) kaarten een goed hulpmiddel. Door het gebruik van kaarten wordt de data verrijkt met positie en afstand tussen posities. Het onderstaande voorbeeld (met Nederlandse gemeentes) laat bijvoorbeeld duidelijk zien dat de meest donkere gebieden voornamelijk in en rond Utrecht liggen. Bij ieder ander diagram zou deze concentratie niet of zeer lastig zichtbaar zijn.



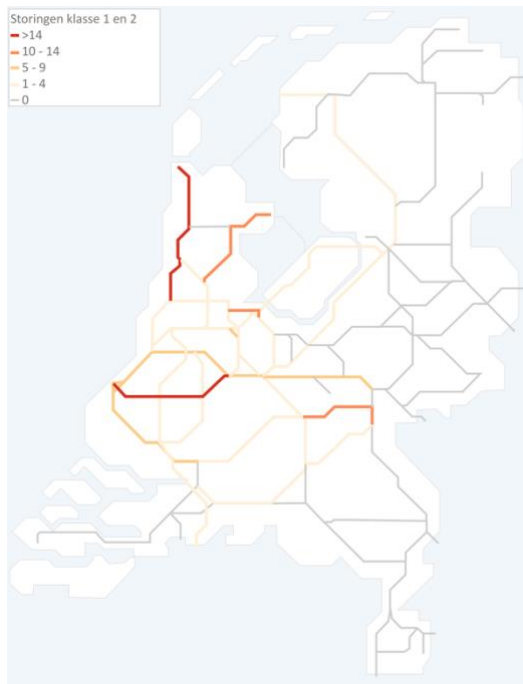
Figuur 66 voorbeeld van een geografische kaart

Er kleeft echter ook een risico aan deze toepassing van een kaart. De omvang van een gebied (hier een gemeente) hoeft geen relatie te hebben met de waarde die getoont wordt met de kleurintensiteit (bijvoorbeeld het aantal patienten). Met andere woorden: een grote gemeente lijkt een grotere waarde te hebben doordat het een grotere oppervlakte van de visualisatie gebruikt (bijvoorbeeld de gemeenten in de flevopolder). Maar de oppervlakte zegt niets over de variable die getoond wordt.

Het op deze manier toepassen van een kaart geeft dus alleen een indicatie van de geografische spreiding van de waarden. Meestal wordt deze vorm dus aangevuld met een andere visualisatie om precisie toe te voegen. Verder is het ook een goed hulpmiddel om uitzonderingen te identificeren (zoals de gemeente in het oosten van Nederland).

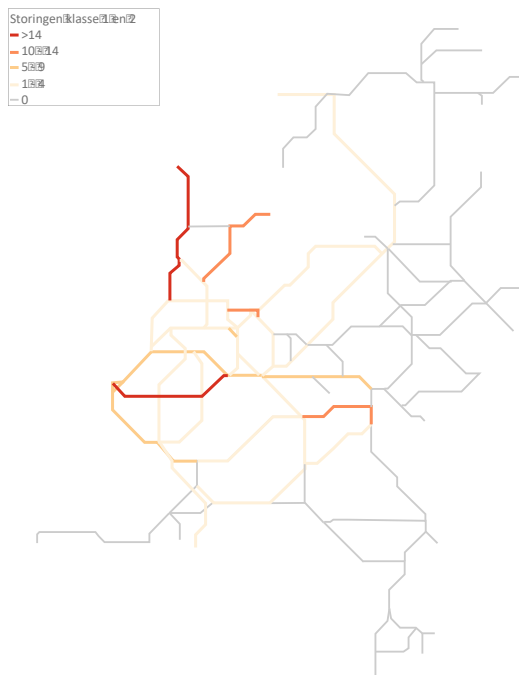
Voor de meeste mensen zit er geen logische volgorde in kleuren, waardoor het gebruiken van verschillende kleuren geen goede optie is. Het beste is om slechts één kleur te gebruiken en voor hoge waarden een donkere variant te gebruiken en voor lagere waarden een lichtere variant.

Naast het gebruik van geografische gebieden kan een kaart ook gebruikt worden om specifieke routes of sporen weer te geven. Ook hierbij speelt de locatie een belangrijke rol, waardoor bepaalde geografische concentraties het beste met een kaart inzichtelijk gemaakt kunnen worden.



Figuur 67 voorbeeld van geografische kaart met sporen

In dit voorbeeld is gekozen om geografische precisie op te offeren voor verbeterde leesbaarheid. Het is namelijk niet van belang om ieder spoor precies weer te geven. In dit voorbeeld zijn de sporen extra dik aangezet om ze beter leesbaar te maken. Tevens is er gekozen om alleen de contouren van de kaart te tonen, want meer details verslechteren alleen maar de leesbaarheid. Aangezien de sporen op zich zelf al een duidelijk geografisch overzicht geven is het ook mogelijk om de kaart helemaal weg te laten en alleen de sporen te tonen, zoals in het volgende voorbeeld.



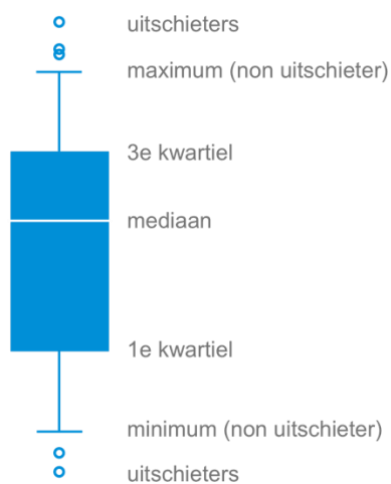
Figuur 68 voorbeeld van geografische kaart met alleen sporen

3.2.2.6 Boxplot

Vaak wordt een verdeling van data afgebeeld als lijndiagram (frequentie polygoon) of staafdiagram (histogram). Deze vormen lenen zich er alleen niet goed voor om verschillende verdelingen met elkaar te vergelijken. Daarvoor kan de boxplot (of snorredoos, doosdiagram of kader-met-staafdiagram⁹) worden gebruikt. Het is echter diagramvorm die niet bij iedereen bekend is, dus alleen bruikbaar voor een publiek dat bekend is met de vijf-getallensamenvatting uit de beschrijvende statistiek¹⁰. Buiten een statistische omgeving kom je de boxplot vrijwel niet tegen.

