

H3 DATA GEBRUIKEN

3.1 Inleiding

In hoofdstuk 1 van deze module lag de nadruk op het basisbegrip rond data en de structuur waarmee je data kunt opslaan. In hoofdstuk 2 hebben we ingezoomd op relationele databases en SQL geleerd om databases op te zetten, te vullen en vooral te bevragen.

In dit hoofdstuk gaan we kijken naar een aantal veelgebruikte toepassingen van data. Dat begint met een verkenning de webapplicatie phpMyAdmin wat één van de populairste tools voor databasemanagement is. Vervolgens kijken we hoe we publieke data kunnen gebruiken en combineren om tot nieuwe informatie en kennis te komen. Een belangrijk aspect daarbij is de presentatie van data door middel van datavisualisatie. Als laatste kijken we naar websites die (dynamisch) gebruik maken van databases met behulp van PHP.

3.2 phpMyAdmin

In deze module heb je kennism gemaakt met verschillende soorten databases. Er zijn speciale applicaties waarin deze databases worden gemaakt en beheerd. Zo'n applicatie heet een **databasemanagementsysteem** (DBMS). De meest gebruikte SQL-databasevariant MySQL wordt vaak beheerd via de webapplicatie **phpMyAdmin**. Dat geldt ook voor de werkomgeving bij deze module.

Deze paragraaf gaat grotendeels over het ontdekken van en leren gebruiken van phpMyAdmin. Het draait hier om vaardigheid en minder om kennis. Je leert de stof door het maken van opdrachten.



FIGUUR 3.1

Opdracht 1 kennismaken met phpMyAdmin

1. Open phpMyAdmin via het startscherm van de databaseomgeving van de module.
2. Links zie je een overzicht van databases. Klik op *mobiel*. Wat zie je nu op het middenscherf?
3. Klik op de tabel(-naam) *vrienden*. Als het goed is krijg je nu het beeld uit figuur 3.2.

Merk op dat bovenaan het centrale scherm de bijbehorende SQL-instructie (`SELECT * FROM `vrienden``) is te lezen.

4. Bij elk record kun je kiezen voor *Wijzigen*. Pas de naam van *Alan* aan naar je eigen naam. Welk SQL-commando verschijnt als je jouw wijziging hebt doorgevoerd?
5. Ga naar de querier en stel vast dat jouw wijzigingen daar zichtbaar zijn (met de juiste query).
6. Ga terug naar phpMyAdmin en klik op *Structuur*. Wat zie je?
7. Klik in het menu bovenaan op *Ontwerper*. (Deze optie kan verborgen staan onder *Meer*.) Hoe worden relaties en sleutels weergegeven?
8. Ga naar de startpagina van de module en kies voor *herstel standaard databases*. Controleer vervolgens in phpMyAdmin dat jouw eigen voornaam weer is vervangen door *Alan*.

	id	naam	plaats
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	1	Alan	Groningen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	2	Bob	Assen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	3	Christel	Heerlen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	4	Daphne	Groningen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	5	Eve	Heerlen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	6	Frits	Delft
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	7	Gonny	Groningen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	8	Hajar	Emmen
<input type="checkbox"/> Wijzigen <input type="checkbox"/> Kopiëren <input type="checkbox"/> Verwijderen	9	Ingo	Assen

FIGUUR 3.2



Opdracht 2 zelf databases maken en vullen

In deze opdracht maak je *H3O02 computerpracticum database maken in MySQL*. Het gelijknamige bestand is te vinden via *bestanden module* in het startscherm van onze werkomgeving.

