

Startbijeenkomst CoP ecologische data-analyse

Datum: 20-05-2021

Tijdstip: 9.00-12.30

Locatie: Online

Aanwezig: 55 deelnemers

Introductie STOWA (door Jasmijn)

Doel CoP: platform voor mensen die met data bezig zijn. Veel mensen hebben vergelijkbare ervaringen, worstelingen, vragen en oplossingen, maar weten elkaar nog lastig te vinden. Met de CoP willen we een platform bieden waar deze inhoudelijke vraagstukken aan de orde komen.

De CoP wordt aangestuurd door het kernteam, bestaande uit Jasmijn Rost (RHDHV), Marc Weeber (Deltares), Johan van Tent (HHSK), Luuk van Gerven (Aa en Maas) en Martin Droog (Dactylis).

Dit is de eerste bijeenkomst, het Kernteam is benieuwd naar wat men verwacht van de CoP, en daar willen we vandaag meer inzicht in krijgen.

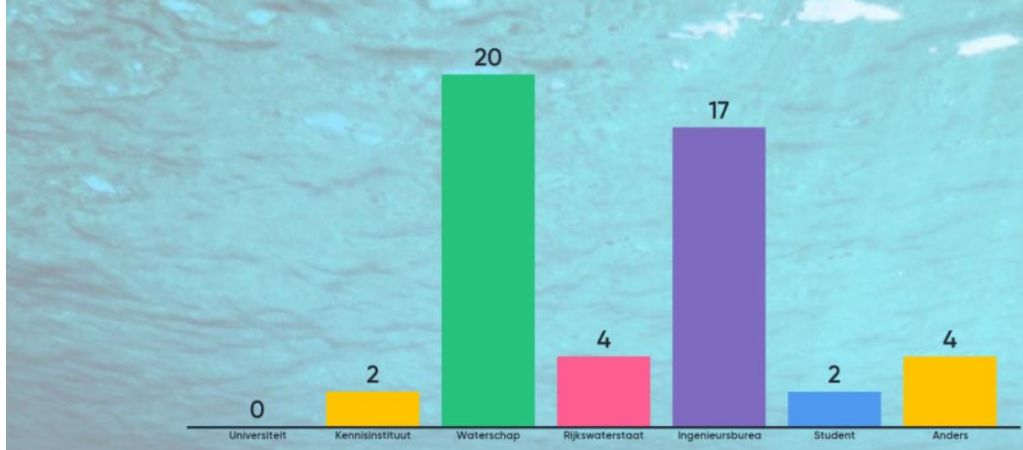
Introductie kernteam en kennismakingsquiz (Marc Weeber)

Via Mentimeter wordt een aantal vragen gesteld aan de aanwezigen.



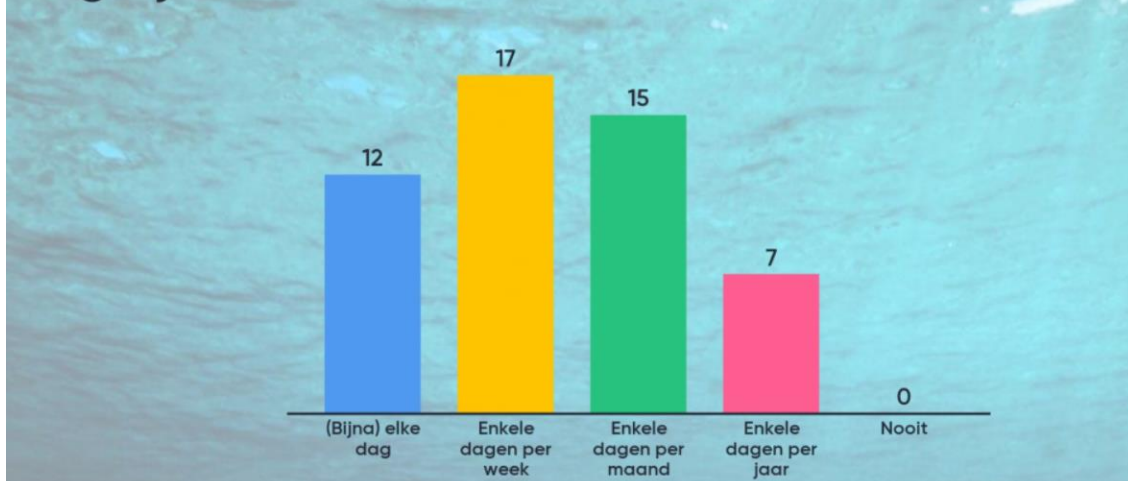
*Minke de Vries (Wetterskip Fryslân) vulde 'anders' in en merkt terecht op dat Fryslân ontbreekt (!). Sorry, Minke!