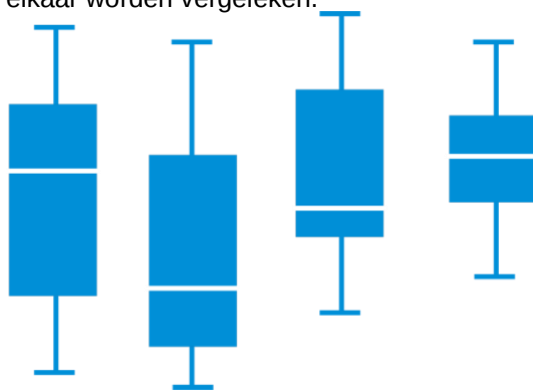
⁹ <https://nl.wikipedia.org/wiki/Boxplot>

¹⁰ <https://nl.wikipedia.org/wiki/Vijf-getallensamenvatting>



Figuur 69 Onderdelen van een boxplot

Met de boxplot ontstaat een vereenvoudigde, maar zeer bruikbare, voorstelling van een gegevensverzameling. Op deze manier kunnen verschillende verzamelingen eenvoudig met elkaar worden vergeleken.



Figuur 70 Vergelijken van verzamelingen met een boxplot

3.2.2.7 KPI Meter

De meeste management dashboards vergelijken een resultaat met een doelstelling of budget. Een gangbare manier om dit te visualiseren is het gebruik maken van een vorm van snelheidsmeter.

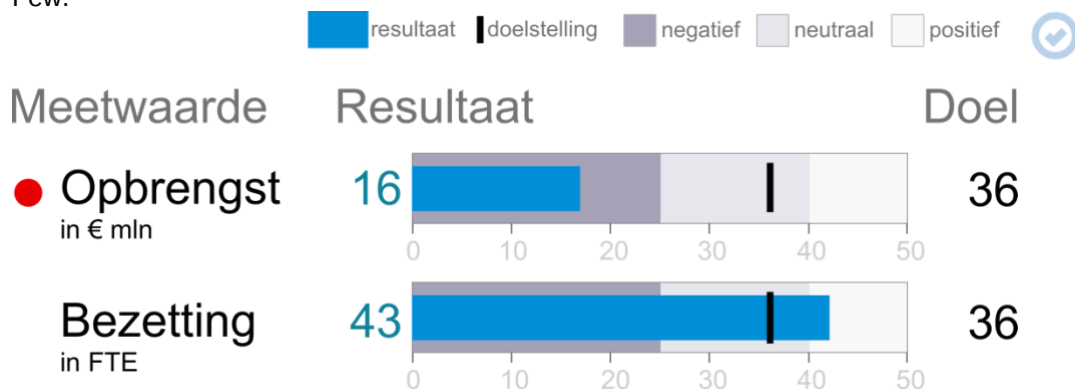


Figuur 71 Voorbeeld van een snelheidsmeter

Deze visualisatievorm heeft echter een aantal grote nadelen:

- Door de ronde vorm neemt de snelheidsmeter veel ruimte in beslag;
- Het resultaat is af te lezen met de hoek waaronder de naald de schaalverdeling aanwijst, maar dit ondersteunt geen vergelijking tussen meerdere snelheidsmeters;
- Over het algemeen hebben de snelheidsmeters zeer weinig context (zoals schaalverdeling en een doel indicator), waardoor het vaak niet meer is dan de afbeelding van één enkele waarde.

Om deze nadelen te ontwijken is de bullet graph¹¹ (KPI meter) ontwikkeld door Stephen Few.



Figuur 72 Bullet graph of KPI meter

Het is wellicht even wennen, maar deze vorm biedt grote voordelen:

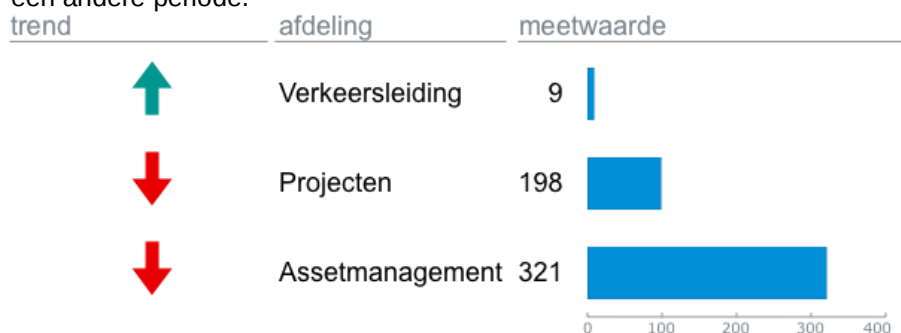
- Neemt weinig ruimte in beslag;
- Door de uitslag in de vorm van een staaf zijn de verschillende waarden goed te vergelijken;
- Alle noodzakelijke context (schaalverdeling en doel) zijn een integraal onderdeel.

De rode punt (of cirkel) wordt hier gebruikt om de aandacht van de lezer te trekken naar een specifieke meetwaarde. Hiermee kun je de lezer helpen in het snel waarnemen van de belangrijkste elementen. Vaak zijn dat de elementen die een grote (negatieve) afwijking vertonen.

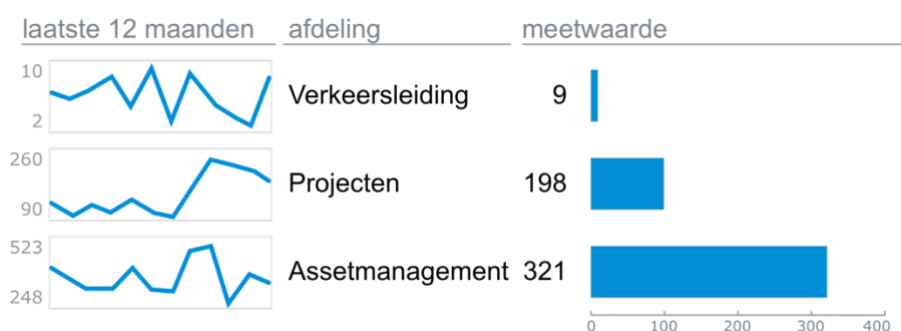
¹¹ https://en.wikipedia.org/wiki/Bullet_graph

3.2.2.8 Sparkline

Een sparkline is een diagram in het formaat van een woord. Meestal zonder assen of een gedetailleerde schaalverdeling. Vaak wordt een sparkline toegepast om snel een “gevoel” te krijgen voor de ontwikkeling van een waarde in de tijd. Het is dus een goed alternatief voor de veelvoorkomende “pijl” die aangeeft of een waarde toe- of afgenomen is ten opzichte van een andere periode.