3.3 Data science en datavisualisatie

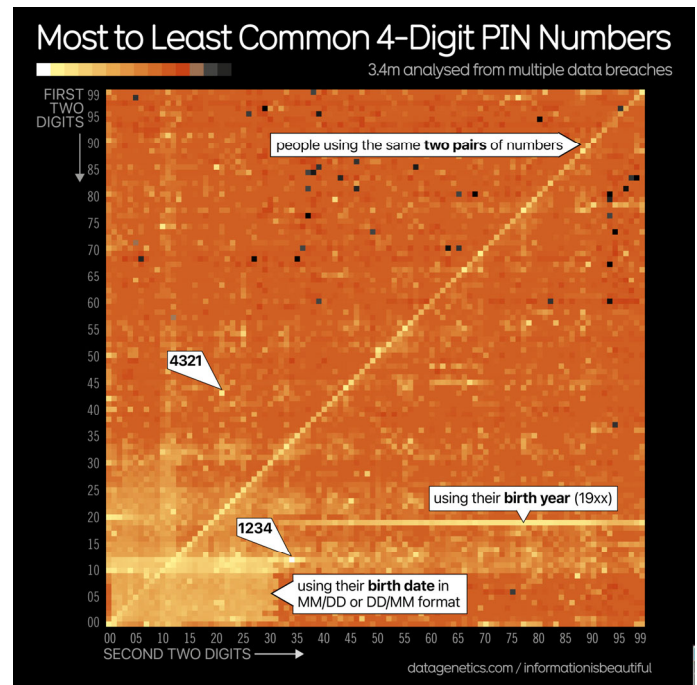
Data science (of datawetenschap) is een multidisciplinair vakgebied dat zich richt op het verkrijgen van inzichten uit data. Datawetenschappers gebruiken verschillende vaardigheden en kennis om waardevolle verbanden te leggen en nieuwe informatie en kennis te destilleren uit grote datasets. Dit kan allerlei doelen hebben: het verbeteren van de besluitvorming, het ontdekken van trends of het identificeren van patronen. Op basis van een data-analyse kunnen dan weer voorspellingen worden gedaan of problemen worden opgelost.

Dit analytische vakgebied is een mix van vakken als wiskunde, statistiek en informatica.

Datawetenschappers maken, behalve van SQL, ook vaak gebruik van programmeertalen en artificiële intelligentie. Je spreekt van **datamining** (gegevensdelving) als het gaat om de patroonherkenning in grote datasets.

Een essentieel onderdeel van *data science* is **datavisualisatie**, wat helpt bij het begrijpelijk maken van bevindingen en inzichten. Veel mensen hebben moeite met de juiste interpretatie van grote hoeveelheden data, wanneer die in een tabel worden gepresenteerd. Door dezelfde data te vertalen naar een beeld – de visualisatie van de data (zoals in figuur 3.3) – is het mogelijk om het achterliggende inzicht te communiceren naar anderen.

Visualiseren kan op talloze manieren. Standaardvoorbeelden zijn grafieken en diagrammen. Vaak worden technieken gecombineerd en zit de kracht van de visualisatie in de eenvoud en de creativiteit waarmee het is gemaakt.



FIGUUR 3.3

Opdracht 3 data science

9. In het vakgebied *data science* wordt de afkorting KDD gebruikt. Gebruik het internet om uit te zoeken waar deze afkorting voor staat en wat de achterliggende betekenis is.
10. In de theorie wordt kort gesproken over datamining. Gebruik AI om tot een uitgebreidere definitie van *datamining* te komen. Zorg er daarbij voor dat in ieder geval wordt benoemd uit welke stappen het proces van datamining doorgaans bestaat.
11. Bedenk drie voorbeelden waar *data science* heeft geholpen of zou kunnen helpen bij het oplossen van een wereldprobleem.

Opdracht 4 eerste blik op visualisaties I

In deze opdracht bekijken we de datavisualisatie in figuur 3.3. Een vergrote versie is te openen via onze werkomgeving (onder *bestanden module* → *hoofdstuk 3*).

12. Wat laat deze visualisatie zien? En was dit je snel duidelijk, of had je daar meer tijd of hulp bij nodig?
13. Wat is de betekenis van de kleurenencoding in deze visualisatie? Vind je dit duidelijk uitgelegd?
14. Geef een voorbeeld van een pincode die statistisch gezien weinig voorkomt.
15. Door de visualisatie worden nieuwe patronen zichtbaar. Benoem er minimaal drie.
16. Bedenk twee manieren waarop deze visualisatie nog sterker of duidelijker zou kunnen zijn.