Wat is je huidige werkgever ?

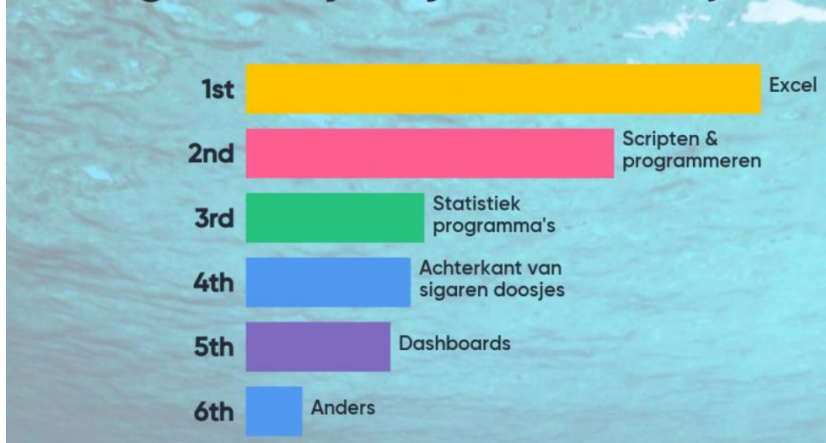


Onder 'anders' vallen onder andere opleidingscoördinator Jouke Heringa en Jako van der Wal en Mieke Moeleker vanuit een hydrobiologisch lab (Aquon).

Hoe vaak maak je gebruik van data-analyse in je dagelijkse werk?



Wat gebruik je bij data-analyse?



Onder 'anders' valt onder andere Notepad ++ (opmerking Ciska).

Wat hoop je uit deze dag te halen?



Mini pitches data-analyse projecten

Jako van der Wal (Aquon): Rekentool (Excel) diatomeeënbeoordeling. Handige spreadsheet die uit diatomeeën-samples de EKR-scores en achterliggende IPS-score berekend en inzichtelijk maakt.

Ciska Overbeek (HDSR): Als databeheerder heeft Ciska datacontrolescripts in R ontwikkeld die de kwaliteit en volledigheid van de biologische gegevens controleert. Zo worden punten en komma's rechtgezet, de volledigheid van meetpunten gecontroleerd etc. De scripts geven een rapportage in Excel-sheets, waardoor helder in beeld wordt gebracht waar fouten zitten en waar (handmatige) controles worden vereist. Er is veel interesse in deze routine en er wordt bij enkele waterschappen aan vergelijkbare methodes gewerkt.

Luuk van Gerven (Aa en Maas): Bij Aa en Maas is een overkoepelend Datalab opgezet. Hierin worden de data beheerd en gecontroleerd, maar er worden ook dashboards gebouwd en waar nodig analyses en voorspellende tools gemaakt. Het aantal data dat dagelijks binnenkomt wordt steeds groter, het datalab moet deze informatie functioneel en inzichtelijk maken. Er wordt gewerkt met R en Python.

