

Toolbox academische vaardigheden: Data analyse en grafieken maken

Het belang van data-analyse

Eén van de meest belangrijke aspecten tijdens het uitvoeren van een onderzoek is het omzetten van de verzamelde data naar bruikbare informatie. Data-analyse kan ons helpen om de oorspronkelijke hypothese van het onderzoek te toetsen, en ook de resultaten van de verzamelde data uit te leggen. Eén van de meest interessante middelen is grafische weergave, aangezien het ons de gelegenheid geeft om data om te zetten in makkelijk te begrijpen informatie. De grafieken worden gebruikt om een dataset te illustreren en presenteren, om zo het begrijpen, vergelijken en analyseren ervan te vergemakkelijken. De grafieken zijn afhankelijk van het type variabelen dat wordt weergegeven.

Type variabelen

Het type grafiek dat gebruikt kan worden, hangt af van het type variabele. De verschillende typen variabelen zijn:

- **Categorische Variabelen (of kwalitatieve variabelen):** Variabelen die een eigenschap weergeven van de onderzochte individuen. Bijvoorbeeld: Gebied (Noord-Zuid), Soort (Amfibie, Vis, Vogel...), Dagdeel (Ochtend, Middag, Avond), ...
- **Numerieke Variabelen (of kwantitatieve variabelen):** Variabelen die de numerieke waarde van de onderzochte individuen weergeven. Bijvoorbeeld: pH, Zuurstofverbruik, Geleidbaarheid, Aantal vogels, Aantal vissen, ...

Sommige variabelen kunnen omgezet worden om nuttige informatie te bieden voor het interpreteren van de data:

- **Data en uren:** De variabelen data en uren kunnen omgezet worden naar eigenschappen die er mee te maken hebben. Bijvoorbeeld: Monster 1 is genomen op 2-12-2020, om 10:40, kan omgezet worden naar verschillende variabelen zoals: Dag van de Week: Maandag, Maand: December, Jaar: 2020, Seizoen: Winter, Dagdeel: Ochtend, ...
- **Georeferenties:** georeferentie punten kunnen omgezet worden naar eigenschappen zoals naam van een gebied (bv Lauwersmeer), positie (bv Noord-Zuid), ...
- **Numerieke variabelen in categorieën:** sommige numerieke variabelen kunnen interessant zijn om weer te geven als eigenschap. Bijvoorbeeld: de pH gemeten in zuur-neutraal-basisch, de hoeveelheid vissen in weinig-voldoende-veel, ...

Data structuur

Eén van de hoofdtaken is om de data op een juiste manier te structureren. De eenvoudigste manier * is om de onderzochte eenheden in de rijen weer te geven (soort, monster, mensen, ...) en om de variabelen in de kolommen weer te geven.

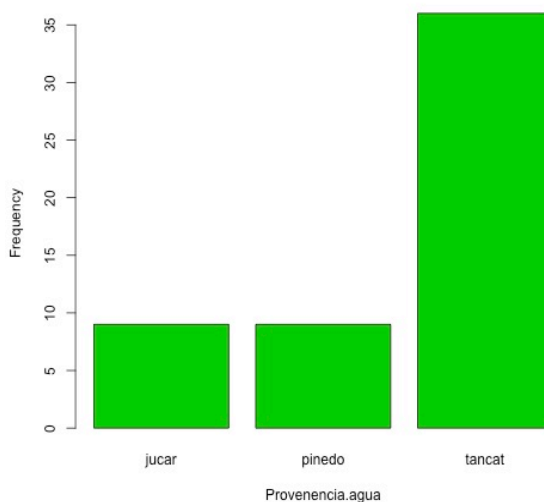
*dit format is correct voor het gebruik van statistische pakketten. Als er gebruik gemaakt wordt van spreadsheets, dan moeten er berekeningen gedaan worden met de data om bepaalde grafieken te verkrijgen.