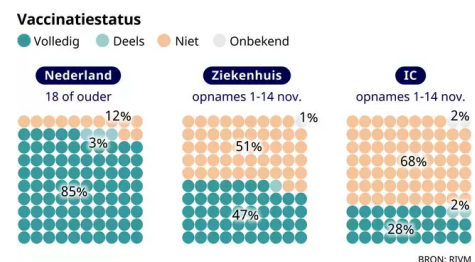
Opdracht 5 misleidende data

Bekijk de video waarin wordt uitgelegd hoe statistiek misleidend kan zijn.

17. In de video wordt gesproken over *Simpsons Paradox*. Beschrijf deze paradox in je eigen woorden.
18. Download het bestand *Opdracht 5 Simpsons Paradox.xlsx* en open dit met Excel.
19. Bestudeer de data (en de metadata). Probeer op basis van logisch redeneren de overige velden van de twee tabellen aan te vullen. LET OP Laat Excel het rekenwerk doen. Vraag eventueel je leraar om hulp, als je nog niet zoveel ervaring met Excel hebt.
20. Hoeveel procent scoort het lokale ziekenhuis beter over alle operaties ten opzichte van het academische ziekenhuis?
21. Klopt het dat het academische ziekenhuis niettemin beter scoort bij zowel de lichte als de zware operaties?
22. Ontwerp (op papier of met Excel of PowerPoint) een eenvoudige datavisualisatie die de conclusies van de vorige twee vragen helder overbrengt door gebruik van beeld.

Opdracht 6 visualiseren van de coronacrisis

Tijdens de coronacrisis werd datavisualisatie ingezet om de bevolking van landen te informeren op basis van grote en complexe datasets. Omdat er ook veel desinformatie werd verspreid, werd veel energie gestoken in het ontkrachten van de desinformatie.



FIGUUR 3.4

23. Wat betekent *desinformatie*? Gebruik eventueel het internet.
24. Bekijk figuur 3.4. Een vergrote versie is te openen via onze werkomgeving (onder *bestanden module* → *hoofdstuk 3*). Wat is de hoofdboodschap van deze visualisatie?
25. Beoordeel deze visualisatie. Wat zijn de sterke punten? Zijn er voor jou onduidelijkheden? Zie je punten ter verbetering?
26. Welke desinformatie zou hiermee ontkracht kunnen worden?

De coronacrisis is internationaal uitzonderlijk goed gedocumenteerd. We bekijken een aantal websites die deze data visualiseren.

27. Open de corona-site met weekcijfers van het [RIVM](#). Bekijk de drie visualisaties.
28. Hoe is de betekenis van de visualisaties duidelijk gemaakt?
 - Wat is de betekenis van de schalen? En van de kleuren?
 - Op welke wijze worden hier meerdere kalenderjaren met elkaar vergeleken?
 - Wat gebeurt er als je met je muis over een grafiek beweegt?

Helaas waren er wereldwijd veel sterfgevallen tijdens de epidemie. We vergelijken twee websites die dit proberen te visualiseren.

29. Open de *Covid 19-dataexplorer* van [ourworldindata.org](#). Klik op *Play time-lapse*.
 - Wat laat deze visualisatie zien?
 - Welke conclusies trek jij uit het zien van deze visualisatie?
30. Open in een tweede tabblad de *Covid Trends* site van [aatishb.com](#). Klik op het afspeelicoontje.
 - Wat laat deze visualisatie zien?
 - Welke conclusies trek jij uit het zien van deze visualisatie?
31. Bespreek overeenkomsten en verschillen tussen beide visualisaties.

Opdracht 7 publieke data inlezen

In deze opdracht maak je *H3O07 publieke data inlezen*. Het gelijknamige bestand is te vinden via *bestanden module* in het startscherm van onze werkomgeving. Je oefent hier met het omzetten van online gepubliceerde data naar een SQL-database.