



USER! 2019 in Toulouse; Over Hexagons & Open Science

Marc Teunis

Dit keer schrijf ik de Zomer Colonette vanuit Frankrijk Toulouse, La Cité Rose. Twee jaar geleden rond dezelfde tijd gaf ik jullie een update vanuit Brussel. Nu zit ik bij hetzelfde congres als toen: USER! Een internationaal congres met meer dan 800 bezoekers, allemaal op de een of andere manier verbonden met de programmeertaal R. Dit keer laat ik geen R-code zien, dus dat valt mee.

Ik ben al een paar dagen op zoek naar hexagons. Dat zijn kleine raatvormige stickers. Een must-have voor de R gekkies. Ik moet er nog eentje, die van de Tidyverse (afb. 1). Maar elke keer als ik bij het kraampje van RStudio kom, zijn de stickers net allemaal voor mijn neus weggekaapt (zie [1] voor meer RStudio hexagons).

Wat opvalt tijdens de verschillende keynotes en presentaties is dat het meer over Open Science gaat. Waarschijnlijk heb je eerder gehoord van Open Access en Open Data. Wellicht zijn er sommige lezers die bekend zijn met het begrip Open Source, maar wat is Open Science? Hier in Toulouse is er bijna geen praatje of keynote lezing die niet hieraan refereert.



Afbeelding 1.

Open Science is een concept waarbij alle aspecten van het uitvoeren van wetenschappelijk onderzoek een open karakter hebben. In dit geval betekent open niet hetzelfde als openbaar of publiek beschikbaar, dat kan wel maar hoeft niet. Nee, open betekent dat je binnen een groep collega-onderzoekers en andere betrokkenen (bijvoorbeeld bio-informatici, statistici, projectleiders etc.) een transparante manier van werken hanteert waarbij alle stappen navolgbaar zijn. Tijdens en na het onderzoek worden er vaak producten publiek beschikbaar gemaakt. >>

Op deze manier is alles traceerbaar en is het onderzoek reproduceerbaar. Het voorkomt duplicatie en zorgt ervoor dat eventuele fouten in analyses kunnen worden opgespoord en gecorrigeerd. De manier waarop we wetenschap bedrijven door de Open Science beweging is radicaal aan het veranderen.

De programmeertaal R speelt een belangrijke rol in de op handen zijnde transformatie in hoe we onderzoeksdata genereren, analyseren en rapporteren [3]. Doordat budgetten voor onderzoek steeds kleiner worden en middelen beperkt, zijn onderzoekers alternatieven gaan zoeken voor bijvoorbeeld dure softwarelicenties en besturingsystemen. Lang is MATLAB de Microsoft Office geweest voor het lab, maar nu niet meer. R is een Open Source programmeeromgeving met een grote en actieve ontwikkel-community. Binnen deze community worden veranderingen binnen de taal en add-ons (zogenaamde packages) via een use-it-or-lose-it principe gepeerreviewed. Als er fouten in source code staan of zaken aanvulling nodig hebben dan wordt dat gefixt, door de community. En omdat deze community heel groot is gaat dat proces snel. Veel sneller dan in software waar grote bedrijven zoals Microsoft achter zitten. Een voorbeeld hiervan zie je in het 3-dimensionale staafdiagram hieronder. Over het algemeen wordt het afgeraden om zo'n grafiek te maken en ik laat hier zien waarom. Als ik zeg dat de waarden op de Y-as 7.75, 11.25 en 12.75 zijn, kun je die dan succesvol aflezen in de figuur? Op welke waarden kom je uit? Vooral de rode staaf geeft een andere waarde (~ 10) dan de in de Excel ingevoerde waarde aan. Deze bug schijnt al zo lang in Excel te zitten als het bestaat. Een dergelijke bug zou in Open Source software heel snel worden gefixt, of het programma zou in onbruik raken.



Open Science kan een boost geven aan het wetenschappelijke onderzoek als geheel. Wanneer we in staat zijn om de krachten te bundelen en de voortgang van de wetenschap te stellen boven persoonlijk of commercieel belang gebeuren en fantastische dingen. Het helpt als er meer voorbeelden komen van hoe het goed werkt. Soms kan geheel open niet, dan is het belangrijk na te gaan waar de win-win zit. Er zijn talloze goede voorbeelden van hoe het wel goed werkt, daar kunnen we van leren. En als iedereen haar (of zijn) steentje bijdraagt, komen we sneller verder [4].

Een fijne Open Zomer!

Afbeelding 2. Een voorbeeld van een 'buggy' bar-chart gemaakt in MS Office Excel (Office 365). De waarden op de Y-as zijn 7.75, 11.25 en 12.75 respectievelijk. Probeer deze waarden eens af te lezen in de grafiek, wat valt op?

Referenties:

- 1 <https://github.com/rstudio/hex-stickers>
- 2 <https://r4ds.had.co.nz/>
- 3 [_ROpenSci; https://ropensci.org/](https://ropensci.org/)
- 4 Masum H, Rao A, Good BM, Todd MH, Edwards AM, et al. (2013) Ten Simple Rules for Cultivating Open Science and Collaborative R&D. *PLoS Comput Biol* 9(9): e1003244. doi:10.1371/journal.pcbi.1003244

«