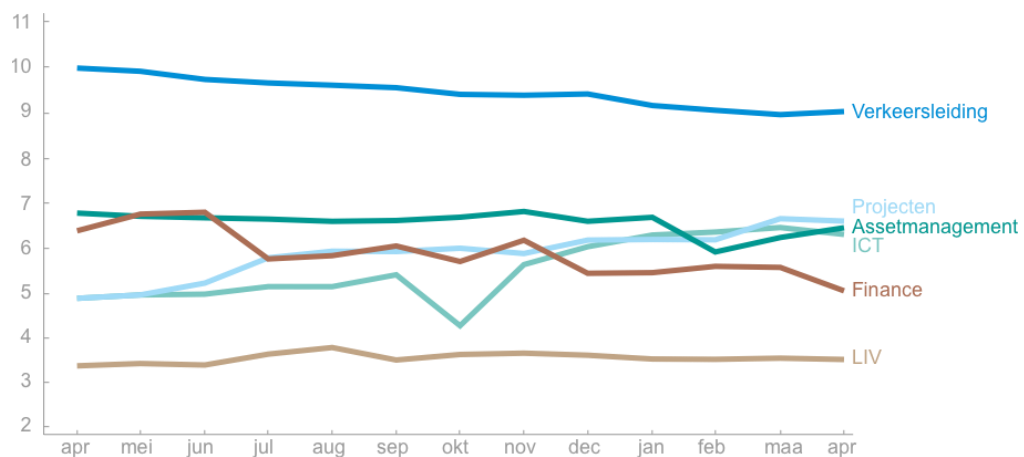
Figuur 73 Hier wordt een pijl gebruikt voor de trend



Figuur 74 Door gebruik te maken van een sparkline geef je veel meer context

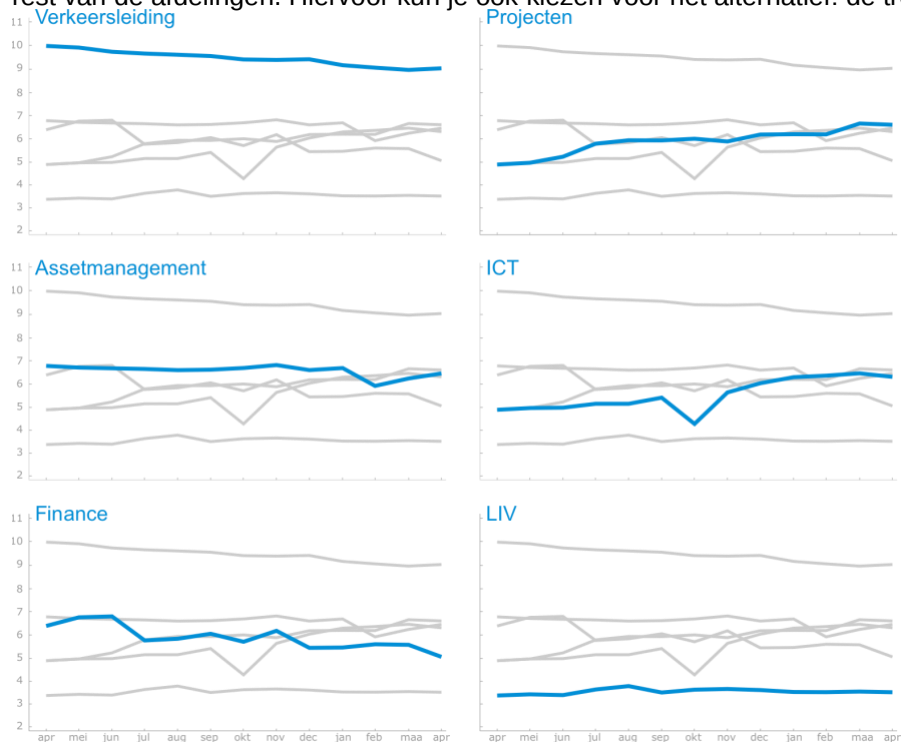
3.2.2.9 Trellis chart (ook small multiples genoemd)

Regelmatig verliest een goed ontworpen visualisatie haar kracht doordat er teveel data-series tegelijk in worden afgebeeld. Om de verschillende series te kunnen onderscheiden moet er bijvoorbeeld gebruik gemaakt worden van extra kleuren.



Figuur 75 Onduidelijkheid door het grote aantal data-series

Vaak is het doel van de het bovenstaande voorbeeld om één afdeling te vergelijken met de rest van de afdelingen. Hiervoor kun je ook kiezen voor het alternatief: de trellis chart.



Figuur 76 Trellis chart voorbeeld

Een trellis chart is een serie van gelijke diagrammen die allen dezelfde schaalverdeling hebben. In ieder individueel diagram wordt een andere lijn geaccentueerd. Zo kun je iedere serie vergelijken met de rest, zonder dat de visualisatie een onduidelijke brei wordt.

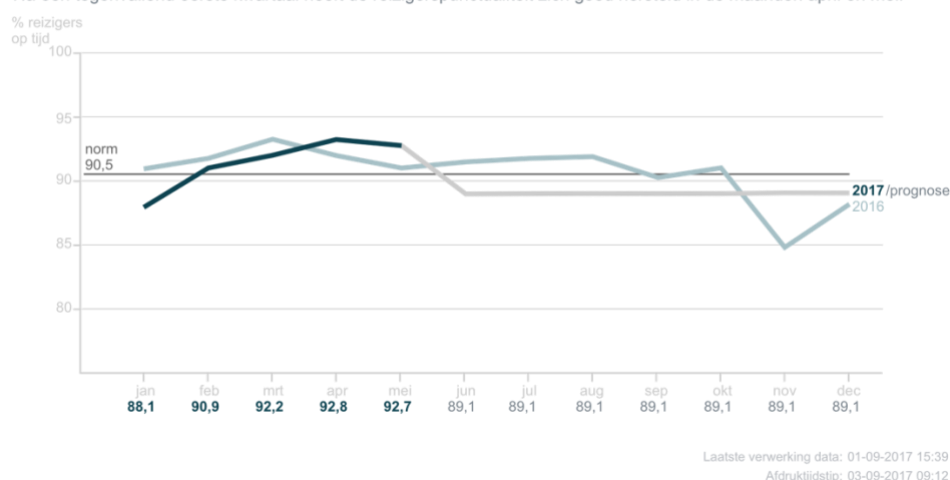
In het bovenstaande voorbeeld valt bijvoorbeeld op dat Team Bedrijfsbureau een continue stijging laat zien. Dit was in het voorbeeld van Figuur 75 een stuk lastiger om waar te nemen.

3.2.3 Ondersteunende elementen

In de voorgaande secties lag de nadruk op het presenteren van de inhoud middels een tabel of diagram. Echter, zonder context of verduidelijking kan de data niet goed tot zijn recht komen. Voor deze benodigde context en verduidelijking hebben we ondersteunende elementen nodig.

Reizigerspunctualiteit HRN (<5 min.)

Na een tegenvallend eerste kwartaal heeft de reizigerspunctualiteit zich goed hersteld in de maanden april en mei.