Presentatie Robuust en herleidbaar inzicht in trends (Martin Droog)

In dit voorbeeldproject is een werkwijze geautomatiseerd (in R) om statistische regressie-analyses uit te voeren op een grote LMF-dataset van de provincie Gelderland. De nadruk ligt hier op het maken van een robuuste werkwijze die de routine makkelijk toepasbaar maakt op vergelijkbare dataset. Het script toetst elke subset op vereisten voor een type regressie en selecteert hiermee het type regressie dat geschikt is.

Er wordt gevraagd om de scripts te delen, dit wordt nagegaan bij de opdrachtgever. In de online chat wordt gediscussieerd over hoe je weet of de data correct zijn. En hoe je weet of het script/de werkwijze goed gaat? Hiervoor zijn controles in te bouwen, foutrapportages en tussentijdse uitdraaien. Voor de ecologische correctheid is echter een interpretatieslag op basis van ecologische kennis vereist, en dit is cruciaal in de vertaalslag naar een zinvolle en goede duiding van de resultaten. Bij statistiek in het bijzonder kunnen toetsen ogenschijnlijk goed gaan, terwijl het niet per se de juiste toets is. Dit blijkt nergens uit, maar moet de ecooloog zelf beoordelen.

Presentatie Reproduceerbaar rapporten: toepassingen in waterbeheer (Willem Stolte)

Willem gaat in op digitale rapportages, die een verrijking zijn ten opzichte van statische .docx of .pdf-documenten. Een rapportage dient in beginsel een goede informatiedrager te zijn. De interactieve opties en de automatische updates maken digitale rapportages op lange termijn kostenefficiënter, eenduidiger en goed up-to-date. Deze via webapplicatie beschikbare rapportages zijn naar verwachting de toekomst, en worden al veel toegepast. Denk bijvoorbeeld aan overzichten zoals een online energie-overzicht of een sport-app zoals Strava.

Willem maakt nadrukkelijk onderscheid tussen een dashboard en een rapportage, waarbij het eerste sneller update, hoofdzakelijk visueel is en op basis van indicatoren enige duiding kan geven, terwijl een rapportage ook minder vaak ververst en interpreterende tekst bevat, die vaak niet automatisch gegenereerd wordt. Tussen deze uitersten bestaan veel hybridevormen.

Als voorbeeld worden twee webpagina's getoond, één van Grevelingen (Deltares) en één van data van HHSK (gemaakt door Johan van Tent). In de discussie worden veel links gezet die helpen om deze toepassingen te leren maken.

Enkele voorbeelden van rekendocumenten:

- Digitale systeemrapportage Grevelingen (Deltares, WMR, SBB, RWS)
<https://www.deltaexpertise.nl/grevelingen/systeemrapportage/index.html>
 - R, Rmarkdown, Bookdown
- Waterkwaliteit HHSK 2020 (Johan van Tent)
https://www.schielandendekrimpenerwaard.nl/kaart/waterkwaliteit/waterkwaliteitsrapportages/wklrap_2020/
 - R, Rmarkdown, Bookdown
- Zeespiegelmonitor (Deltares)
<https://nbviewer.ipython.org/github/openearth/sealevel/blob/master/notebooks/dutch-sea-level-monitor.ipynb>
 - Python, Jupyter notebook

Goed om zelf mee te beginnen:

- <https://github.com/rstudio/bookdown-demo>

Work-in-progress open voorbeeld

- <https://github.com/wstolte/eerstelijnschelde>

Ook wordt het R-package *reticulate* genoemd, waarmee je python kunt draaien in R.

Tot slot wordt besproken dat het eindproduct voor veel opdrachtgevers nog vaak een statische versie is (PDF of docx), maar deze kunnen worden gegenereerd uit beide platforms (R-markdown en Jupyter). Het is ook wenselijk om rapportages op te nemen in de hydrotheek, en daar is als bijlagebestand ook ruimte voor de digitale rapportages.

Pauze

PAST Paleontological Statistics

Presentatie Analyseren kan je leren en introductie opdracht rivierkreeften (Johan van Tent)

Johan geeft een mooie weergave van de reis richting programmeren en (complexe) data-analyse, die voor velen herkenbaar is. Belangrijk hieraan is het gegeven dat iedereen zich deze skills eigen kan maken, en dat het deze CoP daar een centraal platform voor wil bieden.