Individuos	Ubicación	Lugar	Hora	Franja	pH	Oxígeno disuelto	Conductividad	Temperatura
muestra 1	Entrada	Entrada	10:45	Mañana	7,92	117,6	1779	20,4
muestra 2	Centro	Entrada	11:01	Mañana	7,8	117,4	1712	21
muestra 3	Tancat de Mila	Entrada	11:28	Mañana	7,79	114,8	1703	18,3
muestra 4	Oeste	Entrada	12:03	Mediodía	7,78	115	1714	20
muestra 5	G.Perellonet- Salida	Salida	12:44	Mediodía	7,75	115,4	1641	18
muestra 6	G.Puçol- Salida	Salida	13:18	Mediodía	7,86	127,6	1640	18,5
muestra 7	Entrada	Entrada	10:45	Mañana	8,2	116,9	1750	14,5
muestra 8	Centro	Entrada	11:01	Mañana	8,5	117,7	1540	17,5
muestra 9	Tancat de Mila	Entrada	11:28	Mañana	8,5	115,3	1610	17,6
muestra 10	Oeste	Entrada	12:03	Mediodía	8,6	115,8	1560	17,7
muestra 11	G.Perellonet-Salida	Salida	12:44	Mediodía	8,4	115,2	1520	17,6
muestra 12	G.Puçol-Salida	Salida	13:18	Mediodía	8,5	126,9	1520	17,5

Grafieken voor categorische variabelen

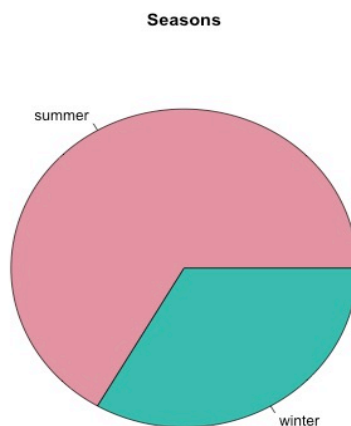
- Staafdiagram

Het aantal elementen of het percentage van elke categorie kan weergegeven worden met het staafdiagram.



- **Cirkeldiagram**

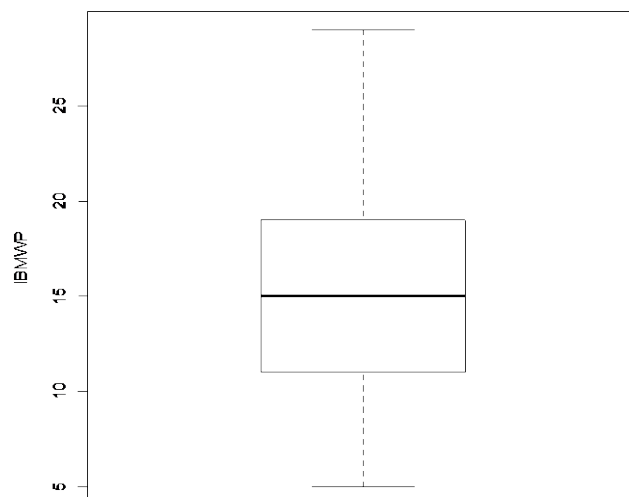
Het cirkeldiagram lijkt op een staafdiagram maar gebruikt een tweedimensionaal coördinaten systeem.



Grafieken voor numerieke variabelen

- **Box-plot**

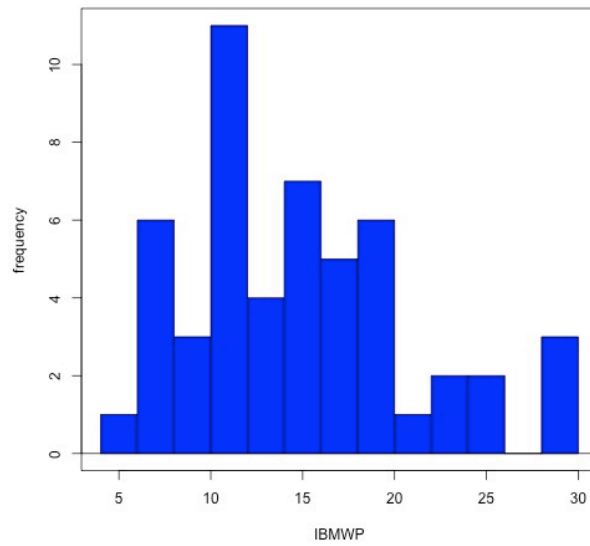
De box-plot is een methode om een grafische weergave te maken van groepen of numerieke data met bijbehorende kwartielen en extreme waarden.