Figuur 77 Voorbeelden van ondersteunende elementen bij een schermvullend diagram

Het eerste ondersteunende element is misschien wel de meest opvallende: de titel en mogelijke sub-titel. Het belangrijkste doel van de titel (vaak in combinatie met de sub-titel) is het aangeven welk onderwerp er getoond wordt. Dit kan aangevuld worden met de belangrijkste vraag die beantwoord wordt. De titel plaats je dus prominent linksboven de visualisatie, vaak in een groter lettertype.

Vervolgens is het belangrijk dat de context van de data duidelijk wordt gemaakt. Meestal is dit het tonen van de gemaakte selecties (bijvoorbeeld welke afdeling) op de data en de bron. Bij de bron is het goed om ook de collectiedatum (en tijd) weer te geven, zodat er geen onduidelijkheid ontstaat bij de lezer hoe “vers” de data is.

In zowel de tabel als het diagram worden labels gebruikt om de betekenis van data te verduidelijken. Het beste is om deze labels zo dicht mogelijk bij de data te plaatsen. In tabellen kan het dus nodig zijn om kopteksten te herhalen als een tabel over meerdere pagina's gaat. In diagrammen is het beste om direct te labelen (zie Figuur 45), maar als dit niet kan gebruik te maken van een legenda. Deze legenda plaats je dan bij voorkeur net onder de titel van het diagram. Op een dashboard staan vaak meerdere diagrammen die een legenda delen. In die gevallen staat onder aan het dashboard de legenda die voor het gehele dashboard geldt.

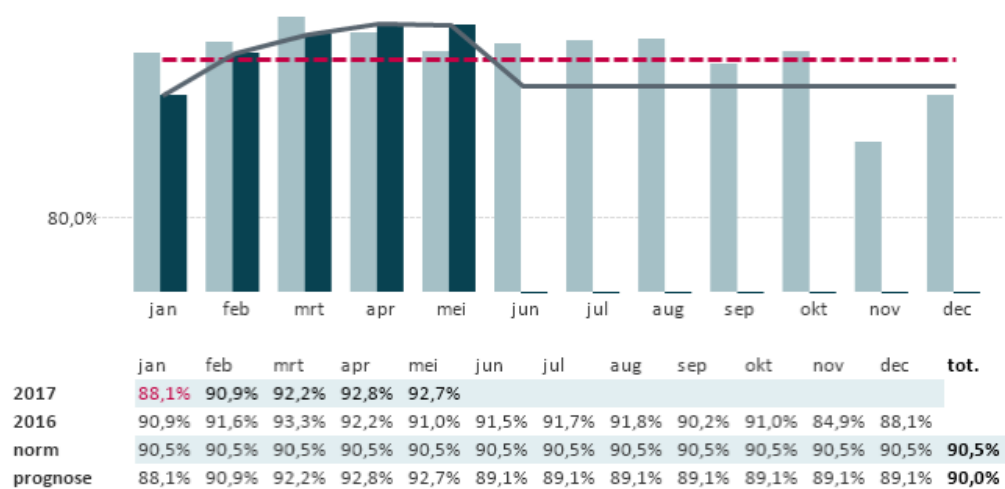
In diagrammen worden vaak rasterlijnen gebruikt om de lezer te helpen bij het bepalen van een waarde in het diagram. Doordat deze geen data vertegenwoordigen moeten deze zo subtiel mogelijk worden afgebeeld. Ze moeten net zichtbaar zijn, maar ook niet meer dan dat. Een goede manier is om hiervoor een lichte grijsstint voor te gebruiken. Let wel op dat lichtgrijs op sommige schermen (ook bij projectie) niet zichtbaar is.

4 Visualisatie voorbeelden

In de voorgaande hoofdstukken zijn de verschillende richtlijnen afzonderlijk benoemd. In dit hoofdstuk brengen we ze allemaal samen door een aantal bestaande visualisaties te herontwerpen. De commentaren en alternatieven houden echter geen rekening met de specifieke context waarbinnen deze datavisualisaties worden gebruikt. Ook zijn de voorbeeld vaak deel van een groter geheel die hier niet zichtbaar is. De uitgewerkte alternatieven zijn “visueel technisch” wel beter, maar daarmee is dus niet gezegd dat ze binnen de gebruikscontext ook beter zijn dan de huidige voorbeeld.

4.1 Reizigerspunctualiteit

Reizigerspunctualiteit HRN (<5 min.)



Figuur 78 Reizigerspunctualiteit (huidig)

Wat direct de aandacht trekt in dit voorbeeld is de rode stippellijn. Deze geeft de norm weer van 90,5%. Het is alleen niet duidelijk of deze norm geldt voor iedere maand of dat het alleen een norm is voor het jaartotaal.

Er is gekozen voor een staafdiagram. Deze helpt vooral als het de bedoeling is om de verschillende series (2017, 2016, norm & prognose) binnen een specifieke maand met elkaar te vergelijken. Inzicht in de ontwikkeling over de tijd van een of meerdere series wordt hierdoor wel minder duidelijk. Probeer maar eens de “vorm” van serie-2016 te bepalen. In de eerste maanden is het lastig om deze vorm te bepalen doordat de overige series er dwars doorheen lopen.

Verder missen er een aantal zaken die de leesbaarheid sterk zouden verbeteren:

- Er is geen legenda aanwezig;
De gebruikte kleuren worden structureel (in meerdere rapporten en dashboards) op deze manier gebruikt, dus voor de meeste gebruikers zal het duidelijk zijn dat de donkerblauwe kleur het huidige jaar is en de lichtblauwe kleur het voorgaande jaar. Maar hoe zit het dan met de donkere lijn? Dit blijkt de prognose lijn te zijn. Een expliciete legenda lijkt wellicht overbodig, maar het is raadzaam deze toch bij te voegen.

- De Schaalverdeling is beperkt tot één waarde en begint niet in 0;
Op de y-as staat alleen de waarde 80,0%. Het is echter niet goed af te lezen bij welke waarde de verschillende staven eindigen. Zonder de tabel is het niet mogelijk de waarden van de verschillende maanden te bepalen. Verder zijn de verschillen zo klein dat het niet aannemelijk is dat de y-as op 0 begint. Hierdoor zijn de staven niet vergelijkbaar.

Ook speelt de prognose een bijzondere rol in zowel de tabel als het diagram. Tot en met de maand mei toont deze serie niet de prognose, maar de realisatie van het huidige jaar (hier: 2017). Dit is onnodig en verwarrend.