De tips zijn:

1. Kennis van mogelijkheden: ontdekken wat er mogelijk is, is een eerste belangrijke stap naar zelf aan de slag gaan. Gebruik daarvoor je netwerk, dit CoP-platform etc. en wissel informatie uit;
2. Programmeren is een superpower. Laat de computer voor je werken, waardoor je veel meer kan dan je dacht (of met handmatig bereikt);
3. Snelheid is belangrijk, zowel in de wil om het te doen als in het vermogen om aanpassingen/suggesties te implementeren;
4. Maak je analyses zichtbaar, zowel voor jezelf (zodat je kunt controleren of het goed gaat en beter inzicht krijgt in de resultaten) als voor anderen (om de resultaten te communiceren);
5. Werk aan een goede, eenduidige structuur in je data. Dit helpt de leesbaarheid, je eigen denkproces en de uitwisseling met anderen.

Tijdens de discussie ging het onder andere over praktische obstakels. Vanwege 'de kracht van programmeren' is het binnen organisaties soms lastig toe te passen, omdat rechten in mappen niet alles toestaan (zoals bijvoorbeeld het gebruik van macro's in Excel).

Daarnaast komt het belang van versiebeheer van packages aan bod. Dit leidt soms tot problemen vanwege verschillen op verschillende computers. In R kun je met het package *renv* een snapshot maken van de versies.

Bespreken ervaringen en wensen in groepen (gezamenlijk)

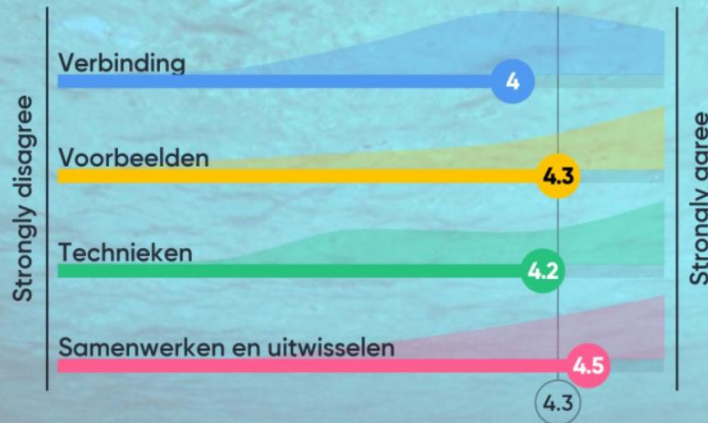
Johan introduceert een dataset over kreeften, waar iedereen zelf mee aan de slag kan. De gedachte is om zelf analyses te bedenken en uit te voeren. Bij de volgende bijeenkomst op **8 juli** wisselen we ervaringen en resultaten uit. Martin suggereert om dit in duo's of kleine groepjes te doen om tijdens het proces van elkaar te leren. Dit kan men naar eigen inzicht organiseren.

https://github.com/RedTent/CoP_EDA/raw/main/data/2021-03-23%20data_kreeften_vegetatie_2020_HHSK_CoP_Ecologische_data_analyse.xlsx

Suggestie Ciska: Gezamenlijke GitHub starten om bestanden uit te wisselen. Het kernteam gaat hiermee aan de slag!

Mentimeter

Hoe belangrijk zijn de doelen voor de CoP voor jou?



Wat zoek je vooral in de CoP Ecologische data-analyse?



Wat moeten we juist NIET doen als CoP Ecologische data-analyse?



Uit de mentimeter volgt een discussie over de afbakening van CoP, vooral over de vraag of we ons alleen bezighouden met de technieken, methodes en skills, of dat we ook de interpretatie en het ecologisch begrip van de materie behandelen. Het kernteam gaat met deze vragen aan de slag om dit concreter af te bakenen.