- **Histogram**

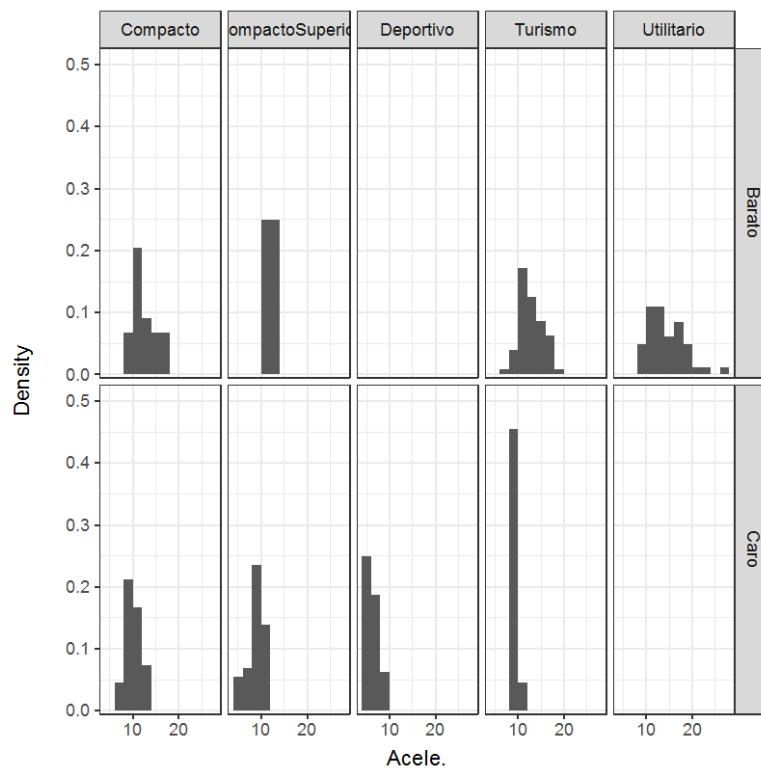
Een histogram is een precieze weergave van de verdeling van numerieke data.



Grafieken met meerdere variabelen

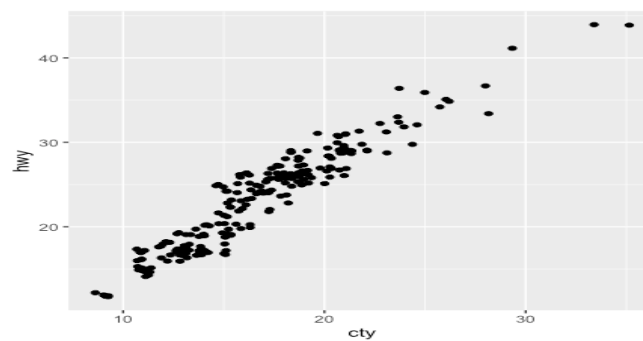
- **Kwantitatieve variabelen vs. kwalitatieve variabelen**

In dit geval kunnen de grafieken voor kwantitatieve variabelen (histogram, staafdiagram) gemaakt worden terwijl er rekening gehouden wordt met verschillende factoren van een kwalitatieve variabele.



- **Kwantitatieve vs. kwantitatieve variabele**

Het spreidingsdiagram is het meest geschikt om twee kwalitatieve variabelen te analyseren en om de relatie tussen beide te bestuderen.



- **Kwantitatieve variabelen vs. tijdsvariabelen**

Tijdsseries worden gebruikt om de verandering van een variabele over een bepaalde tijd te analyseren.