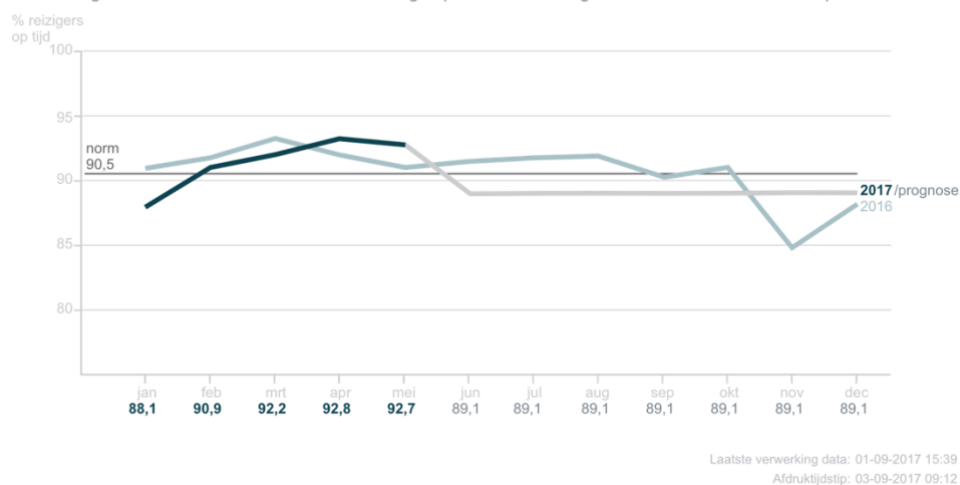
In de tabel vallen ook een paar zaken op. Zo heeft iedere waarde een %-teken, waardoor de waarden zelf minder prominent zichtbaar zijn. Wanneer alle waarden in een tabel dezelfde eenheid hebben (hier: percentage) is het beter om deze eenheid één keer aan te geven in een titel of bijschrift.

De kolomkoppen in de tabel zijn allen links uitgelijnd. In dit voorbeeld is dit (nog) geen probleem doordat alle waarden dezelfde lengte hebben (99,9%). Om te voorkomen dat er problemen ontstaan is het beter om de kolomkoppen uit te lijnen op de inhoud van de kolom. Doordat dit getallen zijn is rechts-uitlijnen een betere benadering.

Tenslotte zijn de tabel en het diagram niet uitgelijnd: de maanden in het diagram staan verder uit elkaar dan in de tabel.

Reizigerspunctualiteit HRN (<5 min.)

Na een tegenvallend eerste kwartaal heeft de reizigerspunctualiteit zich goed hersteld in de maanden april en mei.



Figuur 79 Reizigerspunctualiteit (alternatief)

In *Figuur 79* is een alternatief uitgewerkt voor dezelfde data. Er is gekozen voor een lijndiagram vanwege de volgende redenen:

- Het belangrijkste doel is de ontwikkeling over tijd van verschillende series. Hier leent een lijndiagram zich beter voor;
- De verschillen zijn relatief klein, waardoor een staafdiagram (die in 0 begint) niet gaat werken. Doordat de lezer in een lijndiagram niet direct de lengte vanaf de x-as gebruikt is het niet nodig om in 0 te beginnen. Probeer de schaalverdeling voor dit onderwerp

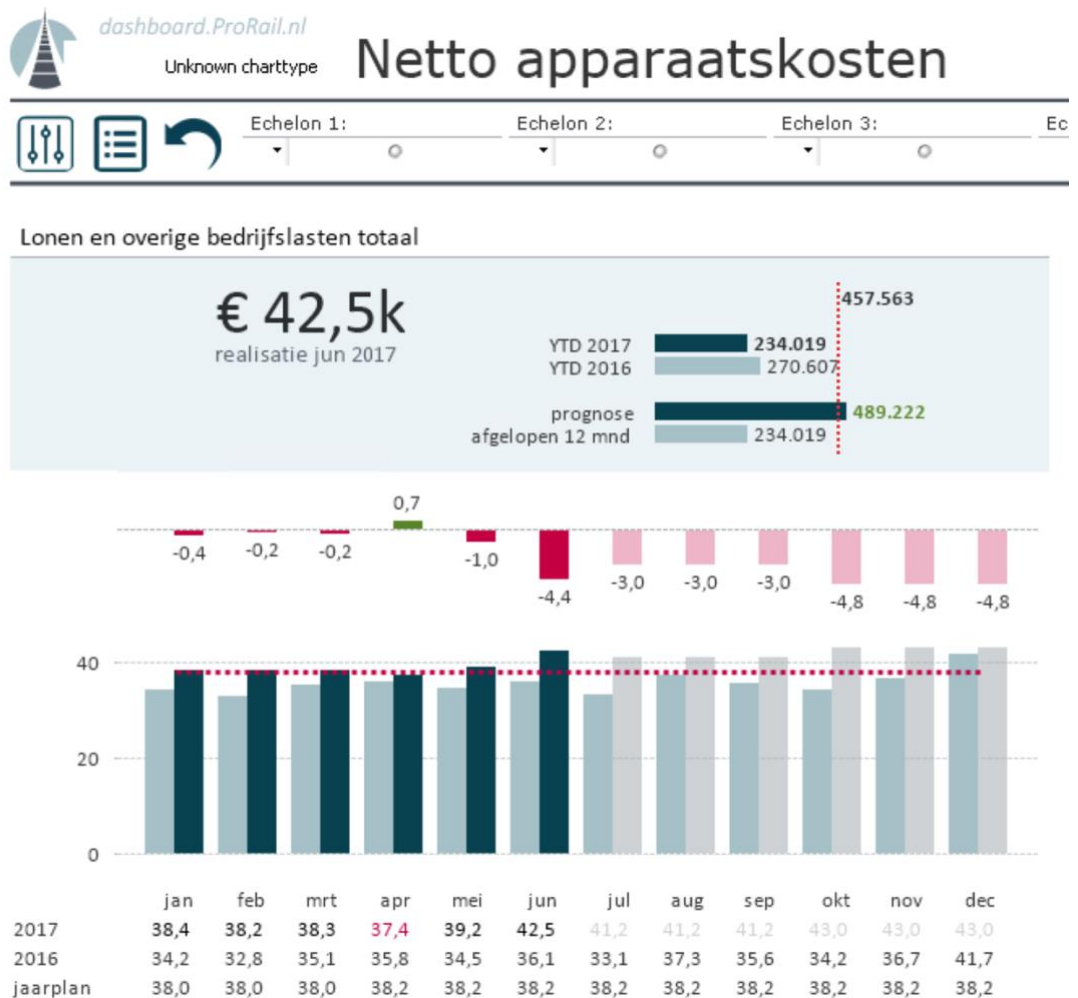
(reizigerspunctualiteit) altijd op dezelfde manier toe te passen (hier: tussen 80 en 100). Hierdoor zijn verschillen altijd hetzelfde en wordt het eenvoudiger om te interpreteren.

Verder is ervoor gekozen om alle series direct te labelen, zodat de lezer niet hoeft te zoeken naar de betekenis.

Doordat 2016 er in staat als referentiewaarde is deze nog verder naar de achtergrond gebracht door er een lichte kleur blauw aan te geven. Ook is de prognose alleen zichtbaar vanaf het moment dat er geen realisatie is. Doordat de prognose een verwachting betreft is deze afgebeeld in lichtgrijs.

Tenslotte zijn de belangrijkste waarden (t/m mei de realisatie en vanaf juni de prognose) voor mogelijke precisie toegevoegd aan de maanden onder aan het diagram.

4.2 Netto Apparaatskosten



Figuur 80 Netto Apparaatskosten (huidig)

Dit voorbeeld is slechts een deel van een dashboard, dus een aantal zaken (zoals de selectie en de legenda) zijn hier niet zichtbaar. Wat hier in het oog springt zijn de rode balken halverwege de visualisatie. Het lijkt dat hier het percentuele verschil tussen 2017 en het jaarplan wordt afgebeeld, maar dat wordt niet gelijk duidelijk gemaakt met bijvoorbeeld een titel.

Verder zien we boven aan deze datavisualisatie een grafische vorm die veelvuldig voorkomt in de ProRail dashboards. We zien daar een belangrijke numerieke waarde afgebeeld (de lonen en overige bedrijfslasten totaal). Daarnaast zien we de cumulatieve (YTD2017, YTD2016 en afgelopen 12 maanden) en een prognose waar we aan het einde van dit jaar verwachten uit te komen. De prognose bevat ook een doelstelling (457.563), maar deze loopt ook over de andere drie staven heen.

Doordat de meeste diagrammen geen schaalverdeling hebben is het lastig om te bepalen wat de omvang van de verschillende waarden is. De staafdiagram bovenaan begint zeker

niet in 0, maar vanwege het missen van een schaal is niet goed vast te stellen waar deze dan wel begint.

Doordat we hier kijken naar lasten valt te verwachten dat hogere lasten (tov jaarplan) niet wenselijk zijn. Alleen is de waarde in april 2017 in de onderste tabel rood, terwijl de waarde onder het jaarplan ligt. De groen/rode staven halverwege de visualisatie laten het tegenovergestelde zien: rood als de lasten hoger zijn en groen als ze lager zijn. Alleen is daar de richting van het verschil omgedraaid: hogere lasten dan het jaarplan worden als negatief verschil getoond.

Tenslotte is het niet duidelijk wat de eenheid is van de visualisatie. Een organisatie als ProRail met bijna 4.000 medewerkers zal zeer waarschijnlijk meer dan €42.500 aan loonkosten hebben per maand. Waarschijnlijk zijn de getoonde waarden duizendtallen, maar dit staat niet vermeld.

Samenvattend wordt er vrij veel ruimte gebruikt om allerlei context te geven aan de uitgaven in één maand, bijvoorbeeld door het lopende boekjaar te tonen. Als de uitgaven van deze maand het belangrijkste gegeven is, dan zou de hoeveelheid ruimte anders ingedeeld kunnen worden. Zeker als we weten dat er nog veel meer kostencategorieën op dit dashboard staan die wellicht in één keer zouden willen beoordelen.

Netto apparaatskosten voor juni 2017



Figuur 81 Netto Apparaatskosten (alternatief)

In Figuur 81 is een alternatief geschetst. Er is hier voor gekozen om de meetwaarde van de huidige maand de belangrijkste rol en daarmee ruimte te geven.

Van links naar rechts zien we op de eerste plaats een sparkline die de ontwikkeling van de meetwaarde over de laatste 12 maanden toont. Vervolgens het label van de meetwaarde direct gevolgd door de realisatie waarde. Tevens is de waarde gevisualiseerd in een dotplot, waarbij de donkerblauwe punt de huidige waarde weergeeft en de lichtblauwe punt die van dezelfde periode vorig jaar.

De lonen en de opbrengsten passen niet op één schaal door het grote verschil in omvang. Echter, door de relatieve afstand tussen de laagste en de hoogste waarde op de schalen gelijk te stellen kunnen we ze toch goed vergelijken. We zien hier dan ook dat de lasten wel gestegen zijn, maar (relatief) veel minder hard dan dat de opbrengsten gedaald zijn.

Tenslotte zijn de afwijkingen getoond ten opzichte van het jaarplan en kunnen we zien dat die verschillen op elkaar lijken, alleen de richting van het verschil is anders.

Dit voorbeeld laat zien dat in veel minder ruimte veel meer informatie aan de lezer kan worden aangeboden.

4.3 Exploitatierekening

Zoals eerder aangeven is de tabel ook een vorm van datavisualisatie. Alleen betekent het niet dat we bij het gebruiken van een tabel alleen nog maar cijfers en labels kunnen tonen. Het volgende voorbeeld is hiervoor illustratief.

Exploitatierekening (x1.000 €)

	realisatie jun	prognose jun	delta
Rijksbijdragen	75.955	68.359	-7.596
Gebruiksvergoeding	27.488	27.662	174
Bijdragen	103.443	96.021	-7.421
Overige bedrijfsopbrengsten	7.185	3.019	-4.167
Opbrengst uren totaal	4.495	9.856	5.361
Diverse bedrijfsopbrengsten	11.680	12.874	1.194
Opbrengsten	115.123	108.896	-6.227
Grootschalig onderhoud	11.011	8.161	2.850
Kleinschalig onderhoud	25.061	24.252	809
Transfer	10.421	5.486	4.935
Beheer en Calamiteitenorganisatie	15.275	10.896	4.379
Overige kosten uitbesteed werk	506	1.532	-1.026
Uitbesteed werk	62.274	50.326	11.948
Lonen & soc.lasten	30.005	26.935	3.070
Inhuur personeel	4.417	3.604	812
Overige personeelskosten	2.404	2.253	151
Huisvesting en werkplekkosten	4.022	3.739	283
Adviseurs / Externe dienstverlening	1.045	2.484	-1.440
Overige Kosten	623	496	127
Lonen en overige bedrijfslasten	42.515	39.512	3.003
Afschrijvingskosten	10.751	10.083	668
Financiële baten en lasten	1.283	1.162	121
Bijzondere kosten	-1.700	17	-1.717
Overig	10.334	11.262	-928
Kosten	115.123	101.100	14.023
Resultaat	-0	7.796	7.796

Figuur 82 Exploitatierekening (huidig)

Deze tabel bevat niet buitensporig veel cellen, maar om door te krijgen wat er aan de hand is rest ons niet anders dan het lezen van alle waarden. Dit geeft ons wel zeer precieze gegevens, maar is jammer genoeg een traag proces.

Een veel voorkomende vraag bij het zien van zo'n tabel is: wat zijn de relatief grote posten? Het antwoord is wel te achterhalen, maar zoals eerder opgemerkt kost dit vrij veel tijd.

In het volgende alternatief is de tabel uitgebreid met staafdiagrammen. De staven geven de relatieve omvang van het getal weer tov het totaal. Hiermee is het binnen één kolom eenvoudig om vast te stellen waar de grote en kleine posten zich bevinden.

delta > 20% rekening	realisatie	prognose	delta	
Rijksbijdragen	75,955	68,359	7,596	
Gebruiksvergoeding	27,488	27,662	-174	
Bijdragen	103,443	96,021	7,422	
Overige bedrijfsopbrengsten	7,185	3,019	4,166	
• Opbrengst uren totaal	4,495	9,856	-5,361	
Diverse bedrijfsopbrengsten	11,680	12,875	-1,195	
Opbrengsten	115,123	108,896	6,227	
• Grootschalig onderhoud	11,011	8,161	2,850	
Kleinschalig onderhoud	25,061	24,252	809	
• Transfer	10,421	5,486	4,935	
• Beheer en Calimiteitenorganisatie	15,275	10,896	4,379	
Overige kosten uitbesteed werk	506	1,532	-1,026	
• Uitbesteed werk	62,274	50,326	11,948	
Lonen & soc.lasten	30,005	26,935	3,070	
• Inhuur personeel	4,417	3,604	813	
Overige personeelskosten	2,404	2,253	151	
Huisvesting en werkplekkosten	4,022	3,739	283	
Adviseurs / Externe dienstverlening	1,045	2,484	-1,440	
• Overige kosten	623	496	127	
Lonen en overige bedrijfslasten	42,515	39,512	3,003	
Afschrijvingskosten	10,751	10,083	668	
Financiële baten en lasten	1,283	1,162	121	
Bijzondere kosten	-1,700	17	-1,717	
Overig	10,334	11,262	-928	
Kosten	115,123	101,100	14,023	
• Resultaat	0	7,796	-7,796	

Figuur 83 Exploitatierekening (alternatief)

Verder is in dit voorbeeld gekozen om de achtergrondkleur van de regels veel lichter te maken, zodat deze niet belangrijker wordt dan de gegevens zelf.

Het ziet er naar uit dat de relatieve verdeling tussen de posten in de realisatie ongeveer gelijk is als die in de prognose, want in het algemeen zijn de staven dezelfde lengte.

Om de lezer nog verder te helpen om sterke afwijkingen te herkennen is er tevens een indicator toegevoegd (rood bolletje in de linker kantlijn). Overal waar deze staat vermeld is er meer dan 20% afwijking tussen de realisatie en de prognose.

5 Bijlage 1: Design Reference Card



Data visualization design

Steps to take into account during the design of data visualizations

This reference card gives an overview of the main topics required to take into account when designing a data visualisation.

1. Objective

- What is the message/objective of the visualization? (incl. Explain, Explore or Exhibit)
- Who will use the visualization? (single person/group, different groups, etc.)
- How will they use this visualization? What questions will they use it to answer? (WHAT, WHY and/or NOW WHAT) What actions will they take in response to these answers?
- What are useful comparisons that will allow the users to see the data in a meaningful context? (targets, historical data, other groups, etc.)
- In what kind of form will the visualization be delivered? (paper, PDF, (interactive) website, excel, PowerPoint, etc.)

2. Data

- What data should be displayed on the visualization? (data items & summary/detail level)
- How often should the data be refreshed?
- Which data items are key in responding the questions?
- How can the data items be grouped?

3. Visual design

- Design overall arrangement of data areas;
- Decide on the support for the principal functions: scan, zoom & detail;
- Decide on table and/or graph types;
- Determine best means to encode values (diagram-selection matrix);
- Determine where to display each data item;
- Determine the best design for the remaining objects (scales, tick marks, narrative, legends, grid lines, etc.);
- Determine if particular data should be featured and how;
- Determine a color scheme.

Nova Silva B.V. Burensedijk 15 4011 KG Zoelen The Netherlands IBAN:NL91 RABO 0128 3688 96 BIC:RABONL2U
Company reg:11070282 VAT:NL8184.31.647.B01 www.novasilva.com info@novasilva.com t +31(0)344 579 120 f +31(0)344 579 121

6 Bijlage 2: Diagram Selectie Matrix



Diagram-selectie-matrix

Relatietype	Objecten om waarden te coderen			
	Punten	Lijnen	Staven	Boxen
	 Spreidingsdiagram Puntendiagram Strip plot	 Lijndiagram Lijndiagram met punten	 Staafdiagram (verticaal) Staafdiagram (horizontaal)	 Box plot (verticaal) Box plot (horizontaal)
Tijdreeksen Waarden geven een verandering in tijd weer (jaarlijks, maandelijks, etc.)	Ja (met een <i>puntendiagram</i> , wanneer er geen waarde is voor ieder tijdsinterval)	Ja (voor het zichtbaar maken van trends en patronen en deze onderling te kunnen vergelijken)	Ja (alleen verticale staven, om nadruk te leggen op individuele waarden en deze individuele waarden onderling te kunnen vergelijken)	Ja (alleen verticale boxen, om weer te geven hoe de verdeling verandert in tijd)
Rangorde Waarden worden geordend naar omvang (oplopend of aflopend)	Ja (met een <i>puntendiagram</i> , met name als de kwantitatieve schaal niet begint bij nul)	Nee	Ja	Ja (voor het weergeven van een op- of aflopende set van verdelingen)
Deel van groter geheel Waarden geven de onderlinge verhoudingen (proporties) weer binnen een geheel (bijvoorbeeld: per regio de verkoop uitgedrukt t.o.v. van de totale verkoop)	Nee	Ja (voor het weergeven van hoe de onderlinge verhoudingen in de tijd veranderd zijn)	Ja	Nee
Afwijking Het verschil tussen twee verzamelingen van waarden (bijvoorbeeld: de variantie tussen werkelijke en verwachte uitgaven)	Ja (met een <i>puntendiagram</i> , met name als de kwantitatieve schaal niet begint bij nul)	Ja (wanneer het in combinatie met een tijdreeks is)	Ja	Nee
Verdeling Frequentie van waarden per interval van laag naar hoog (bijvoorbeeld: frequentie van leeftijdsgroepen met een 10 jaar interval)	Ja (met een <i>strip plot</i> , om individuele waarden af te beelden)	Ja (met een <i>frequentie polygoon</i> , om de vorm van de verdeling weer te geven)	Ja	Ja (bij het vergelijken van verschillende verdelingen)
Correlatie Vergelijken van twee reeksen van waarden (bijvoorbeeld: de hoogte en het gewicht van een groep mensen) om te bepalen of er een relatie bestaat tussen de reeksen	Ja (met een <i>spreidingsdiagram</i>)	Nee	Ja (met een <i>table lens</i> , met name als je publiek niet bekend is met een spreidingsdiagram)	Nee
Geospatial Waarden worden afgebeeld op een kaart om de locatie weer te geven	Ja (met cirkels van verschillende omvang op een kaart)	Ja (om een route op een kaart weer te geven)	Nee	Nee
Nominale vergelijking Eenvoudig vergelijken van ongeordende variabelen (bijvoorbeeld: producten of regio's)	Ja (met een <i>puntendiagram</i> , met name als de kwantitatieve schaal niet begint bij nul)	Nee	Ja	Nee

www.PerceptualEdge.com

Afgeleid van het boek
Show Me the Numbers

© Stephen Few 2004-2012

Vertaald door Nova Silva

7 Bijlage 3: Stijlgids samenvatting

Stijlgids Datavisualisatie samenvatting

Meestgebruikte kleuren

 pr-primary-blue RGB 000/143/215 Primaire data kleur	 pr-secondary-blue-4 RGB 159/218/247 Primaire lichte data kleur	 ProRail Logo Rood RGB 191/018/056 Kleur van logo
 pr-secondary-blue-1 RGB 000/151/144 Secondaire data kleur	 pr-secondary-blue-2 RGB 121/198/192 Secondaire lichte data kleur	 Rood RGB 230/000/004 Signaalkleur
 pr-secondary-brown-1 RGB 171/110/085 Tertiaire data kleur	 pr-secondary-brown-2 RGB 192/164/133 Tertiaire lichte data kleur	 RoodLicht1 RGB 253/071/077 Kwantitatieve indicatie kleur

	RoodLicht2 RGB 253/117/121 Kwantitatieve indicatie kleur
	RoodLicht3 RGB 254/163/166 Kwantitatieve indicatie kleur
	RoodLicht4 RGB 254/209/210 Kwantitatieve indicatie kleur

Lettertype

Primaire titel (Arial 20px Bold)
Caption-grafiek (Arial 16 px)
Algemene tekst (Arial 14px)
Secondaire titel (Arial 12px Bold)
Tertiaire titel (Arial 10px Bold)
AS-waarden (Arial 12px)
Kleine tekst (Arial 9px)

Tips

Tabel

- Onderscheiden rijen/kolommen: begin met witruimte, vervolgens met (subtiel) achtergrondkleur en als dat niet werkt een subtiel gekleurde lijn;
- Accentueren van bijzonderheden: achtergrondkleur, tekstdikte of tekstkleur. Gebruik tot een minimum beperken;
- Alle numerieke waarden hebben een gelijk aantal decimalen en worden rechts uitgelijnd.
- Alle teksten worden links uitgelijnd. Enige uitzondering is een tekstuele kolomkop van een numerieke kolom: de kolomkop wordt op dezelfde manier uitgelijnd als de inhoud van de kolom, dus rechts.
- Noem eenheden (zoals €, %, etc.) zoveel mogelijk in de kop en niet afzonderlijk in iedere cel;
- Om patronen te herkennen kun je een grafische ondersteuning toevoegen (hier in de vorm van staven).

Lijndiagram

- Gebruik niet meer dan 4 series (lees: lijnen) in één diagram;
- Beperk je tot één lijnstijl, het liefste een niet-onderbroken lijn;
- Probeer lijnen direct te labelen, dus geen legenda;
- Maak geen gebruik van "smoothed lines";
- Gebruik bij een numerieke schaalverdeling zo veel mogelijk een natuurlijke stapgrootte: net zoals je zou tellen (1,2,3 of 2,4,6 of 5,10,15 of 25,50,75 etc.);
- Maak geen gebruik van meerdere y-assen. Beter om meerdere losse diagrammen te maken.

Staafdiagram

- De numerieke schaal begint altijd in 0 zonder onderbreking;
- Zorg dat alle labels horizontaal staan;
- Bij lange staaf-labels: gebruik horizontale staafdiagram;
- Gebruik alleen verschillende kleuren als je niet anders kan;
- Gebruik bij een numerieke schaalverdeling zo veel mogelijk een natuurlijke stapgrootte: net zoals je zou tellen (1,2,3 of 2,4,6 of 5,10,15 of 25,50,75 etc.);
- Maak geen gebruik van meerdere y-assen. Beter om meerdere losse diagrammen te maken.

Taartdiagram

- Bij voorkeur niet gebruiken
- Als het niet anders kan dan alleen voor het vergelijken van verdelingen binnen één taartdiagram, niet tussen verschillende taartdiagrammen;
- Maximaal 4 categorieën (taartpunten);
- De som van de taartpunten is altijd 100%;
- Begin op 12-uur met de grootste punt en ga met de klok mee van groot naar klein;
- Label de taartpunten direct en gebruik één kleur (eventueel een tweede kleur voor bijzondere betekenis);

KPI meter (op dit moment nog niet mogelijk in Prorail dashboard)

- Effectief om meerdere KPI's te toetsen aan hun doelstellingen en deze te vergelijken;
- Alleen een staaf gebruiken voor de waarde als de schaal op 0 begint. Gebruik anders een punt;
- Markeer de meetwaarden die aandacht behoeven met een losse indicator (hier rode punt). Markeer niet de staaf/punt.

Sparkline

- Goed alternatief voor de "pijl" die aangeeft of een waarde is toegenomen of afgenomen;
- Kan zowel als lijn- en als staafdiagram toegepast worden;
- Geeft globaal inzicht in de ontwikkeling over tijd van een meetwaarde. Het gaat niet om de details;

Relaties

Opzoeken en vergelijken van individuele waarden waarbij precisie een vereiste is.

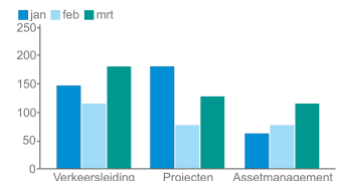
Visualisatie voorbeeld

	Regio A	Regio B
produkt 1	2.267	265
produkt 2	3.365	231
produkt 3	1.554	354
produkt 4	1.357	942
produkt 5	2.987	153
produkt 6	2.582	756
totaal	14.112	2.701

Leent zich bij uitstek goed voor tijdreeksen. Zeker als de trend het belangrijkste doel is



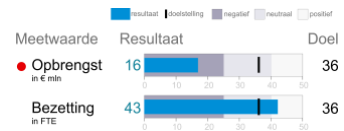
Vooraf sterk in nominaal vergelijk. Bij tijdreeksen helpt het individuele waarden te vergelijken.



Alleen gebruiken bij deel van groter geheel. Vaak is een staafdiagram een betere keuze.



Vooraf veel gebruikt om een mogelijke afwijking tov een doelstelling weer te geven.



Wordt uitsluitend gebruikt in tijdreeksen.

